MASSACHUSETTS INSTITUTE OF TECHNOLOGY
ARTIFICIAL INTELLIGENCE LABORATORY

A.I. Technical Report No. 1541 May, 1995

Parameter Estimation in Chaotic Systems

Elmer S. Hung

This publication can be retrieved by anonymous ftp to publications.ai.mit.edu.

Abstract

This report examines how to estimate the parameters of a chaotic system given noisy observations of the state behavior of the system. Investigating parameter estimation for chaotic systems is interesting because of possible applications for high-precision measurement and for use in other signal processing, communication, and control applications involving chaotic systems.

In this report, we examine theoretical issues regarding parameter estimation in chaotic systems and develop an efficient algorithm to perform parameter estimation. We discover two properties that are helpful for performing parameter estimation on non-structurally stable systems. First, it turns out that most data in a time series of state observations contribute very little information about the underlying parameters of a system, while a few sections of data may be extraordinarily sensitive to parameter changes. Second, for one-parameter families of systems, we demonstrate that there is often a preferred direction in parameter space governing how easily trajectories of one system can "shadow" trajectories of nearby systems. This asymmetry of shadowing behavior in parameter space is proved for certain families of maps of the interval. Numerical evidence indicates that similar results may be true for a wide variety of other systems.

Using the two properties cited above, we devise an algorithm for performing parameter estimation. Standard parameter estimation techniques such as the extended Kalman filter perform poorly on chaotic systems because of divergence problems. The proposed algorithm achieves accuracies several orders of magnitude better than the Kalman filter and has good convergence properties for large data sets. In some systems the algorithm converges at a rate proportional to $\frac{1}{n^2}$ where n is the number of state samples processed. This is significantly better than the $\frac{1}{\sqrt{n}}$ convergence one would expect from nonchaotic oscillators based on purely stochastic considerations.

This report describes research done at the Artificial Intelligence Laboratory of the Massachusetts Institute of Technology. Support for the laboratory's artificial intelligence research is provided in part by the Advanced Research Projects Agency of the Department of Defense under Office of Naval Research contract N00014-92-J-4097. The author was also supported by the Air Force Office of Scientific Research under the Air Force Graduate Fellowship Program.

Acknowledgments

I would first like to thank Gerry Sussman and Jack Wisdom who came up with the idea for this project and did the initial work in investigating the possibility for parameter estimation in chaotic systems.

I would also like to thank my thesis advisors, Gerry Sussman and Hal Abelson, for their advice, encouragement, and support while I have been working on this project.

Also I would like to thank my officemate, Thanos Siapas, for many interesting and helpful discussions about nonlinear dynamics.

Finally, thanks to all the members of Project MAC for providing a stimulating environment to work in. I especially want to thank Stephen Adams, Joe Bank, Peter Beebee, Andy Berlin, Becky Bisbee, Ziggy Blair, Liz Bradley, Natalya Cohen, Daniel Coore, Mike Eisenberg, Arthur Gleckler, Philip Greenspun, Chris Hanson, Kleanthes Koniaris, Brian LaMacchia, Nick Papadakis, Luis Rodriguez, Bill Rozas, Panayotis Skordos, Thanos Siapas, Rajeev Surati, Jason Wilson, Henry Wu, and Feng Zhao for contributing in uncountable ways: from maintaining the computer systems, to providing technical advice, to just being cool people to talk to.

Special thanks to the Air Force Graduate Fellowship Program, the Air Force Office of Scientific Research, SCEEE Services Corporation, and the Phillips Laboratory at Kirtland Air Force Base for providing financial support while I have been working on this report. I especially want to thank Mike Harrison of the Phillips Lab for his hospitality, time, and help in coordinating my fellowship.

This work was also supported in part by the Advanced Research Projects Agency of the Department of Defense under contract number N00014-92-J-4097.

