**CLARKSON UNIVERSITY**

# Comparing Dynamical Systems by Mostly Conjugacy

A dissertation
by

**Jiongxuan Zheng**

Department of Mathematics and Computer Science

Submitted in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy

**(Applied Mathematics)**

**2012**

Accepted by the Graduate School

_____ _____

Date Dean

The undersigned have examined the dissertation entitled

**Comparing Dynamical Systems by Mostly Conjugacy**

presented by

**Jiongxuan Zheng**

a candidate for the degree of

**Doctor of Philosophy (Applied Mathematics)**,

and here by certify that it is worthy of acceptance.

Examing Committee:

| | |
|---|---|
| Date | Erik M. Bollt (Advisor) |
| Date | Joseph D. Skufca (Advisor) |
| Date | Daniel ben-Avraham |
| Date | Jie Sun |
| Date | Adom Giffin |

# Abstract

A primary concern of this thesis is to develop principles and methods to compare dynamical systems when they are not necessarily conjugate (topologically the same). The first main body of this thesis provides an understanding of "mostly conjugacy (mostly homeomorphism)" between "dynamically close" systems, which enables us to measure and interpret the distance from being conjugate. We also generalize this idea from comparing deterministic systems to stochastic systems.

As a second theme, we extend and interpret the concepts of "mostly conjugacy" in symbolic dynamics, where we resort to a variant of the classic Monge-Kantorovich optimization problem to both built a useful change of variables and measure quality of the comparison through the underlying cost called the Wasserstein distance. Later, we build up a bundle structure, visualize as a bundle plot, to show the evolution of symbolic space as we vary a system's parameter. The main object is a specific structure "joint", which happens shortly after bifurcation, implies qualitative changes of system where the kneading points become periodic.

Finally, we apply the above techniques to study time series analysis and modeling on heart rate data, which we have shown to be a one-dimensional nonlinear map that has a stochastic parameter with persistence causing the heart rate and rhythm system to wander about a bifurcation point.

# Acknowledgments

I would like to thank my father (Rufu Zheng) and my mother (Yingqiu Feng) for their encouragement and support, which have been always my motivation. In addition, I cannot find words to express my gratitude to my girlfriend, without whose bad temper I couldn't grow into a patient person both in mathematics and life.

I am heartily thankful to my advisors, Professor Erik M. Bollt and Joseph D. Skufca, whose wisdom, patience and inspiration enable me to develop an understanding of mathematics. I am also thankful to Professor Daniel ben-Avraham, Jie Sun, Takashi Nishikawa, Aaron Luttman, Scott Fulton, Michael Felland and Jack Koplowitz for their education and guidance.

I am indebted to my colleagues and friends Sean Kramer, Stanley R. Huddy, Ranil Basnayake, Tian Ma, Kelum Hewa Gajamannage, Lingchen Bu and Dr. Rana Parshad. I enjoy the harmonious time in our Chaos Lab with all of you and your humor, optimistic and help.

Finally, thank all the committee members Professors Erik M. Bollt, Joseph Skufca, Daniel ben-Avraham, Jie Sun, and Adom Giffin.

# Contents

# List of Figures

# List of Tables

# Chapter 1

# Introduction

## 1.1 Motivation: A Conjugacy between Dynamical Systems

A fundamental question in science is how do we compare systems, or how do we represent data with the best model. Traditional approaches based on normed linear spaces, such as $L^2$, define two objects' distance to be the Euclidean distance. Difference between system-to-system or data-to- system can be measured by looking at the sum/intergration of deviation. Thus by regression analysis, we can obtain the best fit curve, from which we make prediction. The underlying assumption here is close systems should yield similar results, or close. But this concept is brittle in that it will break even for extremely simple systems.

For instance the logistic map, from which people model the population in the physical world. Assume we have two logistic systems, $x_{n+1} = 4x_n(1-x_n)$ and $y_{n+1} = 3.9999x_n(1-x_n)$, where I generate two time series data $x_n$ and $y_n$, with $x_0 = y_0 = 0.88$. From Figure 1.1, the data start acting tremendously different after $n = 14$. So two similar systems with the same formation but slightly different parameters values may behave qualitatively different.

Quantifying model accuracy is to measure how "close" the model is to the original system. In many cases, prediction is the goal. The quality of prediction in dynamical systems

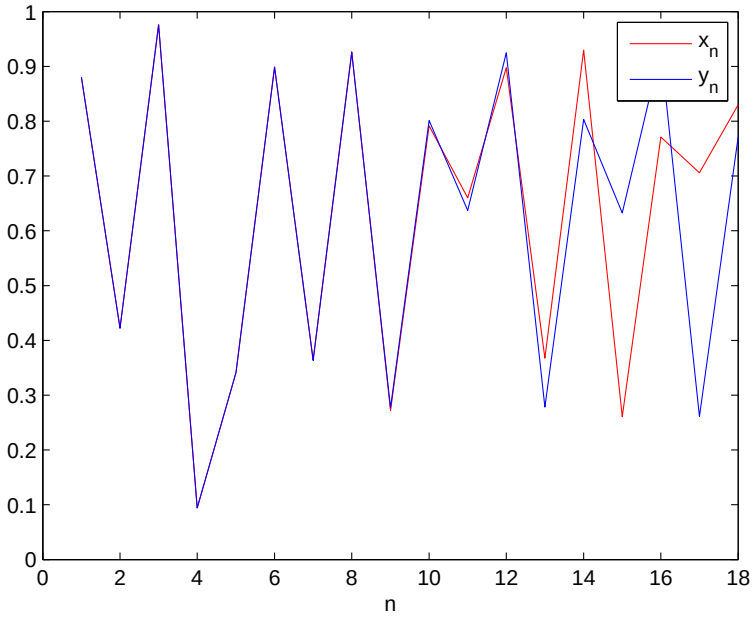


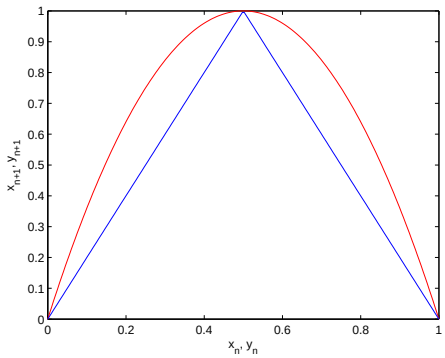

Figure 1.1: Time series data form two slightly different systems: (Red) $\{x_n\}$ is generated by $x_{n+1} = 4x_n(1 - x_n)$; (Blue) $\{y_n\}$ is generated by $y_{n+1} = 3.9999x_n(1 - x_n)$. These two systems differ in the parameter by 0.001 (close in the sense of regression analysis), but the trajectories behave qualitatively different after $n = 14$.

is typically not based on error analysis, like regression methods. Within the dynamical systems framework, the determination of whether two systems are dynamically equivalent is based upon whether or not there is a conjugacy between them. Conjugacy is essentially a smooth change of coordinates from one system to the other, where we say systems are topologically the same. Conjugate systems under this equivalent relationship behave exactly the same, and trajectories are matched via this conjugacy.

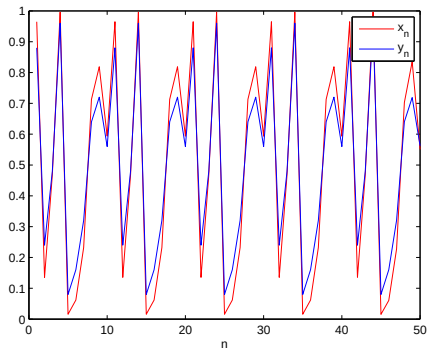
What do we mean by that? Now suppose we have two systems, i.e. the logistic map $x_{n+1} = 4x_n(1 - x_n)$ and the tent map $y_{n+1} = 2y_n \; if \; y_n < 1/2; otherwise \; y_{n+1} = 2(1 - y_n)$, which are shown in Figure 1.2(a). These two systems are in no way similar in the sense of regression analysis. If looking at the time series data from these two systems (Figure 1.2(b)), we can find that even the orbits points are not exactly at the same locations, but are moving following the same pattern. What do we mean by "pattern" here? Let's plot $\{x_n\}$ v.s. $\{y_n\}$ in Figure 1.2(c); we can see that each $x_i$ corresponds to a $y_i$ in a nice way, which forms a smooth function. As a matter of fact, they are related by a conjugacy function $y_n = (1 - cos\pi x_n)/2$, which implies they are considered the same topologically. Thus if given a time series data $x_n$, modeling using the tent map seems to be a better choice than $y_{n+1} = 3.9999y_n(1 - y_n)$. At least we can "describe" the demographic trends correctly. And in principle, we are representing the original data in a different coordinates system only.

Since the beginning of the field of dynamical systems by Henri Poincare [1], characterizing a dynamical system has focused on examining the topological and geometric features of orbits, rather than focusing on the empirical details of the solution of the dynamical system with respect to a specific coordinate system. We seek to understand coordinate independent properties, such as the periodic orbit structure, the count, and stability of periodic orbits.

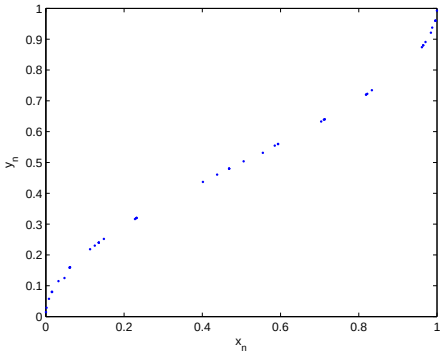
However, most of the time, we are trying to model the "true" system (physical perhaps) by a simpler one, that is "descriptive" of some aspects of the original system, in which case we have to address the issue of quantifying the quality in a mathematically grounded




(a) Tent Map and Logistic Map




(b) Time Series Data




(c) $\{x_n\}$ v.s. $\{y_n\}$

Figure 1.2: Conjugate systems: (a) Data $\{x_n\}$ is generated by the (red) system $x_{n+1} = 4x_n(1-x_n)$. Data $\{y_n\}$ is generated by (blue) system $y_{n+1} = 2y_n \; if \; y_n < 1/2; otherwise \; y_{n+1} = 2(1-y_n)$
; (b) Time series data generated by systems described in (a); (c) A plot of $\{x_n\}$ v.s. $\{y_n\}$. Even though these two systems are in no way the same in the sense of regression analysis, but each $x_i$ corresponds to $y_i$ by a conjugacy $y_n = (1 - cos\pi x_n)/2$, which implies they are considered the same topologically.

manner. In other word, we seek to approximate one thing by another, so we need to have some way of quantifying how much the underlying systems are conjugate.

So what do we mean by "not conjugate" but "approximately"? For our first example where $x_{n+1} = 4x_n(1 - x_n)$ and $y_{n+1} = 3.9999x_n(1 - x_n)$, if we symbol $x_n$ to be 0 when $x_n < 1/2$, otherwise 1, we would have the symbolic sequence 101001010111011101... for $\{x_n\}$ and 10100101011101011... for $\{y_n\}$ (see Figure 1.3). We can see that they have matching symbols until the 15th digit. And the fact is, if the parameter in the $\{y_n\}$ system is 3.99999 or a number closer to 4, we can have more matching symbols. Of course for conjugate systems, we would have a perfect matching. Thus a "good" model provides an "approximate conjugacy" that matches the respective symbolic sequences for as many symbols as possible. This also provides us an optimization criterion to present the "best" model. This is just part of the interesting story.

In this thesis work, we extend the notion "conjugacy" and focus on studying the "approximation" of conjugacy, named mostly conjugacy.

## 1.2 Purposes and Outcomes

In dynamical systems, the usual way to relate two dynamical systems is with the topological notion of conjugacy, related by the following commuting diagram:

$$\begin{array}{ccc} X & \xrightarrow{g_1} & X \\ \downarrow f & & \downarrow f \\ Y & \xrightarrow{g_2} & Y \end{array}$$

A conjugacy $f$ is a homeomorphism, i.e. $f$ is 1-to-1, onto, continuous and has a continuous inverse function, such that $f : X \to Y$ satisfying

$$f \circ g_1 = g_2 \circ f, \tag{1.1}$$

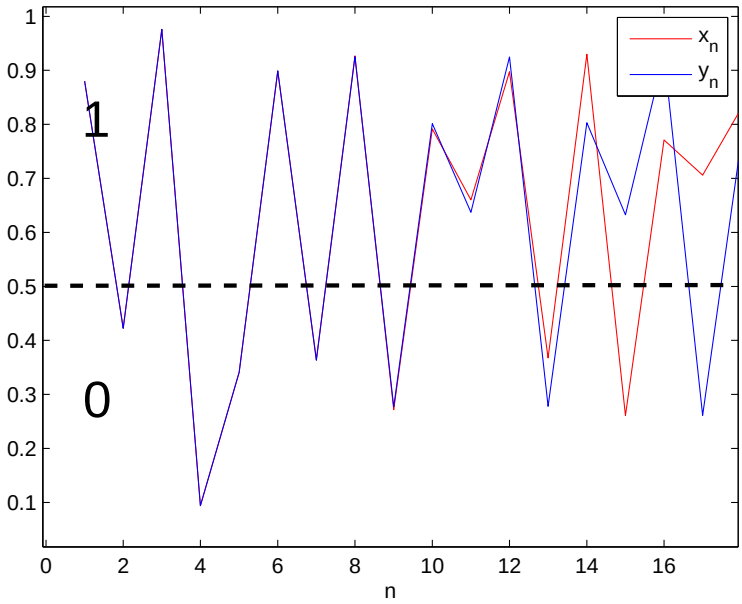



Figure 1.3: If we use the symbol 0 for $x_i < 0.5$ and $y_i < 0.5$; and 1 for $x_i > 0.5$ and $y_i > 0.5$, we would have the symbolic sequence 1010010101110111101... for $\{x_n\}$ and 10100101011101011... for $\{y_n\}$. They have matching symbols until the 15th digit. If the parameter in the $\{y_n\}$ system is 3.99999 or a number closer to 4, we can have more matching symbols. And we would have a perfect matching if we symbolize $\{x_n\}$ and $\{y_n\}$ that are from two conjugate systems.

It is a smooth change of coordinates from system $g_1$ to $g_2$ so that the mappings behave exactly the same in either coordinate system.

While we extend the notion of "conjugacy" by relaxing the condition that $f$ has to be a homeomorphism, which is rarely satisfied, especially when we try to model real data. In particular, we define the function $f : X \to Y$ satisfying $f \circ g_1 = g_2 \circ f$ as a "**commuter**". A commuter is a conjugacy if it is a homeomorphism, but a commuter is any function, not necessary a homeomorphism, that satisfies the commuting relationship.

In our work [6], we define the commuter $f$ to be a function in the $L^p$ space, which enables us to apply classical real analysis tools and measure the "homeomorphic defect". More importantly, we interpret the concept of "distance/defect" in the sense of "**mostly conjugacy**", which is an essential criterion when we do model optimization. Furthermore, we provides some basic regularity properties of the commuter function, and a modeling problem when only given a time series data. In addition, we introduce the algorithm to generate the commuter, and the computation error analysis.

Detail will be discussed in **Chapter 2**.

Later [49], we consider this comparison method within stochastic systems, since oftentimes we are dealing with data or models from stochastic systems rather than the deterministic ones. In this case we seek to extend our understanding of "mostly conjugacy" in a noisy world. Roughly speaking, we consider the underlying stochastic systems to be a "distribution" of deterministic systems, and for each instance we can refer to the establishment that we have. Such extension relies on the "random fixed point theorem [9]", which proves the existence and uniqueness of a "**random commuter** (a distribution version of the deterministic commuter)" for comparison between stochastic systems.

Detail will be discussed in **Chapter 3**.

Later, we realize there are limits in the mathematical technology requiring that the transformations be one-dimensional mappings for construction of the commuters. Further, there are difficulties in numerically computing defects in the more complicated one dimensional

cases, and further limits to higher dimensional problems. Thus, we focus on studying the commuter in the symbolic dynamics perspective. In contrast to studying the commuter as a point-wise matching from one coordinate system to the other, we try to provide a matching between symbol spaces from two systems, like the example of symbolizing trajectories in the previous section.

In **symbolic dynamics**, all allowable trajectories of a system form a shift space. But the concepts of shift space and the comparison between shift space are relatively abstract objects. We consider a shift space as a directed graph, and we applied the graph matching technique (Earth Mover's Distance) to compare shift spaces in the work [44]. We also demonstrate such resulting matching in association with "conjugacy/mostly conjugacy" in the dynamical systems view point.

Detail will be discussed in **Chapter 4**.

A step further, we study a topological dynamics perspective on the full bifurcation unfolding in unimodal mappings. We present a **bundle** structure, and visualize as a bundle plot, to show the evolution of symbolic space as we vary a system's parameter in paper [50]. The bundle plot can be viewed as a limit process of an assignment plot, which are lines assignments between points from two dynamical systems. Such lines assignments are determined by the commuter. The bundle structure is studied with the understanding of implications from system's qualitative changes. A main concern in the bundle plot is a special structure, called "joint", which determines a critical value of a parameter where the kneading point is being a period-$N$ point.

Detail will be discussed in **Chapter 5**.

On the data driven side, we study time series analysis and modeling on heart rate data based on the technique that we have built [43]. As interesting and challenging question in modeling behaviors of human cardiovascular system is how best to provide a simplified representation of both the deterministic and stochastic aspects of heart dynamics. We develop a one-dimensional nonlinear map that describes short term deterministic behavior

in the EKG signals. Our study suggests that there is a stochastic parameter with persistence which causes the heart rate and rhythm system to wander about a bifurcation point. We propose a modified circle map with a jump process noise term as a model which can qualitatively capture such this behavior of low dimensional transient determinism with occasional (stochastically defined) jumps from one deterministic system to another within a one parameter family of deterministic systems.

Detail will be discussed in **Chapter 6**.

Based on the low dimensional structure that we obtain from a persons 24 hours record, we use **Hidden Markov Models (HMM)** to characterize the heart rhythms pattern. Then apply the graph matching method, which we discuss in chapter 4, to compare the models and investigate the difference to a reference model.

Detail will be discussed in **Chapter 7**.

We have our conclusions and future works in **Chapter 8**.

## 1.3 Mathematical Terminology and Definitions

Here we present a brief review of necessary terminology, definitions and theorems, but the list is far from comprehensive.

**Definition 1.3.1** (Dynamical system [37])**.** *A dynamical system consists of a set of possible states, together with a rule that determine the present state in terms of past states. If the rule is applied at discrete times, it is called a discrete-time dynamical system.*

**Definition 1.3.2** (Chaos [39])**.** ***Chaos*** *is aperiodic long-term behavior in a deterministic system that exhibits sensitive dependence on initial conditions.*

*1. "Aperiodic long-term behavior" means that there are trajectories which do not settle down to fixed points, periodic orbits, or quasi-periodic orbits as* $t \to \infty$*. For practical reasons, we should require that such trajectories are not too rare. For instance, we could insist that there be an open set of initial conditions leading to aperiodic trajectories, or*

*perhaps that such trajectories should occur with nonzero probability, given a random initial condition.*

*2. "Deterministic" means that the system has no random or noisy inputs or parameters. The irregular behavior arises from the system's nonlinearity, rather than from noisy driving forces.*

*3. "sensitive dependence on initial conditions" means that nearby trajectories separate exponentially fast, i.e., the system has a positive* ***liapunov exponent****.*

*where we define that the limit to be the Liapunov exponent for the orbit starting at* $x_0$*:*

$$\lambda = \lim_{n\to\infty}\{\tfrac{1}{n}\sum_{i=0}^{n-1}\ln|f'(x_i)|\}.$$

*Note that* $\lambda$ *depends on* $x_0$*. However, it is the same for all* $x_0$ *in the basin of attraction of a given attractor. For stable fixed points and cycles,* $\lambda$ *is negative; for chaotic attractors,* $\lambda$ *is positive.*

**Definition 1.3.3** (Conjugacy)**.** *A conjugacy* $f$ *from system* $g_1$ *to* $g_2$ *is a homeomorphism, i.e.* $f$ *is 1-to-1, onto, continuous and has a continuous inverse function, such that* $f : X \to Y$ *satisfying* $f \circ g_1 = g_2 \circ f$*.*

**Remark 1.3.4** (Regression analysis)**.** *Here we give a brief review of Regression analysis with a simple application using regression line.*

*Suppose we are given samples of pairs* $(x_1, y_1), (x_2, y_2), ..., (x_n, y_n)$*, with all the* $x_i$ *not equal, we want to give a best fit line base on the Least Squares Principle, where the sum of the squares of the distances of those points from the straight line is minimum. Without loss of generality, we assume the distance is measured in the vertical direction (the y-direction).*

*We write the line as* $y = k_0 + k_1 x$*, and the distance to the line as* $|y_j - (k_0 + k_1 x_j)|$*. Thus the sum of the squares of distances is given by* $q = \sum_{j=1}^{n}(y_j - k_0 - k_1 x_j)^2$*. We want to determine* $k_0$ *and* $k_1$ *such that* $q$ *is minimized. We let* $\frac{\partial q}{\partial k_0} = 0$ *and* $\frac{\partial q}{\partial k_1} = 0$*, from which we obtain the best fit line as:*

$$y - \hat{y} = k_1(x - \hat{x}).$$

*where* $\hat{x} = \frac{1}{n}(x_1 + ... + x_n)$*, similar for* $\hat{y}$*, and* $k_1 = \frac{s_{xy}}{s_x^2}$*, where*

$$s_{xy} = \frac{1}{n-1}\sum_{j=1}^{n}(x_j - \hat{x})(y_j - \hat{y}), \text{ and } s_x^2 = \frac{1}{n-1}\sum_{j=1}^{n}(x_j - \hat{x})^2.$$

**Definition 1.3.5** (Graph Isomorphism)**.** *Let* $V(G)$ *be the vertex set of a simple graph and* $E(G)$ *its edge set. Then a graph isomorphism from a simple graph* $G$ *to a simple graph* $H$ *is bijection* $f : V(G) \to V(H)$ *such that* $u, v \in E(G)$ *if and only if* $f(u), f(v) \in E(H)$*.*

**Theorem 1.3.6** (Ulam's Conjecture [20])**.** *1. A finite rank approximation of the Frobenius-Perron operator by*

$$P(j|i) \approx \frac{\#(\{x_k | x_k \in B_i \text{ and } f(x_k) \in G_j\})}{\#(\{x_k \in B_i\})} \tag{1.2}$$

*as Ulam-Galerkin approximation of the Frobenius-Perron operator with a partition* $\Gamma = \{B_i\}$*, which is a finite family of connected sets with nonempty and disjoint interiors that covers the phase space.*

*2. The conjecture that the dominant eigenvector (corresponding to eigenvalue equal to 1 as is necessary for stochastic matrices) weakly approximates the invariant distribution of the Frobenius-Perron operator.*

**Theorem 1.3.7** (Ergodic Theorem)**.** *Let* $T : X \to X$ *be a measure-preserving transformation on a measure space* $(X, \Sigma, \mu)$*, with* $\mu(X) = 1$*. A measure-preserving transformation* $T$ *is ergodic if for every* $E \in \Sigma$ *with* $T^{-1}(E) = E$*, then either* $\mu(E) = 0$ *or* $\mu(E) = 1$*. If* $T$ *is ergodic, then*

$$\lim_{n\to\infty} \frac{1}{n}\sum_{k=0}^{n-1} f(T^k x) = \frac{1}{\mu(X)}\int f d\mu.$$

*where* $f$ *is a* $\mu$*-integrable function, i.e.* $f \in L^1(\mu)$*.*

**Definition 1.3.8** (Entropy of a dynamical system, also known as Metric Entropy, Measure-Theoretic Entropy, Kolmogorov Entropy, Kolmogorov-Sinai Entropy, or just KS entropy)**.** *Let $M$ be the phase space of a discrete time dynamical system, with a probability measure $\mu$. Let $T$ be a measure-preserving transformation of $M$ onto itself. Take a finite partition $\xi = \{C_1, C_2, ..., C_r\}$ of $M$. Let*

$$w_k(x) = j \text{ if } x \in T^{-k}C_j, \ -\infty < k < \infty.$$

*Thus each $x$ generates a random process $w(x) = \{...w_{-n}(x), ..., w_0(x), w_1(x), ..., w_m(x)...\}$. The so-called Shannon-MacMillian theorem tells us there exists the limit*

$$h(T, \xi) = \lim_{n\to\infty} \frac{1}{n} \sum_{i_1,...,i_n} \mu(T^{-1}C_{i_1} \cap ... \cap T^{-n}(C_{i_n}))ln(\mu(T^{-1}C_{i_1} \cap ... \cap T^{-n}(C_{i_n}))).$$

*The **Entropy of a dynamical system** is $h(T) = \sup_\xi h(T, \xi)$ where* sup *is taken over all finite partition $\xi$.*

**Theorem 1.3.9** (Relation between Measure Theoretic Entropy and Topological Entropy)**.**

$$h_{top}(T) = \sup_\mu h_\mu(T).$$

*where the supremum is taken over measures $\mu$ which are $T$-invariant Borel probability measures on $M$, i.e. for every measurable set $A$ in $M$, $\mu(T^{-1}(A)) = \mu(A)$.*

*Note the topological entropy of a dynamical system is the asymptotic growth exponent of the number of finite symbol sequences that the system can generate (given the best state-space partition).*

**Definition 1.3.10** (Symbol dynamics [17])**.** $\Sigma_2 = \{S = (s_0 s_1 s_2...) | s_j = 0 \text{ or } 1\}$. *We called $\Sigma_2$ the sequence space on the two symbols 0 and 1. (More generally, we can consider the space $\Sigma_n$ consisting of infinite sequences of integers between 0 and $n-1$.) Given two sequences $S = (s_0 s_1 s_2...)$ and $T = (t_0 t_1 t_2...)$, define the distance between them by $d(S, T) = \sum_{i=0}^{\infty} \frac{|s_i - t_i|}{2^i}$, which is a metric on $\Sigma_2$. The shift map $\sigma : \Sigma_2 \to \Sigma_2$ is given by $\sigma(s_0 s_1 s_2...) = (s_1 s_2 s_3...)$.*

**Definition 1.3.11** (The kneading sequence [17])**.** *Let $x \in [0, 1]$. The itinerary of $x$ under a unimodal interval mapg is the infinite sequence $S(x) = (s_0 s_1 s_2...)$ where*

$$s_j = \begin{cases} 0, & g^j(x) < x_c, \\ 1, & g^j(x) > x_c, \\ C, & g^j(x) = x_c. \end{cases}$$

*where $x_c$ is the critical point. The kneading sequence $K(g)$ of $g$ is the itinerary of $g(x_c)$, i.e., $K(g) = S(g(x_c))$.*

**Definition 1.3.12** (Gray Code Ordering)**.** *$s$ and $t$ have discrepancy $n$ if $s_i = t_i$ for $0 \leq i < n$, but $s_n \neq t_n$. Let $\tau_n(s)$ denote the number of 1's among $s_0, s_1, ..., s_n$. Suppose $s$ and $t$ have discrepancy $n$. We say $s < t$ if a. $\tau_{n-1}(s)$ is even and $s_n < t_n$; b. $\tau_{n-1}(s)$ is odd and $s_n > t_n$.*

**Theorem 1.3.13** (Intermediate value theorem for kneading sequences)**.** *[51] If a one-parameter family $G_t$ of continuous unimodal maps depends continuously on $t$ and the topological entropy $h(G_t) > 0$ for all $t$ then if $K(G_{t_0}) < K < K(G_{t_1})$ and $K \in \mathscr{M}$ where $\mathscr{M}$ is call the class of sequences which occur as kneading sequences of $G_t$ for all $t$, then there exists $t$ between $t_0$ and $t_1$ with $K(G_t) = K$.*

**Definition 1.3.14** (Hidden Markov Model [47])**.** *A hidden Markov model (HMM) is a statistical Markov model in which the system being modeled is assumed to be a Markov process with unobserved (hidden) states. In particular, A HMM is characterized by the following:*

- *1: number of hidden states $N$. We let $S(t) = \{S_1(t), S_2(t), \cdots, S_N(t)\}$ to be the set of hidden states at time $t$. Note that each of the hidden states in $S$ is a first order Markov process for any time $t$;*
- *2: number of observation states $M$. We let $O = \{O_1, O_2, \cdots, O_M\}$ to be the set of observation states. Note that each of the observation states in $O$ is not necessary a first order Markov process;*

- *3: the hidden states transition probability matrix* $A = [a_{ij}]$*, where* $a_{ij} = P(S_j(t+1)|S_i(t))$*. Note that* $\sum_{j=1}^{N} a_{ij} = 1,\ 1 \le i \le N$*;*

- *4: the emission matrix* $B = [b_{ik}]$*, where* $b_{ik} = P(O_k(t)|S_i(t))$ *for* $1 \le i \le N,\ 1 \le k \le M$*. Note that* $\sum_{k=1}^{M} b_{ik} = 1,\ 1 \le i \le N$*;*

- *5: the initial hidden states probability* $\pi = [\pi_i]$*, where* $\pi_i = P(S_i(0))$*. Note that* $\sum_{i=1}^{N} \pi_i = 1$*;*

*We let* $\lambda = (\pi, A, B)$ *to be the parameter of a hidden Markov model. Given a sequence of observations* $O_1^T$ *of length* $T$*, the goal is to find the best* $\hat{\lambda}$ *to maximize* $P(O_1^T|\lambda)$ *to characterize the signal* $O_1^T$*. For the resulting Markov model, we denote its stationary distribution for the hidden states by* $\Pi = \{\Pi_1, \Pi_2, \cdots, \Pi_N\}$*, where* $\Pi = \pi A$*. And we can obtain the stationary distribution for the observation states that are driven by the hidden dynamics by* $P(O) = \{P(O_1), P(O_2), \cdots, P(O_M)\}$*, where* $P(O_k) = \sum_{i=1}^{N} \Pi_i b_{ik}$*.*

**Definition 1.3.15** (Earth Mover's Distance [21])**.** *EMD is a bipartite network flow problem which can be formalized as a linear programming problem: Let* $I$ *be the set of supplies,* $J$ *be the set of consumers, and* $c_{ij}$ *be the cost to ship a unit from* $i \in I$ *to* $j \in J$*. We want to find a set of flow (matching matrix)* $f_{ij}$ *to minimize the overall cost:*

$$\sum_{i \in I} \sum_{j \in J} c_{ij} f_{ij},$$

*subject to the constrains:*

$$f_{ij} \ge 0,\ i \in I,\ j \in J$$
$$\sum_{i \in I} f_{ij} = y_j,\ j \in J$$
$$\sum_{j \in J} f_{ij} \le x_i,\ i \in I$$

*where* $x_i$ *is the total supply of supplier* $i$*, and* $y_j$ *is the total capacity of consumer* $j$*.*

# Chapter 2

# Mostly Conjugacy: Comparison on Dynamical Systems with Conjugacy Defect Measure

## 2.1 Introduction

Modeling is a fundamental problem whereby we seek to represent a system or data's behaviors. To the extent that science seeks to codify knowledge of the world, a basic tool in science is the model - a simplified representation of the "true" system under consideration, with mathematical models being a particular example. An essential question within this modeling context is "how close is the model to the true phenomena." Where the natural system under consideration is dynamic, with possibly complex behavior, the field of dynamical systems seeks to provide an appropriate framework for study of these systems. Since the inception of the field of dynamical systems by Henri Poincare [1], the fundamental approach has been to examine topological and geometric features of the orbits, rather than focusing on numerical specifics of particular solutions of the dynamical system, as measured in some specific coordinate system. Characterization of the system relies upon

deciphering coordinate independent properties, such as the periodic orbit structure - the count and stability of periodic orbits. Within this dynamical systems framework, the determination of whether two systems are dynamically equivalent is based upon whether or not there is a conjugacy between them. However, in any situation where we seek to approximate one thing by another, we need to have some way of quantifying the error in the approximation. Because "conjugacy" is an equivalence relationship, it can not be in the case where one system is only an "approximation" to the original system.

In the paper written by E. Bollt and J. Skufca [2], the concept of using a commuter to relate non-equivalent systems allows us to measure the topological difference between systems from a different perspective other than traditional methods, like regression analysis. The objective of defect measure provides a criterion to choose the "best" model to the original system or data source. Defining "goodness" in modeling should somehow quantify the quality of achieving our modeling goal, which is prediction in most of the cases. For classical method like regression analysis, comparisons between dynamical systems based on the least square method may fail if the systems turn out to be chaotic. J. Skufca and E. Bollt [2] develop a method of comparing dynamical systems by judging the quality of "matching" by looking at their topological difference (homeomorphic defect). The principle is to investigate how much one system can "smoothly" translate to another system, where smooth transformation means a function which is 1-to-1, onto, continuous and inverse continuous. The measure of deviation from being homeomorphism (homeomorphic defect) quantifies how much the commuter fails to be a homeomorphism. In this chapter, we will study questions like, whether closer systems imply less homeomorphic defect. In addition to E. Bollt's work [2], we extend the results by considering the function space in $L^p$, which provides us a broader view of the "measurable" commuter functions, and homeomorphic defect. Furthermore, we study the regularity of commuter functions. Finaly, we introduce our computation method, which allows one to approximate singular functions with smooth functions, with a computation error analysis given in the end.

This chapter is organized as follow: In section 2, we introduce the concept of mostly conjugacy, and give the example between tent map and logistic map, where the related functions are in $L^p$ space, with $1 \leq p < \infty$. In section 3, we will prove that, if a skew tent map converges to a full tent map, the commuter between this skew tent map and the full tent map converges to the identity function. In section 4, we show that using our scheme to generate commuter functions are measurable if the initial guess function is, and hence the limit of the sequence, i.e. the commuter, is also measurable. Together with the assumption that all the functions that we are discussing are bounded, we conclude that the commuter is in $L^p$. In section 5, we sketch the computation method of generating the commuter, and give the error analysis.

## 2.2 A Contraction Mapping From the Commutative Relationship

In the field of dynamical systems, the concept conjugacy describes an equivalent relation between dynamical systems. More precisely, we have the following definition,

**Definition 2.2.1.** *Let $X$ and $Y$ be topological spaces, and let $g_1 : X \to X$ and $g_2 : Y \to Y$. The dynamical systems $g_1$ and $g_2$ are conjugate if there exists a homeomorphism $h : X \to Y$, such that*

$$g_1(x) = (h^{-1} \circ g_2 \circ h)(x) \tag{2.1}$$

*for all $x \in X$.*

Topologically, $h$ is a function which is continuous, 1-to-1, onto, inverse continuous. If we rewrite (2.1) in a equivalent form

$$(h \circ g_1)(x) = (g_2 \circ h)(x) \tag{2.2}$$

So conjugacy describes a "perfect" matching in the sense of dynamical systems. A typical example is the one we gave in the introduction section, where time series data generated from two conjugate systems behave qualitatively the same. In principle, this is due to the fact that conjugacy is a translation from one coordinates system to another, where during translation we have no lost of dynamics. Here "no lost of dynamics" can be mathematically explained that the transformation is continuous, 1-to-1, onto, inverse continuous, namely homeomorphism. But what if the case is not that perfect? Most often times we are trying to model a complicated world with a simple system which describes the most representative aspects, not all. Or we are given data which are full of noise from unexpected sources which we don't want to model. So we want a reasonable quantifier to show how far this two objects are from being '"perfectly matched", while "closeness" should reflect the dynamical similarity between systems. We can achieve that by relaxing some conditions for the definition of conjugacy.

Here we introduce "mostly conjugacy" in the following definition with a function which satisfies the commutative relationship (2.2), but not necessary a homeomorphism.

**Definition 2.2.2** (Commuter)**.** *Let $X$ and $Y$ be topological spaces, and let $g_1 : X \to X$ and $g_2 : Y \to Y$. If $f : X \to Y$ satisfies the commuting relationship (2.2), then we say $f$ is a commuter relating the dynamical system $g_1$ and $g_2$.*

Notice that this definition doesn't require the function $f$ to be continuous, 1-to-1, onto or inverse continuous. In paper [2], we have iterative scheme to construct commuter functions in the Banach space $B([0,\ 1],\ R)$, the set of all bounded functions from $[0,\ 1]$ to the real numbers, with the norm defined as $\|f\| = \|f\|_\infty := sup|f(x)|$. The main goal of the section is to extend the function space to $L^p$ space, which enables us to do analysis and measure of commuters in a boarder setting.

Even we relax the condition of conjugacy, the existence and uniqueness of commuter function can still be guaranteed under some conditions, to show relationship between systems in some sense. In the following, we use the tent maps and logistic maps as an easy set-

ting to introduce how the commuter, or would be conjugacy, can be explicitly constructed by a fixed-point iteration. In this setting, we will rigorously prove several properties of commuters. We note that for more complicated problems, the methods still work numerically well under some conditions.

Now consider the family of skew tent maps $S(x)$ defined on $[0, 1]$, with:

(1) $S(0) = 0,\ S(1) = 0$;

(2) Peak of tent occurs at $S(a) = b$ with $0 < a < 1$;

(3) To ensure the map is locally expanding, require $max(a, 1 - a) < b \leq 1$.

which define the family of tent maps $\{S_{a,b}\}$

Consider the subset of maps that are symmetric about $x = 1/2$, denote the subfamily as $\tau,\ \tau \subset S,\ \tau_v := S_{1/2,v}$. First, the following lemma gives the existence of a conjugacy:

**Lemma 2.2.3.** *Let $S_{a,b}$ be a particular member of $S$. Then there exists a $v_0$ s.t. $S_{a,b}$ is conjugate to $T(v_0) \in \tau$. ($S \circ h(x) = h \circ T(x)$)*

Secondly, we require the commuter function maps monotone segments of the symmetric tent map to monotone segments of the full tent map. Since in $[0, 1/2]$, the symmetric tent map is $T(x) = 2v_0x$. From (1.2), we have

$$S \circ h(x) = h(2v_0x) \tag{2.3}$$

Notice that $h$ maps the domain $[0, 1/2]$ of of $S(x)$ to the domain $[0, a]$ of $T(x)$. i.e. $h[0, 1/2] = [0, a]$. Since $S(x) = b/ax$, together with (2.3), we get,

$$\frac{b}{a}h(x) = h(2v_0x) \tag{2.4}$$

Similarly, in $(1/2, 1]$, we can have

$$\frac{b}{1-a}(1 - h(x)) = h(2v_0(1 - x)) \tag{2.5}$$

Therefore, the conjugacy function $h(x)$ must satisfy:

$$h(x) = \begin{cases} \frac{a}{b}h(2v_0x), & 0 \le x \le 1/2, \\ 1 - \frac{1-a}{b}h(2v_0(1-x)), & 1/2 < x \le 1. \end{cases}$$

And, a conjugacy should map turning points to turning points, that is

$$h(1/2) = a \tag{2.7}$$

Then at $x = 1/2$, we have

$$h(1/2) = a = \frac{a}{b}h(v_0) \tag{2.8}$$

and

$$h(v_0) = b \tag{2.9}$$

Since the conjugate function $h(x)$ satisfied:

$$h(x) = \begin{cases} \frac{a}{b}h(2v_0x), & 0 \le x \le 1/2, \\ 1 - \frac{1-a}{b}h(2v_0(1-x)), & 1/2 < x \le 1. \end{cases}$$

Now use this equation as a guide, we create an operator whose fixed point will satisfy (2.2).

Consider the space $L^p([0,1])$ with norm $\|f\| = \|f\|_{L^p} := (\int_{[0,1]} |f|^p dx)^{\frac{1}{p}}$, $1 \le p < \infty$, which is a Banach space, complete. More precisely, we give the definition of the $L^p$ space in the following.