Contents

List of Figures

Chapter 1

Introduction

In this report we investigate theoretical limitations and develop computational methods for estimating the parameters of a chaotic system given a noisy time series of state data about the system. There are two primary reasons why we are interested in parameter estimation of chaotic time series. First, there has been considerable interest in recent years regarding signal processing and control applications involving chaotic systems (see e.g., [11], [49], [9]). Parameter estimation has traditionally been an important problem in signal processing and control theory, so in light of recent applications involving chaotic systems, it is important to investigate what happens when the signals and systems involved are chaotic.

Second, it has been suggested that parameter estimation in chaotic systems may have applications for high-precision measurement. In particular the idea is that if a system is chaotic and displays a sensitive dependence on initial conditions, then it can also be sensitive to small changes in parameter values. Consequently, development of successful parameter estimation techniques could make it possible to measure the parameters of a system extremely accurately given a time series of data about the state of the system.

Our goal in this report is to systematically explore the feasibility of parameter estimation in chaotic systems including a theoretical analysis of what accuracies we can reasonably expect to obtain and what factors limit this accuracy. We also present new numerical algorithms for estimating the parameters of chaotic systems and discuss simulations demonstrating the performance of the algorithms.

It turns out that the parameter estimation problem is especially interesting because it is simple enough that one can look carefully at the underlying dynamical mechanisms that affect the feasibility and efficiency of various numerical approaches. This is in contrast with a number of typical research problems involving chaotic time series which are broad enough that heuristics must generally be relied upon to attack the problem numerically. On the other hand, the parameter estimation problem is also complex enough that the results are interesting, and in some cases, quite unexpected. As we shall see, a close examination of the relationship between system dynamics and parameter estimation reveals interesting observations that greatly aid in the development of an efficient numerical approach.

1.1 The problem

Before proceeding further, we should be more explicit about what is meant by "parameter estimation." Basically, the idea is the following: Suppose that we are given a parameterized family of mappings $f_p(x)$, where x is the state vector of the system and p are some invariant parameters of the system. We will assume that $f_p(x)$, varies continuously with x and p. Further, suppose that we are given a sequence of observations,[1] $\{y_n\}$, of a certain state *orbit*,[2] $\{x_n\}$, where:

$$\begin{aligned} x_{n+1} &= f_p(x_n) \\ \text{and} \qquad y_n &= x_n + v_n \end{aligned}$$

for all integer n where v_n represents measurement errors in the data stream, $\{y_n\}$. We are interested in how to estimate the value of p given a stream of data, $\{y_n\}$. Note that we will concentrate the discrete-time formulation, but the results apply analogously to continuous time systems. For example, one might imagine that time is one of the state variables of the system, and that the $y_n's$ represent samples of a continuous-time system.

For analytic purposes it is helpful to assume that, to first approximation, the magnitude of the measurement errors are bounded so that:

$$|v_n| < \epsilon$$

for some $\epsilon > 0$. For purposes of analyzing and evaluating algorithms, it will also be useful later to think of v_n as a random variable with various probability densities.

1.2 Preview of important issues

Parameter estimation and shadowing

Let us now try to get a flavor for some of the important issues that govern the performance of parameter estimation techniques. First of all, given a family of mappings

[1]Instead of writing $\{x_n\}_{n=0}^{\infty}$, we will sometimes write $\{x_n\}$ to denote an infinite sequence of states.

[2]We will refer to a sequence of states, $x_{n}{}_{n=0}^{\infty}$, as an *orbit* of the map f if $x_{n+1} = f(x_n)$ for all integer n. Finite sections of infinite orbits, for example $x_{n}{}_{n=0}^{N}$, for some $N \geq 0$ may also be referred to as orbits.

of the form, f_p, and a noisy stream of state data, $\{y_n\}$, we would like to know which f_p's have orbits that closely *shadow* or follow $\{y_n\}$. We know that $\{y_n\}$ represents an actual orbit of f_p for some value of p, with ϵ magnitude measurement errors added in. Thus, if no orbit of f_p shadows $\{y_n\}$ within ϵ error for a particular parameter value, p_0, then p_0 cannot be the actual parameter value of the system that is being observed. On the other hand, if many systems of the form, f_p, have orbits that closely shadow $\{y_n\}$, then it would be difficult to tell from the state data which of these systems is actually being observed.

It turns out that a significant body of work is available to answer questions like, "what types of systems are insensitive to small perturbations so that orbits of perturbed systems shadow orbits of the original system and vice versa?" However, many of the results in this direction are topological in nature; that is, they answer questions like whether such shadowing orbits must exist or not for certain classes of systems. On the other hand, in order to evaluate the possibilities for parameter estimation, it is also important to know more geometrically-oriented results like, "how closely do shadowing orbits follow each other for nearby systems in parameter space" and "how long do orbits of nearby systems follow each other if the orbits do not shadow each other forever." Such results tend to be more difficult to establish and also depend more specifically on the systems involved.

An example in one dimension

Investigating the geometry of shadowing orbits can yield some interesting results. For example, consider the family of maps:

$$f_p(x) = px(1-x) \tag{1.1}$$

for $x \in [0,1]$ and $p \in [0,4]$. Henceforth we will refer to the family of maps (1.1) as simply the family of quadratic maps.

It is known ([5]) that for a non-negligible set of parameter values, the quadratic maps in (1.1) produce chaotic behavior for almost all initial conditions, meaning that orbits tend to explore intervals in state space, and nearby orbits experience exponential local divergence (i.e., positive Lyapunov exponents). Suppose that we pick $p_0 = 3.9$ and iterate an orbit, $\{x_n\}$, of f_{p_0} starting with the initial condition $x_0 = 0.3$. Numerically, the resulting orbit appears to be chaotic and exhibits the properties cited above, at least for large numbers of iterates. Now consider the question: "What parameter values, p, produce orbits that shadow $\{x_n\}$ for many iterations of f_p?" We can get some idea of the answer to this question by simply picking various values for p near 3.9 and attempting to numerically find orbits that shadow $\{x_n\}$. There are a number of issues (see Chapter 5) about how to do this.[3] However, let us for the moment simply assume that the results we present are at least qualitatively correct.

[3]For example, note that because we cannot numerically iterate the orbit $\{x_n\}$ accurately for many iterations, one could argue that the experiment is dominated by roundoff errors. However, while our

Figures 1.1 and 1.2 show the result of carrying out the described numerical experiment with $p_0 = 3.9$ and $x_0 = 0.3$. For values of p close to p_0, we attempt to find finite orbits of f_p that closely follow the f_{p_0} orbit, $\{x_n = f^n(x_0)\}_{n=0}^N$, for integers $N > 0$.[4] In order to measure how closely maps with different parameters can shadow $\{x_n\}_{n=0}^N$, it is helpful to define $\hat{e}(p, N, x_0, p_0)$ to be the maximal distance between the orbit, $\{x_n\}_{n=0}^N$, and the closest shadowing orbit, $\{f_p^n(z_0)\}_{n=0}^N$, of f_p. In other words, let:

$$\hat{e}(N, p, p_0, x_0) = \inf_{z_0 \in [0,1]} \max_{0 \leq n \leq N} |f_p^n(z_0) - f_{p_0}^n(x_0)|. \tag{1.2}$$

So, for each p and integer $N > 0$, $\hat{e}(N, p, p_0, x_0)$ measures how closely the best possible shadowing orbit of f_p follows the orbit, $\{x_n\}_{n=0}^N$, of f_{p_0}. For the purposes of this particular experiment let $p_0 = 3.9$ and $x_0 = 0.3$ be constant and set $e(N, p) = \hat{e}(N, p, p_0 = 3.9, x_0 = 0.3)$. There is nothing particularly special about our choice of $p_0 = 3.9$ or $x_0 = 0.3$. As we shall see later, many other paramter values and initial conditions yeild similar results.