**Definition 2.2.4** (Definition of $L^p$ Space)**.** *If $E$ is a measurable subset of $R$ and $p$ satisfies*

*$1 \leq p < \infty$, then $L^p(E)$ denotes the collection of measurable $f$ for which $\int_E |f(x)|^p dx$ is finite, that is*

$$L^p(E) = \{f : \int_E |f(x)^p| dx < \infty\},\ 1 \leq p < \infty.$$

*We shall write*

$$\|f\|_{p,E} = (\int_E |f(x)|^p dx)^{\frac{1}{p}},\ 1 \leq p < \infty$$

*thus, $L^p(E)$ is the class of measurable $f$ for which $\|f\|_{p,E}$ is finite.*

From the closed subset $\mathcal{F} \subset B([0,1], R)$

$$\mathcal{F} = \{f | f : [0,1] \to [0,1]\}$$

Then given $(a, b)$ satisfying $max(a, 1-a) < b < 1$ defined a one-parameter family of operators.

$$\mathcal{M}_v : \mathcal{F} \to \mathcal{F} \text{ for } 1/2 < v \leq 1$$

$$\mathcal{M}_v f(x) := \begin{cases} \frac{a}{b} f(2vx), & 0 \leq x \leq 1/2, \\ 1 - \frac{1-a}{b} f(2v(1-x)), & 1/2 < x \leq 1. \end{cases}$$

Consider on $a, b$ and $v$ are required to ensure $\mathcal{F}$ is mapping into itself, also cause the operator to be a contraction.

**Lemma 2.2.5.** *$\mathcal{M}_v$ is a uniform contraction on $\mathcal{F}$, where the contraction is with respect to $\|\cdot\|_p$.*

*Proof.* Define $\lambda = max(\frac{a}{b}, \frac{1-a}{b})$, then $0 \leq \lambda < 1$,

$$\|\mathcal{M}_v f_1 - \mathcal{M}_v f_2\|_{p,[0,1]} = (\int_{[0,1]} |\mathcal{M}_v f_1 - \mathcal{M}_v f_2|^p dx)^{\frac{1}{p}}$$

For $0 \leq x < 1/2$

$$
\begin{aligned}
&(\int_{[0,1/2]} |\mathcal{M}_v f_1 - \mathcal{M}_v f_2|^p dx)^{\frac{1}{p}} \\
&= (\int_{[0,1/2]} |\frac{a}{b}(f_1(2vx) - f_2(2vx)|^p dx)^{\frac{1}{p}} \\
&= \frac{a}{b}(\int_{[0,1]} |f_1(y) - f_2(y)|^p d\frac{y}{2v})^{\frac{1}{p}} \\
&\leq \frac{a}{(2v)^{\frac{1}{p}} b}(\int_{[0,1]} |f_1(y) - f_2(y)|^p dy)^{\frac{1}{p}} \\
&\leq \frac{\lambda}{(2v)^{\frac{1}{p}}} \|f_1 - f_2\|_p
\end{aligned}
$$

Similarly, for $1/2 \leq x < 1$,

$$
\begin{aligned}
&(\int_{[1/2,1]} |\mathcal{M}_v f_1 - \mathcal{M}_v f_2|^p dx)^{\frac{1}{p}} \\
&= (\int_{[1/2,1]} |\frac{1-a}{b}(f_1(2vx) - f_2(2vx)|^p dx)^{\frac{1}{p}} \\
&\leq \frac{1-a}{b}(\int_{[0,1]} |f_1(y) - f_2(y)|^p d\frac{y}{2v})^{\frac{1}{p}} \\
&\leq \frac{1-a}{(2v)^{\frac{1}{p}} b}(\int_{[0,1]} |f_1(y) - f_2(y)|^p dy)^{\frac{1}{p}} \\
&\leq \frac{\lambda}{(2v)^{\frac{1}{p}}} \|f_1 - f_2\|_p
\end{aligned}
$$

Thus, $\|\mathcal{M}_v f_1 - \mathcal{M}_v f_2\|_{p,[0,1]} \leq \lambda \|f_1 - f_2\|_{p,[0,1]}$.

□

So $\mathcal{M}$ is a contraction, with contraction constrain constant $\lambda$. Because $\lambda$ not depends on $v$, the constant is uniform.

**Lemma 2.2.6.** *There is a unique $f_v \in \mathcal{F}$, so that*

$$\mathcal{M}_v f_v = f_v$$

*Moreover, for an arbitrary $f_0 \in \mathcal{F}$, if we define the sequence of functions*

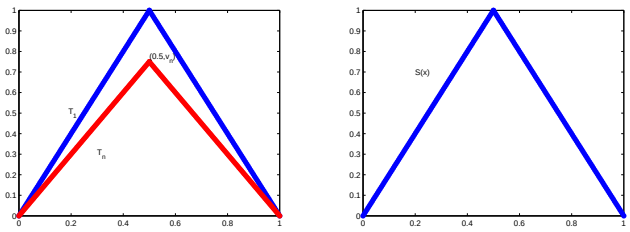
$$f_{n+1} = \mathcal{M}_v f_n$$

*this sequence will converge to* $f_v$

$$f_v := \lim_{n\to\infty} f_n$$

## 2.3 Convergence of Commuter Function

We note that a full tent map is conjugate to itself by the identity function which is a homeomorphism. But as we lower the height of the tent maps, we lose periodic orbits and hence the "dynamics" of the system is reduced. This can also be understood with the tool of the Kneading theory, which we will discuss in detail in later chapters. But now we are curious to the problem like, if we raise the height of a short tent map towards the full one, what will happen to the commuter function. Would the commuter converges to the identity function? Or even more, monotone converges? These questions are essential in the whole commuter topic, since first of all, it validates our motivations and settings in some sense; it also shows the regularity properties of the commuter function, which enables us to do optimization problem for choosing the "best model", at least theoretically.




(a) Shift Tent Maps (b) Full Tent Map

Figure 2.1: **Regularity property.** (Left) A tent map $T_n$ with peak at $(\frac{1}{2}, 1)$(blue), and a tent map $T_n$ with peak at $(\frac{1}{2}, V_n)$(red). (Right) A full tent map $S$. Here $T_n = S$, so the commuter between them is the identity function $I$. As $V_n$ going to 1, the commuter function between $T_n$ and $S$ is going to the identity function $I$ under the norm $\|\cdot\|_{p,[0,1]}$, $1 \leq p < \infty$

**Theorem 2.3.1.** *If the height of tent maps $\{V_n\}$ is a sequence going to 1, then the commuter from the symmetric tent map $T_n$ with peak $V_n$ to the full symmetric tent map $S$, is convergent to the identity function $I(x)$ under the norm $\|\cdot\|_{p,[0,1]}$*

*Proof.* From section 2, it is easy to see the commuter function between $T_n$ and $S$ satisfies the following equations:

$$h_n(x) = \begin{cases} \frac{1}{2}h_n(2v_n x), & 0 \le x \le \frac{1}{2}, \\ 1 - \frac{1}{2}h_n(2v_n(1-x)), & \frac{1}{2} < x \le 1. \end{cases}$$

Note: Just take $a = \frac{1}{2},\ b = 1$ at the original equations.

Also notice that $h_n$ maps monotone segments of the graph of $T_n$ onto monotone segments of $S$. i.e. $h_n$ maps segment on the interval $[0,\ \frac{1}{2}]$ of $T_n$ onto the segment on the interval $[0,\ \frac{1}{2}]$ of $S$. Also notice that the inverse function of $S$ exists on each of these intervals.

First consider $0 \le x \le \frac{1}{2}$,

$$\begin{aligned} \|h_n - I\|_{p,[0,\frac{1}{2}]} &= \|S^{-1} \circ h_n \circ T_n - S^{-1} \circ I \circ S\|_{p,[0,\frac{1}{2}]} \\ &\le L\|h_n \circ T_n - I \circ S\|_{p,[0,\frac{1}{2}]} \end{aligned}$$

Since in this simple example, we are dealing with the system $S(x) = 2x$. Thus it is easy to see that the lipshitz for $S^{-1}(x)$ is $\frac{1}{2}$. However, we will point out the fact in the next part that for more general systems beside $S(x) = 2x$, the above idea of proof with still work.

Thus,

$$\begin{aligned}\|h_n - I\|_{p,[0,\frac{1}{2}]} &\leq \frac{1}{2}\|h_n \circ T_n - I \circ S\|_{p,[0,\frac{1}{2}]} \\ &\leq \frac{1}{2}\|h_n \circ T_n - h_n \circ S\|_{p,[0,\frac{1}{2}]} + \frac{1}{2}\|h_n \circ S - I \circ S\|_{p,[0,\frac{1}{2}]} \\ &= \frac{1}{2}\|h_n \circ T_n - h_n \circ S\|_{p,[0,\frac{1}{2}]} + \frac{1}{2}(\int_0^{\frac{1}{2}} |h_n(2x) - 2x|^p dx)^{\frac{1}{p}}\end{aligned}$$

Let $2x = u$, we can substitute $x = \frac{u}{2}$ on the above equation because of its absolute continuity [2].

So,

$$\begin{aligned}\|h_n - I\|_{p,[0,\frac{1}{2}]} &\leq \frac{1}{2}\|h_n \circ T_n - h_n \circ S\|_{p,[0,\frac{1}{2}]} + \frac{1}{2}(\int_0^{\frac{1}{2}} |h_n(2x) - 2x|^p dx)^{\frac{1}{p}} \\ &= \frac{1}{2}\|h_n \circ T_n - h_n \circ S\|_{p,[0,\frac{1}{2}]} + \frac{1}{2}(\int_0^{1} |h_n(u) - u|^p d\frac{u}{2})^{\frac{1}{p}} \\ &= \frac{1}{2}\|h_n \circ T_n - h_n \circ S\|_{p,[0,\frac{1}{2}]} + \frac{1}{2^{1+\frac{1}{p}}}\|h_n - I\|_{p,[0,1]}\end{aligned} \tag{2.13}$$

On the other hand, for $\frac{1}{2} \leq x \leq 1$,

$$\begin{aligned}\|h_n - I\|_{p,[\frac{1}{2},1]} &\leq L\|h_n \circ T_n - I \circ S\|_{p,[\frac{1}{2},1]} \\ &\leq L\|h_n \circ T_n - h_n \circ S\|_{p,[\frac{1}{2},1]} + L\|h_n \circ S - I \circ S\|_{p,[\frac{1}{2},1]} \\ &= \frac{1}{2}\|h_n \circ T_n - h_n \circ S\|_{p,[\frac{1}{2},1]} \\ &+ \frac{1}{2}(\int_{\frac{1}{2}}^{1} |h_n(2(1-x)) - 2(1-x)|^p dx)^{\frac{1}{p}}\end{aligned}$$

Again, let $2(1 - x) = u$, we can substitute $x = 1 - \frac{u}{2}$ on the above equation because of its absolute continuity [2].

So,

$$\begin{aligned}
\|h_n - I\|_{p,[\frac{1}{2},1]} &\leq \frac{1}{2}\|h_n \circ T_n - h_n \circ S\|_{p,[\frac{1}{2},1]} \\
&\quad + \frac{1}{2}(\int_{\frac{1}{2}}^{1} |h_n(2(1-x)) - 2(1-x)|^p dx)^{\frac{1}{p}} \\
&= \frac{1}{2}\|h_n \circ T_n - h_n \circ S\|_{p,[\frac{1}{2},1]} + \frac{1}{2}(\int_{1}^{0} |h_n(u) - u|^p d(1-\frac{u}{2}))^{\frac{1}{p}} \qquad (2.14) \\
&= \frac{1}{2}\|h_n \circ T_n - h_n \circ S\|_{p,[\frac{1}{2},1]} + \frac{1}{2}(\int_{0}^{1} |h_n(u) - u|^p \frac{1}{2} du)^{\frac{1}{p}} \\
&= \frac{1}{2}\|h_n \circ T_n - h_n \circ S\|_{p,[\frac{1}{2},1]} + \frac{1}{2^{1+\frac{1}{p}}}\|h_n - I\|_{p,[0,1]}
\end{aligned}$$

Since $1 \leq p < \infty$, by raising $p$ power on (1)+(2), we can get

$$\begin{aligned}
\|h_n - I\|^p_{p,[0,\frac{1}{2}]} + \|h_n - I\|^p_{p,[\frac{1}{2},1]} &\leq (\|h_n - I\|_{p,[0,\frac{1}{2}]} + \|h_n - I\|_{p,[\frac{1}{2},1]})^p \\
&\leq (\frac{1}{2}\|h_n \circ T_n - h_n \circ S\|_{p,[0,\frac{1}{2}]} + \frac{1}{2^{1+\frac{1}{p}}}\|h_n - I\|_{p,[0,1]} \\
&\quad + \frac{1}{2}\|h_n \circ T_n - h_n \circ S\|_{p,[\frac{1}{2},1]} + \frac{1}{2^{1+\frac{1}{p}}}\|h_n - I\|_{p,[0,1]})^p
\end{aligned}$$

So,

$$\begin{aligned}
\|h_n - I\|_{p,[0,1]} &\leq \frac{1}{2}\|h_n \circ T_n - h_n \circ S\|_{p,[0,\frac{1}{2}]} + \frac{1}{2}\|h_n \circ T_n - h_n \circ S\|_{p,[\frac{1}{2},1]} \qquad (2.15) \\
&\quad + \frac{1}{2^{\frac{1}{p}}}\|h_n - I\|_{p,[0,1]}
\end{aligned}$$

which implies

$$\begin{aligned}
\|h_n - I\|_{p,[0,1]} &\leq \frac{1}{2(1-\frac{1}{2^{\frac{1}{p}}})}\|h_n \circ T_n - h_n \circ S\|_{p,[0,\frac{1}{2}]} \qquad (2.16) \\
&\quad + \frac{1}{2(1-\frac{1}{2^{\frac{1}{p}}})}\|h_n \circ T_n - h_n \circ S\|_{p,[\frac{1}{2},1]}
\end{aligned}$$

By the continuity in $L^p[2]$, since

$$\|T_n - S\|_{p,[0,\frac{1}{2}]} \to 0,\ \|T_n - S\|_{p,[\frac{1}{2},1]} \to 0\ (V_n \to 1)$$

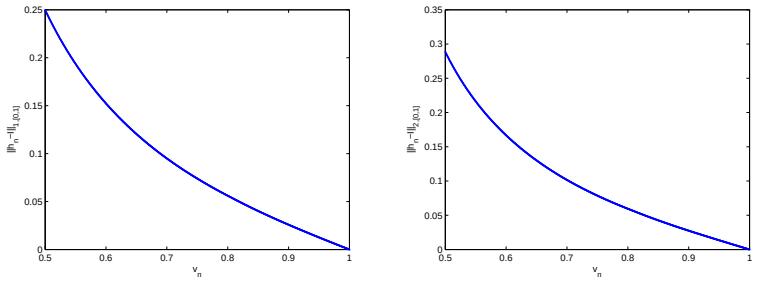
we have

$$\|h_n \circ T_n - h_n \circ S\|_{p,[0,\frac{1}{2}]} \to 0,\ \|h_n \circ T_n - h_n \circ S\|_{p,[\frac{1}{2},1]} \to 0$$

Thus

$$\|h_n - I\|_{p,[0,1]} \to 0\ as\ V_n \to 1,\ 1 \le p < \infty.$$

□




(a) Convergence Analysis with 1-norm

(b) Convergence Analysis with 2-norm

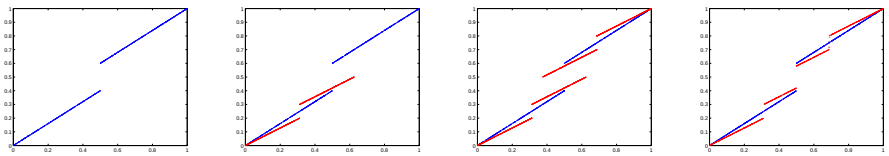
Figure 2.2: **Defect analysis** As $v_n \to 1$, the difference between $h_n$ and $I$ under the norm $\|\cdot\|_{p,[0,1]}$ is going to 0 strictly monotone.

The above theorem says that as $v_n \to 1,\ \|h_n - I\|_{p,[0,1]} \to 0$. In fact, the convergence under this norm is strictly monotone. Before proving this claim, we first look at some numerical evident from Figure 2.2.

As we increasing the values of $v_n$, we construct the commuter between the symmetric tent map with peak $v_n$ and the full symmetric tent map, then we calculate the difference between the commuters and the identity function under 1-norm and 2-norm. We can observe from Figure 2 that the values $\|h_n - I\|_{1,[0,1]}$ and $\|h_n - I\|_{2,[0,1]}$ are decreasing monotone as $v_n$ is increasing.

Please be notice here that, we calculate the "homeomorphic defect" by the difference from the commuter to the identity function in a $L^p([0,1]$. In the very first derivation [2], we define the homeomorphic defect particularly. But the computation is too complicated, and it is different from cases to cases. Thus in the above example, we applied a "surrogate defect measure" to capture the onto-defect, which turns out to be the only homeomorphic defect. In the following discussion, including an application about finding the "best model" for a specific time series data in later section, we still use this surrogate defect measure as the homeomorphic defect, where there only appears onto-defect.

Let's review a little bit about the construction of the commuter functions, which will provide implications for proving the monotone convergence of commuters.



(a) Commuter at $2^{nd}$ Iteration (b) Step 1 (c) Step 2 (d) Commuter at $3^{rd}$ Iteration

Figure 2.3: **Construction Procedure Of Commuter Function**

Generally, suppose we have the commuter function $h(x)$ at the $n^{th}$ step of iteration schema. At the next step, we make a copy of $h(x)$, shrunk it by $1/2$ in the vertical and by $2v$ in the horizontal (Figure 2.3(b)). Take a second copy, scaled the same horizontally, and vertically scaled by $1/2$. Rotate this copy by 180 degrees and place it in the upper right portion of the unit square (Figure 2.3(c)). Then truncate the left copy to the interval $[0, 1/2)$ and the right copy to $[1/2, 1]$. The result (Figure 2.3(d)) return the commuter function $h(x)$

at step $n + 1^{th}$

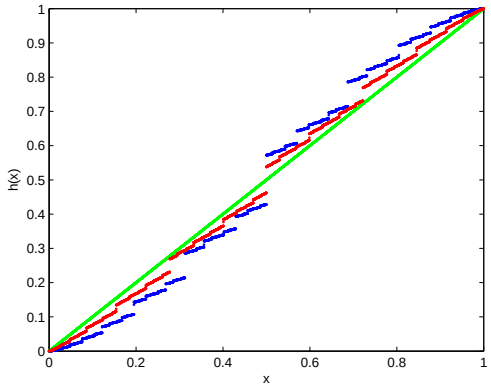



Figure 2.4: **Analysis for the commuter function:** (Green) The identity function. (Red) The commuter between the full tent map and the tent map with peak at $(\frac{1}{2}, 0.9)$. (Blue) The commuter between the full tent map and the tent map with peak at $(\frac{1}{2}, 0.8)$. Notice that on $[0, 0.5]$, every point of the red one is greater than the green one. While on $[0.5, 1]$, every point of the red one is less than the green one.

In Figure 2.4, the red curve is the commuter with the peak higher than the blue ones. And the green line is the identity function. We can see that on $[0, 0.5]$, every point of the red one is greater than the green one. While on $[0.5, 1]$, every point of the red one is less than the green one. This makes the difference between the identity function and the 'red' commuter, i.e.$|I - h_{red}|$, is less than that between the identity function and the 'green' commuter i.e.$|I - h_{blue}|$, which also implies $||h_{red} - I||_{p,[0,1]} < ||h_{blue} - I||_{p,[0,1]},\ 1 \leq p < \infty$.

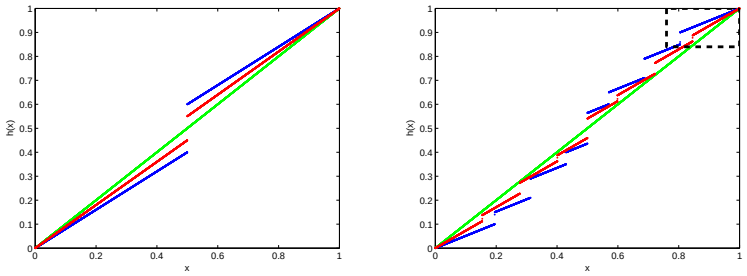
More precisely, we prove the monotone convergence of commuters as follows:

**Theorem 2.3.2.** *If $v_n$ is a sequence going to 1, then the commuter from the symmetric tent map $T_n$ with peak $v_n$ to the full symmetric tent map $S$, is monotone convergent to the identity function $I(x)$ under the norm $\|\cdot\|_{p,[0,1]}$*

*Proof.* Without loss of generality, we let $v_{red}$ be the peak of the higher symmetric tent map, while $v_{blue}$ be peak of the lower symmetric tent map. We are going to show that on the interval $[0, 1/2]$, every point of the commuter from the red symmetric tent map to the full

tent map is greater than the blue one. on the interval $[1/2, 1]$, every point of the commuter from the blue symmetric tent map to the full tent map is greater than the red one. Here we just give the prove for the 1-norm. For arbitrary p-norm, it is also true by the fact of the embedding theorem.

We prove it by induction.



(a) $1^{th}$ Iteration (b) $n^{th}$ iteration

Figure 2.5: **Commuter Functions with Different Peak Values**

For the first iteration, we start with the identity function, see the Green line in Figure 2.5(a). Then we make a copy of $h(x)$, shrunk it by $1/2$ in the vertical and by $\frac{1}{2v_{red}}$, $\frac{1}{2v_{blue}}$ respectively in the horizontal. Take a second copy, scaled the same horizontally, and vertically scaled by $1/2$. Rotate this copy by 180 degrees and place it in the upper right portion of the unit square. Then truncate the left copy to the interval $[0, 1/2)$ and the right copy to $[1/2, 1]$. The result return the commuter function $h_{red}(x)$ and $h_{blue}(x)$ after the first iteration.

We can see that the every red point in the interval $[0, 0.5]$ is greater than the blue one. Every blue point in the interval $[0.5, 1]$ is greater than the red one. That is true because the red one shrunk by $\frac{1}{2v_{red}}$, which is more than what the blue one shrunk. See Figure 2.5(a).

Now we suppose what we need to prove is true in the $n^{th}$ iteration. That is, every red point in the interval $[0, 0.5]$ is greater than the blue one. Every blue point in the interval $[0.5, 1]$ is greater than the red one. At the next step, we still construct the commuter just like the procedure in the first iteration above. Without loss of generality and for simplification, we assume that the commuters look like Figure 2.5(b). In fact this picture is the commuter

after three iterations. In order that at the $n+1^{th}$ iteration, every red point is greater than the blue one in the interval $[0, 0.5]$, we just need to figure out in the most upper right segment, whether the red points shrunk more enough than the blue ones, so that every red point in this segment runs above the blue one. If so, then every point of the whole red commuter will runs above the blue commuter, which is exactly what we want to happen. This can be tell easier if see Figure 2.5(b). Since the blue line segment moves $1/2(1/v_{blue})$ horizontally, the red line segment moves $1/2(1/v_{red})$ horizontally, the blue one moves $1/2(1/v_{blue} - 1/v_{red})$ with respect to the red one. For $n$ sufficiently large, the commuter is going to converge to the final commuter function, the slopes of the blue line segment and the red line segment are going to be 0. At the same time the length of them are going to be 0. So the left end points of the red and blue line segments will be closer and closer as $n \to \infty$, until their distance reaches 0, which is absolutely less than $1/2(1/v_{blue} - 1/v_{red})$. Thus the conclusion is true in $[0, 0.5]$. For $[0.5, 1]$, the argument is similar because of symmetry.

As a result, at the $n+1^{th}$ iteration, the conclusion is also true. This finishes the proof. □

The above statement can be generalized for commuters between arbitrary tent maps, but we may lose monotony of convergence. For example, if we have skew maps with peaks $(a_n, b_n)$ and $(a, 1)$ respectively, see Figure 2.6, and let $h_n$ be the commuter, as the sequence $\{(a_n, b_n)\}$ goes to $(a, 1)$, it can be also showed that $\|h_{a_n,b_n} - I\|_{p,[0,1]} \to 0$. More precisely, we have the following theorem:

**Theorem 2.3.3.** *If $(a_n, b_n) \to (a, 1)$, then the commuter from the skew tent map $T_n$ with peak $(a_n, b_n)$ to the skew tent map $T$ with peak $(a, 1)$, say $h_n$, is convergent to the identity function $I(x)$ under the norm $\|\cdot\|_{p,[0,1]}$, $1 \leq p < \infty$*

*Proof.* From the commutative diagram in section 2, we can have the commuter function between $T_n$ and $T$ satisfies the following equations

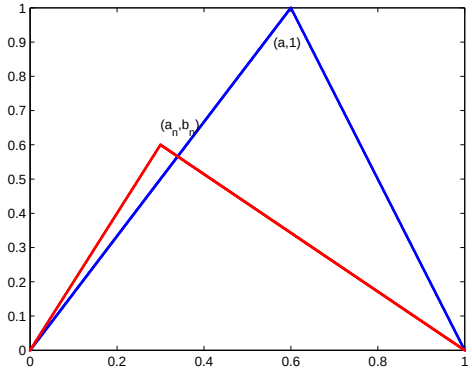



Figure 2.6: **Convergence Analysis for Skew Tent Maps**

$$h_n(x) = \begin{cases} ah_n(\frac{b_n}{a_n}x), & 0 \le x \le a, \\ 1-(1-a)h_n(\frac{b_n}{a_n}(1-x)), & a < x \le 1. \end{cases}$$

Also notice that $h_n$ maps monotone segments of the graph of $T_{a_n,b_n}$ onto monotone segments of $T$. i.e. $h_n$ maps segment on the interval $[0, a_n]$ of $T_{a_n,b_n}$ onto the segment on the interval $[0, a]$ of $T$. Also notice that the inverse function of $T$ exists on each of these intervals.

First consider $0 \le x \le a$,

$$\begin{aligned} \|h_n - I\|_{p,[0,a]} &= \|T^{-1} \circ h_n \circ T_{a_n,b_n} - T^{-1} \circ I \circ T\|_{p,[0,a]} \\ &\le L\|h_n \circ T_{a_n,b_n} - I \circ T\|_{p,[0,a]} \end{aligned}$$

Here, we are dealing with the system $T(x) = 1/ax$. Thus it is easy to see that the lipshitz for $T^{-1}(x)$ is $a$.

Thus,

$$\begin{aligned}
||h_n - I||_{p,[0,a]} &\leq a||h_n \circ T_{a_n,b_n} - I \circ T||_{p,[0,a]} \\
&\leq a||h_n \circ T_{a_n,b_n} - h_n \circ T||_{p,[0,a]} + a||h_n \circ T - I \circ T||_{p,[0,a]} \\
&= a||h_n \circ T_{a_n,b_n} - h_n \circ T||_{p,[0,a]} + a(\int_0^a |h_n(\frac{1}{a}x) - \frac{1}{a}x|^p dx)^{\frac{1}{p}}
\end{aligned}$$

Let $\frac{1}{a}x = u$, we can substitute $x = au$ on the above equation because of its absolute continuity [2].

So,

$$\begin{aligned}
||h_n - I||_{p,[0,a]} &\leq a||h_n \circ T_{a_n,b_n} - h_n \circ T||_{p,[0,a]} + a(\int_0^a |h_n(\frac{1}{a}x) - \frac{1}{a}x|^p dx)^{\frac{1}{p}} \\
&= a||h_n \circ T_{a_n,b_n} - h_n \circ T||_{p,[0,a]} + a(\int_0^1 |h_n(u) - u|^p da u)^{\frac{1}{p}} \\
&= a||h_n \circ T_{a_n,b_n} - h_n \circ T||_{p,[0,a]} + a^{1+\frac{1}{p}}||h_n - I||_{p,[0,1]}
\end{aligned} \tag{2.18}$$

On the other hand, for $a \leq x \leq 1$, we are dealing with the system $T(x) = 1/(1-a)(1-x)$. Thus it is easy to see that the lipshitz for $T^{-1}(x)$ is $(1-a)$.

$$\begin{aligned}
||h_n - I||_{p,[a,1]} &\leq L||h_n \circ T_{a_n,b_n} - I \circ T||_{p,[a,1]} \\
&\leq L||h_n \circ T_{a_n,b_n} - h_n \circ T||_{p,[a,1]} + L||h_n \circ T - I \circ T||_{p,[a,1]} \\
&= (1-a)||h_n \circ T_{a_n,b_n} - h_n \circ T||_{p,[a,1]} \\
&+ (1-a)(\int_a^1 |h_n(\frac{1}{1-a}(1-x)) - \frac{1}{1-a}(1-x)|^p dx)^{\frac{1}{p}}
\end{aligned}$$

Again, let $\frac{1}{1-a}(1-x) = u$, we can substitute $x = 1-(1-a)u$ on the above equation because of its absolute continuity [2].

So,

$$
\begin{aligned}
\|h_n - I\|_{p,[a,1]} &\leq (1-a)\|h_n \circ T_{a_n,b_n} - h_n \circ T\|_{p,[a,1]} \\
&\quad + (1-a)\Big(\int_a^1 |h_n(\frac{1}{1-a}(1-x)) - \frac{1}{1-a}(1-x)|^p dx\Big)^{\frac{1}{p}} \\
&= (1-a)\|h_n \circ T_{a_n,b_n} - h_n \circ T\|_{p,[a,1]} \\
&\quad + (1-a)\Big(\int_1^0 |h_n(u) - u|^p d(1-(1-a)u)\Big)^{\frac{1}{p}} \\
&= (1-a)\|h_n \circ T_{a_n,b_n} - h_n \circ T\|_{p,[a,1]} \\
&\quad + (1-a)\Big(\int_0^1 |h_n(u) - u|^p (1-a)du\Big)^{\frac{1}{p}} \\
&= (1-a)\|h_n \circ T_{a_n,b_n} - h_n \circ T\|_{p,[a,1]} \\
&\quad + (1-a)^{1+\frac{1}{p}}\|h_n - I\|_{p,[0,1]}
\end{aligned}
\tag{2.19}
$$

Since $1 \leq p < \infty$, by raising $p$ power on (16)+(17), we can get

$$
\begin{aligned}
\|h_n - I\|^p_{p,[0,a]} + \|h_n - I\|^p_{p,[a,1]} &\leq (\|h_n - I\|_{p,[0,a]} + \|h_n - I\|_{p,[a,1]})^p \\
&\leq (a\|h_n \circ T_{a_n,b_n} - h_n \circ T\|_{p,[0,a]} + a^{1+\frac{1}{p}}\|h_n - I\|_{p,[0,1]} \\
&\quad + (1-a)\|h_n \circ T_{a_n,b_n} - h_n \circ S\|_{p,[a,1]} \\
&\quad + (1-a)^{1+\frac{1}{p}}\|h_n - I\|_{p,[0,1]})^p
\end{aligned}
$$

So,

$$
\begin{aligned}
\|h_n - I\|_{p,[0,1]} &\leq a\|h_n \circ T_{a_n,b_n} - h_n \circ T\|_{p,[0,a]} + a^{1+\frac{1}{p}}\|h_n - I\|_{p,[0,1]} \\
&\quad + (1-a)\|h_n \circ T_{a_n,b_n} - h_n \circ S\|_{p,[a,1]} \\
&\quad + (1-a)^{1+\frac{1}{p}}\|h_n - I\|_{p,[0,1]}
\end{aligned}
\tag{2.20}
$$

As we assume that $\max\{a, 1-a\} < 1$, so $a < 1$ and $(1-a) < 1$. We also notice that $a^{1+\frac{1}{p}} < a$ and $(1-a)^{1+\frac{1}{p}} < (1-a)$. So here we have $a^{1+\frac{1}{p}} + (1-a)^{1+\frac{1}{p}} < a + (1-a) < 1$.

Rearrange the above equation, we get

$$\|h_n - I\|_{p,[0,1]} \leq \frac{a}{(1 - a^{1+\frac{1}{p}} - (1-a)^{1+\frac{1}{p}})} \|h_n \circ T_{a_n,b_n} - h_n \circ T\|_{p,[0,a]} + \frac{1-a}{(1 - a^{1+\frac{1}{p}} - (1-a)^{1+\frac{1}{p}})} \|h_n \circ T_{a_n,b_n} - h_n \circ T\|_{p,[a,1]}$$

By the continuity in $L^p[2]$, since

$$\|T_{a_n,b_n} - S\|_{p,[0,a]} \to 0,\ \|T_{a_n,b_n} - S\|_{p,[a,1]} \to 0\ ((a_n, b_n) \to (a, 1))$$

we have

$$\|h_n \circ T_{a_n,b_n} - h_n \circ T\|_{p,[0,a]} \to 0,\ \|h_n \circ T_{a_n,b_n} - h_n \circ T\|_{p,[a,1]} \to 0$$

Thus

$$\|h_n - I\|_{p,[0,1]} \to 0\ as\ (a_n, b_n) \to (a, 1),\ 1 \leq p < \infty.$$

□

We note that the above analytic results also applied to general unimodal systems besides tent maps. Later in this chapter, we give an example about modeling a time series data, generated from a short logistic map, by minimizing the defect.

## 2.4 The Measurability of Commuter Functions

The goal of this section is to prove the measurability of the commuter functions, and thus we can measure the homeomorphic defect. We apply most of the lemmas and theorems from classical real analysis, which can be found from [4].

**Theorem 2.4.1.** *For the commuter functions sequence $\{f_n\}$ generated from the iterative schema, if $f_1$, the initial guess commuter, is measurable, i.e. the identity function, the sequence $f_n$ is measurable.*

**Remark 2.4.2.** *$f$ is measurable means $\{x \in E : f(x) > a\}$ is measurable for every finite $a$, $E$ is measurable. Since $f_{n+1}[I_{X_i}] = g_{2i}^{-1} \circ f_n \circ g_1[I_{X_i}]$, $g_{2i}^{-1}$ is continuous on Y. $f_n : X \to Y$. Suppose the range of $g_1[I_{X_i}]$ is a measurable set.*

**Lemma 2.4.3.** *If $\varphi$ is continuous, $f$ is finite a.e. and measurable on $E$, then $\varphi \circ f$ is measurable.*

**Remark 2.4.4.** *Notice that the product of two measurable functions may not be measurable. That's the reason that I force the assumption of the measurability of the range of $g_1[I_{X_i}]$. (More details are in "Proof of Theorem 2.4.1") In fact, this assumption would not be too strong since the dynamical systems $g_1$, $g_2$ in our project are "regular" in some sense.*

*Proof of Theorem 2.4.1.* From the relation that

$$f_2[I_{X_i}] = g_{2i}^{-1} \circ f_1 \circ g_1[I_{X_i}],\ g_{2i}^{-1}$$

Since $f_1$ is measurable on the range of $g_1[I_{X_i}]$ as we suppose, i.e. most of the time we will start with $f_1 = I$, identity function.

So $f_2$ is measurable, so is $f_3 ... f_n$, by **Theorem 2.4.1**. □

**Lemma 2.4.5.** *If $\{f_n\}$is a sequence of measurable functions, then $\sup_k f_k(x)$ and $\inf_k f_k(x)$ are measurable. Here $\{x : \sup_k f_k(x) > a\} = \cup_k \{x : f_k(x) > a\}$*

*Proof.* [3] Since $\inf_k f_k = -\sup_k(-f_k)$, it is enough to prove the result for $\sup_k f_k$. This follow from the fact that $\{\sup_k f_k > a\} = \cup_k\{f_k > a\}$. □

**Theorem 2.4.6.** *From **Theorem 2.4.1**, $f_n$ converges to $f$. We claim that $f$ is also measurable, and hence $f$ is in $L^p$ because we assume the commuter function $f$ is bounded.*

*Proof.* Since

$$\lim_{k\to\infty} \sup f_k = \inf_j\{\sup_{k\geq j} f_k\}$$
$$\lim_{k\to\infty} \inf f_k = \sup_j\{\inf_{k\geq j} f_k\}$$

and $\lim_{k\to\infty} f_k$ exists, so it equals to $\lim_{k\to\infty} \sup f_k$ and $\lim_{k\to\infty} \inf f_k$, hence $f$ is measurable. □

## 2.5 Homeomorphic Defect

In this section, we briefly review the definition and computation of homeomorphic defect in paper [2], with a slight modification.

Suppose we have two dynamical systems, $g_1 : X \to X$ and $g_2 : Y \to Y$, and suppose we have the commuter $f$ such that $f \circ g_1 = g_2 \circ f$, we denote

$\lambda_O(f) = \{$*amount that $f$ is not onto*$\}$,

$\lambda_{1-1}(f) = \{$*amount that $f$ is not* $1-1\}$,

$\lambda_C(f) = \{$*amount that $f$ is not continuous*$\}$,

$\lambda_{C^{-1}} = \{$*amount that $f^{-1}$ is not continous*$\}$,

where we note that $f^{-1}$ may not be well defined. Then we define the homeomorphic defect of $f$, denoted $\lambda(f)$, as a convex combination

$$\lambda(f) = \alpha_1\lambda_O(f) + \alpha_2\lambda_{1-1}(f) + \alpha_3\lambda_C(f) + \alpha_4\lambda_{C^{-1}}(f), \tag{2.21}$$

where $0 \leq \alpha \leq 1$, and $\sum \alpha_i = 1$.

Theoretically, $\lambda(f) \geq 0$, with equality when $f$ is a homeomorphism. We note that the converse argument does not hold since the definitions are measure based, but the goal here is to maintain the flexibility of the definitions to allow broader applicability.

Detail definitions and interpretations of each defect can be found in paper [2]. In the following, I provide my computation method for each of the defects for unimodal systems, which differs a little bit from the one in paper [2], but with an easier implementation and a more convenient algorithm.

For the onto defect, a typical feature in the commuter is the appearance of vertical gaps. I simply measure the onto defect $\lambda_O(f)$ by measuring such gaps, which is given by

$$\lambda_O(f) = \int_x |f(x + \delta x) - f(x)| dx, \tag{2.22}$$

We note that in the numerical implementation, this computation will over estimate the onto defect, since even for a homeomorphism it will have a non-zero onto defect. But the goal of measuring the defect is to find the best model among a family of candidates (typically with the same formation) by minimizing the defect. So computationally, we prefer to use this surrogate defect measure to do approximation, but at the same time retain the regularity properties of commuters. Since the commuter has the fractal structure, most of the time we choose to only take the largest vertical gap as a surrogate onto defect measure.

For the 1-1 defect, we try to find the measure of $x \in X$ such that $f(x) = f(x')$ but $x \neq x'$. Thus we calculate the 1-1 defect as

$$\lambda_{1-1}(f) = \frac{sizeof(x : f(x) \; is \; not \; unique)}{sizeof(x)}, \tag{2.23}$$

We note that we use the *unique* function in matlab to help implement this estimation.

For the continuous defect, we measure it by:

$$\lambda_C(f) = \frac{sizeof(x : |f(x) - f(x + \delta x)| > \epsilon)}{sizeof(x)} \tag{2.24}$$

with $\epsilon$ pre-determined as a small value.

While calculating the inverse continuous defect, we simply rotate the function $f$, then compute the continuous defect.

And finally, we should consider the scaling of each defect, where we can adjust it by setting each weight $\alpha_i$.