Figure 1.1 shows the result of numerically computing $e(N, p)$ with respect to p for three values of N. The three v-shaped traces in the figure represents a plot of $e(N, p)$ for $N = 61$, $N = 250$, and $N = 1000$. $e(N, p)$ is plotted on the $y-$axis, while $p - p_0$, the difference in parameter value, p, from the original parameter value, p_0, is labeled on the $x-$axis. Note the distinct asymmetry of the graph between values of p greater than and less than $p_0 = 3.9$. In fact, for $N = 250$ and $N = 1000$ the graph is so steep for $p < p_0$ that it looks coincident with the vertical line demarking $p - p_0 = 0$. It seems that at least in this case, systems with parameter values, p, less than p_0 do not shadow the orbit, $\{x_n\}$, nearly as easily as those systems with parameter values greater than p_0. In some sense, it seems that orbits for higher parameter values are more flexible, or have a greater degree of freedom than do orbits for slightly lower parameter values.

This phenonmenon of asymmetrical shadowing may seem counterintuitive. If an orbit, $O(p_0)$, of paramteer value p_0 is shadowed by an orbit, $O(p_0 + \delta)$, of a slightly parameter value, $p_0 + \delta$, then given the orbit, $O(p_0 + \delta)$, of parameter $p_0 + \delta$, isn't $O(p_0 + \delta)$, shadowed by the orbit, $O(p_0)$, of a lower parameter value, p_0? Yes, but as we shall see, it may be that the set of orbits of $f_{p_0+\delta}$ that are shadowed by an orbit of f_{p_0} is actually vanishingly small. That is, if an orbit of $f_{p_0+\delta}$ is generated by choosing an inital condition at random, we would find that the probability that that orbit is shadowed by an orbit of p_0 is zero.

Returning to the example at hand, we find that the asymmetry in parameter space is even more apparent if we consider how $e(N, p)$ varies with N. Basically we want to

particular numerically-generated starting orbit may not look like the actual orbit, $\{x_n\}$, with initial condition $x_0 = 0.3$ for large values n, we will later see that qualitatively the pictures we present are similar.

[4]Here we let $f^{n+1} = f(f^n)$, so that the function, f^n, refers to the composition of f with itself n times (define f^0 to be simply the identity function). Note that if $x_{n+1} = f(x_n)$ for all integer n, then $x_n = f^n(x_0)$ for all n.

keep track of how the curves in figure 1.1 move inward toward the vertical line, $p = p_0$, as N increases. We can do this by fixing a constant, e_0, and keeping track of which parameter values, p, satisfy $e(N,p) < e_0$ for varying values of N. For example, for a particular value of e_0, suppose that I_N is the maximal interval in parameter space such that $p_0 \in I_N$ and $e(N,p) < e_0$ for all $p \in I_N$. We are interested in what fraction of the interval, I_N, is greater than or less than the original parameter value, p_0. To keep track of this let $I_N = I_N^- \cup I_N^+$ so that $I_N^- = [p_N^-, p_0]$ and $I_N^+ = [p_0, p_N^+]$ where $p_N^- \leq p_0$ and $p_N^+ \geq p_0$. Let $a(N)$ be the length of I_N^- and let $b(N)$ be the length of I_N^+. Figure 1.2 shows graphs of $a(N)$ and $b(N)$ with respect to N as computed numerically for $e_0 = 0.01$. Note that the scale for $a(N)$ and $b(N)$ on the y-axis is logarithmic so that $a(N)$ is several orders of magnitude smaller than $b(N)$ for larger values of N, reflecting the asymmetry in parameter space. Also, we see that $a(N)$ and $b(N)$ both appear approximately constant for large stretches of N except where $a(N)$ decreases in large increments over a small number of iterates. We will later see that these decreases in $a(N)$ occur along short stretches of the orbit, $\{x_n\}$, where small differences in the parameter value of the system can easily be distinguished by even noisy state data.

Applying theory to develop estimation algorithms

Figures 1.1 and 1.2 illustrate two interesting properties for the quadratic map example: (1) there is an asymmetry in the shadowing behavior of maps in parameter space, and (2) most iterates of a specific orbit are apparently not very sensitive to small changes in parameter values, while a few special iterates may be especially sensitive to parameter changes. It turns out that these two properties can be extremely helpful in developing an algorithm to do parameter estimation.