We remark that in the previous section, we show the "monotone property" for the onto defect. Similar conclusions should be given for the other types of defects, which we reserve as future works. But we point out that $\lambda(f)$, a convex combination of convex functions, may not be convex. So the optimization process may result in non-unique local minimizers, which requires the modeler to judge which one is more likely to work.

## 2.6 Modeling Time Series Data by Minimizing Defects

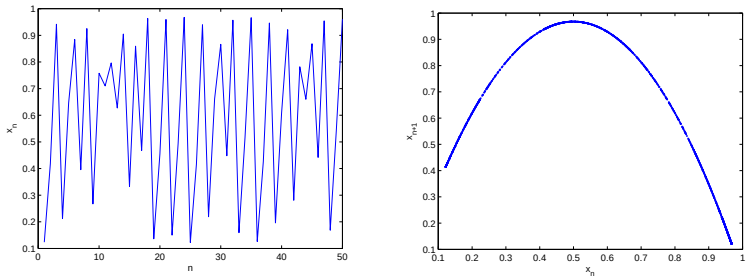
Now that we have established sufficient analysis for commuters, the next goal is to apply it to modeling problems. In particular, given a time series data generated by a deterministic system, how can we give the "best model" in the sense of mostly conjugacy.

First, we choose the family of candidate models with a parameter that we want to estimate, where the commuter between the data and the best model has the least defect among the family. In the implementation, we can write down the general algorithm as follows:
(1) Choose the family of candidate models $\{g_{2b}\}$, with parameter $b$;
(2) Compute the commuters $f_b$ from the data to each model $g_{2b}$, and the corresponding defect $\lambda(f_b)$;
(3) Choose the best model $g_{2b'}$ such that $b'$ minimizes $\lambda(f_b)$.

Note we achieve steps (2) and (3) by the bisection method.

Figure 2.7(a) shows the first 50th data points out of 2000 from a time series data generated by a short logistic map $g_1(x) = rx(1 - x)$, where $r = 3.87$, and the initial value is randomly picked. So given this time series data only, our goal is to determine the parameter

in the candidate models by mostly conjugacy. For the entire data set, which is of length 2000, we plot the time delay embedding $x_n$ versus $x_{n+1}$ in Figure 2.7(b). We note Figure 2.7(b) is the invariant set of the logistic system $g_1$.




(a) Time Series Data (b) $x_n$ versus $x_{n+1}$

Figure 2.7: **Model Optimization**

We choose our candidate models $\{g_{2b}\}$ to be defined as follows:

$$y_{n+1} = g_{2b}(y_n) = \begin{cases} \dfrac{b - y_{left}}{a - x_{left}}(y - a) + b, & y_n \leq 1/2, \\ \dfrac{b - y_{right}}{a - x_{right}}(y - a) + b, & y_n > 1/2. \end{cases}$$

which is the invariant set of the symmetric tent map with height $b$, the left fulcrum at $(x_{left}, y_{left})$, and the right fulcrum at $(x_{right}, y_{right})$ (see Figure 2.8 red tent map). We note that $y_{right} = g_{2b}(b)$ and $y_{left} = g_{2b}^2(b)$. So essentially the candidate model needs only 1 parameter $b$ to build up. Now let's look at the model $g_{2b}$ with $b = 0.9$ in Figure 2.9(a), where the commuter is presented in Figure 2.9(b). We can see the commuter exhibits the onto defect with vertical gaps observed. This tells that our model at this height is over estimate since the onto defect means there are some dynamics from the model that can not be represented by the data. On the other hand, if we take the model $g_{2b}$ with $b = 0.8$ in Figure 2.9(c), where the commuter is presented in Figure 2.9(d). We can see the commuter exhibits the 1-1 defect with flat spot observed. This tells that our model at this height is under estimate since the 1-1 defect means there are some dynamics from the data that

can not be represented by the model. We write down our defect function in the following equation (2.26), where we put 0 weight on the continuous defect and inverse continuous defect. This particular choice of the weights relies on the fact that the continuous defect overlaps the onto defect; while the inverse continuous defect overlaps the 1-1 defect. But we point out that in principle, modelers have their own weights on each of the defect terms in order to achieve their specific modeling purpose.

$$\lambda(b) = 0.5\lambda_{1-1}(b) + 0.5\lambda_{onto} + 0\lambda_C + 0\lambda_{C^{-1}}. \tag{2.26}$$

So the next step is to try to find the best model by minimizing the defect across $\{g_{2b}\}$ where $b \in [0.7, 1]$. And here, we define the surrogate onto defect measure $\lambda_{onto}(b)$ to be the largest vertical gaps in each commuter; and the surrogate 1-1 defect measure $\lambda_{1-1}(b)$ to be the largest flat spot in each commuter.

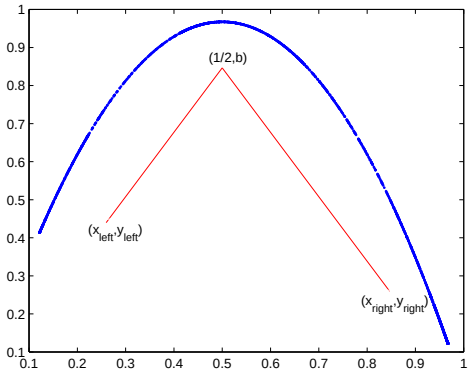
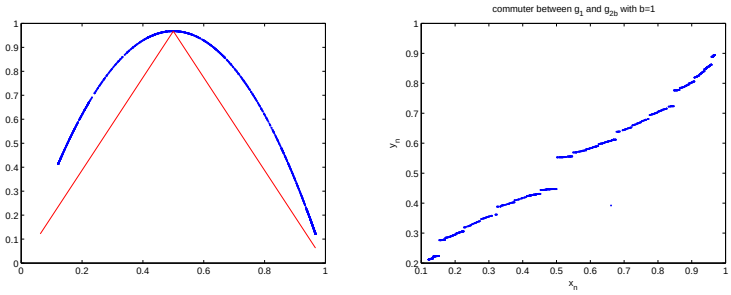


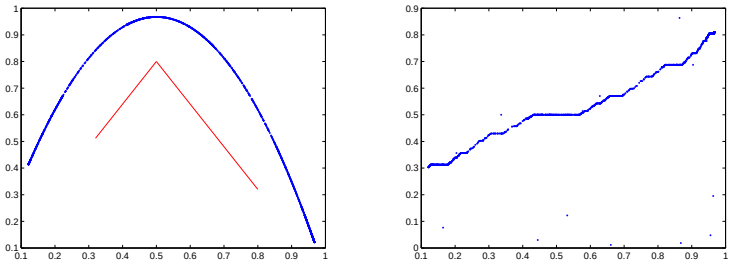

Figure 2.8: **Candidate Model**

By implementing our algorithm, we obtain the function $\lambda(b)$ in Figure 2.10(a), where we can achieve the minimum of $\lambda(b)$ at $b' = 0.8242424$. We plot the corresponding model $g_{2b'}$ in Figure 2.10(b) with red coloring. We can see that the best tent map is significantly shorter than the data (the logistic map). This makes sense since the tent map would exhibit more dynamics than the logistic map with the same height, unless they all reach height 1. In Figure 2.10(c), we also provide the commuter between the data and our best model, from




(a) Over height Model

(b) A commuter between $g_1$ and $g_{2b}$ with $b = 0.9$



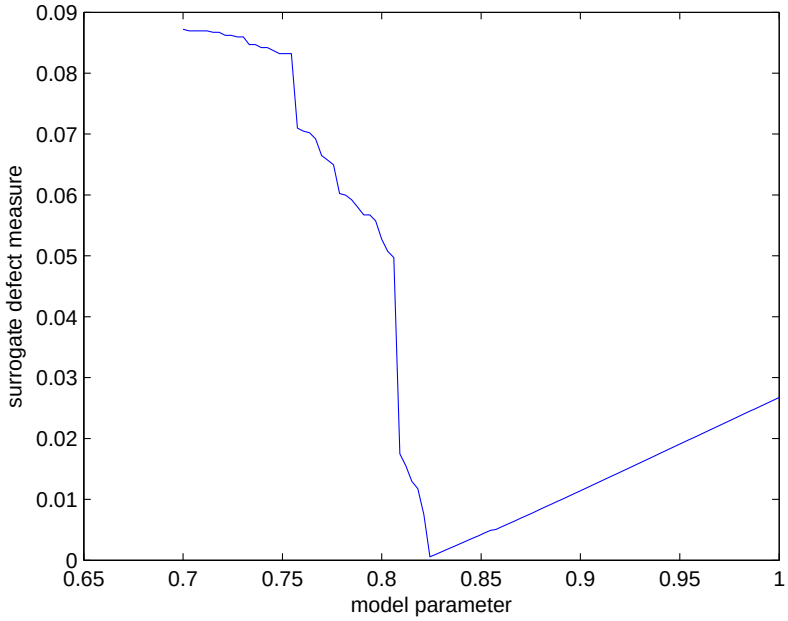
(c) Under height Model

(d) A commuter between $g_1$ and $g_{2b}$ with $b = 0.8$

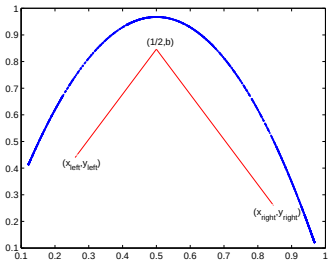
Figure 2.9: **Different Models**

which we rarely observe 1-1 defect or onto defect since, in theory, since these two systems are conjugate.

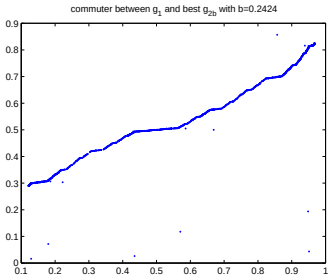
Now, how is our best model doing? One of the major goals of modeling is prediction, where accuracy is quantified by the "matching" between the time series data and the ones from model, which we plot it in Figure 2.11 for our example. As we discussed in the introduction, mostly conjugacy, as with regression analysis, seeks to match systems behaviors as much as possible. From Figure 2.11, the data emitted from the best model matches the data qualitative behaviors. In principle, they should have a perfect match for any length, but there might be inaccuracy which comes from computational aspects, which we will discuss in the next section. Nevertheless, we should presume that in most of the cases when we do modeling, we have no knowledge of what kinds of candidate models we should choose. And thus, we shouldn't expect that the best model is conjugate to the data with a 0 defect measure. In other words, it is only the best among the candidates, the one which captures




(a) Defect function




(b) best model




(c) Commuter between data and the best model at $b = 0.8242424$

Figure 2.10: **Surrogate defect measure**

the most dynamics of the data.

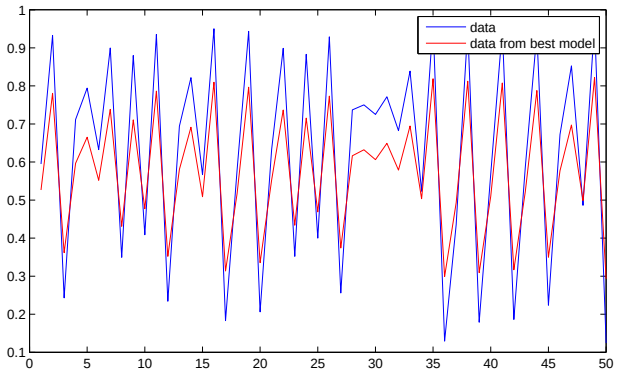



Figure 2.11: **Comparison between data and data from the best model**

At last we also want to note that, although the optimization process is straight forwards with the mostly conjugacy setting, the implementation can be very difficult for general systems especially non-unimodal systems. The main problem is, the defect of commuters may not be a piecewise monotone function with respect to the parameter. This due to the fact that we may have various types of defects at the same time, i.e. 1-to-1, onto, continuous and inverse continuous, which may not result in any convexity of $\lambda(b)$.

## 2.7 An Improvement On Piecewise Interpolation

As we mentioned in the previous section, we may have some computation aspects which affects the accuracy of the "best" model. Besides computational precision, the main sources of error come from the construction of commuters, which we always use piecewise step functions to interpolate and approximate. In particular, we compute the actual commuter function $f$ by $\hat{f}$, which is a piecewise interpolation function of the smoothed function of $f$, as showed in the following diagram:

$$f \xrightarrow{Blur} f_\epsilon \xrightarrow{PiecewiseInterpolating} \hat{f}$$

Where $f$ is the original commuter, in $L^p$ space, $1 \leq p < \infty$. While $f_\epsilon = f * K_\epsilon = \int f(x-t)K_\epsilon(t)dt$ and $\hat{f}$ is the piecewise interpolation to $f_\epsilon$.

**Remark 2.7.1.** *Define* $K_\epsilon(x) = \epsilon^{-n}K(\frac{x}{\epsilon}) = \epsilon^{-n}K(\frac{x_1}{\epsilon}, \frac{x_2}{\epsilon}, ..., \frac{x_n}{\epsilon})$ *with* $K \in L^1(R^n),\ \epsilon > 0$*, then it follows,*

$$(i)\ \int_{R^n} K_\epsilon = \int_{R^n} K = 1;$$
$$(ii)\ \int_{|x|>\delta} |K_\epsilon| \to 0\ as\ \epsilon \to 0,\ for\ any\ fixed\ \delta > 0$$

For example, if $K(x) = \chi_{\{|x|<1\}}(x)$, then $K_\epsilon(x) = \epsilon^{-n}\chi_{\{|x|<\epsilon\}}(x)$. In the following, we provide a error analysis on such approximation, and of course, to show that the approximated commuter can approach the actual one once we increase computational resolution.

**Theorem 2.7.2.** *If* $f \in L^p(R^n),\ 1 \leq p < \infty$*, then* $\forall \eta > 0,\ \exists \delta > 0,\ s.t.$

$$\begin{aligned} \|f - \hat{f}\|_p &\leq \|f - f_\epsilon\|_p + \|f_\epsilon - \hat{f}\|_p \\ &\leq \|K\|_1^{\frac{p}{p'}} \cdot [\eta\|K\|_1 + (2\|f\|_p)^p \int_{|t|\geq\delta} |K_\epsilon(t)|dt] + \frac{L_1L_2C_1\|f_\epsilon^{(n+1)}\|_p}{(n+1)!}h^{n+1} \end{aligned} \tag{2.27}$$

*Where h is the mesh size for* $x_n,\ L_1,\ L_2$ *is the Lipchiz constant,* $C_1$ *is a constant, n is the degree of interpolation degree, p and* $p'$ *are conjugate components.*

*Proof.* For the first term,

$$\begin{aligned} \|\hat{f} - f_\epsilon\|_p &\leq L_1 \cdot \|\hat{f} \circ g_1(x) - f_\epsilon \circ g_1(x)\|_p \\ &\leq \frac{L_1C_1\|f_\epsilon^{(n+1)}\|_p}{(n+1)!} \max\{g_1(x_i) - g_1(x_{i-1}), i = 1, 2...N\}^{n+1} \\ &\leq \frac{L_1L_2C_1\|f_\epsilon^{(n+1)}\|_p}{(n+1)!}h^{n+1} \end{aligned}$$

Notice that in the second inequality, we just apply the basic error estimate for piecewise interpolation. This can be found in many books about *Numerical Analysis* [3].

This finishes the first term. For the second term, we apply the analysis in book [4],

$$\hat{f}(x) = \hat{f}(x) \int_{R^n} K_\epsilon(t)dt = \int_{R^n} \hat{f}(x)K_\epsilon(t)dt$$

So

$$\begin{aligned} |\hat{f}(x) - f_\epsilon| &= | \int_{R^n} [\hat{f}(x-t) - \hat{f}(x)]K_\epsilon(t)dt \\ &\leq \int_{R^n} |\hat{f}(x-t) - \hat{f}(x)| \cdot |K_\epsilon(t)|^{\frac{1}{p}} \cdot |K_\epsilon(t)|^{\frac{1}{p'}} dt \end{aligned} \tag{2.28}$$

In 2.28, we apply Holder's Theorem for the term $\int_{R^n} K_\epsilon(t)dt$ with conjugate components $p$ and $p'$. Now we raise 2.28 to $p^{th}$ power and integrate with respect to $x$, we get

$$\begin{aligned} \int_{R^n} |\hat{f}(x) - f_\epsilon(x)|^p dt &\leq ||K||_1^{\frac{p}{p'}} [\int_{R^n} \int_{R^n} |\hat{f}(x-t) - \hat{f}(x)|^p |K_\epsilon(t)|dt]dx \\ &= ||K||_1^{\frac{p}{p'}} \cdot \int_{R^n} |K_\epsilon(t) \cdot [\int_{R^n} |\hat{f}(x-t) - \hat{f}(x)|^p |dx]dt \end{aligned} \tag{2.29}$$

So

$$||f_\epsilon - \hat{f}(x)||_p^p \leq ||K||_1^{\frac{p}{p'}} \cdot \int_{R^n} |K_\epsilon(t) \cdot ||\hat{f}(x-t) - \hat{f}(x)||_p^p dt \tag{2.30}$$

Notice that in 2.30, we can change the order of integration because the function $(f(x-t) - f(x))^p(K_\epsilon(t))$ is in $L^p(R^n \times R^n)$. And this comes from the fact that both $f(x)$ and $K_\epsilon(t)$ are in $L^p(R^n \times R^n)$.

For $\delta > 0$, write

$$\begin{aligned} \int_{R^n} |K_\epsilon(t)| \cdot ||\hat{f}(x-t) - \hat{f}(x)||_p^p dt &= \int_{|t|<\delta} + \int_{|t|\geq\delta} \\ &= A_{\epsilon,\delta} + B_{\epsilon,\delta} \end{aligned}$$

Continuity of $L^p$ space says: If $f \in L^p,\ 1 \leq p < \infty$, then

$$\lim_{|h|\to 0} \|f(x+h) - f(x)\|_p = 0$$

So $\forall \eta > 0,\ \exists \delta > 0,\ s.t.\ if\ |t| < \delta,\ \|f(x+h) - f(x)\|_p^p < \eta.$

Then

$$A_{\epsilon,\delta} \leq \eta \int_{|t|<\delta} |K_\epsilon(t)|dt \leq \eta \cdot \|K\|_1\ for\ all\ \epsilon \tag{2.31}$$

Moreover, $\|\hat{f}(x-t) - \hat{f}(x)\|_p^p$ is a bounded function. By Minkowski's Inequality,

$$\begin{aligned} \|\hat{f}(x-t) - \hat{f}(x)\|_p^p &\leq (\|\hat{f}(x-t)\|_p + \|\hat{f}(x)\|_p)^p \\ &= (2\|\hat{f}(x)\|_p)^p \end{aligned} \tag{2.32}$$

So,

$$B_{\epsilon,\delta} \leq (2\|\hat{f}(x)\|_p)^p \cdot \int_{|t|\geq\delta} |K_\epsilon(t)|dt \to 0\ as\ \epsilon \to 0\ for\ some\ fixed\ \delta > 0. \tag{2.33}$$

From 2.32 and 2.33,

$$\|f_\epsilon - \hat{f}\|_p \leq \|K\|_1^{\frac{p}{p'}} \cdot [\eta\|K\|_1 + (2\|f\|_p)^p \int_{|t|\geq\delta} |K_\epsilon(t)|dt]$$

This finishes the second term. □

## 2.8 Discussion

In summary, we develop a contraction mapping from the commutative relation between dynamical systems in $L^p$ spaces. We derive regularity work for the commuter function in $L^p$ spaces. Then we apply mostly conjugacy to do a modeling problem, where we provide the best model to a given time series data by minimizing the commuter defect. In addition, we give an error analysis on our computation of commuters, which allows us to approximate singular functions with smooth functions of arbitrary degree of differentiability.

In the next chapter, we will study the case where the underlying dynamical systems are stochastic. We would similarly develop a so-call "random commuter" to describe the random change of coordinates between two stochastic systems. Regularity property would also be studied.

# Chapter 3

# Randomly and Mostly Conjugacy: Comparison on Stochastically Perturbed Dynamical Systems

A central issue in the applied sciences is that of reduced modelling. It is often difficult if not impossible to differentiate a seemingly good model from an actually good one. The questions a modeler has to keep in mind are manyfold. What are the salient features of the physical process that should be captured by the model? What dynamical properties of the underlying process should be mimicked? What can perhaps be discarded? In the dynamical systems literature two systems are considered the same if there exists a conjugacy between them. This is essentially a homeomorphic change of coordinates that allows us to change freely between the two systems. However, for most real world applications this is seldom observed. No pragmatic model captures all the features of the process it is representative of. Thus a physical process and a "representative" model for it, are never quite conjugate. Once you accept that this discrepancy will almost always occur in most real world modelling applications, it becomes imperative to investigate the issues therein. This endevour seems to be important not only from a dynamical systems point of view, but from the more general

standpoint of modelling in science.

The work of Bollt and Skufca, [2], has investigated some of these issue in great detail. They develop a systematic methodology to compare systems which are not quite conjugate, thus coining the phrase "almost conjugacy". They lay down rigorously the notion of the commuter function, in such a setting. The commuter now departs from having the standard properties of a homeomorphism. Bollt and Skufca thus associate to this object a defect measure associated with how much it fails to be a homeomorphism. They prove the convergence of this object and derive a robust algorithm for its numerical computation. In [5] they give a symbolic dynamic interpretation to the commuter. This extends results form [2], confined mostly to one dimensional maps, to a higher dimensional setting. In my work et al., [6], rigorously justifies these computations by proving various regularity results concerning the commuter functions considered therein. I also makes various error estimates concerning the computation of the commuter function via the algorithm presented in [6].

A central theme of the above mentioned works is that they have all focused on a deterministic setting. In the current manuscript we will study the conjugacy between two dynamical systems, when one of these is perturbed via a stochastic term. Our hope is to lay down a rigorous first step in the direction of "mostly conjugacy" to stochastically perturbed dynamical systems. We endevour to thus extend the works of Bollt and Skufca in a random setting. Thus the foremost goal of this manuscript is to address the question, "How do you compare noisy systems, in the sense of conjugacy?". We hope to provide a reasonable answer to this question, detailing the nuances involved in the process, and discussing how we overcome them.

The outline of this chapter is as follows. In section 2 we outline certain heuristics as concerns the derivation of the Frobenius-Perron operator in some simple cases. This hopefully will provide the reader with sufficient intuition to grasp the analysis that follows. We also work out explicitly the evolution of densities via two dynamical systems (without any external noise) which are known to be conjugate. Here the exact form of the commuter

is known. We also present a definition of the random commuter. In section 3 we extend this analysis to cases where there is external noise. We deal with the case of uniform noise and normal noise and present our convergence results therein. In section 4 we present an alternate view of the commuter function. We present results when the commuter is viewed as a **random** operator. In section 5 we present some concluding remarks. We also provide appendicies where some detailed calculations are made to support some of the hypothesis made in the text.

## 3.1 Introduction

Consider the following dynamical system

$$x_{n+1} = S(x_n) \tag{3.1}$$

We can ask the question as to how densities of initial conditions evolve under this system as opposed to a particular initial condition. There is a standard methodology to this end. Note, no external noise is needed for this procedure. We can view the "randomness" as coming from the initial data. That is the initial data is unknown exactly apriori, we only know its distribution. We present certain heuristic ideas to lay down the frame work for the forthcoming analysis. For details the reader is referred to [7].

Given our dynamical system $f$ and N initial conditions $x_1^0, x_2^0, ..., x_N^0$, we apply $S$ on them to obtain N new states

$$x_1^1 = S(x_1^0), x_2^1 = S(x_2^0), ..., x_N^1 = S(x_N^0) \tag{3.2}$$

Given the N initial conditions, $x_1^0, x_2^0, ..., x_N^0$. We can perform an experiment in which we throw these onto the interval [0,1]. We could then count what fraction of these landed in say some predecided (not too small) interval $\Delta \subset [0, 1]$. This probability/quantity is easily

represented by the following sum

$$\frac{1}{N}\sum_{j=1}^{N} 1_{\Delta}(x_j^0) \tag{3.3}$$

Suppose now that we have a density function $f_0(x)$ for these initial states. This function governs the probabilities/likelyhood of these states falling inside $\Delta$. Heuristically it must be true that

$$\int_{\Delta} f_0(u)du = \frac{1}{N}\sum_{j=1}^{N} 1_{\Delta}(x_j^0) \tag{3.4}$$

proceeding exactly as before we can say map these initial states via $f$ onto the interval. The density function $f_1(x)$ for the states $x_1^1, x_2^1, ..., x_N^1$ should then satisfy

$$\int_{\Delta} f_1(u)du = \frac{1}{N}\sum_{j=1}^{N} 1_{\Delta}(x_j^1) \tag{3.5}$$

Our aim is to derive a relation between $f_0$ and $f_1$. To this end we define the pre-image of the interval $\delta$

$$S^{-1}(\Delta) = \{x : S(x) \in \Delta\} \tag{3.6}$$

Of course $x_j^1 \in \Delta$ if and only if $x_j^0 \in S^{-1}(\Delta)$. With this in mind we can explicitly write down a relation between $f_0$ and $f_1$.

$$\int_{\Delta} f_1(u)du = \int_{S^{-1}(\Delta)} f_0(u)du \tag{3.7}$$

We can choose $\Delta = [0, x]$ to yield

$$\int_{a}^{x} f_1(u)du = \int_{S^{-1}([0,x])} f_0(u)du \tag{3.8}$$

We can differentiate the above with respect to x as a parameter in the limits of integra-

tion to obtain

$$f_1(x) = \frac{d}{dx}\int_{S^{-1}([0,x])} f_0(u)du \tag{3.9}$$

We drop the subscripts in the above and define the **Frobenius-Perron operator** $P$ corresponding to the map S as

$$Pf(x) = \frac{d}{dx}\int_{S^{-1}([0,x])} f(u)du \tag{3.10}$$

This operator now gives us a rule to evolve densities. Although derived heuristically here, we will make extensive use of this operator in the forthcoming analysis. Equation (3.67) can be applied to the tent and logistic maps to see what the form of the Frobenius-Perron operator in this case is. Consider the full tent map

$$g(x) = 2x,\ 0 \leq x \leq \frac{1}{2} \tag{3.11}$$

$$g(x) = 2(1-x),\ \frac{1}{2} \leq x \leq 1 \tag{3.12}$$

Now we can invert g piecewise to obtain

$$g^{-1}([0,x]) = [0, \frac{x}{2}] \cup [1 - \frac{x}{2}, 1] \tag{3.13}$$

so using (3.67) we have

$$
\begin{aligned}
& Pg(x) \\
= & \frac{d}{dx}\int_{S^{-1}([0,x])} g(u)du \\
= & \frac{d}{dx}\left(\int_0^{\frac{x}{2}} g(u)du + \int_{1-\frac{x}{2}}^{1} g(u)du\right) \\
= & \frac{1}{2}g(\frac{x}{2}) + \frac{1}{2}g(1-\frac{x}{2})
\end{aligned} \tag{3.14}
$$

Now we use an iterative trick and assume $f(u) = 1$. Usually plugging this form into the above would give and expression for $P(Pg)$, and we'd proceed iteratively. Here however we have

$$Pg(x) = \frac{1}{2}(1) + \frac{1}{2}(1) = 1 \tag{3.15}$$

If we plug this in again for g we get

$$P(Pg(x)) = \frac{1}{2}(1) + \frac{1}{2}(1) = 1 \tag{3.16}$$

Thus we obtain

$$Pg(x) = \frac{1}{2}g(\frac{x}{2}) + \frac{1}{2}g(1-\frac{x}{2}) \tag{3.17}$$

We can apply a similar analysis to the logistic map

$$g(x) = 4x(1-x),\ 0 \le x \le 1, \tag{3.18}$$

to yield the Perron-Frobenius operator in this case.

$$Pg(x) = \frac{1}{2}f(\frac{1}{2} - \frac{1}{2}\sqrt{1-x}) + \frac{1}{2}f(\frac{1}{2} + \frac{1}{2}\sqrt{1-x}) \tag{3.19}$$

We will use equations (3.17) and (3.19) extensively.

### 3.1.1 Preliminaries

Consider the following dynamical systems

$$g_1(x) : X \to X \tag{3.20}$$

$$g_2(y) : Y \to Y \tag{3.21}$$

Recall that the dynamical systems are called conjugate if there exists a homeomorphism

$$f : X \to Y \tag{3.22}$$

Such that the following diagram commutes

$$\begin{array}{ccc} X & \stackrel{g_1}{\to} & X \\ \downarrow f & & \downarrow f \\ Y & \stackrel{g_2}{\to} & Y \end{array}$$

This in particular implies that the following equation holds pointwise

$$f(g_1(x)) = g_2(f(x)). \tag{3.23}$$

Our aim in the current manuscript is to adopt the above methodology to the case where one of the dynamical systems is forced by noise. It is of interest to see how the above plays out when we wish to evolve densities. This will lead naturally to our definition of a random

commuter which is the next order of business. We assume now that instead of $X$ our phase space is given by

$$L^1(X) \tag{3.24}$$

where $X = [0, 1]$. Formally we will confine our attention to

$$D = \left\{ \rho \in L^1(X) : \int_X |\rho(x)| dx = 1, \rho \geq 0 \right\} \tag{3.25}$$

$D$ is a space consisting of densities on $X$. Furthermore instead of $Y$ our phase space will be given by

$$L^1(Y) \tag{3.26}$$

where $Y = [0, 1]$. Again we are essentially looking at

$$D^{'} = \left\{ \rho \in L^1(Y) : \int_Y |\rho(y)| dy = 1, \rho \geq 0 \right\} \tag{3.27}$$

Now instead of points $x \in X$ moving under the action of the dynamics of $g_1$ and points $y \in Y$ moving under the action of the dynamics of $g_2$, we have densities $\rho(x) \in D$, and $\rho(y) \in D^{'}$ that are evolved under the action of transfer operators. These operators are given formally by the Frobenius-Perron operators

$$P_{g_1}\rho(x) : L^1(X \times \Omega) \to L^1(X) \tag{3.28}$$

$$P_{g_2}\rho(y) : L^1(Y) \to L^1(Y) \tag{3.29}$$

With this in mind we can make the following definition.

**Definition 3.1.1.** *Consider the following dynamical systems*

$$g_1(x_n) = S_1(x_n) \tag{3.30}$$

$$g_2(x_n) = S_2(x_n) + \xi_n \tag{3.31}$$

*Where $\xi_n$ is a stochastic forcing term. Furthermore assume the Frobenius-Perron operators for the systems are given by*

$$P_{g_1}\rho(x) : L^1(X \times \Omega) \to L^1(X) \tag{3.32}$$

$$P_{g_2}\rho(y) : L^1(Y) \to L^1(Y) \tag{3.33}$$

*The commuter function between the two dynamical systems is a density point-wise, in the sense that the following holds*

$$f(P_{g_1}\rho(x)) = P_{g_2}f(\rho(x)) \tag{3.34}$$

*This can be viewed as the following diagram commuting*

$$\begin{array}{ccc} L^1(X \times \Omega) & \overset{P_{g_1}}{\to} & L^1(X) \\ \downarrow f & & \downarrow\downarrow f \\ L^1(Y) & \overset{P_{g_2}}{\to} & L^1(Y) \end{array}$$

The two arrows on the right are a representation for any number of arrows each for a particular realisation of the noise. So in essence, for any particular realisation of the noise $\omega$, the Frobenius-Perron operator $P_{g_1}$ acts on a density $\rho(x, .) \in L^1(X \times \Omega)$ to produce a new density $P_{g_1}(\rho(x, .)) \in L^1(X)$. This density is carried by the commuter f to a new density $f(P_{g_1}(\rho(x, .))) \in L^Y$.

Along the other leg the commuter f acts on the density $\rho(x, \omega)$ and carries it to $L^1(Y)$. Here it is picked up by the Frobenius-Perron operator $P_{g_2}$ and carried to $L^1(y)$. The action of the commuter f can be viewed as a change of coordinates, following the standard theory of change of variables for densities. Here the jacobian of the transformation is essentially the Radon-Nikodym derivative. The next subsection will highlight this.

### 3.1.2 The case without external forcing noise

We will next work out explicitly what (3.23) turns out to be with two conjugate dynamical systems. However we will use (3.34). This will demonstarte that equation (3.34) is just a generalisation of equation (3.23), when we are evolving densities instead of individual points.

For purposes of testing we will assume $g_2$ to be the logistic map with $r = 4$ and $g_1$ to be the full tent map. In this case the form of the commuter $f(x)$ is well known. It is explicitly given as,

$$f(x) = \frac{1 - cos(\pi x)}{2} \tag{3.35}$$

Also the invariant densities under $g_1(x)$ and $g_2(x)$ are known. For the full tent map the invariant density is

$$\rho(x) = 1_{[0,1]}(x) \tag{3.36}$$

For the full logistic map it is

$$\rho(x) = \frac{1}{\pi \sqrt{x(1-x)}} \tag{3.37}$$

Furthermore the explicit form of the Frobenius-Perron operators are also known, see equations (3.19), (3.17), for quick derivations. In any event, we can compute

$$P_{g_1}\rho(x) = \frac{1}{2}\rho(\frac{x}{2}) + \frac{1}{2}\rho(1 - \frac{x}{2}) \tag{3.38}$$

$$P_{g_2}\rho(x) = \frac{1}{4\sqrt{1-x}}\left\{\rho(\frac{1}{2} - \frac{1}{2}\sqrt{1-x}) + \rho(\frac{1}{2} + \frac{1}{2}\sqrt{1-x})\right\} \tag{3.39}$$

Also note the change of variable formula that allows one to get at a density $\rho(y)$ of a random variable y, under the change of variable

$$y = f(x) \tag{3.40}$$

If the density under the random variable x, say $\rho(x)$ is known. The change of variable formula reads

$$\rho(y) = \left|\frac{1}{f'(f^{-1}(y))}\right|\rho(f^{-1}(y)) \tag{3.41}$$

We will now attempt to work out equation (3.34) explicitly for the case when our 2 systems are the full tent map and the logistic map, with parameter value $r = 4$. Thus given a $\rho(x) \in D$, $f(\rho(x))$ can be viewed as the new density under a change of variable already known to us. Essentially

$$y = f(x) = \frac{1 - cos(\pi x)}{2} \tag{3.42}$$

Thus

$$f^{'}(x) = \pi\frac{sin(\pi x)}{2} \tag{3.43}$$

and

$$f^{-1}(x) = \frac{1}{\pi}cos^{-1}(1 - 2x) \tag{3.44}$$

Using the above, consider the action of the change of variable on a density $\rho(x) \in D$

$$f(\rho(x)) = \rho(y) = \frac{1}{\pi\sqrt{y(1-y)}}\rho(\frac{1}{\pi}cos^{-1}(1-2y)) \tag{3.45}$$

Now under the action of the Frobenius-Perron operator $P_{g_2}$ the above is carried to

$$P_{g_2}f(\rho(x)) = \frac{1}{2\pi\sqrt{y(1-y)}}\left(\rho(\frac{1}{\pi}cos^{-1}(\sqrt{1-y})) + \rho(\frac{1}{\pi}cos^{-1}(-\sqrt{1-y}))\right) \tag{3.46}$$

Similarly we can compute

$$\begin{aligned} & f(P_{g_1}(\rho(x))) \\ = \; & \frac{1}{\pi\sqrt{y(1-y)}}\frac{1}{4\sqrt{1-\frac{2}{\pi}cos^{-1}(1-y)}}(\rho(\frac{1}{2}-\frac{1}{2}\sqrt{1-\frac{2}{\pi}cos^{-1}(1-y)})) \\ + \; & (\frac{1}{\pi\sqrt{y(1-y)}}\frac{1}{4\sqrt{1-\frac{2}{\pi}cos^{-1}(1-y)}})\rho(\frac{1}{2}+\frac{1}{2}\sqrt{1-\frac{2}{\pi}cos^{-1}(1-y)})) \end{aligned} \tag{3.47}$$

If we make a change of variable

$$y' = \frac{2}{\pi}cos^{-1}(1-y) \tag{3.48}$$

and then change

$$y' = y \tag{3.49}$$

treating $y'$ as a dummy variable, then the following will hold

$$f(P_{g_1}\rho(x)) = P_{g_2}f(\rho(x)), \tag{3.50}$$

if it is true that that

$$\rho(y) = \frac{1}{2\sqrt{1-\frac{2}{\pi}y}}(\rho(1-\sqrt{1-y})) \tag{3.51}$$

This follows easily by making the change of variable

$$y = 1-(1-y^{'})^{2} \tag{3.52}$$

to yield

$$\rho(y) = \frac{1}{2\sqrt{1-\frac{2}{\pi}y^{'}}}(\rho(1-\sqrt{1-y^{'}})) \tag{3.53}$$

but then changing $y^{'} = y$ it follows that

$$\frac{1}{2\sqrt{1-\frac{2}{\pi}y^{'}}}(\rho(1-\sqrt{1-y^{'}})) = \frac{1}{2\sqrt{1-\frac{2}{\pi}y}}(\rho(1-\sqrt{1-y})) \tag{3.54}$$

Thus we have demonstrated that the following holds

$$f(P_{g_1}\rho(x)) = P_{g_2}f(\rho(x)), \tag{3.55}$$

## 3.2 *The case with noise

**This section was written by the first author Rana D. Parshad in the paper [49], which mainly analysis conjugacy between stochastic systems with different types of external noise. I keep this section in my thesis for the completeness of our discussion.**

### 3.2.1 Uniformly distributed noise

We now consider the following two dynamical systems $g_1$ and $g_2$ as defined below.

$$\left\{\begin{array}{l} g_1(x,\xi_x) = 2(1-2\epsilon)x + (\xi_x+1)\epsilon, 0 \le x < 1/2, \\ \quad = 2(1-2\epsilon)(1-x) + (\xi_x+1)\epsilon, 1/2 \le x \le 1. \end{array}\right\} \tag{3.56}$$

$$\left\{\begin{array}{l} g_2(y) = 2y, 0 \le y < 1/2, \\ \quad 2(1-y), 1/2 \le y \le 1. \end{array}\right\} \tag{3.57}$$

We assume that $\xi_x$ is uniformly distributed in the interval unit interval $[0, 1]$. We wish to show the convergence of the Frobenius-Perron operator under $g_1$. Note the operator under consideration now contains a stochastic kernel. This can be written out explicitly following [7]. Let $(X, A, \mu)$ be a measure space. For any $f \in D$ we have

$$Pf(x) = \int_X K(x,y)f(y)dy \tag{3.58}$$

Here $K(x, y)$ is a stochastic kernel that satisfies

$$K(x,y) \ge 0, \tag{3.59}$$

and

$$\int_X K(x,y)dx = 1 \tag{3.60}$$

For our purposes

$$K(x,y) = g(x - S(y)) \tag{3.61}$$

where

$$g(x) = 1_{[0,1]}(x) \text{ is the density of } \xi. \tag{3.62}$$

We begin by introducing some definition and then recalling a theorem form [7].

**Definition 3.2.1.** *Let $(X, A, \mu)$ be a measure space and $P : L^1(X) \to L^1(X)$ a Markov*

*operator. Then* $\{P^n\}$ *is said to be* ***asymptotically stable*** *if there exists a unique* $f_* \in D$ *such that* $Pf_* = f_*$ *and*

$$\lim_{n\to\infty} \|P^n f - f_*\| = 0 \text{ for every } f \in D \tag{3.63}$$

**Theorem 3.2.2** (Lasota & Mackey, 1991)**.** *Let* $(X, A, \mu)$ *be a finite measure space and* $P : L^1(X) \to L^1(X)$ *a Markov operator. Assume there is a* $p > 1$ *and* $K > 0$ *such that for every density* $f \in D$ *we have* $P^n \in L^p$ *for sufficiently large n, and*

$$\limsup_{n\to\infty} \|P^n f\|_p \leq K \tag{3.64}$$

*Then P is constrictive*

**Definition 3.2.3.** *A constrictive operator is asymptotically stable.*

Note that $P$ being constrictive implies the convergence of $\{P^n\}$. We next state our result via the following Lemma,

**Lemma 3.2.4.** *consider the dynamical system perturbed stochastically*

$$\left\{\begin{array}{l} g_1(x, \xi_x) = 2(1-2\epsilon)x + (\xi_x + 1)\epsilon, 0 \leq x < 1/2, \\ \quad = 2(1-2\epsilon)(1-x) + (\xi_x + 1)\epsilon, 1/2 \leq x \leq 1. \end{array}\right\} \tag{3.65}$$

*where* $\xi_n$ *is i.i.d uniformly distributed in [0,1]. The stochastic Frobenius-Perron operator for this dynamical system converges to the stationary density of the system. That is*

$$\lim_{n\to\infty} \|P^n f - f_*\| = 0 \text{ for every } f \in D \tag{3.66}$$

*Proof.* The form of the Frobenius-Perron operator is easily constructed via techniques from [7]. It is given by the joint density of $f_n$ and $\xi$. That is

$$f_{n+1}(x) = Pf_n(x) = \int_{\mathbb{R}} f_n(y) g(x - S(y)) dy \tag{3.67}$$

Here $g(x - S(y))$ is the density of the random variable $\epsilon\xi$. Since $\xi$ is uniformly distributed in [0,1], equation (3.67) reduces to

$$\begin{aligned} & Pf_n(x) \\ = & \int_{\mathbb{R}} f_n(y)\epsilon\ 1_{[S^{-1}(x),S^{-1}(x-1)]}(x - S(y))dy \\ = & \epsilon \int_{S^{-1}(x)}^{S^{-1}(x-1)} f_n(y)dy \end{aligned} \tag{3.68}$$

Thus it follows that

$$\begin{aligned} & \limsup_{n\to\infty} |Pf_n(x)|_2^2 \\ = & \limsup_{n\to\infty} |\int_{\mathbb{R}} f_n(y)\epsilon\ 1_{[S^{-1}(x),S^{-1}(x-1)]}(x - S(y))dy|_2^2 \\ = & |\epsilon \int_{S^{-1}(x)}^{S^{-1}(x-1)} \limsup_{n\to\infty} f_n(y)dy|_2^2 \text{ via Lebesgue dominated convergence theorem} \end{aligned} \tag{3.69}$$

Note, that markov maps are dense in the space of piecewise linear maps, [8]. Thus it follows that there exists an $N$ such that for $n \geq N$

$$|f_n(y) - f^*|_2 \leq \epsilon_1 \tag{3.70}$$

Where $f^*$ is the density of a markov map, thus is given by a piecewise constant function

$$\limsup_{n\to\infty} \sum_{i=1}^{k} C_i 1_{[a_i,b_i]}(x) \tag{3.71}$$

Note the following can also be obtained via approximating $f_n$ by a step function, since $f_n \in L^1(X)$ by definition. In any event we obtain,

$$
\begin{aligned}
& |\epsilon \int_{S^{-1}(x)}^{S^{-1}(x-1)} (\limsup_{n\to\infty} f_n(y) - \sum_{i=1}^{k} C_i 1_{[a_i,b_i]}(y))dy|_2^2 \\
\leq\ & \epsilon| \int_{S^{-1}(x)}^{S^{-1}(x-1)} \sum_{i=1}^{k} C_i 1_{[a_i,b_i]}(x)|_2^2 + \epsilon_1 \\
\leq\ & \epsilon C| \int_{S^{-1}(x)}^{S^{-1}(x-1)} dy|_2^2 + \epsilon_1 \\
=\ & \epsilon C|S^{-1}(x-1) - S^{-1}(x)|_0^1 + \epsilon_1 \\
\leq\ & K
\end{aligned} \tag{3.72}
$$

Thus we can conclude that

$$
\limsup_{n\to\infty} |Pf_n(x)|_2^2 \leq K, \tag{3.73}
$$

Thus the Lemma is proved via application of the earlier cited Theorem.

□

A 2-d figure of a random commuter is included below.

It is interesting to note that although a uniform noise is used to generate the above, we could also have used a smooth sequence of increasing or decreasing maps. This is due to the ergodicity property being present in the system given by (3.56). For details the reader is referred to Appendix A.

### 3.2.2 The case of normal noise

Consider the following dynamical system

$$
\left\{
\begin{array}{l}
g_1(x,\xi_x) = 2(1-2\epsilon)x + (\xi_x+1)\epsilon, 0 \leq x < 1/2, \\
\quad = 2(1-2\epsilon)(1-x) + (\xi_x+1)\epsilon, 1/2 \leq x \leq 1.
\end{array}
\right\} \tag{3.74}
$$

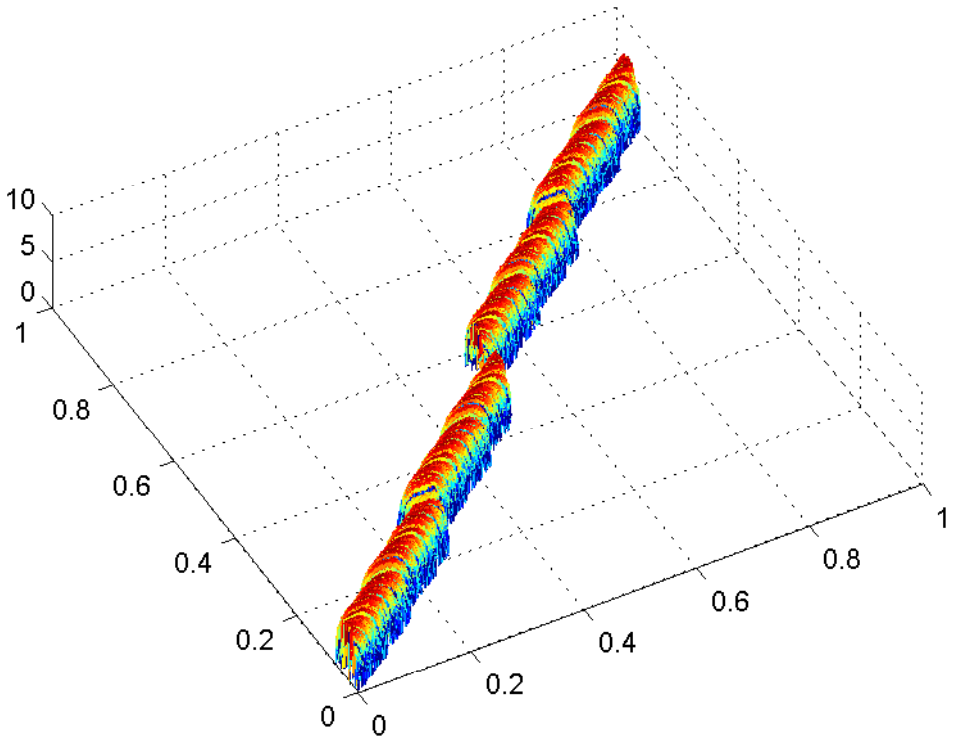



Figure 3.1: **Convergence Analysis for Skew Tent Maps**

We assume that $g_1$ is identically zero outside the unit interval. In this section we consider the case where $\xi$ has a normal distribution. Since $g_1$ is identically zero outside the unit interval we have circumvented the problem of orbits wandering of to infinity. The Perron-Frobenius operator $P : L^1(X) \to L^1(X)$ for the system is computed to be

$$Pf(x) = \int_X K(x, y) f(y) dy \tag{3.75}$$

Here $K(x, y)$ is the stochastic kernel defined by

$$K(x, y) = \frac{1}{\sqrt{2\pi}} e^{\frac{(x-S(y))^2}{2}} \tag{3.76}$$

We begin with the following definitions.

**Definition 3.2.5.** *the sequence of iterates of the Frobenius-Perron operator $\{P^n\}$ are said to be asymptotically stable if there $\exists$ a unique $f^* \in D$ such that*

$$Pf^* = f^* \tag{3.77}$$

*and*

$$\lim_{n\to\infty} ||P^n f - f|| = 0 \tag{3.78}$$

**Definition 3.2.6.** *A function $V(x)$ such that*

$$\lim_{x\to\infty} V(x) = \infty \tag{3.79}$$

*is called a Liapunov function.*

We will use the following Theorem to demonstrate our result

**Theorem 3.2.7.** *Let $K(x, y)$ be a stochastic Kernel. Consider the Frobenius-Perron operator defined by*

$$Pf(x) = \int_X f(y)K(x,y)dy \tag{3.80}$$

*If there exists a liapunov function $V(x)$ such that for an unbounded domain $G \subset X$,*

$$\int_G V(x)K(x,y)dx \leq \alpha V(y) + \beta,\ 0 \leq \alpha < 1,\ \beta > 0 \tag{3.81}$$

*then $\{P^n\}$ is asymptotically stable*

We now state our result in the following Lemma

**Lemma 3.2.8.** *Consider the dynamical system defined by*

$$\left\{\begin{array}{l} g_1(x,\xi_x) = 2(1-2\epsilon)x + (\xi_x+1)\epsilon, 0 \leq x < 1/2, \\ \quad = 2(1-2\epsilon)(1-x) + (\xi_x+1)\epsilon, 1/2 \leq x \leq 1. \end{array}\right\} \tag{3.82}$$

*where $\xi$ has a normal distribution. Also assume $g_1$ is identically zero outside the unit interval. The stochastic Frobenius-Perron operator for this dynamical system is asymptotically compact, hence converges to a stationary density.*

*Proof.* It follows that

$$\int_G V(x)K(x,y)dx = \int_G V(x)g(x-S(y))dx = \frac{1}{\sqrt{2\pi}}\int_G V(x)e^{-\frac{(x-S(y))^2}{2}}dx \tag{3.83}$$

Note we have an explicit form for the density $g$ now as normal noise is being considered. We next set $G = [0,\infty)$, and define the following functional

$$F(y) = \frac{1}{\sqrt{2\pi}}\int_0^\infty V(x)e^{-\frac{(x-S(y))^2}{2}}dx \tag{3.84}$$

We take the derivative of the above

$$
\begin{aligned}
& \frac{d}{dy}F(y) \\
= \; & \frac{d}{dy}\left(\frac{1}{\sqrt{2\pi}}\int_G V(x)e^{-\frac{(x-S(y))^2}{2}}dx\right) \\
= \; & \frac{1}{\sqrt{2\pi}}\int_0^\infty \frac{d}{dy}V(x)e^{-\frac{(x-S(y))^2}{2}}dx \\
= \; & \frac{1}{\sqrt{2\pi}}\int_0^\infty V(x)(x-S(y))(S'(y))e^{-\frac{(x-S(y))^2}{2}}dx \\
= \; & \frac{1}{\sqrt{2\pi}}\int_0^\infty V(x)xS'(y)e^{-\frac{(x-S(y))^2}{2}}dx - \frac{1}{\sqrt{2\pi}}\int_0^\infty V(x)S(y)S'(y)e^{-\frac{(x-S(y))^2}{2}}dx \\
= \; & S'(y)\frac{1}{\sqrt{2\pi}}\int_0^\infty V(x)xe^{-\frac{(x-S(y))^2}{2}}dx - S(y)S'(y)F(y)
\end{aligned} \tag{3.85}
$$

Since we are dealing with a phase space $Y = [0, 1]$, with $S(y)$ having different dynamics on $[0, \frac{1}{2}]$ than $[\frac{1}{2}, 1]$, we will break our analysis into 2 cases. We first consider $y \in [0, \frac{1}{2}]$. Recall the identity

$$
S(y)S'(y) = \frac{d}{dy}\left(\frac{S(y)^2}{2}\right) \tag{3.86}
$$

From the form of $S(y)$ in $[0, \frac{1}{2}]$, it is easily seen that $(S(y))^2 \nearrow$. Thus we have that

$$
S(y)S'(y) = \frac{d}{dy}\left(\frac{S(y)^2}{2}\right) > C_1 > 0 \tag{3.87}
$$

We have a strict inequality here as $S(y)$ is never constant. Furthermore we restrict ourselves to the class of Liapunov functions that grow polynomially, so are $O(x^n)$. Then it is seen that

$$
S'(y)\frac{1}{\sqrt{2\pi}}\int_0^\infty V(x)xe^{-\frac{(x-S(y))^2}{2}}dx \leq C\int_0^\infty x^{n+1}e^{-\frac{(x-1)^2}{2}}dx \leq C_2 \tag{3.88}
$$

This yields the following inequality

$$\frac{d}{dy}F(y) \leq -C_1F(y) + C_2 \tag{3.89}$$

It now follows via Gronwall's Lemma that

$$F(y) \leq e^{-C_1y}(F(0) + C_2y) \leq e^{-C_1}(F(0) + C_2) \leq \alpha V(y) + \beta \tag{3.90}$$

For say $\alpha = \frac{1}{2}$ and $\beta = e^{-C_1}(F(0) + C_2)$

For the case of $y \in [\frac{1}{2}, 1]$ the analysis is similar. Thus we obtain

$$\frac{d}{dy}F(y) \leq C_1F(y) + C_2 \tag{3.91}$$

Gronwall's Lemma now yields

$$F(y) \leq e^{C_1y}(F(0) + C_2y) \leq e^{C_1}(F(0) + C_2) \leq \alpha V(y) + \beta \tag{3.92}$$

For say $\alpha = \frac{1}{2}$ and $\beta = e^{C_1}(F(0) + C_2)$ Thus the lemma is proved by application of the earlier stated corollary. □

We perform some simulations to view the above object when the systems are a logistic map with $r = 3.75$ perturbed by a uniform noise and a tent map with $r = 0.9$. This yields the following

## 3.3 Random Fixed Point Theorem

We now change direction slightly and give yet another interpretation of the commuter function between randomly forced dynamical systems. It is possible to view the action of the commuter on densities as the action of a random operator. There is a vast literature on probabilistic functional analysis, [9], [10], and the use of random operator theory. We begin via the following definitions

**Definition 3.3.1.** *Let $(X, A, \mu)$ be a measure space. A mapping $T : X \times \omega \to X$ is called a random operator if for each fixed $x \in X$ the mapping the map $T(., x) : \omega \to X$ is measurable.*

**Definition 3.3.2.** *A measurable map $\xi : \Omega \to X$ is a random fixed point of a random operator $T : X \times \omega \to X$ if*

$$T(\omega, \xi(\omega)) = \xi(\omega), \text{ for each } \omega \in \Omega \tag{3.93}$$

We recall the following theorem, as best suited for our application.

**Theorem 3.3.3** (Random Fixed Point Theorem)**.** *Let $T$ be a continuous random operator on $L^1([0,1] \times \Omega)$ to $L^1[0,1]$. Let $\lambda_\xi$ be a real-valued random variable such that $\lambda_\xi < 1$ almost surely and*

$$\|Tf_1 - \mathcal{T}f_2\|_1 \leq \lambda_\xi \|f_1 - f_2\|_1 \tag{3.94}$$

*for every two functions $f_1,\ f_2 \in L^1[0,1]$. Then there exists an $L^1[0,1]$-value random function $f_\xi$, which is the unique fixed point of $T$, i.e.*

$$Tf_\xi = f_\xi \tag{3.95}$$

Our goal is to use the above to prove that the commuter function that we have considerd earlier, when viewed as a random operator posesses a fixed point, thus converges. This reinforces our result about the convergence of this object for various classes of noise, via stochastic Frobenius-Perron operator methods, that we considered in the previous section. Recall again the commutative diagram in our random setting

$$\begin{array}{ccc} L^1(X \times \Omega) & \overset{P_{g_1}}{\to} & L^1(X) \\ \downarrow f & & \downarrow\downarrow f \\ L^1(Y) & \overset{P_{g_2}}{\to} & L^1(Y) \end{array}$$

The above implies that the following holds

$$f(P_{g_1}(\rho(x))) = P_{g_2}(f(\rho(x))) \tag{3.96}$$

Clearly the commuter acts on densities in both legs of the diagram. Thus it is natural to view it as a random map, acting on densities, so as to bring in the afforementioned random fixed point theory. We first carefully define the right random operator in our setting. We take our cue from earlier works, [**?**]. Recall in the deterministic setting $f$ satisfies

$$f(g_1(x)) = g_2(f(x)) \tag{3.97}$$

Since we know what $g_1$ and $g_2$ are, a functional equation for f can be set up explicitly. We adopt the same methodology to the random setting. Consider

$$\left\{\begin{array}{l} g_1(x,\xi_x) = 2(1-2\epsilon)x + (\xi_x+1)\epsilon, 0 \le x < 1/2, \\ \quad = 2(1-2\epsilon)(1-x) + (\xi_x+1)\epsilon, 1/2 \le x \le 1. \end{array}\right\} \tag{3.98}$$

$$\left\{\begin{array}{l} g_2(y) = 2y, 0 \le y < 1/2, \\ \quad 2(1-y), 1/2 \le y \le 1. \end{array}\right\} \tag{3.99}$$

For any fixed density $\xi$, the following equation holds

$$f_\xi(g_1(x,\xi_x)) = g_2(f_\xi(x)) \tag{3.100}$$

where $f_\xi : L^1(X \times \Omega) \to L^1(Y)$

This yields

$$\left\{\begin{array}{c} 2f_\xi(x) = f_\xi(2(1-2\epsilon)x + (\xi_x+1)\epsilon,\ 0 \le x < 1/2 \\ 2(1-f_\xi(x)) = f_\xi(2(1-2\epsilon)(1-x) + (\xi_x+1)\epsilon,\ 1/2 < x \le 1 \end{array}\right. \tag{3.101}$$

Thus we have a functional equation for $f_\xi(x)$,

$$\begin{cases} f_{\xi}(x) = \dfrac{1}{2} f_{\xi}(2(1-2\epsilon)x + (\xi_x + 1)\epsilon,\ 0 \leq x < 1/2 \\ f_{\xi}(x) = 1 - \dfrac{1}{2} f_{\xi}(2(1-2\epsilon)(1-x) + (\xi_x + 1)\epsilon,\ 1/2 < x \leq 1 \end{cases} \tag{3.102}$$

We use the above to define a random operator as follows

$$\begin{cases} Tf_{\xi}(x) = \dfrac{1}{2} f_{\xi}(2(1-2\epsilon)x + (\xi_x + 1)\epsilon,\ 0 \leq x < 1/2, \\ \qquad\quad = 1 - \dfrac{1}{2} f_{\xi}(2(1-2\epsilon)(1-x) + (\xi_x + 1)\epsilon,\ 1/2 < x \leq 1. \end{cases} \tag{3.103}$$

This follows $\forall f_1, f_2 \in L^1[0,1], 1 \leq p < \infty$. For the above defined random operator we can state the following result

**Lemma 3.3.4.** *Consider the following dynamical system*

$$\left\{ \begin{array}{l} g_1(x, \xi_x) = 2(1-2\epsilon)x + (\xi_x + 1)\epsilon, 0 \leq x < 1/2, \\ \quad = 2(1-2\epsilon)(1-x) + (\xi_x + 1)\epsilon, 1/2 \leq x \leq 1. \end{array} \right\} \tag{3.104}$$

$$\left\{ \begin{array}{l} g_2(y) = 2y, 0 \leq y < 1/2, \\ \quad 2(1-y), 1/2 \leq y \leq 1. \end{array} \right\} \tag{3.105}$$

*The commuter function $f : L^1(X \times \Omega) \to L^1(Y)$ between the systems above, when viewed as an appropriate random operator, such as defined in* (3.103)*, posesses a random fixed point.*

*Proof.* We consider the difference of $Tf_1$ and $Tf_2$ piecewise on $[0, 1/2]$ and $[1/2, 1]$ as

defined.

$$\begin{aligned}
\|Tf_1 - Tf_2\|_{[0,1/2]} &= (\int_{[0,1/2]} |Tf_1(x) - Tf_2(x)|dx) \\
&= (\int_{[0,1/2]} |\frac{1}{2}f_1(2(1-2\epsilon)x + (\xi_x+1)\epsilon) - \frac{1}{2}f_2(2(1-2\epsilon)x + (\xi_x+1)\epsilon)|dx) \\
&= (\int_{(\xi_x+1)\epsilon}^{(1-2\epsilon)+(\xi_x+1)\epsilon} |\frac{1}{2}f_1(y) - \frac{1}{2}f_2(y)|\frac{1}{2(1-2\epsilon)}dy) \\
&= \frac{1}{2}(\frac{1}{2(1-2\epsilon)})\|f_1 - f_2\|_{[(\xi_x+1)\epsilon,(1-2\epsilon)+(\xi_x+1)\epsilon]}.
\end{aligned}$$

or

$$\|Tf_1 - Tf_2\|_{[0,1/2]} = \frac{1}{2}(\frac{1}{2(1-2\epsilon)})\|f_1 - f_2\|_{[(\xi_x+1)\epsilon,(1-2\epsilon)+(\xi_x+1)\epsilon]} \tag{3.106}$$

We now set

$$y = 2(1-2\epsilon)x + (\xi_x+1)\epsilon, \tag{3.107}$$

in the above.

On the other hand, when $1/2 < x \leq 1$, we have

$$\begin{aligned}
\|Tf_1 - Tf_2\|_{[1/2,1]} &= \int_{[1/2,1]} |Tf_1(x) - Tf_2(x)|dx \\
&= \frac{1}{2}(\int_{(1-2\epsilon)+(\xi_x+1)\epsilon}^{(\xi_x+1)\epsilon} |\frac{1}{2}f_1(y) - \frac{1}{2}f_2(y)|\frac{1}{-2(1-2\epsilon)}dy) \\
&= \frac{1}{2}(\int_{(\xi_x+1)\epsilon}^{(1-2\epsilon)+(\xi_x+1)\epsilon} |\frac{1}{2}f_1(y) - \frac{1}{2}f_2(y)|\frac{1}{2(1-2\epsilon)}dy) \\
&= \frac{1}{2}(\frac{1}{2(1-2\epsilon)})\|f_1 - f_2\|_{[(\xi_x+1)\epsilon,(1-2\epsilon)+(\xi_x+1)\epsilon]}.
\end{aligned}$$

or

$$\|Tf_1 - Tf_2\|_{[1/2,1]} = \frac{1}{2}(\frac{1}{2(1-2\epsilon)})\|f_1 - f_2\|_{[(\xi_x+1)\epsilon,(1-2\epsilon)+(\xi_x+1)\epsilon]} \tag{3.108}$$

Again we set

$$y = 2(1-2\epsilon)(1-x) + (\xi_x + 1)\epsilon \tag{3.109}$$

in the above equation.

We now add (3.106) to (3.108) to obtain

$$\begin{aligned} \|Tf_1 - Tf_2\|_{[0,1]} &= \frac{1}{2}(\frac{1}{1-2\epsilon})\|f_1 - f_2\|_{[(\xi_x+1)\epsilon,(1-2\epsilon)+(\xi_x+1)\epsilon]} \\ &\leq \frac{1}{2}(\frac{1}{1-2\epsilon})\|f_1 - f_2\|_{[0,1]}. \end{aligned}$$

Thus we have

$$\|Tf_1 - Tf_2\|_{[0,1]} \leq \lambda\|f_1 - f_2\|_{[0,1]}, \tag{3.110}$$

where

$$\lambda = \frac{1}{2}(\frac{1}{1-2\epsilon}). \tag{3.111}$$

We can choose $\epsilon < 1/4$, so $(\frac{1}{1-2\epsilon}) < 2$. Thus we obtain

$$\lambda = \frac{1}{2}(\frac{1}{1-2\epsilon}) < 1. \tag{3.112}$$

This completes the proof of the Lemma. □

## 3.4 conclusion

In conclusion we have laid down a rigorous definition of a random commuter. We have established its convergence via use of the Frobenius Perron theory, under pertubation of uniformly distributed noise and normally distributed noise. This result is seen via our numerical simulations also. Our results concerning the commuter when viewed as a random operator reinforce these earlier derived convergence results. However various questions

remain open at this point. Most of our analysis is confined to the case of one dimensional dynamical systems represented as maps. It would be interesting to consider how our methodologies apply when the underlying dynamical systems are flows. White noise perturbation in such a setting cam also be considered. This would lead us to investigate the SDE case. The continuous version of the Frobenius-Perron operator might provide a valuable tool in this case. It is also interesting to try some of the above mentioned techniques when the commuters are not quite homeomorphic, which is getting at the true spirit of "almost conjugacy". To this end we provide some details in Appendix B, where some of the associated difficulties can be circumvented via bringing in appropriately defined delta functions. These and related questions are the subject of current investigation.

# Chapter 4

# Comparing Dynamical Systems by Graph Matching Method: A Graph Matching Interpretation of Mostly Conjugacy

In this chapter, we consider comparing dynamical systems by using graph matching method either between the graphs representing the underlying symbolic dynamics, or graphs approximating the action of the systems on a fine but otherwise non generating partition. For conjugate systems, the graphs are isomorphic and we show that the permutation matrices that relate the adjacency matrices coincides with the solution of Monge's mass transport problem. We use the underlying Earth Mover's Distance (EMD) to generate the "approximate" matching matrix to illustrate the association of graphs which are derived from equal-distance partitioning of the phase spaces of systems. In addition, for one system which embeds into the other, we show that the comparison of these two systems by our method is an issue of subgraph matching.

## 4.1 Introduction

A basic question in science and in dynamical systems models of scientific problems is how to compare two systems. For an equivalence relation on dynamical systems, standard practice is to use the notion of *conjugacy*. The problem in practice is that it can be difficult in the general scenario of more than one dimension to find such a homeomorphic change of variables that relates the two systems, even if one exists. Furthermore if no such exact equivalence exists, the question becomes "is one system a good approximation of another?" In [2, 11], we explore this isuue by generating a change of variables between the systems, and then we measured the quality of the comparison by discussing the deviation of the resulting change of variables from a homeomorphism, which we called homeomorphic defect. The problem is the contraction mapping methods we used to identify the change of variables do not easily extend to the general scenario of multi-variate dynamical systems. Thus the need to develop alternative methods in the spirit of our goal [2, 11] of relaxing conjugacy to compare somewhat related dynamical systems. Here we will resort to a variant of the classic Monge-Kantorovich optimization problem [12, 13] to both built a useful change of variables and measure quality of the comparison through the underlying cost called the Wasserstein distance, in a spirit such as the related work for time series in [14].

In symbolic dynamics [15, 16, 18], a shift space is a set of infinite symbol streams representing the trajectory of a dynamical system. Shifts spaces with a finite set of forbidden blocks are called shifts of finite type. Such shifts spaces have a simple representation as a special case of so-called sophic shifts, that is a directed graph, where nodes are symbols. Furthermore we may define edges weights as the probability of a transition from one symbolic word to another. As such, the directed graph becomes a discrete representation of the Frobenius-Perron operator, [19] through an approximation method related to the Ulam's method, [20].

Within the usual dynamical systems framework, the determination of whether two systems are dynamically equivalent is based upon whether or not there is conjugacy between

them, [16]. In terms of a symbolic dynamics representation [18, 15] of the underlying systems, the conjugacy matches symbol sequence from each symbolic space. System comparison within the concept of "conjugacy" would be straightforward, since all we need is to check is the existence of a similarity transformation that relates the adjacency matrices (one for each system) which generates the grammars of the underlying subshifts. Here we will demonstrate that such a matrix is equivalent to finding the Monge solution [13] to a specific mass transport problem. Because of well understood difficulties in finding Monge solutions, we instead choose to solve for the relaxed "matching matrix", which is actually a Kantorovich solution to relating the adjacency matrices of the dynamical systems. We use the earth mover's distance (EMD) [21] to generate such matching matrices which we show serves as a suitable commuter for the dynamical system. In addition, we prove that if one system embeds into the other, the comparison becomes a subgraph matching problem.

The outline of this chapter is as follows: in section 2, we introduce the idea of our graph matching method to compare dynamical systems; In section 3, we give some theoretical results regarding the comparison between conjugate systems, with a particular example; In section 4, we consider the error analysis, with a numerical result to study the regularity property of our algorithm; in section 5, we extend our methods to compare systems where one embeds into the other, with both theoretical and numerical results given; In section 6, we show that our method can be easily extended to compare higher dimensional dynamical systems. We summarize our results in section 7.

## 4.2 Graph Representations of Dynamical systems

We adopt our notations from Lind and Marcus [18]: A graph $G$ consists of a finite set $v = v(G)$ of vertices, and with a finite set $\varepsilon = \varepsilon(G)$ of edges. Each edge $e \in \varepsilon(G)$ starts at a vertex denoted by $i(e) \in \varepsilon(G)$ and ends at a vertex $t(e) \in \varepsilon(G)$. The weighted adjacency/stochastic matrix of $G$, denoted as $A_G$, is a conditional probability matrix where the $(i, j)$ entry assigns

the probability $P(j|i)$ (or $P(i \to j|i)$), the conditional probability from node $i$ to $j$ given the current state is in node $i$. The stochastic matrix is essentially a matrix description of a Markov chain on the graph $G = (\nu, \varepsilon)$ [18]. We note that some references may distinguish a stochastic matrix and the graph adjacency matrix, defining the latter as a binary matrix which assigns 1 to connected nodes, otherwise 0. In this paper, the transition matrix and stochastic matrix are used interchangeably, and (unless specifically stated otherwise) when we refer to "the adjacency matrix," we mean a *weighted* adjancency matrix with weights assigned to describe the appropriate transition probabilities.

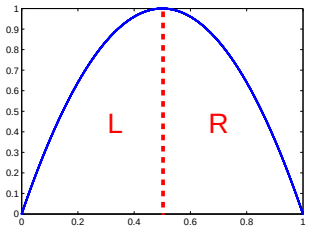
In practice, given a test orbit $\{x_k\}_{k=1}^{N}$ s.t. $x_{k+1} = f(x_k)$ from dynamical system $f$, then the transition probabilities can be estimated as follows,

$$P(j|i) \approx \frac{\#(\{x_k | x_k \in B_i \; and \; f(x_k) \in B_j\})}{\#(\{x_k \in B_i\})} \tag{4.1}$$

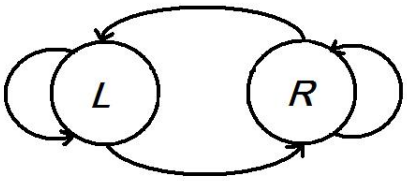
as Ulam-Galerkin approximation of the Frobenius-Perron operator with a partition $\Gamma = \{B_i\}$, which is a finite family of connected sets with nonempty and disjoint interiors that covers the phase space [20, 52].

As example, Figure 4.1 shows a partition on the logistic map. We label the interval $[0, 1/2]$ with $L$, and $(1/2, 1]$ with $R$. Start from any initial point that yields a chaotic trajectory, record $L$ or $R$ in each iteration based on the evolution of dynamics. Then a weighted directed graph is built based on the recorded link list. In this example, the transition matrix would be a 2 by 2 matrix where entries are all 1/2.

The weighted directed graph is a simple representation of the dynamical system in that the adjacency matrix of the graph records all the allowable bi-infinite walks. Consequently, we assert that one way to relate the dynamics of two different systems is to perform a graph matching operation on the associated graph representations of the symbol dynamics. A key benefit of using a graph matching approach is that the study of graph matching techniques




(a) Logistic Map




(b) Graph for Logistic Map

Figure 4.1: Graph

and algorithms is well developed in the field. Among many of the efficient graph matching methods, we use the Earth Mover's Distance, which essentially solves the discrete case of the Kantorovich's problem. What we hope to answer in this paper is,"what does the matching of the graphs using EMD tell us about the similarity of the dynamical systems?" As preliminary explanation, we briefly recall the Monge-Kantorovich problem.

### 4.2.1 Monge-Kantorovich Problem

The Monge-Kantorovich problem is an optimal transportation problem [13]: Specially, given two distributions $\mu$ for space $X$ and $\nu$ for $Y$, one is required to find the optimal transportation plan within a given some metric describing the transportation cost. The Monge's problem is formalized as:

$$Minimize \;\; I[T] := \int_X c(x,\ T(x)))d\mu(x)$$

for all measurable map $T : X \to Y$ such that $\nu$ is a push-forward [13] of $\mu$ by $T$. The term $c(x,\ T(x))$ defines the cost to transport a unit from $x$ to $T(x)$.

On the other hand, the Kontorovich's problem is a relaxed version of the Monge's one. It allows each point in $X$ to be associated to different points in $Y$. We can no longer use a map $T$ to describe this association. We shall use a probability measures $\pi$, the transference plan, on the product space $X \times Y$, where informally $d\pi(x, y)$ measures the amount of mass transferred from location $x$ to $y$. We denote the set of all such probability measures by

$$\Pi(\mu, \nu) = \{\pi \in P(X \times Y) : \pi[A \times Y] = \mu[A],\ \pi[X \times B] = \nu[B]\ for\ all\ measurable\ A, B\}$$

And the Kontorovich's optimal transportation problem can be formalized as follow:

$$Minimize\ \ I[\pi] := \int_{X \times Y} c(x, y) d\pi(x, y)$$

for $\pi \in \Pi(\mu, \nu)$, where $\Pi(\mu, \nu)$ (defined above) is called the space of transference plans from $X$ to $Y$.

In the later discussion of this paper, we focus on the discrete version of the Monge-Kantorovich problem, with solution given as a matrix $F$ that expresses the optimal transportation in that matrix entry $f_{ij}$ gives the amount of mass to be moved from discrete location $x_i$ to location $y_j$. In that setting, the special case of a Monge solution would then be a permutation matrix, because the Monge problem requires that the full mass at location $x_i$ be moved to a single site $y_j$.

We will compare two systems by studying the Monge-Kantorovich solution between the first eigenvectors of their associated adjacency matrices relying on Ulam's conjecture. Ulam's conjecture was proposed by S. Ulam [20] and referred that Ulam-Gelerkin matrix determined by Eq. 4.1 has dominant state/eigenvector that converges weakly to invariant measure/distribution of the Frobenius-Perron operator, which has been proven for certain special dynamical systems.

## 4.3 Analysis

In this section, we relate graph matching and conjugacy. Particularly, we prove an interesting relationship between the matching matrix $F$ that describes the Monge solution to a particular problem, and the adjacency matrices, $A_{G_1}$ and $A_{G_2}$, that describe the dynamics of the two systems:

$$F \cdot A_{G_1} = A_{G_2} \cdot F.$$

We first introduce the notation: A partition $\Gamma = \{x_0,\ x_1, \cdots,\ x_m\}$ on interval $[a, b]$ is a set of points where $a = x_1 < x_1 < \cdots < x_m = b$. We denote $|\Gamma|$ as the number of subintervals after partitioning, which we call the cardinality of partition. Let $G$ be a graph with edge set $\varepsilon$ and adjacency matrix $A_G$. The edge shift $X_G$ or $X_A$ is the shift space over the alphabet $\mathscr{A} = \varepsilon$ specified by $X_G = X_A = \{\xi = (\xi_i)_{i\in\mathbb{Z}} \in \varepsilon^{\mathbb{Z}} : t(\xi_i) = i(\xi_{i+1}) \text{ for all } i \in \mathbb{Z}\}$. The shift map on $X_G$ or $X_{A_G}$ is called the edge shift map and is denoted by $\sigma_G$ or $\sigma_{A_G}$.

### 4.3.1 Conjugate Systems

Two dynamical systems are said to be *topologically conjugate* if there exists a homeomorphism which describes the change of coordinates between systems. Consequently, when we associate these systems to dynamics on graphs, we may use that homeomorphism to relate the partition in one system to the partion in the other. Because trajectories in each system are matched (via the homeomorphism), allowable paths through the graphs must also be matched, and the graphs of the two systems would be. So in this case, the associated graph adjacency matrices, say $A$ (for one system) and $B$ (for the other), are *similar,* and related by a permutation matrix $P$. Essentially, $P$ tells how the nodes of the compared graphs are associated, in a 1-to-1 way, so that $A = P^{-1}BP$. In addition, $P$ relates the eigenvectors of $A$ and $B$ as a linear isomorphism [18]. And, we can identify conjugate edge shifts from these conjugate systems (Theorem 4.7.1 in Appendix), which also determines a so-called "strong shift equivalent" relationship. More precisely, we have the following definition:

**Definition 4.3.1** ([18])**.** *Let $A$ and $B$ be nonnegative integral matrices. An elementary equivalence from $A$ to $B$ is a pair $(R, S)$ of rectangular nonnegative integral matrices satisfying*

$$A = RS \text{ and } B = SR$$

*In this case we write $(R, S) : A \cong B$. A strong shift equivalence of lag $l$ from $A$ to $B$ is a sequence of $l$ elementary equivalences*

$$(R_1, S_1) : A = A_0 \cong A_1,\ (R_2, S_2) : A_1 \cong A_2 \cdots (R_l, S_l) : A_{l-1} \cong A_l = B$$

*In this case we write $A \approx B$ (lag $l$). We say that $A$ is strong shift equivalent to $B$ (and write $A \approx B$) if there is a strong shift equivalence of some lag from $A$ to $B$.*

In the following we show that, for isomorphic graphs, which we obtain from conjugate systems by an appropriate partitioning, the permutation matrix $P$ which provides the similarity transform between the associated adjacency matrices is the same matrix identified by the Monge solution to an appropriately defined transport problem.

**Theorem 4.3.2.** *Suppose we are given two conjugate systems $SYS_1$ and $SYS_2$, and partitions $\Gamma_i$, $i = 1, 2$, with $|\Gamma_i| = M$, such that the resulting edge shifts $X_{G_1}$ is conjugate to $SYS_1$, $X_{G_2}$ is conjugate to $SYS_2$. Let $A_{G_1}$ be the adjacency matrix of the graph generated from $SYS_1$ with this partition, and similarly $A_{G_2}$ for $SYS_2$. Suppose $A_{G_1}$ is isomorphic to $A_{G_2}$. Let $F$ be the Monge solution between dominant eigenvectors of adjacency matrices* [1] *of graphs from $SYS_1$ and $SYS_2$, then $F$ satisfies*

$$F \cdot A_{G_1} = A_{G_2} \cdot F.$$

*Proof.* Since $SYS_1$ is conjugate to $SYS_2$, $X_{G_1} \cong X_{G_2}$. By Theorem 4.7.1, $A_{G_1} \approx A_{G_2}$ (strong shift equivalent). By Definition 4.7.2 and Theorem 4.7.3, $A_{G_1} \sim A_{G_2}$ (shift equivalent). By Definition 4.7.4 and Theorem 4.7.5, $A_{G_1} \sim_{\mathbb{Z}} A_{G_2}$ (shift equivalent over $\mathbb{Z}$). On the other hand, since $A_{G_1}$ is isomorphic to $A_{G_2}$, there exists a permutation matrix $P$ such that

[1]

**Remark 4.3.3.** *Note that in the proof, we apply theorems and definitions from the appendix. Some of the fundamental theorems and definitions that we are using require matrices to be over $\mathbb{Z}$ or $\mathbb{Z}^+$. But as we define before in section 4.2, we generally consider adjacency matrices to be stochastic matrices. The stochastic matrix setting makes our method easy to be implemented, as will be illustrated in section* **??**. *Because these stochastic weights are determined by counting transitions in a long trajectory, and then normalizing each row (to row sum one), each of these stochastic matrices has rational entries, and appropriate scaling yields an integer matrix. This renormalized matrix is similarly interpretable as a weighted adjacency matrix. Consequently, the proofs that rely on integer matrix theorems remain valid under such a renormalization. The Monge's solution would be interpreted as showing how to move "counts" instead of "densities." Even though some references [18, 23] discuss shift equivalence issues for irreducible shift of finite type with a Markov measure, we prefer not to introduce additional theorems and definitions to address this technical detail, relying simply on the frequency counting to justify our approach.*

$A_{G_1} = P^{-1}A_{G_2}P$. Now if we let $S = P$, $R = P^{-1}A_{G_2}$, then $A_{G_1} \sim_{\mathbb{Z}} A_{G_2}(lag\ 1)$ according to Definition 4.3.1. By Theorem 4.7.6 and Theorem 4.7.7, $w = Sv \neq 0$ where $v$ and $w$ are Perron eigenvectors of $A_{G_1}$ and $A_{G_2}$, respectively, and $S$ satisfies $S \cdot A_{G_1} = A_{G_2} \cdot S$. In addition, $S$ is a permutation matrix, hence $S$ is invertible.