First of all, the asymmetry illustrated in figure 1.1 can be quite helpful. For instance, in the example we just considered, few maps, f_p, with parameter values lower than p_0 have orbits that can shadow the given orbit of f_{p_0}. Suppose that we are given noisy measurements of a state orbit, $\{x_n\}$. If we find that only maps from a certain interval in parameter space can shadow the observed data, then the real parameter value should be close to the lower endpoint of this parameter range. Thus, to first order, if e_0 is the magnitude of measurement error, the error in the parameter estimate is approximately governed by either $a(N)$ or $b(N)$, whichever one happens to be smaller.

In addition, we will see later that figure 1.2 reflects the fact that a few sections of the observed state data stream contribute greatly to our knowledge of the parameters of the system, while much of the rest of the data contributes almost no new information. Thus, if we can quickly sift through all the useless data and examine the critical data very carefully, we should be able to vastly improve a parameter estimation technique.

The key to this is whether or not physically interesting systems have the properties described above. A major objective of this report will be to investigate the relevant mechanisms behind the two properties and explore what types of systems might exhibit these properties. We will then investigate how to take advantage of these two properties

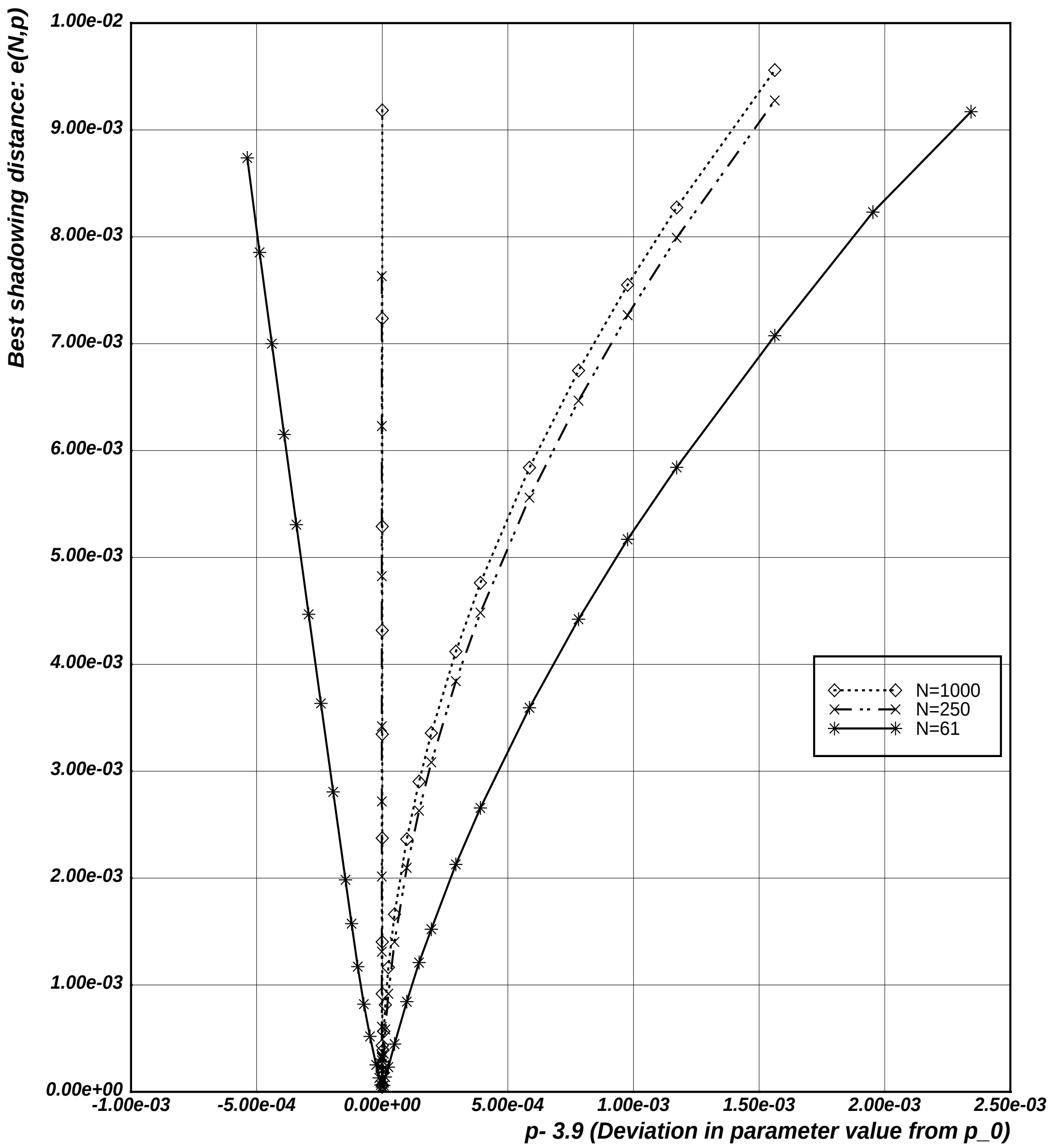

Figure 1.1: Graph of best shadowing distance, $e(N,p)$, with respect to p, for $f_p = px(1-x)$, $p_0 = 3.9$ and $x_0 = 0.3$. $e(N,p)$ measures how closely an orbit of f_p can shadow a fixed orbit, $\{x_n\}_{n=0}^{N}$, of f_{p_0}. On the $x-$axis of the graph, p is labeled as $p - p_0$, the difference in parameter value from the parameter used to generate $\{x_n\}_{n=0}^{N}$. $e(N,p)$ is plotted on the $y-$axis with N held constant for three different values of N. The three v-shaped curves represent $e(N,p)$ for $N = 61$, $N = 250$, and $N = 1000$. Note the distinct asymmetry in how well orbits of f_p track $\{x_n\}_{n=0}^{N}$ for $p > p_0$ and $p < p_0$.