On the other hand, let $F$ be the Monge solution between $v$ and $w$, where $v$ and $w$ are dominating eigenvectors of $A_{G_1}$ and $A_{G_2}$, then $F$ satisfies $w = Fv \neq 0$. We claim that $F = S$.

By *Choquet's theorem* and *Birkhoff's theorem* [13], $F$ is a permutation matrix, which implies $F$ is invertible. Substitute $v = F^{-1}w$ into $w = Sv$, we have

$$w = S \cdot F^{-1}w, \tag{4.2}$$

where $S \cdot F^{-1}$ is also always a permutation matrix, regardless of the cardinality of partition. We note that (4.2) is not sufficient in that $S \cdot F^{-1}$ could be a rotation matrix other than identity matrix that satisfies $w = S \cdot F^{-1}w$.

We prove by induction that, $S \cdot F^{-1}$ is an identity matrix. In the following, we let $E_N = S \cdot F^{-1}$, where the cardinality of partition $M = N$, where we intent to use induction on $N$.

When $N = M = 2$, the only permutation matrices that $S \cdot F^{-1}$ can take are

$$E_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

or

$$\hat{E_2} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

but $|\hat{E_2}| = -1$, so $\hat{E_2}$ can not be a rotation matrix. Thus $S \cdot F^{-1} = E_2$.

Now we suppose when $M = N$, $E_N$ is an $N$ by $N$ identity matrix, so that $w = S \cdot F^{-1}w$.

When $M = N + 1$, in order that the resulting edge shifts with cardinality of partition $N+1$ is conjugate to the one with cardinality of partition $N$, the graph with $N+1$ nodes has

to be a splitting of the one with $N$ nodes [18]. Without loss of generality, we can assume the first node of the graph splits into two nodes, from cardinality of partition $N$ to $N+1$. Then the first bin of dominate eigenvector $w$ of case $N$ would break into two for case $N+1$, with the rest of bins unchanged. For example, from Figure 4.2(a) to Figure 4.2(b), *bin* 1 breaks into *bin* $1a$ and *bin* $1b$.

Regardless of how *bin* $1a$ and *bin* $1b$ match each other, we claim that the rest bins of $w$, like bins from 2 to 5 in Figure 4.2(b), match themselves identically as before. Because otherwise, if we amalgamate *bin* $1a$ and *bin* $1b$ back into *bin* 1, or say from case $N+1$ back to case $N$, the bins of $w$ would not match themself identically. This contradicts to our assumption of the matching of bins of $w$ when $M = N$.

Also, since $S \cdot F^{-1}$ is a permutation matrix, $E_{N+1}$ would has the forms of

$$E_{N+1} = \begin{bmatrix} E_2 & 0 \\ 0 & E_{N-1} \end{bmatrix}$$

or

$$\hat{E_{N+1}} = \begin{bmatrix} \hat{E_2} & 0 \\ 0 & E_{N-1} \end{bmatrix}$$

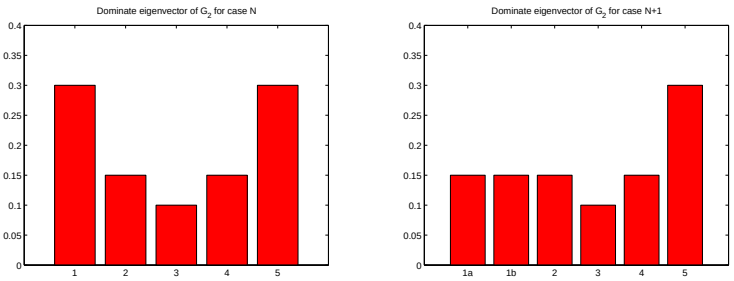
where $E_{N-1}$ is a $N-1$ by $N-1$ identity matrix. Thus, by the calculation of determinant of block matrix, we know that $|\hat{E_{N+1}}| = -1$, which can not be a rotation matrix. Thus, $S \cdot F^{-1} = E_{N+1}$, which is a $N+1$ by $N+1$ identity matrix.

Similar for the case $v = F \cdot S^{-1}v$. Thus, we have $S = F$.

□

So we can consider finding the similarity matrix $P$ as a process of choosing the association matrices that minimize the transportation cost.

However, there is not a general algorithm to find the similarity matrix $P$. Although we prove the equivalence between finding the similarity matrix and finding Monge's solution, that equivalence does not mean that either problem is easily solved. In fact, Monge's prob-




(a) Dominate eigenvector $w$ for case $N$

(b) Dominate eigenvector $w$ for case $N+1$

Figure 4.2: Dominate eigenvectors of graph for case $N$ and $N+1$

lem is difficult because the metric that defines the topology of the transport problem could be very degenerate from the point of view of convexity properties [13]. For non-conjugate systems (and non-isomorphic graphs), there still exists a Monge solution which associates nodes of the two graphs in a 1-to-1 fashion in a way that minimizes the "cost of association" with respect to a particular choice of norm to measure the transportation cost. We observe that the Monge solution yields a linear transformation from the dominate eigenvector of one graph to the other. On the other hand, the Monge solution, does *not* provide a similar linear transformation between the adjacency matrices when the underlying graphs are not isomorphic. Consequently, the relaxed version, i.e. Kantorovich's problem, provides a more natural and implementable method to find an "optimal matching", where optimality is with respect to the relaxed transport problem.

### 4.3.2 Partitioning Issues

As additional reasons for framing this "system matching" as a Kantorovich problem is that framing as a Monge problem requires that we our partition is such that the edge shift is conjugate to the original dynamics. In most practical cases, it is hard or costly to find an appropriate partition (i.e. Markov partition) so that the resulting graph can fully describe the original dynamical system. Rather, we seek a method that can be applied to comparing general dynamical systems *without* solving the "hard" problem of finding a Markov parti-

tion. As alternative, we will use an equal-distance partition. We have to be very careful here, because in general the equal-distance partition won't give a conjugate shift space. So, in principle, if we proceed our method by comparing dynamical systems via equal-distance partition, we are actually comparing different systems than what we really would like to compare. However, this issue can be addressed as an "approximation problem." If we refine our equal-distance partition to sufficiently small cells, then the resulting shift space can approximate the one that would be observed from a Markov partition.

As example of this approximation via grid refinement, L. Billing and E. Bollt proved that the family of skew tent map is Markov for a dense set of parameters in the chaotic region [24]. In particular, in any given neighborhood of a non-Markov map in this chaotic region, there exists a Markov map that uniformly approximates the non-Markov one. J. Zheng proved that this approximation implies approximation in the sense of conjugacy [11]. In addition, H. Teramoto and T. Komatsuzaki showed that if a Markov partition has a certain relationship called "map-refinement of the other Markov partition", the shift spaces corresponding to these two Markov partitions are topologically the same [25]. Thus, an adequately refined equidistant partition can approximate a Markov partition in the sense of conjugacy. In section 4.4.1, we examine this numerical error estimate for a tent map example.

There are several cases where the Kantorovich's solution coincides with the Monge's solution. When we consider the structure of the space with metric, or say the cost function, defined with a regular $p > 1$ norm, the strict convexity of the metric guarantees there is a unique solution to the Kantorovich problem, which turns out to be also the solution to the Monge problem [13].

However, because of the partitioning issues that we have discussed before, the graphs representations of the systems are somehow an "approximation". So the adjacency matrices from conjugate systems are no longer similarly related due to inexact partition of phase space. However, we seek to relax the problem in the way of allowing partial matching,

which is to find the Kantororich's solution, to describe the association.

In probability theory, the Earth Mover's Distance (EMD) is a measure of distance between two probability distributions. EMD can be viewed as the solution of the discrete case of Kantorovich problem. It is a bipartite network flow problem which can be formalized as a linear programming problem [21]: Let $I$ be the set of supplies, $J$ be the set of consumers, and $c_{ij}$ be the cost to ship a unit from $i \in I$ to $j \in J$. We want to find a set of flow (matching matrix) $f_{ij}$ to minimize the overall cost:

$$\sum_{i \in I} \sum_{j \in J} c_{ij} f_{ij},$$

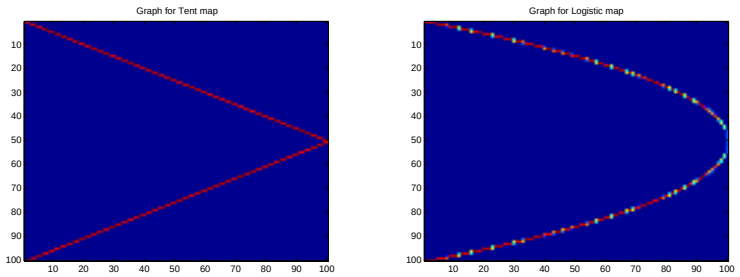
subject to the constrains:

$$f_{ij} \geq 0,\ i \in I,\ j \in J$$
$$\sum_{i \in I} f_{ij} = y_j,\ j \in J$$
$$\sum_{j \in J} f_{ij} \leq x_i,\ i \in I$$

where $x_i$ is the total supply of supplier $i$, and $y_j$ is the total capacity of consumer $j$.

Here we consider the cost function $c_{ij} = |x - y|^2$, with the metric to be the regular 2 norm. In this case, the optimal matching/flow is the gradients of the convex functions, which are monotone and orientation preserving. On the other hand, we consider the matrix norm to be the induced 2 norm, since with induced norm, the matrices can be viewed as operators. And if two systems are conjugate, they are isomorphic as linear transformations if we restrict the operators/matrices to their eventual range [18].

We give the example about how to create the matching matrix between systems: suppose we are given two dynamical systems, i.e. the logistic map and tent map, we make a equidistant partition on the phase space with partitionning numbers $M$ and $N$, respectively. We randomly choose an initial starting point, and then iterate it on each system, record the interval/node that the trajectory passes, and built the graphs based on the recording link list. Figure 4.3 presents the graphs of logistic map and tent map.




(a) Adjacency Matrix of Tent Map (b) Adjacency Matrix of Logistic Map

Figure 4.3: Graph

The matching matrix (Figure 4.4) generated by Earth Mover's distance associates the nodes between tent graph and logistic graph in the way of minimizing the overall cost determined by predefined topology. Notice the matching itself is a matrix. In Figure 4.4, the x-axis are nodes for tent graph, y-axis are for logistic graph. Color intensity implies similarity of the matching. We also plot the homeomorphsim function of tent and logistic in blue curve. We can see that both the matching matrix and the conjugacy describe the dirt of the tent distribution is moved in the direction so that it can match the logistic map distribution.

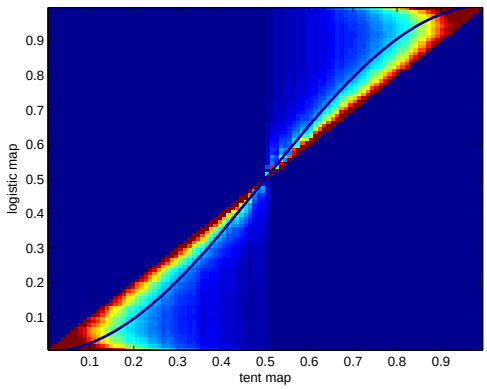



Figure 4.4: Graph matching between Logistic map and Tent map.

### 4.3.3 Further Discussion about the Measure on Adjacency Matrices

In Figure 4.4, we present the matching matrix between tent map and logistic map, which are conjugate systems. Note we apply the Earth Mover's Distance on their adjacency matrices whose entries are assigned the conditional probability measure. We can think of the matching matrix as a different representation of conjugacy, since the EMD is actually giving the transformation between the first eigenvectors of the tent map and logistic map.

This matching method may be confusing sometimes, since there are systems that are not conjugate but generating a "conjugate like" matching matrix. For example, if we are comparing a two hump tent map, instead of the one hump tent map, to the logistic map, our method would still give a matching matrix as shown in Figure 4.4. This comes from the fact that, different systems may have the same invariant measures.

Thus we also present an alternative but similar approach for our matching method. Instead of assigning the conditional probability measure to the entries of the adjacency matrices of underlying systems, we equalize these non-zero weights of each row of the adjacency matrices.

Figure 4.5 shows the matching matrix between the tent map and the logistic map with equalized weights on their adjacency matrices. This focuses more on the topological relation between systems. In the following section, we will discuss regularity properties when using the adjacency matrices with probability measure.

## 4.4 Error Analysis and Regularity Property Analysis

The graph matching method to compare dynamical systems that we discussed in the previous section suffers from both numerical and theoretical errors: the partition errors that we mentioned in section 4.3.2; the statistical errors from comparing the graphs which are constructed by a particular finite trajectory; the linear program solver error from the Earth Mover's distance algorithm, which can be reduced by resetting the stopping criteria stricter.

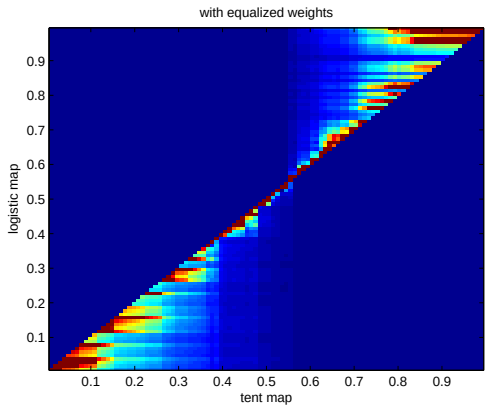


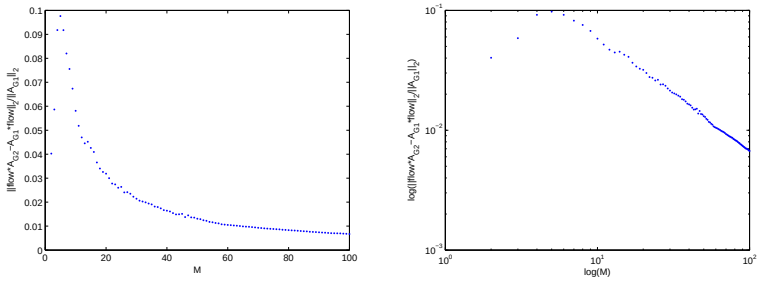

Figure 4.5: Matching Matrix using Equalized weights on the Adjacency Matrices

In this section, we will show, numerically, that these errors can be reduced by refining the partition and increasing sampling size.

On the other hand, we will consider the regularity of our method, i.e. we will show whether "closer" systems gives a smaller number of "dissimilarity" for matching. As our goal of this work is to compare general dynamical systems by graph matching method. We seek to study the metric of "dissimilarity" as the measure of distance from being conjugacy, or isomorphism for the associate graphs.

### 4.4.1 Partition Error

As we have discussed the partitioning issue in section 4.3.2, we can use the equidistant partition to approximate the Markov partition with more and more cardinality of partition $M$. Continuing the example in previous section, we consider the logistic map $T_1$ and the symmetric full tent map $T_2$, and the value $\frac{\|flow*A_{T2}-A_{T1}*flow\|_2}{\|A_{T1}\|_2}$, a relative error of being isomorphic, when increasing $M$. Figure 4.6 suggests the error is decreasing approximately polynomially as $M$ increases. In particular, when cardinality of partition is 100, i.e. $M = 100$, $\frac{\|flow*A_{T2}-A_{T1}*flow\|_2}{\|A_{T1}\|_2} = 0.46\%$.




(a) Partition error for refining partition

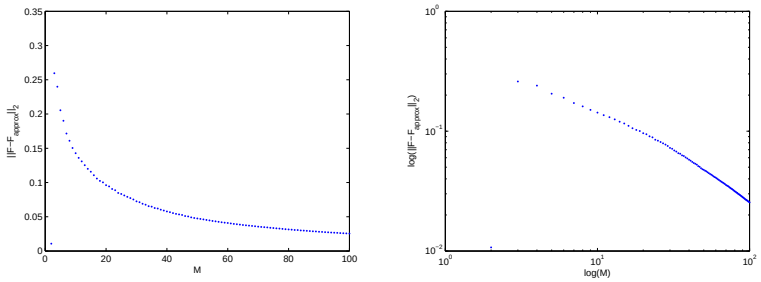
(b) Partition error for refining partition in *log* scale

Figure 4.6: Partition error $\frac{||flow*A_{T2}-A_{T1}*flow||_2}{||A_{T1}||_2}$ with number of partition $M$ increasing.

### 4.4.2 Statistical Error

We apply the Earth Mover's distance to compare graphs which are from taking a particular finite trajectory of systems. If we let $F$ be the matching matrix between the exact graphs from systems, which can be solved analytically in this example, and let $F_{approximate}$ be the matching matrix between the approximate graphs from finite trajectories, and $||F - F_{approximate}||_2$ be the error. Then by taking longer iterations for the trajectory and transient, we have a more precise statistical results, which gives a smaller number of error.

On the other hand, as shown in Figure 4.7, if we let the cardinality $M$ increase, the error is decreasing approximately polynomially.




(a) Error analysis for refining partition

(b) Error analysis for refining partition in *log* scale

Figure 4.7: Error analysis for refining partition

### 4.4.3 Regularity Property for the Metric

We have discussed the issue for conjugate systems, which allows perfect matching for the corresponding graphs. What we are going to consider is, whether "closer" systems implies smaller measure of the dissimilarity, which we regard as a regularity issue of the method. Here we consider an example with a logistic map $L$ and a symmetric full tent map with additive noise $T$. We let the measure of dissimilarity to be $\|F * A_T - A_L * F\|_2$, where $F$ is the matching matrix by Earth Mover's distance. As shown in Figure 4.8, when we decrease the amplitude of the additive noise, the dissimilarity value $\|F * A_T - A_L * F\|_2$ is decreasing to a certain value monotonously.

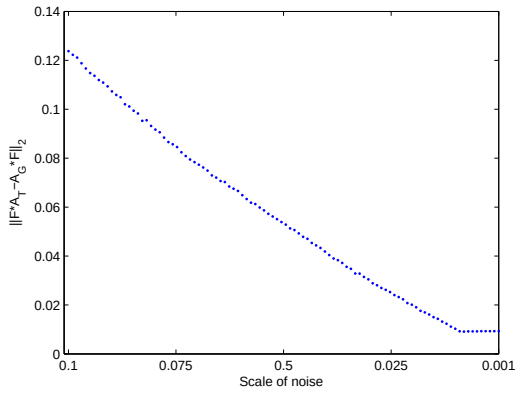



Figure 4.8: Measure of dissimilarity $\|flow * A_T - A_L * flow\|_2$ with decreasing scale of noise for the noisy tent map.

## 4.5 For Embedding Systems

We have established the graph matching method to compare conjugate systems. However, in most of the cases, the underlying systems are not perfectly homeomorphic. It is our goal of this topic to develop the graph matching method to compare general dynamical systems. But the difficulty is, even we can follow the same procedure to present the matching matrix like Figure 4.4 for general non-conjugate systems, the theoretical explanation and justifica-

tion stand in our way. But at this stage, we have shown that, if one system embeds into the other, the issue would become a subgraph matching problem.

We have the following definitions regarding to embedding:

**Definition 4.5.1.** *If $\phi : X \to Y$ is one-to-one, then $\phi$ is called an embedding of $X$ into $Y$.*

**Definition 4.5.2.** *$P(X)$ denote the set of all periodic points in $X$. $q_n(X)$ is the number of points with least periodic $n$, i.e. $q_n(X) = |\{x : \phi^n(x) = x,\ \phi^k(x) \neq x,\ 0 < k < n\}|$.*

Note that if $\phi : X \to Y$ is an embedding, then $\phi$ restricts to a one-to-one shift-commuting mapping from $P(X)$ into $P(Y)$. The existence of such a mapping is equivalent to the condition $q_n(X) \leq q_n(Y)$ for all $n \geq 1$.We denote these equivalent conditions by $P(X) \hookrightarrow P(Y)$, and call this the embedding periodic point condition.

In addition, we refer "the induced subgraph" as:

**Definition 4.5.3.** *Let $H$ be a graph with vertex set $v$. For each subset $w$ of $v$ define the induced subgraph of $H$ from $w$ to have vertex set $w$ and edge set the collection of all edges in $H$ that start and end in $w$. An induced subgraph of $H$ is one that is induced by some subset of $v$,*

Note that if $H$ has adjacency matrix $B$, and $G$ is an induced subgraph of $H$, then $A_G$ is a principal submatrix of $B$; i.e., $A_G$ is obtained from $B$ by deleting the $j$th row and $j$ column for a certain set of $j$'s.

We would first give the proof of Theorem 4.5.4, which is similar to the proof of *Masking Lemma* (Theorem 4.7.10). But the difference is, our proof is like "cutting edges", while the proof of *Masking Lemma* is "adding edges".

**Theorem 4.5.4.** *Let $G$ and $H$ be graphs. Suppose that $X_G$ embeds into $X_H$. Then there is a induced subgraph $K$ of $H$, such that $X_G \cong X_K$.*

*Proof.* Let $G'$ and $H'$ be graphs constructed in Theorem 4.7.9. Then $A_H \cong A_{H'}$. So there is a sequence of graphs

$$H' = H_0,\ H_1 \cdots,\ H_k = H$$

such that $A_{H_i} \cong A_{H_{i+1}}$ for $0 \leq i \leq k-1$. By the next Lemma that will be proved, we will find a sequence of graphs

$$G' = G_0,\ G_1 \cdots,\ G_k$$

such that each $G_i$ is an induced subgraph of $H_i$, and the elementary equivalence from $A_{H_i}$ to $A_{H_{i+1}}$ extends to one from $A_{G_i}$ to $A_{G_{i+1}}$. We set $K = G_k$. Then $A_K \cong A_{G'} \cong A_G$ and $K$ is an induced subgraph $H$. □

**Lemma 4.5.5** ("cutting edges")**.** *Let $G_1,\ H_1,\ H_2$ be graphs such that $A_{H_1} \cong A_{H_2}$ and $G_2$ is an induced subgraph of $H_1$. Then there is a graph $G_2$ such that $A_{G_1} \cong A_{G_2}$ and $G_2$ is an induced subgraph of $H_2$.*

*Proof.* Let $A = A_{H_1}$ and $B = A_{H_2}$. Let $(R, S) : A \cong B$ so that $RS = A$ and $SR = B$. Let $G_{R,S}$ be the auxiliary graph. This graph has vertex set which is the disjoint union of $v(H_1)$ and $v(H_2)$. It contains a copy of $H_1$, whose edges are called $A-edges$, and a copy of $G_2$, whose edges are $B-edges$. For each $I \in v(H_1)$ and $J \in v(H_2)$, there are $R_{IJ}$ edges from $I$ to $J$, called $R-edges$, and $S_{JI}$ edges from $J$ to $I$, called $S-edges$.

We use $G_1$ to reduce $G_{R,S}$, forming a new graph $\bar{G}$ as follows. Let $v = v(H_1) \setminus v(G_1)$ and $\varepsilon = \varepsilon(H_1) \setminus \varepsilon(G_1)$. Subtract $v \cup \varepsilon$ to the vertex set of $G_{R,S}$ to form the vertex set of $\bar{G}$. To avoid notational confusion, let $I_e$ denote the subtracted vertex corresponding to $e \in \varepsilon$. It will be helpful to think of vertices in $v$ as being subtracted to the "$H_1-part$" of $G_{R,S}$. And vertices $I_e$, for each $e \in \varepsilon$, as being subtracted to the "$H_2-part$" of $G_{R,S}$. For each $e \in \varepsilon$, subtract a $R-edges\ r(e)$ from the initial state of $e$ to $I_e$, and $S-edge\ s(e)$ from $I_e$ to the terminal state of $e$. For each $R,S-path$ from $I$ to $I'$ that this cuts, subtract to $\bar{G}$ an $A-edge$ from $I$ to $I'$; similarly, for each $S,R-path$ from $J$ to $J'$ that this cuts, subtract to $\bar{G}$ a $B-edge$ from $J$ to $J'$ (see Figure 4.9). This completes the construction of $\bar{G}$.

Each $e \in \varepsilon$ determines a cut $R,S-path\ r(e)s(e)$ from $G_{R,S}$ with the same initial and terminal states in $G_1$ as in $\bar{G}$, and these are the only subtracted $R,S-paths$. Hence the

graph with vertices $v(H_1) \setminus v$ and $A-edges$ of $\bar{G}$ is isomorphic to $G_1$. Let $G_2$ be the graph with vertices $v(H_2) \setminus \{I_e : e \in \varepsilon\}$ and all $B-edges$ of $\bar{G}$. Let $\bar{R}$ and $\bar{S}$ record the incidences of $R-edges$ and $S-edges$ in $\bar{G}$. The $\bar{G}$ is the auxiliary graph $G_{\bar{R},\bar{S}}$ corresponding to the elementary equivalence $(\bar{R}, \bar{S}) : A_{G_1} \cong A_{G_2}$. Furthermore, $G_2$ is the subgraph of $H_2$ induced from the vertex subset $v_{G_2}$. □

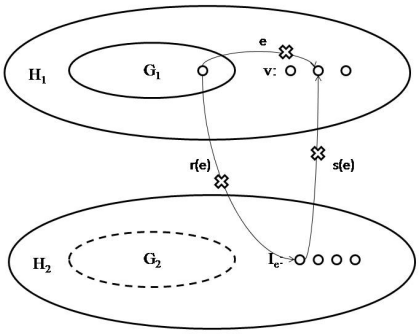


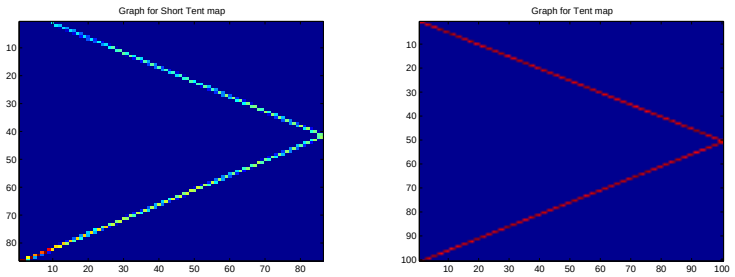

Figure 4.9: Auxiliary graph

Combine Theorem 4.5.4 and Theorem 4.3.2, we have the following theorem, which establish the theory for non-conjugate systems where one embeds into the other.

**Theorem 4.5.6.** *Suppose we are given two non-conjugate systems $SYS_1$ and $SYS_2$, and a partition $\Gamma$ with $|\Gamma| = M$, such that the resulting edge shifts $X_{G_1}$ is conjugate to $SYS_1$, $X_{G_2}$ is conjugate to $SYS_2$, where $A_{G_1}$ is the adjacency matrix of graph that generated from $SYS_1$ with this partition, and $A_{G_2}$ is from $SYS_2$. Suppose $X_{G_1}$ embeds into $X_{G_2}$. Then there is a induced subgraph $G_2'$ of $G_2$, such that there is a matrix $F$ (the Earth Mover's Distance matching matrix between graphs from $X_{G_1}$ and $X_{G_2'}$), satisfies*

$$F \cdot A_{G_1} = A_{G_2'} \cdot F.$$

A simple illustration would be the short symmetric tent map and full tent map, which are not conjugate systems, and the short one embeds into the full one.




(a) Adjacency Matrix for short tent map

(b) Adjacency Matrix for full tent map

Figure 4.10: Induced Subgraph

Figure 4.10(a), denote $A_G$, is the graph for short symmetric tent map, whose peak is at 0.95, while Figure 4.10(b), denote $A_K$, is the graph for short symmetric tent map. We claim that Figure 4.10(a) is a induced subgraph of Figure 4.10(b), or say Figure 4.10(a) can be enlarged to Figure 4.10(b).

## 4.6 For Higher Dimensional Dynamical Systems

The graph matching method for higher dimensional dynamical systems can be easily extended from the 1-d case. For 2-d dynamical systems, we partition the phase by equal sized boxes. Then the graphs are determined by investigating the trajectory of a randomly chosen starting point, like the procedure for 1-d maps. And EMD can be applied to obtain the matching matrix.

As a example, we use the graph matching method to compare two conjugate Henon map (blue and red Henon maps in Figure 4.11). We choose the blue Henon map as:

$$\begin{cases} x_{n+1} = 1 - ax_n^2 + y_n \\ y_{n+1} = bx_n \end{cases} \tag{4.3}$$

with parameters $a = 1.4,\ b = 0.3$, and the red Henon map, which is a cubic transformation of the blue one, i.e. $x_{red} = x_{blue}^3;\ y_{red} = y_{blue}^3$, which is a smooth change of coordinate.

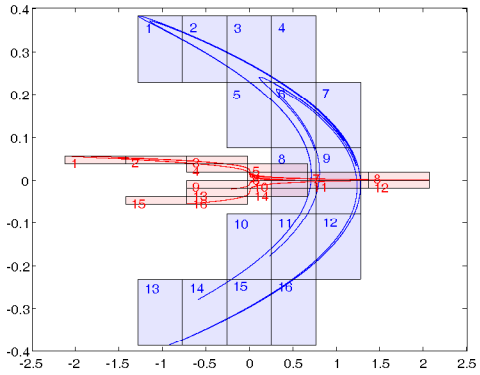
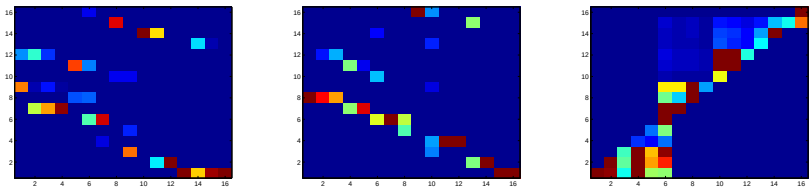


Figure 4.11: Conjugate Henon Maps

Figure 4.12(a) and Figure 4.12(b) show the graphs of the blue Henon and red Henon, respectively. Applying EMD on this two graphs gives us the matching matrix Figure 4.12(c). Note that due to the coarse partition, which is good for visualization, the matching matrix only suggests a rough association between the boxes in Figure 4.11.



(a) Blue Henon Maps (b) Red Henon Map (c) Matching Matrix

Figure 4.12: Graph Matching for 2D Dynamical Systems

## 4.7 Conclusion

In this paper, we compare dynamical systems by using graph matching method. For conjugate systems, which have isomorphic graphs, we show that the permutation matrix that relates the adjacency matrices of the graphs is the same as the solution of Monge's problem. We use Earth Mover's Distance (EMD) to generate the "approximated" matching matrix to illustrate the association of graphs which are derived from equal-distance partitioning of

the phase spaces of systems. In addition, for one system embedding into the other, we show that the comparison of these two systems by our method is an issue of subgraph matching. In the end we show how to extend the method to compare higher dimensional systems.

For our ultimate objective and future work, we seek to theoretically extend the graph matching method to compare more general dynamical systems, which are not necessary conjugate or embedding. On the other hand, the regularity property of our method should be justified. In particular, "closer" systems in the metric of graphs' distance should imply closer dynamical relationships, as considered in paper [2]. And by this setting, finding an optimal system to model a time series data or approximate a system can be achieved by minimizing the "dissimilarity" of matching.

# Appendix

Here we list the theorems and definitions that we used in our proof. The reader can refer to Lind [18] for further details.

**Theorem 4.7.1** ([18])**.** *The edge shifts $X_A$ and $X_B$ are conjugate if and only if matrices A and B are strong shift equivalent, we write*

$$X_A \cong X_B \Longleftrightarrow A \approx B$$

**Definition 4.7.2** ([18])**.** *Let A and B be nonnegative integral matrices and $l \geq 1$. A shift equivalence of lag l from A to B is a pair $(R, S)$ of rectangular nonnegative integral matrices satisfying the four shift equivalence equations*

$$\begin{aligned} &(i) AR = RB, SA = BS \\ &(ii) A^l = RS, B^l = SR \end{aligned} \tag{4.4}$$

*Denote by $(R, S) : A \sim B$ (lag $l$). We say that A is shift equivalent to B, written $A \sim B$, if there us a shift equivalence from A to B of some lag.*

**Theorem 4.7.3** ([18]). *Strong shift equivalence implies shift equivalence, we write*

$$A \approx B \Longrightarrow A \sim B.$$

**Definition 4.7.4** ([18]). *Let A and B be integral matrices. Then A and B are shift equivalent over $\mathbb{Z}$ with lag l, written $A \sim_{\mathbb{Z}} B$ (lag l), if there are rectangular integral matrices R and S satisfying the shift equivalence equations Definition 4.7.2 (i) and (ii). In this case we write $(R, S) : A \sim_{\mathbb{Z}} B$ (lag l).*

**Theorem 4.7.5** ([18]). *$A \sim B \Longrightarrow A \sim_{\mathbb{Z}} B$. If $A,\ B$ are primitive, $A \sim_{\mathbb{Z}} B \Longrightarrow A \sim B$.*

**Theorem 4.7.6** ([18]). *Let A and B be integral matrices, and suppose that R and S are rectangular integral matrices so that $(R, S) : A \sim_{\mathbb{Z}} B$. If $\mu \neq 0$ and $v \neq 0$ are such that $Av = \mu v$, then $w = Sv \neq 0$ and $Bw = \mu w$. Hence A and B have the same set of nonzero eigenvalues, so that $\lambda_A = \lambda_B$.*

**Theorem 4.7.7** (Perron-Frobenius Theorem, [18]). *Let $A \neq 0$ be an irreducible matrix. Then A has a positive eigenvector $v_A$ with corresponding eigenvalue $\lambda_A > 0$ that is both geometrically and algebraically simple. If $\mu$ is another eigenvalue for A, the $|\mu| < \lambda_A$. Any positive eigenvector for A is a positive multiple of $v_A$. We call $\lambda_A$ the Perron eigenvalue of A, and $v_A$ a Perron eigenvector of A.*

**Theorem 4.7.8** (Embedding Theorem, [18]). *Let X and Y be irreducible shifts of finite type. Then there is a proper embedding of X into Y if and only if $h(X) < h(Y)$ and $P(X) \hookrightarrow P(Y)$.*

**Theorem 4.7.9.** *([18]) Suppose that $X_G$ embeds into $X_H$. Then there are graphs $G'$ and $H'$ so that $X_{G'} \cong X_G,\ X_{H'} \cong X_H$, and $G'$ is a subgraph of $H'$.*

Note that the Masking Lemma, state below, strengthens this result by showing that we only need to modify $H$, leaving $G$ exactly as it is, and that $G$ can be realized as a particular kind of subgraph of $H'$.

**Theorem 4.7.10** (Masking Lemma, [18]). *Let G and H be graphs. Suppose that $X_G$ embeds into $X_H$. Then there is a graph K such that $X_K \cong X_H$ and G is an induced subgraph of K.*

# Chapter 5

# The Bundle plot: A presentation of symbolic space's evolution under the system's parameter changes

In this chapter, we will study a topological dynamics perspective on the full bifurcation unfolding in unimodal mappings. We present a bundle structure, visualized as a bundle plot, to show the evolution of symbolic space as we vary a system parameter. The bundle plot can be viewed as a limit process of an assignment plot, which are lines assignments between points from two dynamical systems. Such line assignments are determined by a commuter, which is a coordinates transformation function that satisfies a commuting relationship but not necessary a homeomorphism. The bundle structure is studied with understanding of implication from system's qualitative changes. In addition, the case of the bundle plot with higher dimensional parameter variation is also considered. A main concern in the bundle plot is a special structure, called "joint", which determines a critical value of the parameter where the kneading sequence becomes periodic.

## 5.1 Introduction

An interesting and fundamental question in science and dynamical systems is how do we describe evolution of systems due to parameters changes. Bifurcation theory characterizes qualitative changes in the way of tracking fixed points' creation, or destruction, or stability changes. A bifurcation occurs when a small change made to a system's parameter causes a topological difference in its behavior. So two similar systems with the same formation but slightly different parameters values may behave qualitatively different. Some "difference measurements" such as least square method fail to reveal the underlying topological distinction. Bollt and Skufca introduced a concept of a commuter to extent the equivalence relationship of conjugacy to measure the distance from being conjugate , which can also be thought of as the degree of "matching" between trajectories[2, 6]. Later, they extended and interpreted the concepts of "commuter" and "defect measure" to symbol space, and introduced a new visualization technique, called assignment plot, to show matchings between symbol sequences of two topological spaces [5].

Based on this perspective, we consider a "limit process" of assembling assignment plots from a set of different system's parameter values. We extend here the idea of studying the qualitative difference between two systems to that within a family of systems, where we give the name "bundle plot" to the resulting structure. Compared to traditional Bifurcation theory, instead of studying the qualitative changes of fixed points, we picture the symbolic space's evolution due to continuous changes of systems' parameter. The bundle structure implies trajectories, or say symbolic sequences, creation or destruction, and the "speed" of such changes. In addition, a special structure "joint" implies the kneading sequence is periodic, which happens shortly after periodic window's opening.

This chapter is organized as follows: in section 5.2, we review the concept of "mostly conjugacy" and "commuter; in section 5.3, we extended and interpreted the concepts in a symbolic dynamics perspective; in section 5.4, we extend the assignment plot to the bundle plot; we consider the bundle plot in higher dimensional parametric changes in section 5.5;

then we compare our result to that of bifurcation theory in section 5.6; we finish our paper with discussion and future work in section 5.7.

## 5.2 Comparing dynamical systems: review of non-homeomorphic commuter

An essential question within the modeling context is "how close is the models to the true phenomena." Where the natural system under consideration is dynamic, with possibly complex behavior. The field of dynamical systems seeks to provide an appropriate framework for study of these systems. Since the inception of the field of dynamical systems by Henri Poincare [1], the fundamental approach has been to examine topological and geometric features of orbits, rather than focusing on numerical specifics of particular solutions of the dynamical system, as measured in some specific coordinate system. Characterization of the system relies upon deciphering coordinate independent properties, such as the periodic orbit structure - the count and stability of periodic orbits.

In dynamical systems, the usual way to relate two dynamical systems is with the topological notion of conjugacy, related by the following commuting diagram,

$$\begin{array}{ccc} X & \xrightarrow{g_{b_1}} & X \\ \downarrow{\scriptstyle f_{b_1 b_2}} & & \downarrow{\scriptstyle f_{b_1 b_2}} \\ Y & \xrightarrow{g_{b_2}} & Y \end{array}$$

The conjugacy function $f_{b_1 b_2} : X \to Y$ satisfies the commuting relationship

$$f_{b_1 b_2} \circ g_{b_1} = g_{b_2} \circ f_{b_1 b_2}, \tag{5.1}$$

with $b_i$ the parameter of system $g_{b_i}$. The conjugacy $f_{b_1 b_2}$ is a homeomorphism from system $g_{b_1}$ to $g_{b_2}$, i.e. $f_{b_1 b_2}$ is 1-to-1, onto, continuous and has a continuous inverse function. It

is the change of coordinates that the mappings behave exactly the same in either coordinate system. However, a commuter is an arbitrary function, not necessary a homeomorphism, that satisfies the commuting relationship, in which case it is a non-homeomorphic change of coordinates translating between dissimilar systems. Note the commuter provides a matching between trajectories for $g_{b_1}$ and $g_{b_2}$ over- and/or under-representations are reflected as 1-to-1 and onto problems in $f$, while trajectories that permit matching only for finite time are related to discontinuities in $f$. We therefore developed measures of commuters $f$ to quantify "how much" $f$ fail to be a "perfect match", i.e. homeomorphism [2, 6]. For two topological conjugate systems which are related by a homeomorphism, the dynamics of one system completely describe the dynamics of the other. Thus, the notion of "distance to conjugacy" should provide a means of determining the extent to which the dynamics are similar. Paper [2] defined and studied the measure of the deviation from homeomorphism, named homeomorphic defect, which provides a weighted average based on measurements of possible failure of $f$ being onto, 1-1, continuous and inverse continuous. This defect measure is proven to have certain regularity properties [6], which supports the our definition of "distance".

On construction of commuter $f_{b_1b_2}$, we define a commuter operator $f^{n+1}_{b_1b_2}(x) = \mathfrak{C}^{g_{b_2}}_{g_{b_1}} f^n_{b_1b_2}(x) = g^{-1}_{b_2} \circ f^n_{b_1b_2} \circ g_{b_1}(x)$, with $f^1_{b_1b_2} = I$. The subscript of $f^n_{b_1b_2}$ means this is a $n$-iterate commuter from system $g_{b_1}$ to $g_{b_2}$. We require $g^{-1}_{b_2}$ to be a well defined inverse function and piecewise Lipshitz continuous with constant $L < 1$, which guarantees there is a unique fixed point, say $f_{b_1b_2}$, of the commutation operator $\mathfrak{C}^{g_{b_2}}_{g_{b_1}}$. The left column of Figure 5.1 shows example of commuters between a short logistic map and skew full tent map, and between skew full tent map and short logistic map. Notice the vertical gaps of the first commuter shows that, there are some intervals of the full tent can not be matched by the short logistic.

## 5.3 A symbolic dynamics interpretation of commuter functions

The commuter function in the previous section describes a point-wise matching from one coordinate system to the other. We discuss the idea of a commuter in a symbolic dynamics interpretation, which illustrates the non-homeomorphism in a broader setting. In particular, each point $x$ in $g_{b_1}$ is represented in a different coordinate system $g_{b_2}$ by $y = f_{b_1 b_2}(x)$. The degree to which the commuter fails to be a homeomorphism defines what we call a homeomorphic defect. However, there were limits in the mathematical technology requiring that the transformations be one-dimensional mappings for rigorous construction of the construction of the commuters by fixed point iteration. Further, there are difficulties in numerically computing defects in the more complicated one dimensional cases, and further limits to higher dimensional problems. Interpretation of dynamical systems through a symbolic representation has become the standard tool for identifying key dynamical structures and behaviors, particularly when studying chaotic systems [**?**]. Therefore, Bollt and Skufca et [5] extend the "commuter" theory to a symbolic dynamics setting, which allows for multivariate transformations, with construction methods separate from the fixed point iteration, and new methods to compute defect. In addition, we introduce assignment mappings/plots, which is a new visualization technique of commuters, to understand and illustrate commuters in a broader perspective (see Figure 5.1).

In fact, the symbolic commuter matches points in $X$ to points in $Y$ such that the respective symbolic sequences will match for as many symbols as possible, which is proven in paper [5]. This also provides us an optimization criterion to construct commuters for higher dimensional dynamical systems. In principle, we assume that the dynamical systems under consideration are presented to us with a known symbolic dynamic partitioning. In practise, if necessary, we can use the uniform partitioning to approximate the "actual" partition that can fully describe the original system, which we concern in a later work [44].

To define symbol dynamics for systems, we assume a shift space (or say grammarr) on $\Sigma_n$, with integer symbols $1, ..., n$. We associate symbol $i$ with interval $I_{X_i}$ for the dynamics on $X$ and with $[y_{i-1}, y_i]$ for the dynamics on $Y$. A trajectory of systems $g_{b_1}$, given by $\{x, g_{b_1}(x), g_{b_1}^2(x), ...\}$ has an associated symbolic trajectory $S(x) = s_0 s_1, ...$, where $g_{b_1}^j(x) \in I_{X_i} \Rightarrow s_j = i$. Similarly, a trajectory of system $g_{b_2}$, given by $\{y, g_{b_2}(y), g_{b_2}^2(y), ...\}$ has an associated symbolic trajectory $S(y) = s_0 s_1, ...$, where $g_{b_2}^j(y) \in [y_{i-1}, y_i] \Rightarrow s_j = i$. We remark that because the closed partition on $Y$ gives an overlap at endpoints, the symbolic trajectory is non-unique for any preimage of any element of $E_y$. We denote $\Sigma^y$ as the subshift of all possible symbolic sequences for $y$.

The algorithm to construct symbolic commuter, which is given in detail in paper [5], yields a unique commuter after one has assigned an appropriate partitioning from the spaces $X$ and $Y$. An assignment plot shows line segments from representative points in $X$, under $g_{12}$ to their images in $Y$. Figure 5.1 illustrates the characterization of the commuter as providing a matching between points/symbolic sequences/trajectories via lines.

In Figure 5.1, we show assignment plots (blue lines) taken from the commuters between the maps on the left side of the diagram. The top assignment plot is taken from the commuter between a subshift logistic map $g_{b_1}$ and a full shift skew-tent map $g_{b_2}$. The vertical gap of the commuter says $g_{b_2}$ has some dynamics that $g_{b_1}$ doesn't. For example, $g_{b_1}$ admits neither *abaa* nor *bbaa*, which can be observed from the first assignment plot, associated to the largest vertical gap in the first commuter. Similarly for the second assignment plot, $g_{b_2}$ is a full shift on two symbols, the words *abaa* and *bbaa* exist in the dynamics of $g_{b_2}$, but cannot be "matched" to a depth of four symbols with any point of $g_{b_1}$. Those intervals associate to the largest horizontal portion of the second commuter. The commuter maps those points to 0.5, on the boundary between symbols *a* and *b* of the $g_{b_1}$ dynamics, yielding a match to a depth of three symbols, either *aba* or *bba* as appropriate.

In the next section, we generalize the idea of assignment plot for comparison among a family of systems, named a bundle plot. A bundle plot can be regarded as an evolution of

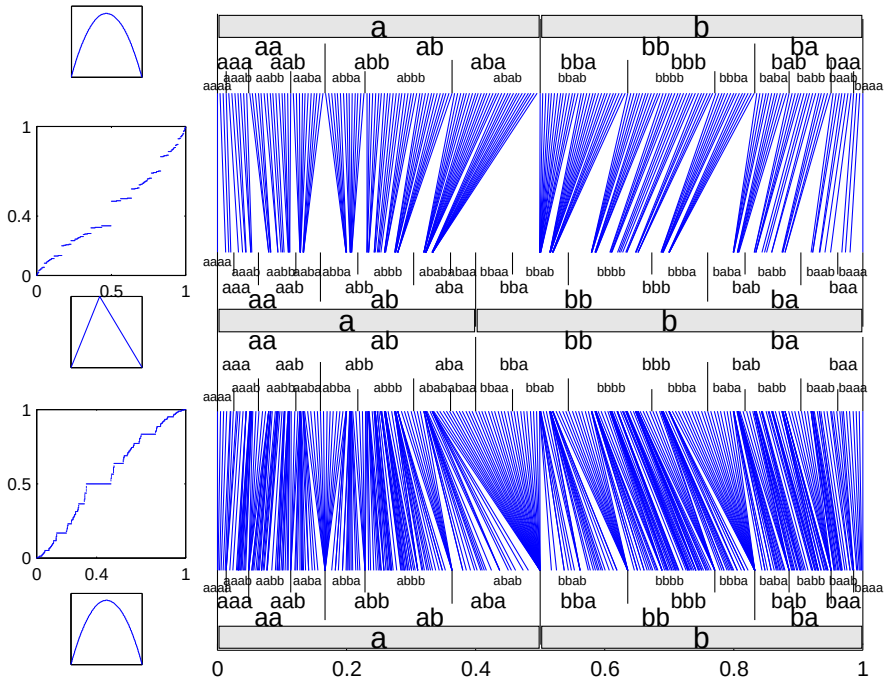



Figure 5.1: **Commuter as the matching of symbol dynamics.** (Left column) A sequence of maps: (1) $g_{b_1} = 3.6x(1-x)$, a subshift logistic map; (2) the commuter between $g_{b_1}$ and (3) the full shift skew-tent map $g_{b_2}$; (4) the commuter from $g_{b_2}$ back (5) map $g_{b_1}$. (Right column) The symbol dynamic partition of the unit interval for maps $g_{b_1}$ (at top and bottom) and $g_{b_2}$ (in the middle), where we show all intervals for word length of four or fewer symbols. The assignment plot (blue lines) is taken from the commuters between the maps (as computed from the fixed point iteration), and shows that the commuter also gives a symbolic dynamic match between the two systems. *Observe*: (1) Map $g_{b_1}$ admits neither *abaa* nor *bbaa*, associated to the largest vertical gap in the first commuter. (2) Because $g_{b_2}$ is a full shift on two symbols, the words *abaa* and *bbaa* exist in the dynamics of $g_{b_2}$, but cannot be "matched" to a depth of four symbols with any point of $g_{b_1}$. Those intervals associate to the largest horizontal portion of the second commuter. The commuter maps those points to 0.5, on the boundary between symbols *a* and *b* of the $g_{b_1}$ dynamics, yielding a match to a depth of three symbols, either *aba* or *bba* as appropriate.

the shift space of a particular system, under a variation of a parameter. It shows how the system gains or loses dynamics/grammar when we increase or decrease the parameter.

## 5.4 The bundle: assignment plots of commuters for a family of maps

For a discrete dynamical system $x_{n+1} = g(x_n, b_i) := g_{b_i}(x_n)$, where $b_i$ is the parameter of the system, we want to study the creation or destruction of allowable symbolic sequences/trajectories of the system as we vary $b_i$. We let $g_{b_1}$ be the base system with a particular fixed parameter $b_1$, and compare it to mappings $\{g_{b_i}\}$. Denote $f_{b_1 b_i}$ to be the commuter from $g_{b_1}$ to $g_{b_i}$. Denote $f^n_{b_1 b_i}$ be the $n-step$ commuter between $g_{b_1}$ and $g_{b_i}$ in the iteration scheme that generates the commuter. We let $f^1_{b_1 b_i}$, i.e. the initial guess, to be $I$.

For each $x$ in the base space $X$, we line up all the range points of $\{f_{b_1 b_i}(x)\}$, with $\{b_i\}$ values in an interval. Figure 5.2 illustrates such association from the family of tent maps across $.5 \leq b_i \leq 1$, with $b_1 = 1$.

$$x_{n+1} = g_{b_i}(x_n) = \begin{cases} 2b_i x_n, & 0 \leq x_n \leq 1/2, \\ 2b_i(1 - x_n), & 1/2 < x_n \leq 1. \end{cases}$$

We call $g_1$, the full ($b_1 = 1$) symmetric tent map, the base system. Each curve describes the matching between each $x$ and $\{f_{1 b_i}(x)\}$. We name the entire structure as "bundle", and each curve as a "fiber". The names of "bundle" and "fiber" here are consistent with the concept of bundles in Topology, as we cite the definition below:

**Definition 5.4.1** (Fiber bundle [50])**.** *A fiber bundle (also called simply a bundle) with fiber $F$ is a map $f : E \to B$ where $E$ is called the total space of the fiber bundle and $B$ the base space of the fiber bundle. The main condition for the map to be a fiber bundle is that every point in the base space $b \in B$ has a neighborhood $U$ such that $f^{-1}(U)$ is homeomorphic to*

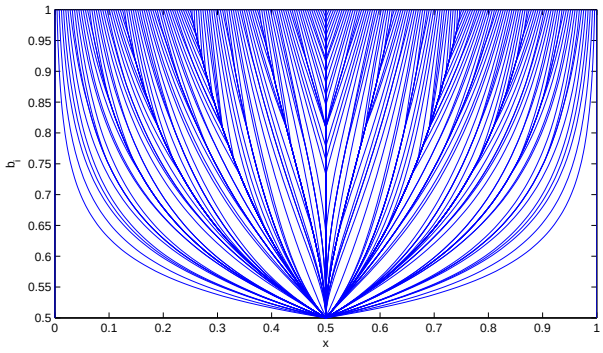



Figure 5.2: **A Bundle Plot** of one dimensional mappings $f_{1b_i} : X \to Y$, with the domain $X = [0, 1]$ shown at the top, and the range $Y \subset [0, 1]$. Line segments show assignments from $X$ to each $Y$ of a uniform grid of sample points representing the commuters $f_{1b_i}$ from $g_{b_1}$ and $g_{b_i}$, where $g_{b_1}$ is the full symmetric tent map and $g_{b_i}$ is short symmetric tent map with height $b_i \in [1/2, 1]$.

*$U \times F$ in a special way. Namely, if*

$$h : f^{-1}(U) \to U \times F \tag{5.3}$$

*is the homeomorphism, then*

$$proj_U \circ h = f_{|f^{-1}(U)|}, \tag{5.4}$$

*where the map $proj_U$ means projection onto the U component. The homeomorphisms h which "commute with projection" are called local trivializations for the fiber bundle $f$. In other words, E looks like the product $B \times F$ (at least locally), except that the fibers $f^{-1}(x)$ for $x \in B$ may be a bit "twisted".*

So following this definition, if we let $B = \{b_i\},\ F = \{f_{1b_i}(x)\}$, then $E = B \times F$ is a trivial fiber bundle.

In Figure 5.2, we can see that some of the fibers merge to singletons as we decrease the height of the tent map. A similar story happens for family of symmetric logistic maps (Figure 5.3), i.e. $\{g_{b_i}(x) : x_{n+1} = g_{b_i}(x_n) = 4b_i x_n(1 - x_n)\}$ with the height $0.89 < b_i < 1$ and

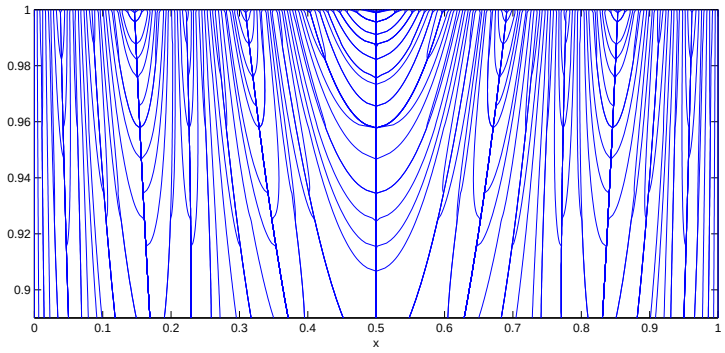



Figure 5.3: **A Bundle Plot** of one dimensional mappings $f_{1b_i} : X \to Y$, with the domain $X = [0, 1]$ shown at the top, and the range $Y \subset [0, 1]$. Line segments show assignments from $X$ to each $Y$ of a uniform grid of sample points representing the commuters $f_{1b_i}$ from $g_{b_1}$ and $g_{b_i}$, where $g_{b_1}$ is the full logistic map and $g_{b_i}$ is short logistic map with height $b_i \in [0.89, 1]$.

$b_1 = 1$. Note $b_i > 0.89$ guarantees the existence of the commuter between $g_{b_1}$ and $g_{b_i}$.

Here we only concern about "symbolic fibers" $\{f_{1b_i}(x)\}$ where $x$ are the generating partition boundary points. In particular,

**Definition 5.4.2.** *Let $\mathscr{P} = \{x|\ g^n_{b_1}(x) = 1/2\}$, the set of all pre-image of 1/2 of $g_{b_1}(x)$. The symbolic fibers is the set $F = \{\{f_{1b_i}(x)\}|x \in \mathscr{P}\}$.*

A singleton from mergence of symbolic fibers at $b$ simply implies $\exists x_1, x_2 \in \mathscr{P},\ x_1 \neq x_2$ with $\{f_{1b_i}(x_1)\} = \{f_{1b_i}(x_2)\},\ \forall b_i \geq b$. The fibers from non-generating partitioning end points are "bounded" by two symbolic fibers and do not cross each other.

In the following, we study the bundle and those symbolic fibers in terms of symbolic dynamics. We shall start our discussion of bundles in terms of the tent map and hopefully it is apparent that other maps can be handled similarly. As can be observed from Figure 5.2, we identify some basic "branch" structures which assemble the entire bundle plot. As roughly sketch in Figure 5.4, we investigate the "branch" structure as 3 elemental structures, which demonstrate in Figure 5.4 top. Top left shows one symbolic fiber on each side of the mid vertical fiber join together; Top middle only have one symbolic fiber from one side joins the mid fiber; while top right has two symbolic fibers on each side of mid

fiber join together. The whole bundle structure is a fractal structure such that any elemental structure repeats others include itself. In the following, we theoretically interpret the elemental structures in terms of symbolic dynamics.

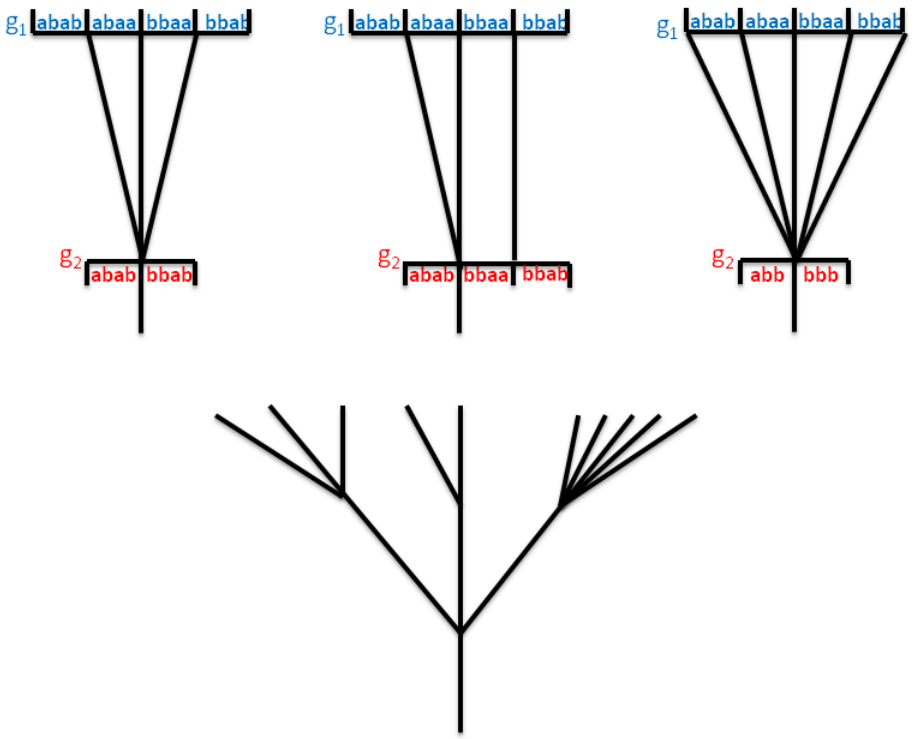
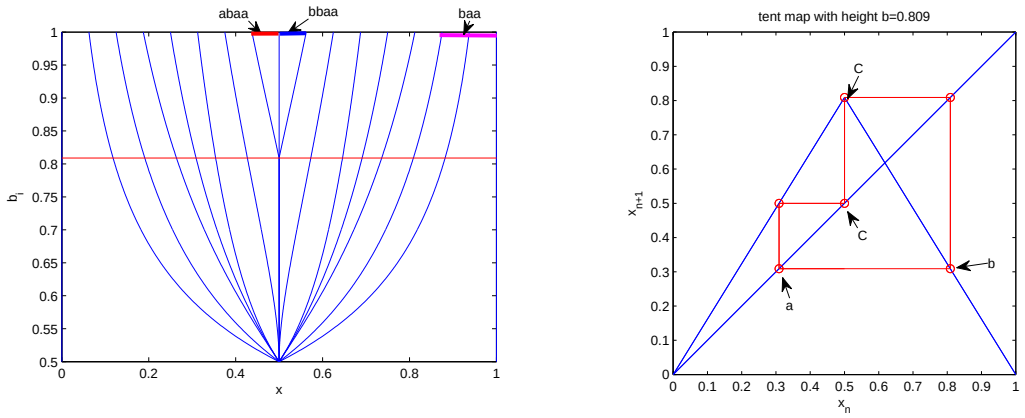



Figure 5.4: Elemental structure of a bundle: Suppose we have a generating partition, which consists of these four intervals *abab*, *abaa*, *bbaa*, *bbab* in $g_{b_1}$ system. (Top left) For the system $g_{b_2}$, suppose the sequences *abaa* and *bbaa* are not allowable, the intervals *abaa* and *bbaa* in $g_{b_1}$ map to a singleton of $g_{b_2}$. We called this case to be "losing 2 words of length 4"; (Top Middle) only an interval on the left side of the mid fiber lose 1 word, while the right side still has a perfect matching; (Top right) *bbaa*, *bbab*, *abaa* and *abab* (or say the entire *aba* and *bba*) are lost at the same time, which describes a "lost of 4 words of length 4, or lost of 2 words of length 3". (Bottom) The entire bundle plot is assembled by the top three elemental structures.

We first focus on the top left elemental structure in Figure 5.4. As we decrease the height of the tent map, the system loses dynamics by losing allowable symbolic sequences. Suppose we have a generating partition, which gives four intervals *abab*, *abaa*, *bbaa*, *bbab*

in $g_{b_1}$ system. Then, for instance, the points in the interval *abab* has an infinitely long sequence and starting with the word *abab*. Now suppose we compare $g_{b_1}$ to itself (Figure 5.6 left column). The commuter is trivially the identity maps, meaning every symbolic sequence matches itself perfectly ("perfect" means the commuter is a homeomorphism); But as we lower the height of $g_{b_2}$ below 1 (Figure 5.6 mid column, the red tent map is lower), the *abaa* and *bbaa* intervals are compressed, since they lose allowable symbolic sequences starting with *abaa* or *bbaa* due to loss of dynamics of the $g_{b_2}$ system; Until a critical value of $b$, the *abaa* and *bbaa* intervals completely disappears, and the *abaa* and *bbaa* of $g_{b_1}$ map to a singleton of $g_{b_2}$. We simply called this case to be "losing 2 words of length 4", in the sense that we lose the word *abaa* and *bbaa* at the same time; In fact, in this case, we lose the word *baa*. To see this, we notice when we lower the height of tent map $g_{b_2}$, the invariant set is smaller and exclude the interval *baa*, which appears as the most right interval under the generating partition (see Figure 5.5(a)). Thus the system loses the ability to generate any word with *baa* included, like *abaa* and *bbaa*. Therefore the words *abaa* and *bbaa* are not allowable in the grammar. What's more important is, we lose both words simultaneously. A step further, the lost of *baa* implies the associated kneading sequence is periodic (which will be proven in section 5.6). For example, Figure 5.5(a) shows the tent map bundle plot on the generating partition up to 4 symbols, with the red line at around 0.809 crossing a singleton, which is from the fibers of the left end point of *abaa* and right end point of *bbaa*. Consider the tent map with that associate height (around 0.809), whose kneading sequence happens to be period-3 with $\overline{Cba}$ (or say $g_{0.809}(g_{0.809}(g_{0.809}(.5))) = .5$). If we start with an initial point slightly greater than 0.5 (see Figure 5.5(b)), we can only have the symbolic sequence *bbab*.... In other words, *bbaa*... is impossible. Similarly if we start with an initial point slightly less than 0.5, we only have *abab*. But if we have the tent map higher than 0.809, both *abaa* and *bbaa* are allowable. For the family of symmetric tent map, the lost of words with certain length can be in pairs.

The top right structure in Figure 5.4, however, describes a "lost of 4 words of length




(a) Tent map bundle plot on the generating partition up to 4 symbols

(b) The tent map with height 0.809

Figure 5.5: **Illustration of lost of 2 words of length 4**: (a) The tent map bundle plot on the generating partition up to 4 symbols, with the red line at around 0.809 crossing a singleton, which is from the fibers of the left end point of *abaa* and right end point of *bbaa*. The singleton across the red line implies the lost of words *abaa* and *bbaa*, which further implies the lost of the word *baa*. This is due to the fact that the invariant set of tent map at height 0.809 excludes the interval *baa*, and thus *abaa* and *bbaa* are not allowable in the grammar; (b) The tent map with that associate height (around 0.809), whose kneading sequence happens to period-3 with $\overline{Cba}$ (or say $g_{0.809}(g_{0.809}(g_{0.809}(.5))) = .5$). Note both *abaa* and *bbaa*... are not allowable under the system.

4", which due to the fact that *abab*, *abaa*, *bbaa* and *bbab* are lost at the same time. This correspond to a "faster" lost of words comparing to the case of "losing 2 words of length 4".

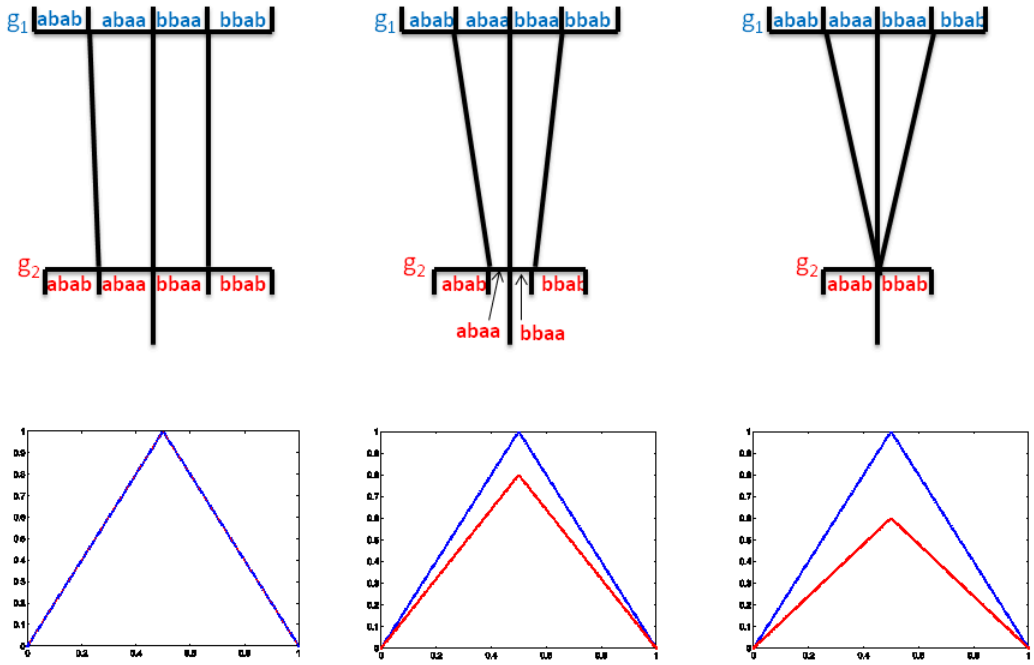



Figure 5.6: Degeneration of intervals: (Left column) suppose we compare $g_{b_1}$ to itself, the commuter is trivially the identity maps, meaning every symbolic sequence matches itself perfectly; (Middle column) as we lower the height of $g_{b_2}$ below 1, the *abaa* and *bbaa* intervals are compressed; (Right column) at a critical value of $b$, the *abaa* and *bbaa* intervals completely disappear, the *abaa* and *bbaa* in $g_{b_1}$ map to a singleton of $g_{b_2}$.

Based on these observations and analysis, we call the singleton a "joint", and define its associate parameter $b$ as follow:

**Definition 5.4.3** (Joint)**.** *A joint parameter is denoted as*

$$b_{joint}(\omega) = \sup\{b \mid \textit{the grammar } \Sigma \textit{ loses a word } \omega \textit{ of length } N\}. \tag{5.5}$$

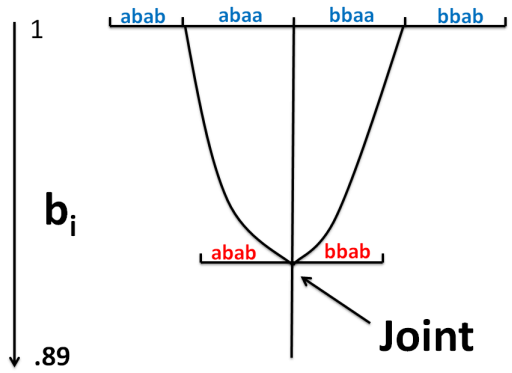



Figure 5.7: Joint: the existence of a joint implies lost of a word/words

As long as there is no confusion, we talk joint as the topological behavior and the associated parameter values. Whenever there is no need to distinguish $b_{joint}(\omega_1)$ from $b_{joint}(\omega_2)$, we simplify the notation $b_{joint}(\omega)$ as $b_{joint}$. In this paper, we concern problems about the joint: When does joint happen? Is there any relationship between the joints and the bifurcation? Generally speaking, when we decrease the height of the tent maps or logistic maps, the complexity and entropy of the systems would decrease, due to the loss of periodic orbits. For instance, in Figure 5.8, we calculate the total number of fibers for each $b_i$, which, as expected, gives a monotone increasing function, as we increase $b_i$ towards 1. As we increase the size of the grid points in $X$ space and $b$ space, the "flat spot" and "jump" in Figure 5.8 still exist, because once a particular symbolic sequence is lost, all the symbolic sequences of its pre-image disappear simultaneously. Thus, as we decrease the value of $b_i$, we may lose symbolic sequences that already lost.

In particular, the following theorem proves that the lost of symbolic sequences are actually non-smooth. More precisely, it can be represented as a Lebesgue singular function. In other words, the derivative of commuters with respect to parameter $b$ is not a smooth function, but 0 almost everywhere (though it is monotone non-decreasing from 0 to 1). We introduce one theorem and two lemma to prove this claim.

**Theorem 5.4.4.** *The derivative of commuter with respect to the model parameter b exists*

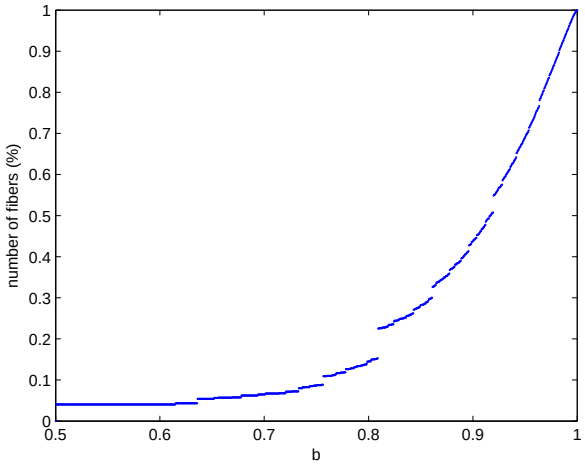



Figure 5.8: fibers' number: As we decrease the height of the tent map, the system loses dynamics in the way of losing allowable symbolic sequences. We calculate the total number of fibers for each $b_i$.

*almost every where.*

*Proof.* Since the commuter function is a piecewise monotone function with respect to parameter $b$ [2, 6], and hence of bounded variation. Thus its derivative exists a.e. [4]. □

**Lemma 5.4.5.** $f^n_{1b}(x_0) \to f_b(x_0)$ *for all* $x_0 \in [0,1]$ *as* $n \to \infty$.

The proof of this lemma refer to our former work [2]. It shows the existence and uniqueness of the commuter function by contraction mapping theorem.

**Lemma 5.4.6.** $f^n_{1b_i}(x_0) \to f^n_{1b}(x_0)$ *uniformly for* $x_0 \in [0,1]$ *as* $b_i \to b$.

*Proof.* For $x_0 \in [0, 1/2]$, we want to show, for $\forall\epsilon$, $\exists\delta$, for all $|b_i - b| < \delta$, we have $|f^n_{1b_i}(x_0) - f^n_{1b}(x_0)| < \epsilon$.

We show by induction:

For $n = 1$, since $f^1_{1b_i} = f^1_{1b} = I$, the uniform convergence holds;

Now assume the case $n$ is true, that is $\forall\epsilon$, $\exists\delta$, for all $|b_i - b| < \delta$, we have $|f^n_{1b_i}(x_0) - f^n_{1b}(x_0)| < \epsilon$.

For the case $n+1$, since $f_{1b_i}^{n+1}(x_0) = g_{b_i}^{-1} \circ f_{1b_i}^n \circ g_1(x_0) = \frac{1}{2b_i} f_{1b_i}^n(2x_0)$, for the above $\epsilon$ in case $n$, $\exists\, \delta = \epsilon$, so that:

$$
\begin{aligned}
&|f_{1b}^{n+1}(x_0) - f_{1b_i}^{n+1}(x_0)| \\
=&|\frac{1}{2b} f_{1b}^n(2x_0) - \frac{1}{2b_{i+1}} f_{1b_i}^n(2x_0)| \\
\leq&\frac{1}{2b}|f_{1b}^n(2x_0) - f_{1b_i}^n(2x_0)| + |\frac{1}{2b} - \frac{1}{2b_i}|f_{1b_i}^n(2x_0) \\
<&\frac{\epsilon}{2b} + \frac{\epsilon}{2bb_i} = (\frac{1}{2b} + \frac{1}{2bb_i})\epsilon < 3\epsilon.
\end{aligned}
$$

For $x_0 \in [1/2, 1]$, the proof is similar. □

Note that we are considering $b_i,\ b \in [1/2, 1]$.

**Theorem 5.4.7.** *The derivative of the commuter with respect to the parameter b is 0 whenever it exists.*

*Proof.* For $x_0 \in (0, \frac{1}{2})$, we let $H' = \lim_{b_2 \to b_3} \frac{|f_{1b_2}(x_0) - f_{1b_3}(x_0)|}{|b_2 - b_3|}$. We want to show $H' = 0$ whenever it exists.

Denote $(H^n)' = \lim_{b_2 \to b_3} \frac{|f_{1b_2}^n(x_0) - f_{1b_3}^n(x_0)|}{|b_2 - b_3|}$, then by Lemma 5.4.5 and Lemma 5.4.6, we can interchange the limit and have

$$
H' = \lim_{b_2 \to b_3} \lim_{n \to \infty} \frac{|f_{1b_2}^n(x_0) - f_{1b_3}^n(x_0)|}{|b_2 - b_3|} = \lim_{n \to \infty} \lim_{b_2 \to b_3} \frac{|f_{1b_2}^n(x_0) - f_{1b_3}^n(x_0)|}{|b_2 - b_3|} = \lim_{n \to \infty} (H^n)'. \tag{5.6}
$$

We show $(H^n)' = 0$ by induction.

For that case when $n = 1$, since $f_{1b_2}^1 = f_{1b_3}^1 = I$, the conclusion holds;

Assume the case $n$ is true, that is $(H^n)' = 0$ whenever it exists;

For the case $n+1$, $(H^{n+1})' = \lim_{b_2 \to b_3} \frac{|f_{1b_2}^n(x_0) - f_{1b_3}^n(x_0)|}{|b_2 - b_3|}$, $x_0 \in [0, 1]$. Since $f_{1i}^{n+1} = g_{b_i}^{-1} \circ f_{1b_i}^n \circ g_{b_1}$, we have

$$(H^{n+1})' = \lim_{b_2 \to b_3} \frac{|\frac{1}{2b_2} f^n_{1b_2}(x_0) - \frac{1}{2b_3} f^n_{1b_3}(x_0)|}{|b_2 - b_3|},\ x_0 \in [0, 1/2]. \tag{5.7}$$

Note the proof for $x_0 \in [1/2, 1]$ is similar.

Without loss of generality, we assume $\frac{1}{2b_2} f^n_{b_1 b_2}(x_0) > \frac{1}{2b_3} f^n_{b_1 b_3}(x_0)$ and $b_2 > b_3$, then we have

$$\frac{\frac{1}{2b_2}(f^n_{1b_2}(2x_0) - f^n_{1b_3}(2x_0))}{b_2 - b_3} \leq \frac{\frac{1}{2b_2} f^n_{1b_2}(2x_0) - \frac{1}{2b_3} f^n_{1b_3}(2x_0)}{b_2 - b_3} \leq \frac{\frac{1}{2b_3}(f^n_{1b_2}(2x_0) - f^n_{1b_3}(2x_0))}{b_2 - b_3}$$

We take the limit of $b_2 \to b_3$ on the above inequality, and by assumption of case $n$, we have $(H^{n+1})' = 0$. We note that for the assumption in case $n$, the conclusion is true for $x_0 \in [0, 1]$. While in case $n + 1$, we are actually considering $(H^{n+1})'(2x_0),\ x_0 \in [0, 1/2]$. But the measure of $x_0$ for $(H^n)' \neq 0$ is 0 by Theorem 5.4.4. So in other words, $(H^n)' = 0$ a.e. for $x_0 \in [0, 1]$, which is what we want to show.

□

Now we have proven that the variation of system parameter leads to a non-smooth reduction of the system's dynamics. A typical symptom is the joint of fibers, which indicates the "loss of words" of some length. This gives rise to homeomorphic defect since some symbolic sequences of the base system can not be matched. In the next section, we give another type of bundle, which we call "conjugate/skew bundle". Still, we vary the parameter of the system, but it only exhibits a homeomorphic change of the system. In other words, the bundle plot draws associations within the conjugate classes of the system, with no joint and no symbolic sequences lost.

### 5.4.1 Skew bundle

In the previous section, we consider the loss of symbolic sequences as we vary one parameter of the system. In this section, we are considering the bundle plot for a family of equivalent systems where all the systems are conjugate. We use the tent map as an example. It is known [2], that skew tent maps with height 1 are conjugate systems. We fixed

the heights of the tent maps $b_i \equiv 1$, while change the peak point's $x$-axis $a_i$ from 0 to 1. In particular, the family of skew tent maps is given by the following:

$$x_{n+1} = g_{a_i}(x_n) = \begin{cases} \dfrac{1}{a_i} x_n, & 0 \le x_n \le 1/2, \\ \dfrac{1}{1-a_i}(1-x_n), & 1/2 < x_n \le 1. \end{cases} \tag{5.8}$$

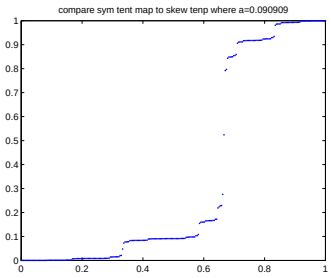
As shown in Figure 5.9, the skew bundle exhibits a discontinuity that said colloquially, reminds us of a "side-parted hairstyle", with a single hair separating two sides. The fibers don't intersect each other, which indicate there is not lost of symbolic sequences of the system. Notice that the bundle plot in Figure 5.9 is not symmetric and, unlike the case in the previous section, the vertical gaps of commuters (see Figure 5.9 top) do not mean there are homeomorphic defects. In fact, in this case the commuters are singular functions [2].

There are no joints except at $a_i = 0$ and $a_i = 1$, whose tent maps have only one leg. For $0 < a_i < 1$, the tent map changes under conjugacies, as Figure 5.9(a) and Figure 5.9(b) show two particular conjugacies, which are between the symmetric tent map and skews tent maps.