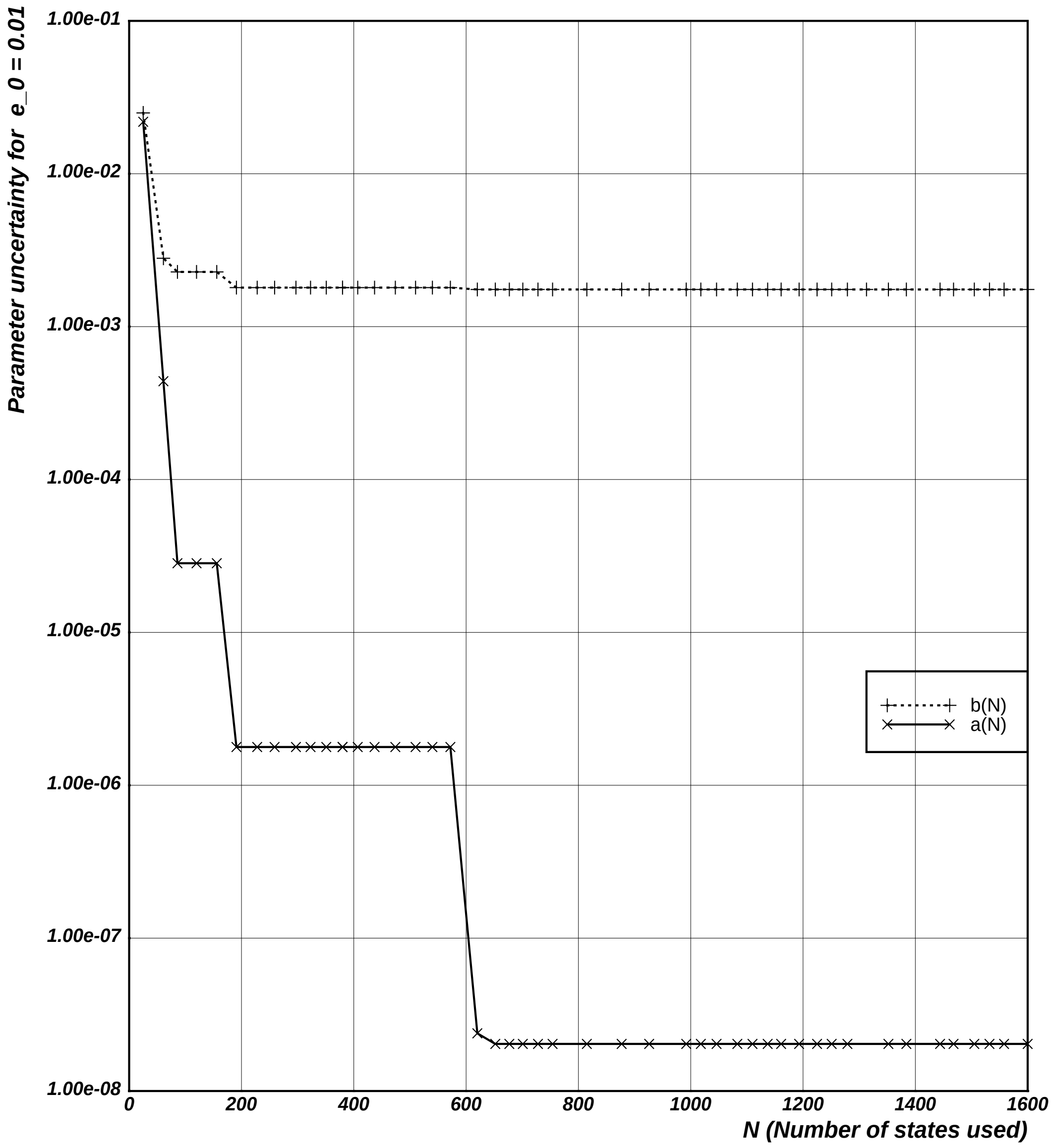

Figure 1.2: Graph of $a(N)$ and $b(N)$ with respect to N for $e_0 = 0.01$. $a(N)$ is a measure of the number of parameter values, $p < p_0$, such that there exists an orbit of f_p that can shadow the orbit, $\{x_n\}_{n=0}^N$, of f_{p_0} with less than e_0 error. Similarly $b(N)$ measures the number of parameter values, $p > p_0$, such that f_p that can shadow the orbit, $\{x_n\}_{n=0}^N$, with less than e_0 error.

to produce superior parameter estimation algorithms.

1.3 New results and previous work

1.3.1 Dynamical theory and shadowing orbits

There has not been much work directly attacking the parameter estimation problem for chaotic systems. However, as we saw in the previous section, the feasibility of parameter estimation is closely related to the concept of shadowing orbits.

For uniformly hyperbolic systems, it is well know that orbits of perturbed systems shadow orbits of the original system forever ([7],[4]). Applying this result to parameter estimation, we find that one cannot expect to get accurate information about the parameters of a hyperbolic system based on state data, since it is difficult to distinguish orbits from systems with two nearby parameter values.