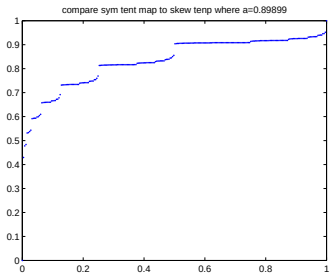
The non-symmetric structure of the bundle plot in Figure 5.9 is due to the stable fixed point's location. We solve the non-trivial fixed point from $x^* = \frac{1}{1-a}(1-x^*)$, and have

$$a = 2 - \frac{1}{x^*},\ x^* > 1/2. \tag{5.9}$$

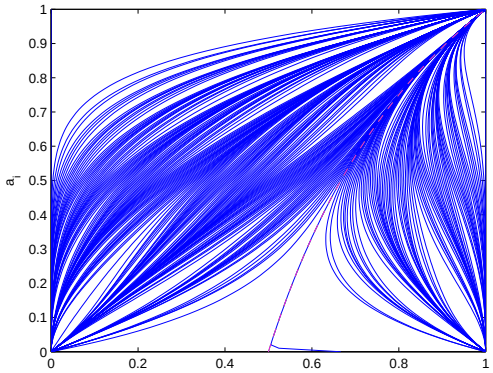
We plot it with red dash line in Figure 5.9(c), which is exactly the "single hair". Points $x < 1/2$ of $g_{a_1}$ match points $x < a_i$ of $\{g_{a_i}\}$; $1/2 < x < x^*$ of $g_{a_1}$ match points $a_i < x < x^*$ of $\{g_{a_i}\}$; $x > x^*$ of $g_{a_1}$ match those of $\{g_{a_i}\}$; While the fixed point $x^*$ of $g_{a_1}$ matches to those of $\{g_{a_i}\}$. As $a_i$ moves from 0 to 1, the stable fixed point moves from 1/2 to 1, which skews the bundle to the right.




(a) Commuter between symmetric tent map and skew(left) tent map




(b) Commuter between symmetric tent map and skew(right) tent map




(c) Bundle for Skew Tent Maps

Figure 5.9: **A Bundle Plot for skew tent maps**: Line segments show assignments from $X$ to each $Y$ of a uniform grid of sample points representing the commuters $f_{1a_i}$ from $g_{a_1}$ to $g_{a_i}$, where $g_{a_1}$ is the full symmetric ($a_1 = 1/2$) tent map and $g_{a_i}$ is skew tent maps with peak points at $(a_i, 1)$. (Top left) The commuter between the full symmetric tent map and a skew tent map with peak at $(0.090909, 1)$; (Top right) The commuter between the full symmetric tent map and a skew tent map with peak at $(0.890909, 1)$; (Bottom) the skew bundle, which exhibits a "side-parted hairstyle", with a single hair, which is the stable fixed points' location for different $g_{a_i}$.

## 5.5 Bundle for higher dimensional parameter space

Here we consider the bundle plot where we change two parameters of the system simultaneously, which gives a 3 dimensional bundle. For instance, Figure 5.10 and Figure 5.11 show bundle plots for comparing full symmetric tent map and tent maps with different fixed $b_i$ while varying $a_i$ of the tent map, with peak location $(a_i, b_i)$.

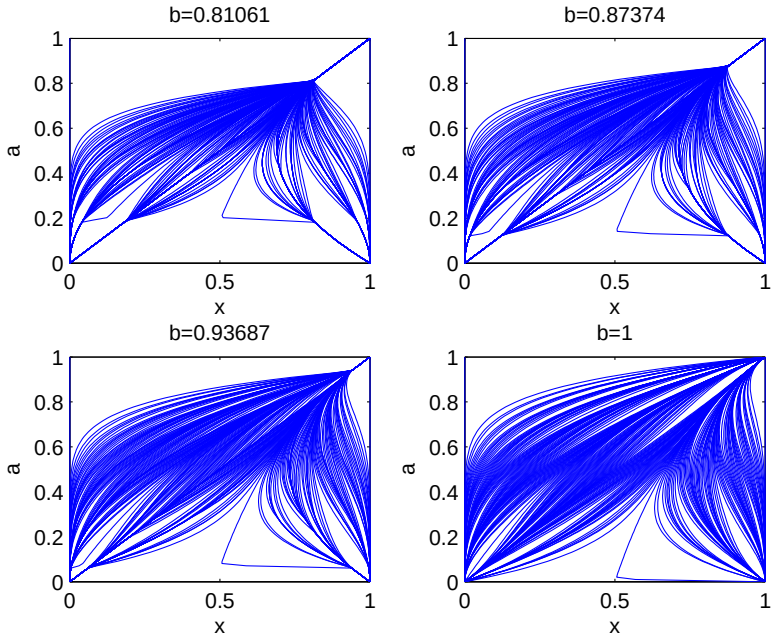



Figure 5.10: Bundle plots for comparing full symmetric tent map and tent maps with different fixed $b_i$ while varying $a_i$ of the tent map with peak location $(a_i, b_i)$.

## 5.6 Bundle plot v.s. bifurcation plot

The bifurcation plot keeps track on the changes of the stable fixed points. Bifurcation happens due to creation, or destruction, or stability changes of fixed points. On the other hand, we have "joint" where symbolic fibers merge (Figure 5.7). As a joint exists, we lose a word/words of some length. In terms of mostly conjugacy, we say that a homeomorphic defect is born. So both bifurcation and the bundle plot seem to provide information about

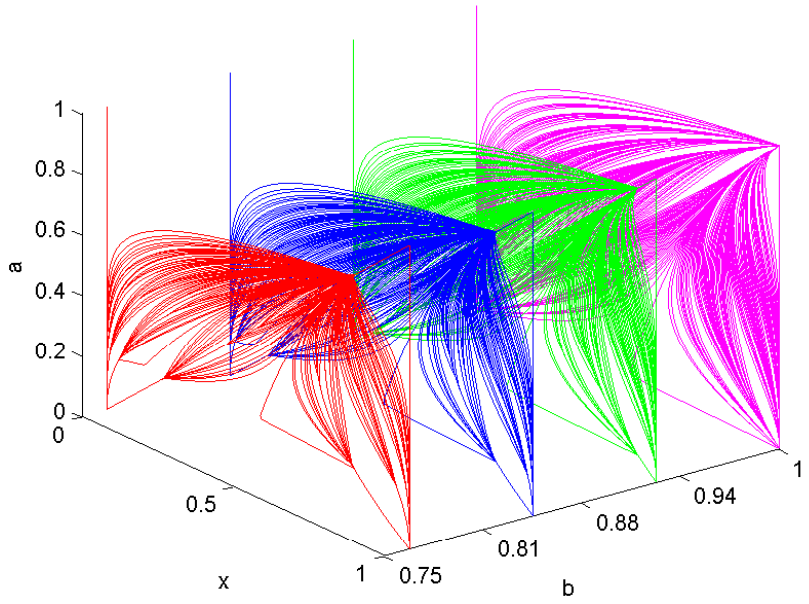



Figure 5.11: A 3-d bundle plots for comparing full symmetric tent map and tent maps with different fixed $b_i$ while varying $a_i$ of the tent map with peak location $(a_i, b_i)$.

qualitative changes of systems, but in a different perspective. In order to relate and compare these two objects, we seek to answer questions like, for which parameter does joint occur? Does a joint occur when a bifurcation does? It turns out that joint happens shortly after bifurcation. In fact, the joint describes the qualitative changes of the kneading sequence as we vary a system parameter. In this section, we use the kneading theory to study the bundle's joints.

The kneading sequence is defined as follow:

**Definition 5.6.1.** *[53] Let* $x \in [0, 1]$. *The itinerary of* $x$ *under* $g$ *is the infinite sequence* $S(x) = (s_0 s_1 s_2 ...)$ *where*

$$s_j = \begin{cases} 0, & g^j(x) < 1/2, \\ 1, & g^j(x) > 1/2, \\ C, & g^j(x) = 1/2. \end{cases}$$

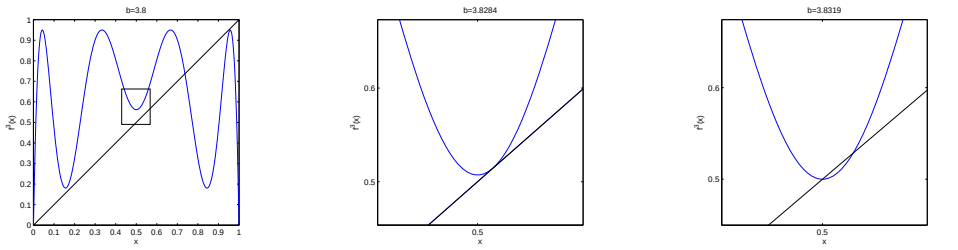
*The kneading sequence $K(g)$ of $g$ is the itinerary of $g(1/2)$, i.e., $K(g) = S(g(1/2))$.*

The Kneading theory provides a critical understanding of the dynamics of unimodal maps. The kneading sequence determines which periodic orbits exist. We refer readers to Devaney's book [17] and Milnor's [53] for more details about the Kneading theory, where the underlying system $g$ is a general unimodal map. Here we cite an important theorem, the Intermediate value theorem for kneading sequences[51], which we will use in later proof.

**Theorem 5.6.2.** *[Intermediate value theorem for kneading sequences][51] If a one-parameter family $g_b$ of continuous unimodal maps depends continuously on $b$ and the topological entropy $h(g_b) > 0$ for all $b$, then if $K(g_{b_0}) < K < K(g_{b_1})$ and $K \in \mathscr{M}$ where $\mathscr{M}$ is call the class of sequences which occur as kneading sequences of $g_b$ for all $b$, then there exists $b'$ between $b_0$ and $b_1$ with $K(g_{b'}) = K$.*

Notice if we consider a full family of system $\{g_b\}$ where $b \in [b_0, b_1]$, the class of kneading sequences $\mathscr{M}$ would consists of all possible symbolic sequences starting with symbol $C$ between $K(g_0)$ and $K(g_{b_1})$.

The opening of a periodic window, for instance the logistic map $x_{n+1} = g_{b_i}(x_n) = 4b_i x_n(1-x_n)$, can be specified by looking at the local structure of $g_{b_i}^N$. Figure 5.12 shows the graphs of $g_{b_i}^3 = g_{b_i}(g_{b_i}(g_{b_i}))$. As $b_i$ increases, the "hump" inside the box of Figure 5.12(a), which resembles the "upside-down" of the original quadratic map, grows until it is tangent to the identity line. Then for this $b_i$, period-3 window is open. As we further increase $b_i$ a little bit, the identity line cross the kneading point, which implies $g_{b_i}^3(1/2) = 1/2$. The kneading point becomes a point of period-3 and super-stable. In Figure 5.13, we plot the opening of period-3 with red horizontal line, and the period-3 kneading point with green horizontal line. We can see that there is a joint shortly after the bifurcation point, and close to the period-3 kneading point. Is this joint at the same location as the period-3 kneading point? Or more generally, what is the relationship between the joints and the periodic kneading points? The following theorem states that the joint happens shortly after bifurcation, and implies the kneading sequence is periodic.




(a) $f^3(x)$ (b) Bifurcation (c) Kneading intersection

Figure 5.12: Bifurcation points and Kneading points: (Left) the "hump" of $f^3(x)$ inside the box resembles the "upside-down" of the original quadratic map; (Middle) as $b_i$ increases, the "hump" grows until it is tangent to the identity line, which indicate the opening of period-3 window; (Right) As we further increase $b_i$ a little bit, the identity line cross the kneading point, which implies $g^3_{b_i}(1/2) = 1/2$.

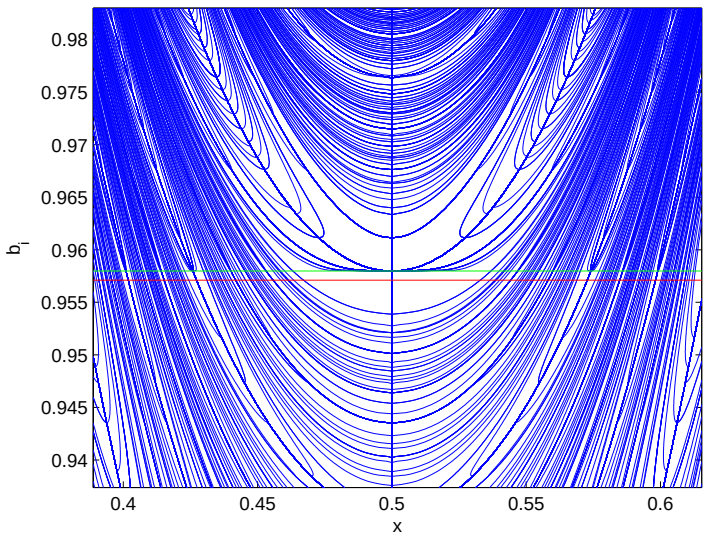



Figure 5.13: The Kneading point and Bifurcation point: we plot the bifurcation point (the opening of period-3) with red horizontal line, and the period-3 kneading point with green horizontal line. As Theorem 5.6.3 proves, the joint (green line) happens shortly after bifurcation, and implies a period-$N$ kneading point.

**Theorem 5.6.3.** *The existence of a joint implies the kneading sequence is periodic.*

*Proof.* Suppose we have a joint. By definition, WOLOG, assume we have $b_{joint}$ s.t. we lose a word $\omega = \underbrace{\times\times\times...\times\times\times\mathbf{1}}_{\text{N}}$ of length $N$ from the grammar. Assume at $b_{higher} > b_{joint}$, we have not lost the word yet. Assume

$$K(g_{b_{higher}}) = C***...***\underbrace{\times\times\times...\times\times\times\mathbf{1}}_{\text{N}}***..., \tag{5.11}$$

and suppose we have lost the word $\omega = \underbrace{\times\times\times...\times\times\times\mathbf{1}}_{\text{N}}$ of length $N$ at $b_{lower} < b_{joint}$, then either

$$K(g_{b_{lower}}) = C***...***\underbrace{\times\times\times...\times\times\times\mathbf{C}}_{\text{N}}***... = \overline{C***...***\underbrace{\times\times\times...\times\times\times}_{\text{N-1}}}, \tag{5.12}$$

or

$$K(g_{b_{lower}}) = C***...***\underbrace{\times\times\times...\times\times\times\mathbf{0}}_{\text{N}}***..., \tag{5.13}$$

If it is in case 1 with $K(g_{b_{lower}})$ given by equation (5.12), then $b_{joint} = b_{lower}$ since there is no $b'$ s.t. $K(g_{b_{lower}}) < K(g_{b'}) < K(g_{b_{joint}})$, $b_{lower} < b' < b_{joint}$and we lose $\omega$ at $b'$. In other words, $b_{lower}$ is the supreme of $b$ that loses $\omega$. In this case, $K(g_{b_{joint}})$ is periodic.

If it is in case 2 with $K(g_{b_{lower}})$ given by equation (5.13), then for $K = C***...***\underbrace{\times\times\times...\times\times\times\mathbf{C}}_{\text{N}}***... = \overline{C***...***\underbrace{\times\times\times...\times\times\times}_{\text{N-1}}}$, we have $K \in \mathscr{M}$ and

$$K(g_{b_{lower}}) < K < K(g_{b_{joint}}), \tag{5.14}$$

Then by the Intermediate Value Theorem for Kneading Sequences (Theorem 5.6.2), there exists $b'$ s.t. $b_{lower} < b' < b_{joint}$, $K(g_{b'}) = K$ and we lose $\omega$ at $b'$. So in this case, $b_{joint} = b'$ and $K(g_{b_{joint}})$ is periodic. This finishes our proof.

□

We give the bundle plot and the bifurcation of logistic map in Figure 5.14. The vertical red lines are positions of some joints, which are shortly after the bifurcation points. On the other hand, as a well-known phenomenon, “Period Three Implies Chaos” [37] proves the fact that the existence of a period-three orbit implies the existence of a large set of sensitive points. The can actually be suggested in the bundle plot. In Figure 5.13, the red line indicates the period-3 opening. Later, the green line crosses a joint. Notice there are multiple fibers join at this joint. As we discussed before (see Figure 5.6 top right), the more symbolic fibers merge at one joint, the more words we lose at one time. Through our computation, we realize this joint is a special one in that it always has more fibers joins than others, within a given grid size of $X$ space and $\{b_i\}$. In other words, when the parameter $b_i$ is at this joint (green line), the symbolic sequences are gained/lost more than others, which says the system dynamics exhibits significantly ”faster” qualitative changes at this moment.

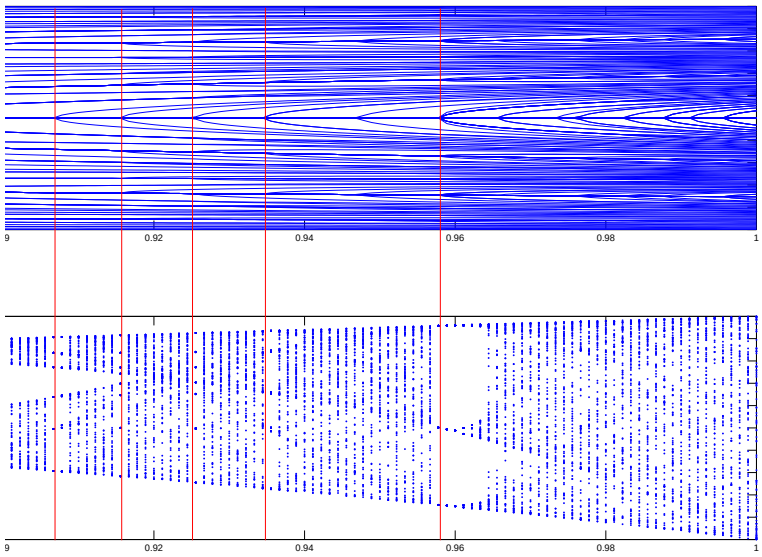



Figure 5.14: Comparison between Bundle Plot and Bifurcation Plot: the vertical red lines are positions of some joints, which are shortly after the bifurcation points.

## 5.7 Discussion

In this chapter, we consider a different perspective of studying a system's qualitative changes due to parameter variation. We provide a bundle plot, which is from commuters, to picture such evolution of the symbolic space. The joint implies qualitative changes where the kneading points become periodic. And it happens shortly after bifurcation.

The construction of bundle plots is applicable to other unimodal systems. Generally speaking, we only require the family of maps $\{g_{b_i}\}$ piecewise invertible and

- $\hat{g}_{b_i}^{-1}$ is piecewise continuous on $Y$;
- $\hat{g}_{b_i}^{-1}$ is piecewise Lipshitz continuous, with Lipshitz continuous $L < 1$.

which guarantees there is an unique commuter. We believe that the family maps $g_{b_i}$ can be generalized to be maps with positive Lyapunov exponents, since such "average" contraction of $g_{b_i}^{-1}$ can also result in a unique commuter, which gives rise to the bundle plot. On the other hand, we are also trying to extend this method to higher dimensional systems, where we assume that the dynamical systems under consideration are presented to us with a known symbolic dynamic partitioning.

We also note that the study in this chapter is based on a given system. In fact, in the construction of bundle, the base function $g_{b_1}$ can be given by a time series data. In this case, we are to provide a family of system $\{g_{b_i}\}$, and see how these systems match the original data set.

# Chapter 6

# Heart Rate Variability: Time Series Analysis on EKG as Determinism with Jump Stochastic Parameters

In this chapter, we use measured heart rate information (RR intervals) to develop a one-dimensional nonlinear map that describes short term deterministic behavior in the data. Our study suggests that there is a stochastic parameter with persistence which causes the heart rate and rhythm system to wander about a bifurcation point. We propose a modified circle map with a jump process noise term as a model which can qualitatively capture such this behavior of low dimensional transient determinism with occasional (stochastically defined) jumps from one deterministic system to another within a one parameter family of deterministic systems.

## 6.1 Introduction

Modeling the behavior of human cardiovascular system is an interesting problem which draws extensive attention from researchers. A fundamental and challenging question is

how best to provide a simplified representation of both the deterministic and stochastic aspects of heart dynamics. Suder et al. [26] recorded RR intervals while restricting the paced respiration cycle lengths above 8 s, from which they observe that heart rate variability obeyed a dynamic rule that can be expressed by a one-dimensional, nonlinear circle map. Later, Jason et al. [27] showed that even during spontaneous breathing (with subject at rest), the angle component of RR-interval still has a highly deterministic structure after filtering out low frequency components. Shiau et al. [28] argued that the deterministic characteristic could result from sympathetic activation and thermo-regulation, which are primarily evidenced in the low frequency component of the RR-interval signal. They used a simple nonlinear noise-reduction method [29] to remove the high frequency component and used the next angle map to reconstruct a deterministic attractor.

Our approach is based on Shiau's work, but we use cubic smoothing spline to remove the high frequency component of RR-intervals time series data. This filter approach (via spline) was selected (1) for ease of implementation[1], but also (2) because the smoothing spline enforces a continuity, consistent with our expectation that heart rate variability should vary continuously . We find the resultant data can be easily related to a modified sine circle map, but with an interesting noise behavior which we describe as a stochastic parametric factor with persistence. Here we define 'persistence' as piecewise constant in time, with occasional random jumps. The stochastic behavior that we find is similar in character to that identified in Lerma's recent study on the stochastic aspects of Cardiac Arrhythmias [30], which argued that in the neighborhood of bifurcation points, the fluctuations induced by the stochastic opening and closing of individual ion channels in the cell membrane results in membrane noise that may lead to randomness in the observed dynamics of cardiac rhythm systems. In [31], Kuusela et al. give a simple one-dimensional Langevin-type stochastic difference equation which can model the heart rate fluctuations in a time scale from minutes to hours. The similarity between to our work is that both

[1]Smoothing spline requires choice of a single smoothing parameter, while Schreiber's noise-reduction methods requires optimizing over a pair of parameter values.

provide a stochastic model which aims at uncovering the interaction of determinism and stochastic control of cardiac dynamics. However, we focus our study on the low frequency component of RR-intervals data, which relates to sympathetic and vagal activity of heart, to reveal the stochastic jump with persistence around a bifurcation point.

The paper is organized as follows: Section 2 discusses our filtering approach and our decision to focus on the low frequency (LF) component of the Heart Rate Variability signal (HRV) in the phase space reconstruction. Section 3 discusses the phase reconstruction from the ECG data and discusses the complex behaviors observed in the data. In Section 4, we give a reasonable model to simulate these observed behaviors, with conclusions in section 5.

## 6.2 Spectral Components of HRV

Over the past two decades, the general body of research has recognized a significant relationship between the autonomic nervous system and cardiovascular mortality, including sudden cardiac death. Although cardiac automaticity is intrinsic to various pacemaker tissues, heart rate and the rhythm are largely under control of the autonomic nervous system [32]. Different frequency ranges of the HRV have been related to various physiological phenomenons [32]. Studies of spectral components of short term recordings of HRV show that the efferent vagal activity is a major contributor to the high frequency (HF) component, while the LF component is considered as a marker for sympathetic modulations, with some studies also suggesting that LF reflects both sympathetic and vagal activity [32].

In Suder's experimental method [26], they obtain a one-dimensional, nonlinear deterministic observable from the HF component by restricting the respiration-cycle to greater than $8s$. Their theoretical foundation is that controlling respiration can induce an increase in the HF signal. After high pass filtering, they identify a one-dimensional deterministic process. Similarly, Janson's work [27] obtains a low-dimensional structure by extracting

the HF component of HRV, but without control on respiration. In contrast, Shiau [28] obtains a one-dimensional deterministic process from the LF component under spontaneous breathing conditions. Regardless of the frequency domain being studied, there is evidence showing that low-dimensional deterministic processes describing the autonomic nervous system can be observed, with the low-dimensional models providing insight into heart rate dynamics [33, 34]. In this paper, we focus on the LF component while generating a one-dimensional map, which corresponds to sympathetic and vagal activity [32]. A primary reason for choosing to analyze the LF component is that it is easier to reduce the effects of noise introduced by measurement error.

## 6.3 Electrocardiogram Data

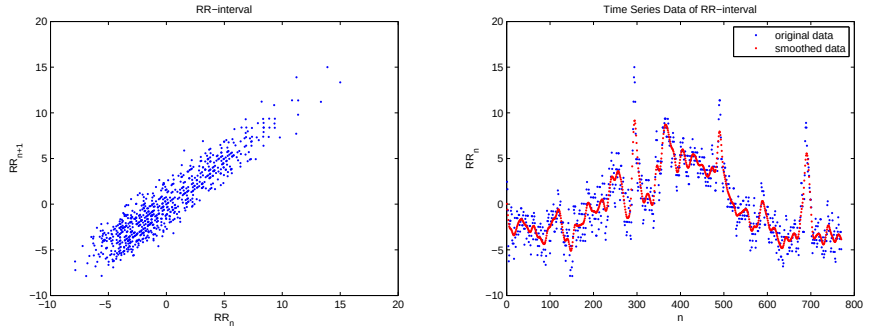
From an open source repository (physionet.org), we obtained RR-intervals data taken from volunteers who were supine and asked to breathe at a fixed rate of 0.25 Hz for 10 min. (Data originally from [35].) The two-dimensional embedding of the raw data, shown in Figure 6.1(a), has no apparent low dimensional structure. To identify a low-dimensional model, we apply the following sequence of processing steps: First, we filter out the HF component by applying a cubic smoothing spline interpolation to the RR-intervals data, as illustrated in Figure 6.1(b), an alternative filtering technique to that of [29].

As second step, for each filtered data point, we compute an angular coordinate representation $\phi_n$ measured relative to the centroid of the data set, $\overline{RR}$ :

$$\phi_n = \frac{1}{\pi} \arctan\left(\frac{RR_{n+1} - \overline{RR}}{RR_n - \overline{RR}}\right) \bmod 1. \tag{6.1}$$

We then construct a time delay embedding of the angular coordinates to produce the "next angle map."

Figure 6.2(a) shows the time delay embedding representation of the next angle map (6.1) data, which appears to be reasonably well described by a one dimensional curve, but




(a) RR-interval

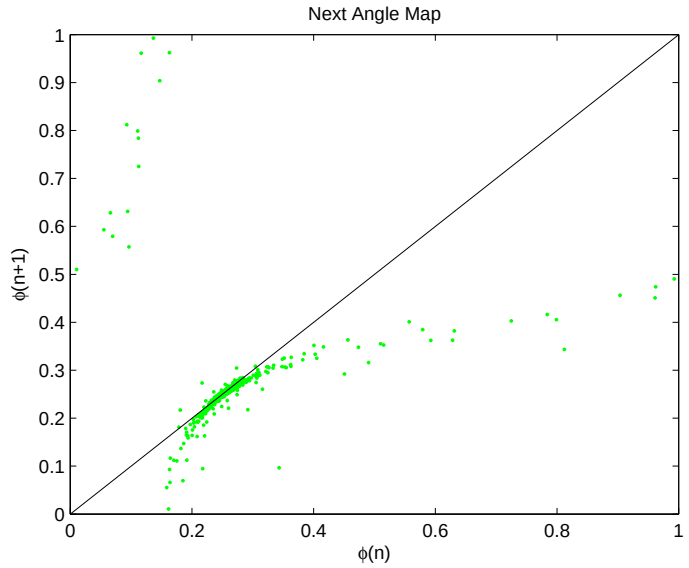
(b) RR-interval after cubic smoothing spline

Figure 6.1: ECG Data. (Left) Time delay embedding of raw RR-interval data has no apparent low-dimensional deterministic structure. (Right) A cubic smoothing spline is applied to filter out the high frequency component of series data of *RR*. The remaining low frequency component would lead to a low-dimensional structure after we extract the angle part in polar coordinates.

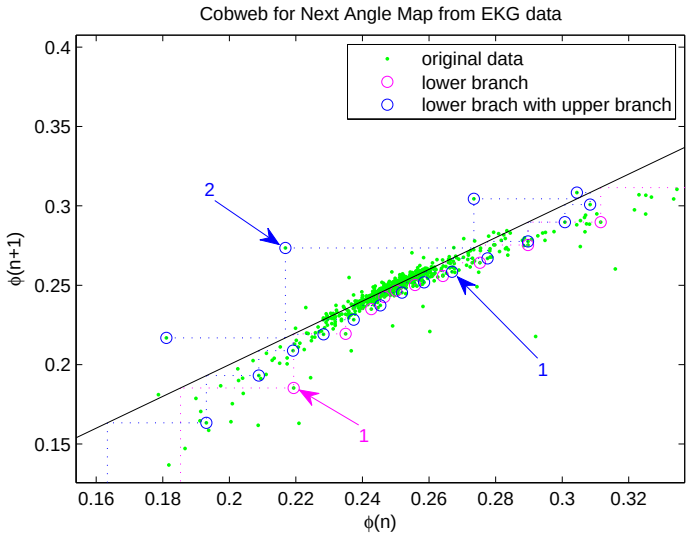
with data lying nearly on, but both above and below, the identity line on that graph. We find that there exist two typical cases for the portion of the trajectory that lies near the intersection between identity line and the data:

- Case 1: The trajectory evolves along a lower branch, below the identity line, resulting in decreasing values;
- Case 2: The trajectory evolves along an upper branch, which intersects the identity line with two fixed points, generating increasing values.

Figure 6.2(b)) illustrates these two behavior. We observe from our data that the trajectory appears to persist along a branch (either upper or lower) for several iterations before "jumping" to the other branch. *Of special importance is that although the full data set may appear as a "cloud" as if it were noisy observations, any particular trajectory follows a smooth (seemingly deterministic) path for several iterates. The noisy cloud appearance results from overlaying numerous deterministic partial trajectories, each with slightly different parameters.* We describe this behavior as "persistence," and view it as parametric




(a) Two-dimensional embedding of angles




(b) Cobweb follows different stochastic branches

Figure 6.2: Next Angle Map data. (a) A time delay embedding of data computed from next angle map (6.1) indicates that a one-dimensional representation may be a reasonable approximation. (b) Cobweb along typical data trajectories. Highlighting three trajectory segments near the line $\phi_{n+1} = \phi_n$. We note that the trajectory sometimes travels following a lower branch (below the line and decreasing — case 1) while at other times, it follows an upper branch (above the line and increasing— case 2).

noise. Describing this phenomena as *noise* is not meant to imply that the underlying process is truly stochastic, but is simply a recognition that even if it is deterministic, we are not modeling the complex control system behavior, which would be affected by the body condition, circumstance, emotion,etcetera.

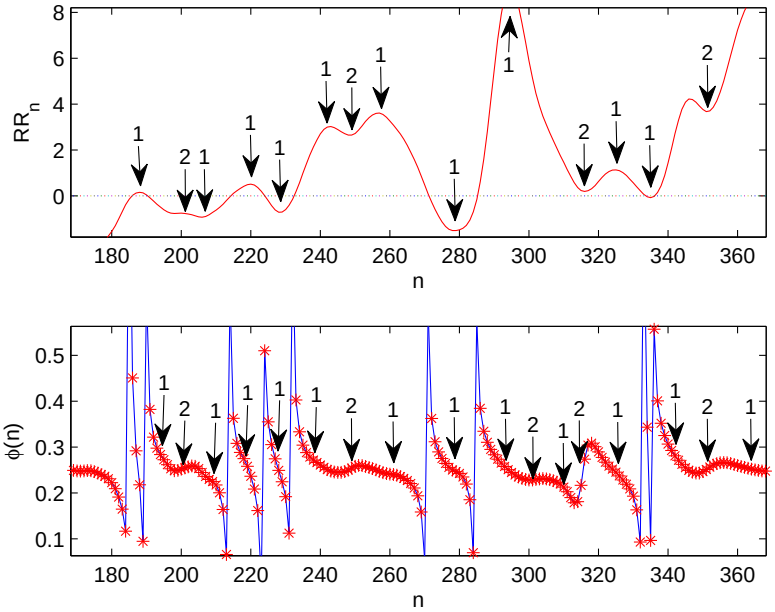



Figure 6.3: Data trajectories. (Top) Demeaned *RR* data ($RR_n - \overline{RR}$). We label '1' to the case when the trajectory follows the lower branch, and '2' to the case when the trajectory follows the upper branch. (Bottom) Series data of $\phi$. Decreasing $\phi_n$ relates to '1'; increasing $\phi_n$ relates to '2.' For both graphs, the data is discreet, with the curve drawn for clarity of illustration.

For convenience, the following discussion uses '1' as shorthand notation for case 1 trajectories, and '2' for case 2. In 6.3, we plot the sequence data of $RR_n$ and $\phi(n)$. From (6.1), we can understand the relationship between these data for each monotone interval of $\phi(n)$ :

Case 1 — $\phi(n+1) < \phi(n)$, or $\phi(n)$ is decreasing, with data below the identity line of Fig. 6.1(b) . The corresponding interval for $RR_n$ would be the one where $RR_n$ is concave down above the average line or concave up and below average.

Case 2 — $\phi(n+1) > \phi(n)$, or $\phi(n)$ is increasing, data above the identity line in

Fig. 6.1(b), corresponding to $RR_n$ intervals of downward cavity when below average or upward concavity when above average.

These observed behaviors lead to a natural symbolic labeling and transition graph representation using symbols '1' and '2.' In this context, the "persistence" of the system to stay on a particular branched for several iterates would be evidenced in the transition diagram (Figure 6.4) as $p_{ii} > 1/2$, with larger values indicating greater persistence. We remark that our data and case description admits some ambiguity that might appear in longer dataset: (1) it is possible that the trajectory following some branch may experience a parameter 'jump', but land on a new branch that is the same *case* as before the jump, and that jump would not result in a transition to the other node, (2) the trajectory might jump to a Case 2 branch (with a stable fixed point) while the system state is above the fixed point, in which case the trajectory would decrease toward the stable fixed point, with this behavior essentially indistinguishable from a Case 1 trajectory.

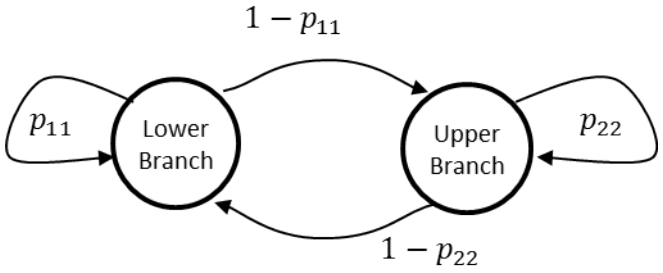



Figure 6.4: Transition relationship between upper and lower branch trajectories, as observed in the data. When $p_{ii}$ is large, the system will tend to "persist" along a either the upper or lower branch.

In this paper, we are not trying to provide a physiological explanation for this behavior. Rather, we simply have *observed* the phenomena. In the next section, we build a model based on the circle map [36, 38] which can mimic the features discussed above through representation as a stochastic process with an interesting form of 'noise' in that process.

## 6.4 The Circle Map

*The circle map* is a generic term describing a family of dynamical systems whose state space can be interpreted as angles of a circle. A simple example of circle map using the modulo function and is given by $\phi_{n+1} = f(\phi_n) = \omega + \phi_n \ (mod\ T)$, where $\omega$ and $T$ are constant. A second example is the *sine circle map*, which was introduced by Kolmogorov and well studied by Vladimir Arnold [36, 38], is defined by

$$\phi_{n+1} = (\phi_n + \omega + k \sin \phi_n) \ modulo\ 2\pi \tag{6.2}$$

where $\omega$ and $k$ are constants. Figure 6.5 plots the Sine Circle Map with $\omega = 5$ and $k = 1.25$ as the red curve, while the plotted points (green) are from the data. The circle map provides a prototypical model for systems that are controlled by two pacemakers, with the term $k \sin \phi$ describing the effect of the nonlinear oscillator coupling [38], which is why it has been previously studied as a model for the heart [36].

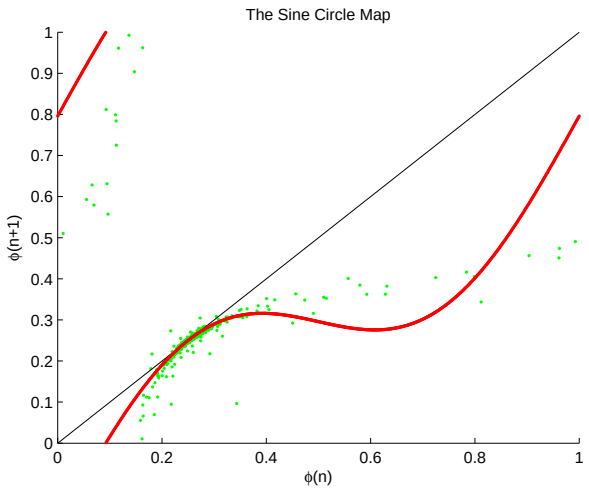



Figure 6.5: The Sine Circle Map. A scaled version of the Sine Circle Map (red) overlaid on our data from the two-dimensional embedding of $\phi_n$ (green), with $\omega = 5$ and $k = 1.25$. The circle map is not intended to interpolate the points, but the family of such curves does provide *qualitatively* similar dynamics. (color online)

Figure 6.6(a) shows the sine circle map with $\omega = 5$ and $k = 1.2, 1.275$ and $1.35$. When

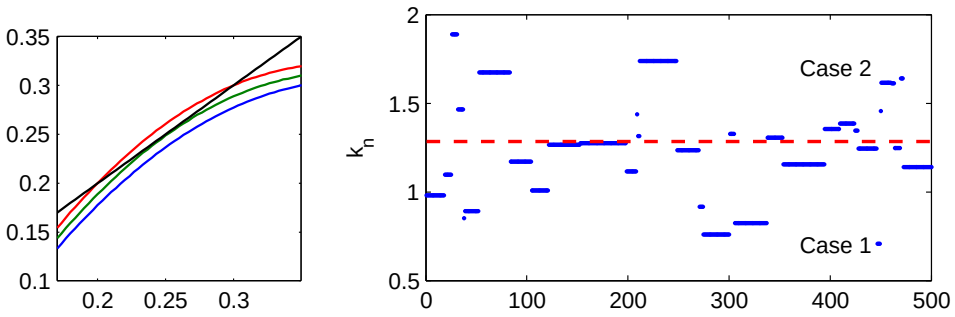
$k = 1.2$ trajectories would mimic the Case 1 trajectories of our data. As $k$ is increased, the map gets closer to the identity line, yields trajectories with slow passage between the identity line and the curve [39], a ghost of the fixed point associated to the saddle node bifurcation that occurs at approximately $k = 1.283$. When $k = 1.35$, the curve has crossed through the identity line. Two fixed points are created, one is stable, the other is unstable. Trajectories on this map would mimic Case 2 behavior during the transient approach toward the stable fixed point.

In order to simulate similar behaviors of our ECG data using the circle map, we proceed as follows: We assume that $k$ is piecewise constant (the persistence) with changes to $k$ occurring as Poisson arrivals. When $k$ changes, the new value for $k$ is chosen as an i.i.d. variable from a Gaussian distribution:

$$k_n = \begin{cases} 1.1938 + 0.25\epsilon_n, & p_n < p_c, \\ k_{n-1}, & \text{else}, \end{cases} \tag{6.3a}$$

where $0 \leq p_c \leq 1$, $p_n \sim U[0,1]$, and $\epsilon_n \sim N[0,1]$. Critical value $p_c$ governs how often $k_n$ changes its value. (The Gaussian scale parameters as well as the choice $p_c = 1/17$ were selected heuristically to provide a good visual match to the observed data trajectory.) A visualization of such $k_n$ is given in Figure 6.6(b). Whenever $k_n < 1.283$ (below the red horizontal line in Figure 6.6(b)) the dynamics are similar to case 1 of our data. For $k_n > 1.283$, the behavior is like case 2.

Figure 6.7(a) shows an example of a typical trajectory generated by our model. The cobweb around the crossover section is plotted in Figure 6.7(b) for a few trajectory segments, illustrative reasonable qualitative agreement with our data. We note that long sequences of '2' behavior results in the nearly horizontal part of $\phi_n$ in Figure 6.7(a). We similarly give a graph representation for the modified circle map in Figure 6.8. For our model, we can compute the transition probabilities as follows: Define a $k$-refresh as the condition that $k_n \neq k_{n-1}$, the situation where the system has jumped to a new parameter value. Then the




(a) zoomed circle map (b) $k_n$

Figure 6.6: The Sine Circle Map. (Left) Zoomed to the portion of the map where saddle-node bifurcation first occurs. We fix $\omega = 5$, and show the curves for $k = 1.2, 1.275, 1.35$, with the identity line $y = x$ plotted in black. For $k = 1.35$, (the upper red curve), the curve has pushed through the identity line to create a stable fixed point. (Right) A simulation of sequence $\{k_n\}$, using the stochastic process defined in (6.3a). Whenever $k_n$ is above the red horizontal line, the sine circle map is behaves similarly to case 2 of our data; whenever $k_n$ is below the red horizontal line, the sine circle map is behaving similarly to case 1 of our data;

conditional probability that refresh results in an "upper branch" (Case 2) behavior is given by

$$p_u := P(k_n > 1.283 \,|\, k_n \neq k_{n-1}), \tag{6.4}$$

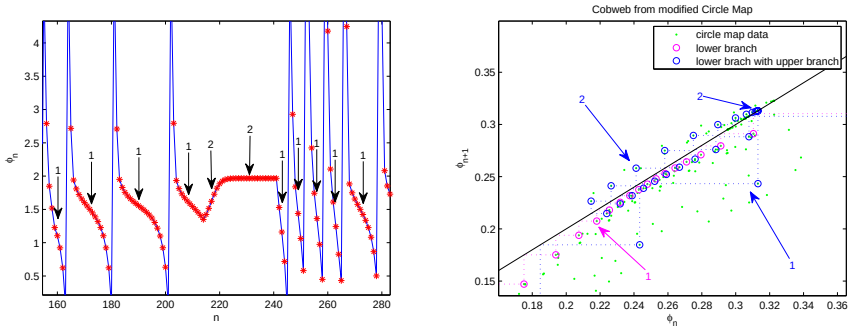
is independent of system state and is easily computed from the normal distribution function. For the particular parameters described above, this yields $p_u \approx 0.361$. The transition from '1' to '1' is due to either no refresh of parameter or a refresh, but with new $k_n < 1.283$, so that

$$p_{11} = (1 - p_c) + p_c(1 - p_u) \approx 0.979.$$

The other transition probabilities can be easily computed in a similar fashion.

## 6.5 Conclusion and Discussion

In summary, we apply cubic smoothing spline to the RR-intervals data to remove the HF component, and then we extract the angles coordinate. A delay embedding of the data in-




(a) Simulated Data using Modified Circle Map (b) Cobweb From Modified Circle Map

Figure 6.7: The Modified Circle Map. (Left) For our modified circle map, we label 1 to the interval of $\phi_n$ when the model's curve is below the identical line; 2 to the interval when the model's curve is above the identical line and intersects it with two fixed points. (Right) Cobweb representation of trajectory segments near the intersection of the circle map and the identity line.

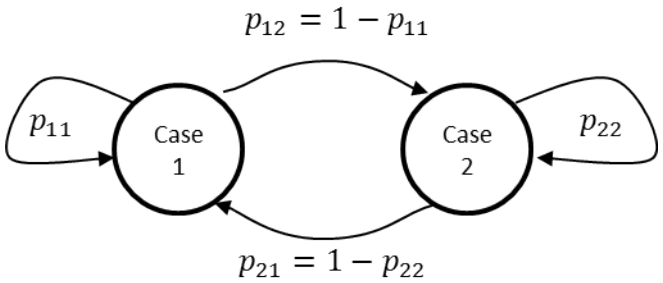



Figure 6.8: Transition relationship between Case 1 and Case 2 trajectories in Modified Circle Map, with transition probabilities labeled.

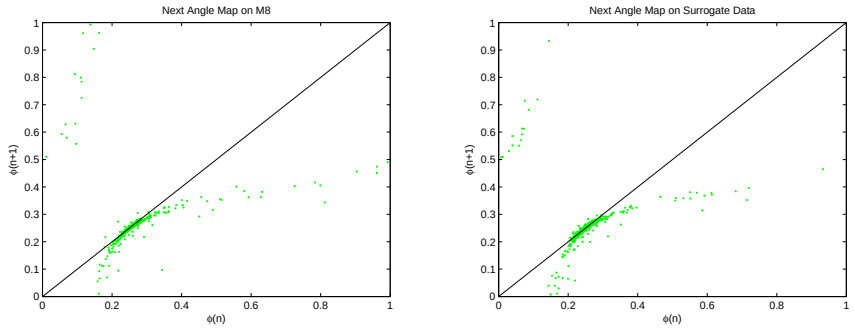
dicates that a low-dimensional representation is reasonable, but analysis of structure in the time series indicates that a better representation can be achieved by viewing the heart rate as driven at a parameter point near a saddle node bifurcation. Stochastic fluctuations (with persistence) cause the system to wander back and forth through that bifurcation. To model this behavior which derives from combination of deterministic and stochastic factors, we propose a stochastically perturbed modified circle map. The model shows qualitative agreement with the two-dimensional embedding of angles from heart rate data in the sense of simulating the stochastic behaviors. We remark that our choice to model the stochastic behavior of parameter $k$ as piecewise constant with Poisson arrival of jumps was based on the simplicity of the approach. It is certainly reasonable to also consider the case where $k$ experiences small fluctuations (rather than constant behavior between jumps), in which case a standard jump-diffusion model might be appropriate. However, our noisy data did not appear to be sufficient to resolve that low level diffusion.

In this paper, we do not address the question why this modified circle is "good" to model this heart rate data, leaving as open question the issues of what physiological effects might generate such behavior. Moreover, the question of "goodness," from a mathematical framework, requires that we have some way to quantify deviation of model and system. The degree to which a "toy model" might be representative of a more complicated system is a fundamental issue from dynamical systems that is not easily resolved. For example, comparisons between dynamical systems based on $L^p$ spaces may fail to be a good judgment if the systems turn out to be chaotic. Mostly Conjugacy [2] appears to be a promising method to address this problem, because it compares dynamical systems in a way of judging the quality of "matching" by looking at their topological difference (homeomorphic defect). Additionally, our choice of specific model parameters was based primarily on a qualitative assessment. Using tools from the theory of commuters [2], tailored to this stochastic setting, may allow for a reasoned way to choose parameters to best match system dynamics. Comparing the models and the heart rate data by using Mostly Conjugacy method

will be our next work on this topic. We suggest that applying these parameter estimation techniques to data collected during regular physical examination of heart could provide a new method for automated "change detection" with respect to cardiac function, where one might hope that detection of such changes in dynamics might have clinical relevance.

## 6.6 Appendix: Test for Determinism using Surrogate Data.

A primary concern when conducting analysis of processed data is that observed structures may be an artifact of the *processing,* with no relevance to the true phenomena under study. To show that the low dimensional structure of successive angles does not simply result from the low-pass filtering, we compare our observations with surrogate data created by linear Gaussian stochastic process which preserves both the spectrum and the histogram of the empirical RR-interval data [41]. We process that surrogate data using the same filtering methods, as described in Section 6.3. Figure 6.9 plots the Angle map based on data from the experimentally measured data [35] compared with processed surrogate data.




(a) Angle map from measured data (b) Angle map from surrogate data

Figure 6.9: Test for nonlinearity. (Left) Angle map based on data [35] after low-pass filtering. (Right) Angle map based on surrogate data using linear Gaussian stochastic process with the matching amplitude spectrum and distribution as the data.

In the "eyeball" metric, treating the data as a cloud of points in $\mathbb{R}^2$, we observe lit-

tle qualitative difference between measured and surrogate data. However, as a dynamic process, we see distinguishing features. As described in Section 6.4, in our data, we observe that trajectories (in delay coordinates shown in Figure 6.9) that are near the line $\phi(n) = \phi(n+1)$ that either track upward (when above the diagonal) or downward (when below) with trajectories only infrequently "crossing" the line. Examining the surrogate data, generated from a Gaussian stochastic process with the matching spectrum and distribution as the data, we find *frequent* crossings of that main diagonal after processing the surrogate data, as represented by the illustration in Figure 6.10 .

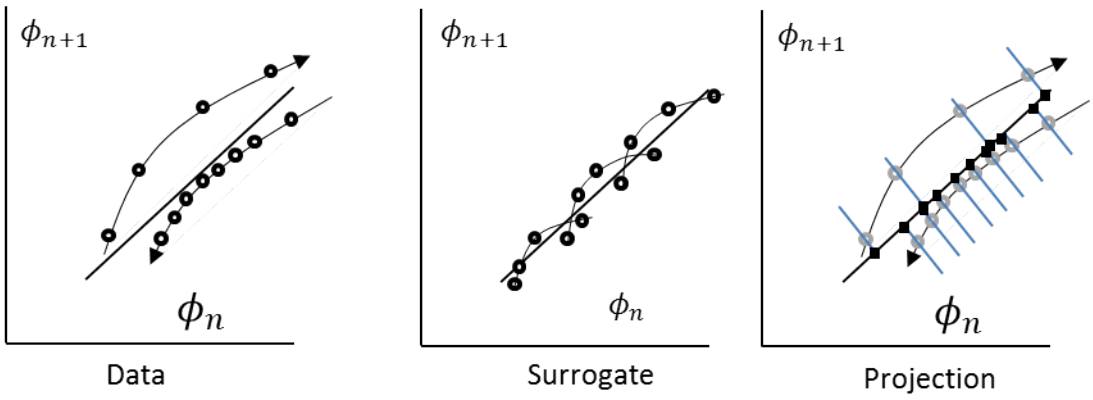



Figure 6.10: Data vs. Surrogate trajectories. Cartoon illustration of (L) Data trajectories, and (C) Surrogate trajectories, illustrating the difference in character. To build a test statistic, (R) we project data onto the main diagonal, and compare distributions.

In particular, if focus on that portion of trajectories lying near the diagonal (in delay space) where $\phi_{n+1} = \phi_n$, and project data onto that line (as illustrated Figure 6.10(R)). We test as follow: We take as null hypothesis that the time series of RR-interval is generated by a linear stochastic process. Let sampled data $c$ be the projection of the selected (select the points around near the main diagonal in phase space) onto the diagonal. We then generate 10 surrogate data sets $\{s1, s2, ...s10\}$ for the *RR* interval data, where we model as a guassian process with the same power spectrum and distribution as the empirical RR-interval data,

using method $iAAFT$ [42]. Let $cc = [s1\ s2\ ...\ s10]$ bet the concatenation of all 10 surrogate data sets. Then $cc$ can be viewed as the "average" behavior of surrogate data sets. We perform a $Kolmogorov - Smirnov\ test$ on $c$ and $cc$. The test results in the rejection of the null hypothesis at the 5% significance level, with $p - value = 3.4037e - 004$.

Our test is meant to establish that the low dimensional structure of RR-interval data does not come from a Gaussian stochastic process. In some sense, this test seems to argue for greater structure to the underlying process, reasonably captured by our model of deterministic transient behavior, though the formal result is simply rejection of that null hypothesis.

# Chapter 7

# Uncovering the Dynamics of Heart Rhythms by Graph Matching Method

In this chapter, we continue our study on heart rate variable data analysis, but focus more on uncovering the dynamics of the heart rhythms. Based on the low dimensional structure that we obtain from a person's 24 hour-record, we use Hidden Marko Models (HMM) to characterize the pattern of heart rhythms. Then we apply the graph matching method to compare the models to investigate the difference to a reference model.

## 7.1 Introduction

An interesting question in Cardiac Arrhythmias is how do we model and study the dynamics of the cardiac system. Once the raw ECG or RR intervals data are given on different dates, how can we tell the person's cardiological conditional changes from day to day. The heart rate rhythm is largely under control of the autonomic nervous system, which fluctuates due to both intrinsic and extrinsic factors. Thus a first concern is whether we can have a simplified representation of both the deterministic and stochastic aspects of the cardiac system. Our previous work [43] suggested that a low dimensional nonlinear map, which exhibits a stochastic parametric jumps with persistence, can be constructed within the low

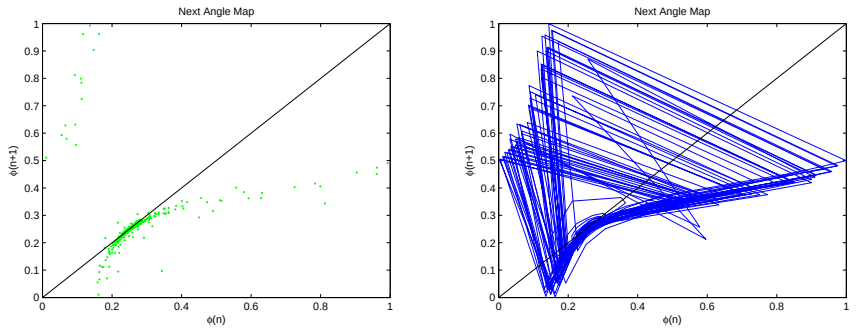
frequency component of RR-interval time series data. This complicated stochastic behavior occasionally pushes the system across a bifurcation back and forth. As the second trend, which is the primary interest of this work, we produce reasonable chronological models and compare them where the "dissimilarity" can quantify how much the systems fail to be the same within dynamical systems perspective.

In our recent study [44], we built Markov models from deterministic systems by constructing a weighted directed graph with equal-distance partitioning the invariance set. Then we used Earth Mover's Distance (EMD) [21] to compare associated graphs. The matching matrix is shown to coincide with the solution of the corresponding Kantorovich problem. In addition, the dissimilarity measures the distance between systems, in the sense of "distance" from being conjugate [2, 11]. But the problem of direct application of the method is that, in theory, we need a sufficiently large number of partition number so that the resulting symbolic representation can well translate the original dynamics. But the cost would be the computation time. On the other hand, the particular stochastic behavior of the low dimensional representation of HRV that we derived may not be appropriate to be modeled by a first order Markov model. As was consider by Silipo R. et [45, 46], where the minimum order of the Markov model charaterizes the complexity of the underlying nonlinear structure of RR-interval signals, which is typically not first order.

We therefore employ Hidden Markov Models (HMM), which is a "hidden" first order Markov model that governs the observable signals, so that the resulting "dissimilarity" from Graph Matching Method applied on HMMs enables us to interpret the dynamical variance due to actual changes, hopefully.

## 7.2 Low Dimensional Deterministic Process of Electrocardiogram Data

Given a time series data of RR-interval, we develop a method to reduce the data to a low dimensional curve in our previous work [43]. We first use a cubic smoothing spine to extract the low frequency component of RR-interval data, which corresponds to sympathetic and vagal activity [32], then we compute the angular coordinate via the next angle map. We observe that the time delay embedding representation of the next angle map data suggests a low dimensional "deterministic" curve", which we show in Figure 7.1(a). If we connect all the successive data points in Figure 7.1(a), we have another Figure 7.1(b), which suggests a small circle is embedded in a large circle in the phase space.




(a) Low dimensional representation (b) Low dimensional representation

Figure 7.1: Low dimensional map obtained from RR-interval data

We also observe a phenomenon that, the trajectories lie nearly on, but both above and below, the identity line on the figure. We find that there exist two typical cases for the portion of the trajectory that lies near the intersection between identity line and the data:

- Case 1: The trajectory evolves along a lower branch, below the identity line, resulting in decreasing values;
- Case 2: The trajectory evolves along an upper branch, which intersects the identity line with two fixed points, generating increasing values.

In addition, the trajectory appears to persist along a branch (either upper or lower) for several iterations before "jumping" to the other branch. This behavior is a common nonlinear phenomena called "intermittency" [42, 38]. We roughly sketch the idea of case 1 and case 2 in Figure 7.2.

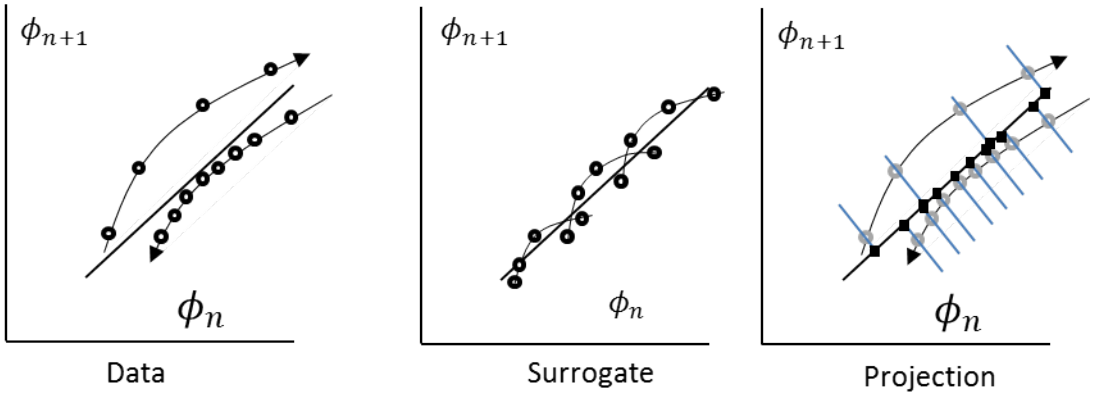



Figure 7.2: Stochastic behavior

## 7.3 Compare Dynamical Systems via Graph Matching Method

The idea of comparing dynamical systems via graph matching method [44] is: 1. partition the phase space into finite subintervals/nodes; 2. construct a directed weighed graph based on the record of the subintervals/nodes that a sufficiently long trajectory visits; 3. compare graphs by Earth Mover's distance (EMD) [21], which is essentially comparing the "approximated" invariant density distributions of the systems. The resulting matching matrix, which we have shown, is an "relaxed" matrix to the permutation matrix for the isomorphic graphs constructed from conjugate systems. The advantage of such "relaxation" allows us to numerically compare non-perfect conjugate systems where the "dissimilarity" implies how far the systems are from being homeomorphic.

Direct application of the Graph Matching method is problematic, since the equal-distance partitioning for the low dimensional map from the empirical RR-interval data might not produce a Markov models. As a matter of fact, even for deterministic map, if the underlying partitioning is not Markov, the symbols generated may not necessarily describe the evolution of signals using the right order of memory of the Markov model. We can use a sufficient large number of partitioning number to obtain an "approximated" Markov model [44]. However, this would highly reduce the efficiency.

We thereby employ Hidden Markov Models to address this issue.

### 7.3.1 The Hidden Markov Model

A hidden Markov model (HMM) [47] is a first order Markov model governing unobserved hidden states, while generating sequences of symbols which are observable. HMM is good at classifying and predicting signals. It has a wide range of application in human behaviors like speech recognition, finance like stock market forecasting and credit card fraud detection, bioinformatics like DNA sequence analysis.

A HMM is characterized by the following:

- 1: number of hidden states $N$. We let $S(t) = \{S_1(t), S_2(t), \cdots, S_N(t)\}$ to be the set of hidden states at time $t$. Note that each of the hidden states in $S$ is a first order Markov process for any time $t$;

- 2: number of observation states $M$. We let $O = \{O_1, O_2, \cdots, O_M\}$ to be the set of observation states. Note that each of the observation states in $O$ is not necessary a first order Markov process;

- 3: the hidden states transition probability matrix $A = [a_{ij}]$, where $a_{ij} = P(S_j(t+1)|S_i(t))$. Note that $\sum_{j=1}^{N} a_{ij} = 1,\ 1 \leq i \leq N$;

- 4: the emission matrix $B = [b_{ik}]$, where $b_{ik} = P(O_k(t)|S_i(t))$ for $1 \leq i \leq N,\ 1 \leq k \leq M$. Note that $\sum_{k=1}^{M} b_{ik} = 1,\ 1 \leq i \leq N$;

- 5: the initial hidden states probability $\pi = [\pi_i]$, where $\pi_i = P(S_i(0))$. Note that $\sum_{i=1}^{N} \pi_i = 1$;

We let $\lambda = (\pi, A, B)$ to be the parameter of a hidden Markov model. Given a sequence of observations $O_1^T$ of length $T$, the goal is to find the best $\hat{\lambda}$ to maximize $P(O_1^T|\lambda)$ to characterize the signal $O_1^T$. For the resulting Markov model, we denote its stationary distribution for the hidden states by $\Pi = \{\Pi_1, \Pi_2, \cdots, \Pi_N\}$, where $\Pi = \pi A$. And we can obtain the stationary distribution for the observation states that are driven by the hidden dynamics by $P(O) = \{P(O_1), P(O_2), \cdots, P(O_M)\}$, where $P(O_k) = \sum_{i=1}^{N} \Pi_i b_{ik}$.

As we mentioned before, it is inappropriate if we directly model the observations by a first order Markov model and obtain the "approximated" stationary distribution by looking at the dominating eigenvector, which is different from $P(O)$, since the underlying structure of the signals may not be a first order Markov process. But $P(O)$ can well describe the stationary distribution of observation, since it is obtained from the linear transformation of the stationary distribution of hidden states.

A widely used method to compare HMM, or more precisely the stationary distributions $P(O)$, is to use Kullback-Leibler (KL), which is a measure based on the relative entropy between distributions. But KL does not define a metric since it fails to satisfy the triangle inequality. Here we apply the Earth Mover's Distance (EMD), which is the solution of the discrete case of Kantorovich's problem. As EMD is also recognized as Wasserstein distances [13], which is a well defined metric when we consider the underlying topology is a $L^p$ space with $p \geq 1$, the resulting dissimilarity implies the distance from being conjugacy within the dynamical systems framework [44].

In particular, EMD is a bipartite network flow problem which can be formalized as a linear programming problem [21]: Let $I$ be the set of supplies, $J$ be the set of consumers, and $c_{ij}$ be the cost to ship a unit from $i \in I$ to $j \in J$. We want to find a set of flow (matching matrix) $f_{ij}$ to minimize the overall cost:

$$\sum_{i \in I} \sum_{j \in J} c_{ij} f_{ij},$$

subject to the constrains:

$$f_{ij} \geq 0,\ i \in I,\ j \in J$$
$$\sum_{i \in I} f_{ij} = y_j,\ j \in J$$
$$\sum_{j \in J} f_{ij} \leq x_i,\ i \in I$$

where $x_i$ is the total supply of supplier $i$, and $y_j$ is the total capacity of consumer $j$.

Here we consider the cost function $c_{ij} = |x - y|^2$, with the metric to be the regular 2-norm. In this case, the optimal matching/flow is the gradients of the convex functions, which are monotone and orientation preserving. On the other hand, we consider the matrix norm to be the induced 2 norm, since with induced norm, the matrices can be viewed as operators. And if two systems are conjugate, they are isomorphic as linear transformations if we restrict the operators/matrices to their eventual range [18].

## 7.4 Analysis of the Heart Rate Variability for 24-Hour RR-interval data

24-hour of RR-interval records are analyzed from randomly selected subjects in the category "MIT-BIH Normal Sinus Rhythm Database" in PhysioNet.org [48]. The subjects in this category are healthy and had no significant arrhythmias; they include 5 men, aged 26 to 45, and 13 women, aged 20 to 50. The ECG records are sampled at 128 Hz with 12 bits resolution. We pick two subjects with full analysis process which discussed before to show the performance, comparing to some widely used indices as reference. RR-interval record from subject 16265 starts from 8:04 to 9:31 in the next day, with total length of 25.5 hours and 11,730,944 intervals; RR-interval record from subject 17453 starts from 9:20 to 9:42 in the next day, with total length of 24.37 hours and 11,235,328 intervals.

As we don't have a detailed note of the subjects' activity within the whole day to qualify our method's performance, we study some widely used indices, HR and pNN30, as

references. The HR, average heart rate, quantifies subjects average heart beat per minute (bmp) over a period of time. NN30 is the number if successive RR intervals that differ by more than 30 ms, and pNN30 is the proportion of NN30 divided by the total number of RR intervals. The pNN30 is a good descriptor of long-term HR variations. Generally speaking, HR is lower and pNN30 is larger at night, especially during sleeping.

### 7.4.1 Modeling

We separate each 24-hour RR-interval data into 48 time segments, so each segment would last for approximated half an hour. We use the first 2,000 RR-interval, which is around 20 min of each segment, as windows to construct the hidden markov models with 10 observation states and 15 hidden states. And we take 60 runs for each window to result in the best hidden markov model, since the EM algorithm only gives a local optimal point for a particular initial guess. This process then provide us with 48 hidden markov models. Meanwhile, we compute the associate HR and pNN30 for each window.

We then set the last model as reference, compare it the rest of 47 models and itself using the method described in section 7.3, which result in distance numbers. The figures for time series of distances, HR, and pNN30 for subjects 16265 and 17454 are given in Figure 7.3.

### 7.4.2 Discussion

Now we examine 24-hour dynamics of cardiac rhythm using the modified Graph Matching Method and compare it to the HR and pNN30 trends.

The HRV's dynamics of subject 16265 is illustrated on the left column of Figure 7.3. As we can see from the HR and pNN30 figures, the subject begins sleeping at around 21:00 till the end of the record, where a large decrease of HR and increase of pNN30 are observed. During this time, the HMM distance is relatively stable and smooth comparing to the one at day time, indicating that there is no large fluctuation to the sympathetic and vagal activity of heart during sleep time. Similar observation can be found from subject 17453 (right

column of Figure 7.3), where the subject clearly sleep at around 01:00. A sharp decrease of HR and increase of pNN30 at 01:00 are corresponding to the sharp decrease of the HMM distance. And during the sleep time, the stable HMM distance also suggests controls of sympathetic and vagal activity of heart are stable.

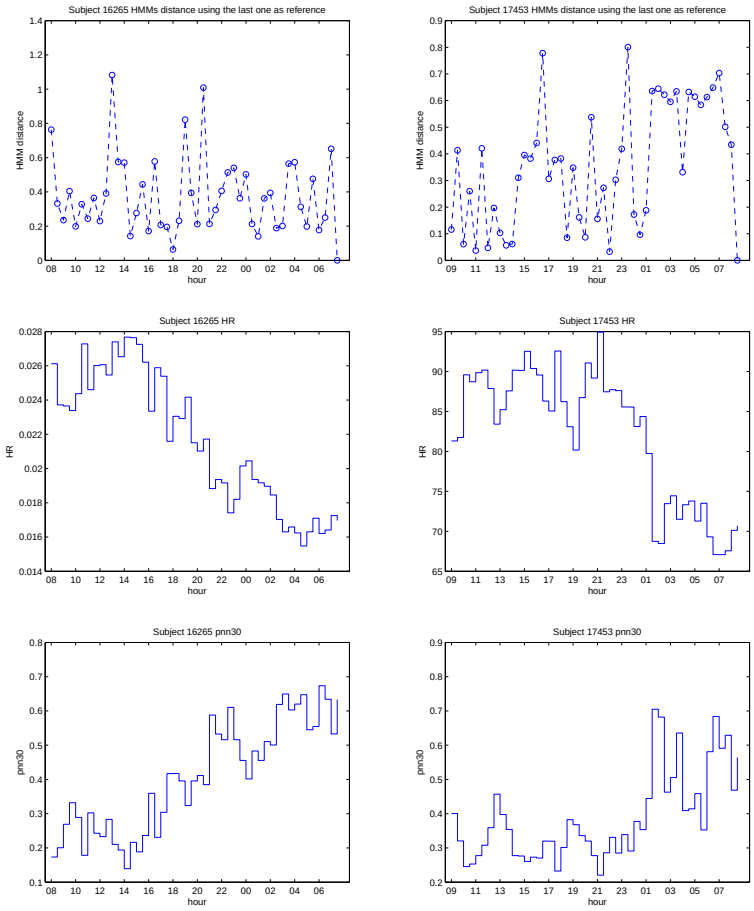



Figure 7.3: 24-hour ECG records

# Chapter 8

# Conclusion

A primary concern of this thesis is to develop principles and methods to compare dynamical systems when they are not necessarily conjugate (topologically the same). The first main body of this thesis provides an understanding of "mostly conjugacy (mostly homeomorphism)" between dissimilar systems in $L^p$ space, which enables us to measure and interpret the distance from being conjugate. We also generalize this idea from comparing deterministic systems to stochastic systems.

As a second theme, we extending and interpreting the concepts of "mostly conjugacy" in symbolic dynamics, where we resort to a variant of the classic Monge-Kantorovich optimization problem to both built a useful change of variables and measure quality of the comparison through the underlying cost called the Wasserstein distance. Later, we build up a bundle structure, visualize as a bundle plot, to show the evolution of symbolic space as we vary a system's parameter. The main object is a specific structure "joint", which happens shortly after bifurcation, implies qualitative changes of system where the kneading points become periodic.

Finally, we apply the above techniques to study time series analysis and modeling on heart rate data, which we have shown to be a one-dimensional nonlinear map that has a stochastic parameter with persistence causing the heart rate and rhythm system to wander

about a bifurcation point.

## 8.1 Contributions

Modeling is a fundamental problem whereby we seek to represent a system or data's behaviors. In this thesis, we introduce a brand new method "mostly conjugacy" to compare dynamical systems by judging the quality of "matching" by looking at their topological difference (homeomorphic defect). Comparing to the traditional regression's method, which based on normed linear spaces such as $L^2$, "mostly conjugacy" focuses more on systems' orbit structures. We provide solid analysis to address the meaning of the degree to which a simple model might be representative of a more complicated one. In addition, we apply the techniques and spirit to study actual data, i.e. the heart rate data, which provides interesting implications on its own right.

## 8.2 Future works

The thesis focuses on theoretical construction of "mostly conjugacy", with a lot of examples and applications are on well formulated systems. It is our goal to apply "mostly conjugacy" to various data sources, and provide both dynamical perspectives and practical implications to the actual data, like the topic in section 2.6 and section 7. On the other hand, it is always our objective to improve the method and algorithm of "mostly conjugacy", such that it can be applicable to more general classes of systems.

# Bibliography


[1] H. Poincare and R. Magini, II Nuovo Cimento. Les mthodes nouvelles de la mcanique cleste. (1895-1900), 10, 128, 1899.

[2] Joseph D. Skufca and Erik M. Bollt. A concept of homeomorphic defect for defining mostly conjugate dynamical systems. *Chaos*, 013118, 2008.

[3] Allen Isaacson. Numerical Analysis for Applied Science. *Wiley-Interscience, New York*, 1997.

[4] Richard L. Wheeden and Antoni Zygumd. Measure and Integral. *CRC, New York*, 1997.

[5] Joseph D. Skufca and Erik M. Bollt. On comparing dunamical systems by defective conjugacy: A symbolic dynamicas interpretation of commuter functions. *Physica D*, 2010.

[6] Jiongxuan Zheng, Joseph D. Skufca and Erik M. Bollt. Regularity of commuter functions for homeomorphic defect measure in dynamical systems model comparison. *DCDS-A Discrete and Continuous Dynamical Systems*, 2010.

[7] A. Lasota and M.Mackey. Chaos, Fractals and Noise: Stochastic aspects of dynamics. *Applied Mathematical Sciences*, 1999.

[8] L.Billings and E.Bollt. Probability density functions of some skew tent maps. *Chaos, Solitons and Fractals*, 2001.

[9] A.Reid. Random Integral Equations. *Academic Press, New York*, 1972.

[10] N.Shahzad and S.Latif. Random Fixed Points for Several Classes of 1-Ball-Contractive and 1-Set-Contractive Random Maps. *Journal of mathematical analysis and applications*, 1999.

[11] Jiongxuan Zheng, Erik M Bollt and Joseph D Skufca. Regularity of commuter functions for homeomorphic defect measure in dynamical systems model comparison. *DCDS-A Discrete and Continuous Dynamical Systems*, 2010.

[12] L. C. Evans, W. Gangbo. Differential Equations Methods for the Monge-Kantorovich Mass Transfer-Problem. *American Mathematical Society*, 1999.

[13] C. Villani. Topics in Optimal Transportation. *American Mathematical Society*, 2003.

[14] Michael Muskulus and Sjoerd Verduyn-Lunel. Wasserstein distances in the analysis of time series and dynamical systems. *Physica D*, 240(2011),45-58.

[15] M. Morse and G. A. Hedlund. Symbolic dynamics. *Amer. J. Math*, 60 (1938), 815-866.

[16] Robinson, C. Dynamical systems: Stability, symbolic dynamics, and chaos. *CRC Press*, 1995.

[17] Robert L. Devaney. An introduction to chaos dyanmical systems. *second edition, Westview Press*, 2003.

[18] Douglas Lind and Brian Marcus. An Introduaction to Symbolic Dynamics and Coding. *First ed., Cambridge University Press*, 1995.

[19] A. Lasota and M. C.Mackey. Chaos, fractals, and noise, Stochastic Aspects of Dynamics. *Springer*, 1994.

[20] S. M. Ulam. Problems in Modern Mathematics. *Wiley*, 1970.

[21] Yossi Rubner, Carlo Tomasi and Leonidas. J. Guibas. A Metric for Distributions with Applications to Image Databases. *IEEE International Conference*, (1998).

[22] N. Pytheas Fogg. Substitutions in Dynamics, Arithmetics and Combinatorics. *Springer*, 2002.

[23] B. Marcus and S. Tuncel. The weight-per-symbol polytope and scaffolds of invariants associated with Markov chains. *Erg. Th. and Dyn. Syst.*, 11 (1991), 129-180 (2010).

[24] L. Billings and E. M. Bollt. Probability density functions of some skew tent maps. *Chaos, Solitions and Fractals.*, 12 (2001).

[25] Hiroshi Teramoto and Tamiki Komatsuzaki. How does a choice of Markov partition affect the resultant symbolic dynamics? *Chaos, Solitions and Fractals.*, 20, 037113 (2010).

[26] Katrin Suder, Friedhelm R. Drepper, Michael Schiek and Hans-Henning Abel. One-dimensional, nonlinear determinism characterizes heart rate pattern during paced respiration *Am J Phiciol Heart Circ Physiol.*, 275., 162 1998, pp. 1092–1102..

[27] N.B. Janson, A.G. Balanov, V.S. Anishchenko and P.V.E. McClintock. Modelling the dynamics of angles of human R-R intervals *Physiol. Meas.*, 22: 565-579 (2001).

[28] Yuo-Hsien Shiau, Shu-Shya Hseu and Huey-Wen Yien. One-dimensional Deterministic Process Extracted from Noisy R-R Intervals under Spontaneous Breathing Conditions *Journal of Medical and Biological Engineering.*, 26(3): 121-124 (2001).

[29] T. Schreiber. Extremely simple nonlinear noise-reduction method *Phys.Rev.E.*, 47: 2401-2404 (1993).

[30] Claudia Lerma, Trine Krogh-Madsen, Micheal Guevara, and Leon Glass. Stochastic Aspects of Cardiac Arrhymias. *Journal Statistical Physics.*, 128: 347-374 (2007).

[31] Tom Kuusela, Tony Shepherd, Jarmo Hietarinta. Stochastic model for heart-rate fluctuations *Phys.Rev.E.*, 67: 061904 (2003).

[32] Task Force of the European Socirty of Cariology and the North American Scirty of Pacing and Electrophysiology. Heart rate variability: standards of measurement, physiological interpretation and clinical use *European Heart Journal.*, 17: 354-381 (1996).

[33] Pomeranz M, Macaulay RJB, and Caudill MA. Assessment of autonomic function in humans by heart rate spectral analysis. *Am J Physiol.*, 248: H151-3 (1985).

[34] Pagani M, Lombardi F, and F, Guzzetti S. Power spetral analysis of heart rate and arterial pressure variablities as a marker of sympatho-vagal interaction in man and conscious dog. *Circ Res.*, 59: 178-93 (1986).

[35] Peng C-K, Mietus JE, Liu Y, Khalsa G, Douglas PS, Benson H and Goldberger AL. Exaggerated Heart Rate Oscillations During Two Meditation Techniques *International Journal of Cardiology.*, 70: 101-107 (1999).

[36] Leon Glass. Cardiac arrhythmias and circles maps–A classical problem *Chaos: An Interdisciplinary Journal of Nonlinear Science 1.*, (1991), pp. 13–19.

[37] Kathleen T. Alligood, Tim D. Sauer and James A. Yorke. Chaos: An introduction to dynamical systems *Third ed., Springer Press.*, (2000).

[38] Edward Ott. Chaos in Dynamical systems *First ed., Cambridge University Press.*, (1993).

[39] Steven H. Strogatz. Nonlinear Dynamics and Chaos, with Applications to Physics, Biology, Chemistry and Engineering *First ed., Westview Press.*, (2000).

[40] Otakar Fojt and Jiri Holcik. Applying Nonlinear Dynamics to ECG Signal Processing. *IEEE Engineering in Medicine and Biology*, (1998).

[41] Schreiber, T., and A. Schmitz. Improved surrogate data for nonlinearity tests. *Phys. Rev. Lett.*, 77: 635-638 (1996).

[42] Holger Kantz and Thomas Schreiber. Nonlinear Time Series Analysis *Second ed., Cambridge Press.*, (2004).

[43] Jiongxuan Zheng, Erik M. Bollt and Joe D. Sckufca. Heart Rate Variability as Determinism with Jump Stochastic Parameters *submitted to MBE.*, (2012).

[44] Jiongxuan Zheng, Erik M. Bollt and Joe D. Sckufca. Comparing Dynamical Systems by Graph Matching Method *sumitted to Physica D.*, (2012).

[45] Silipo R, Deco G. Vergassola R. and Gremigni C. A Characterization of HRV's Nonlinear Hidden Dynamics by Means of Markov Models *IEEE Trans Biomedical Eng.*, 1998;79:15-27.

[46] Silipo R, Deco G. Schurmann B. Vergassola R. and Gremigni C. Investigating the underlying Markovian dynamics of ECG rthythms by information flow *Chaos, Solitions and Fractals.*, 12(2001) 2877-2888.

[47] Andrew M. Fraser. Hidden Markov Models and Dynamical Systems *SIAM.*, (2008).

[48] Goldberger AL, Amaral LAN, Glass L, Hausdorrf JM, Ivanov PCh, Mark RG, Mietus JE, Moody GB, Peng C-K, Stanley HE. PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiological Signals *Circulation 101(23):e215-e220 [Circulation Electronic Pages; http://circ.ahajournals.org/cgi/content/full/101/23/e215].*, 2000 (June 13).

[49] Rana D. Parshad, Jiongxuan Zheng, Erik M. Bollt and Joseph D. Skufca. Mostly Conjugacy of Stochastically Perturbed Dynamical Systems *Int. J. Contemp. Math. Sciences*, Vol. 7, no. 9, 395 - 412 (2012).

[50] Jiongxuan Zheng, Erik M. Bollt and Joseph D. Skufca. The Bundle plot: A presentation of symbolic space's evolution under the system's parameter changes *to appear in International Jounal of Bifurcation and Chaos*, (2012).

[51] Misiurewicz M. Jumps of entropy in one dimension *Fund Math*, (1989;132:215-26).

[52] Erik M. Bollt and Naratip Santitissadeekorn. Applied and Computational Measurable Dynamics *submitted to SIAM*, (2012).

[53] J. Milnor and W. Thurston. On iterated maps of the interval Dynamical systems *Lecture Notes in Math.*, (1988).