However, most physically interesting chaotic systems are not in fact hyperbolic. In general[5], one can only expect so-called subexponentially hyperbolic behavior (see eg., [52]), so that hyperbolicity on a state orbit is available on a local scale, but is not uniform over an infinite orbit. The result is that most finite *pseudo-orbits*[6] of a system can be shadowed closely by a real orbit of that system. This observation was made in [24], where attempts were also made to establish bounds on the shadowing behavior of finite orbits in nonhyperbolic systems by using linearization to exploit the locally hyperbolic behavior along a typical state orbit. Such work received interest because shadowing was thought of as a helpful property that lends credibility to computer-generated orbits with roundoff error.

In the case of parameter estimation, the hyperbolic degeneracies that prevent shadowing behavior are in fact the focus of most of the interest. This is unlike past work involving shadowing orbits, because in order to investigate the feasibility of parameter estimation, it is important to specifically examine the mechanism behind the lack of shadowing behavior in nonhyperbolic systems. In addition, it is also necessary to examine carefully how orbits for one parameter value can shadow orbits for systems with a continuum of different parameter values.

The result is that we find that most measurements of the state of a system contain comparatively little information about the parameters of the system except for those iterates where the hyperbolic behavior of a system becomes degenerate. This is the phenomenon we observed with the quadratic map.

[5] for example, for almost all C^2 diffeomorphisms

[6] A pseudo-orbit of a map, g, is a sequence of states $\{z_n\}$ such that $z_{n+1} = f(z_n) + v_n$ for all n, where the magnitude of the noise, $|v_n|$, is assumed to be small.

In this report, we discuss how the lack of shadowing behavior seems to be the result of a mechanism which shall be referred to as *folding* in state space. It also seems that this folding behavior tends naturally to result in one-sided shadowing behavior in parameter space, making it possible to effectively distinguish parameter values near areas where folding occurs.

For one dimensional maps like the quadratic map, we have been able to characterize the results quantitatively. For example, for the quadratic map, $f_p(x) = px(1-x)$, we show that the following is true:

Proposition: Let

$$\tilde{e}(p, p_0, x_0) = lim_{N\to\infty}\hat{e}(N, p, p_0, x_0)$$

where $\hat{e}(p, p_0, x_0)$ is as given in (1.2). There exist constants $\delta > 0$, $C > 0$, and $K > 0$ such that the following is true: For any $\gamma \in (0, 1)$, there is a set, $E(\gamma) \subset [0, 4]$, of positive Lebesgue measure such that if $p_0 \in E(\gamma), then:$

(1) For $x_0 \in [0, 4]$,

$$\tilde{e}(p, p_0, x_0) < C|p - p_0|^{\frac{1}{3}}$$

for all $p \in (p_0, p_0 + \delta)$.

(2) For almost all $x_0 \in [0, 4]$,

$$\tilde{e}(p, p_0, x_0) > K(p - p_0)^{\gamma}$$

for all $p \in (p_0 - \delta, p_0)$.

This follows from Theorem 3.4.2.

From the proposition we see that there can in fact be a pronounced asymmetry in the shadowing behavior of orbits in parameter space and that this phenomenon is quite prevalent. For the quadratic maps (1.1) with positive Lyapunov exponents, it can also be shown that the asymmetry always favors one particular direction in parameter space for maps. That is, it is always easier for orbits of maps with slightly higher parameters to shadow orbits of maps with slightly lower parameters.

For more complicated systems, like systems in higher dimensions, it is more difficult to establish definite analytical results. However we present numerical results that demonstrate that surprisingly many systems have the properties discussed, namely that (1) a small fraction of the data contains most of the information about the parameters of the system, and (2) there is an asymmetry in the behavior of shadowing orbits in parameter space.

1.3.2 Parameter estimation techniques

Traditionally, parameter estimation is carried out numerically using algorithms like the extended Kalman filter. However, we will demonstrate that algorithms like the extended Kalman filter that linearize state and parameter space around a certain trajectory actually perform worse than one might expect simply from linearization errors. This is basically because most of the information about the parameters are contained in a small number of data points, the very data points where nonlinear folding behavior is most important. Techniques like the extended Kalman filter have a difficult time modeling the folding behavior of state space with these data points, along with the local exponential expansion and contraction properties of state space in chaotic systems. The result is that these algorithms typically diverge. In other words, the algorithm's estimate of the error in its parameter estimate quickly becomes much less than the actual error, so that the algorithm ends up converging to the wrong parameter value.

In this report, we describe a new algorithm for performing parameter estimation on chaotic systems and show numerical results demonstrating the effectiveness of the new algorithm and comparing the algorithm with traditional techniques. The new algorithm attempts to sift through most of the data quickly, concentrating on the measurements that are most sensitive to parameter values. The algorithm then uses a technique, based on a Monte Carlo method, to pick out a parameter estimate by taking advantage of the fact that shadowing behavior tends to be asymmetrical in parameter space.

1.4 Overview

This report is divided into two major parts. The first part, which includes chapters 2-4, discusses theoretical results concerning parameter estimation in chaotic systems. In particular, we are interested in questions like: (1) What possible constraints are there to the accuracy of parameter estimates, and what kind of accuracy can one expect given large amounts of data? (2) How is the accuracy of a parameter estimate likely to depend on the magnitude of the measurement error and the number of state measurements available? (3) What types of systems exhibit the most sensitivity to small parameter changes, and what types of systems are likely to produce the most (and least) accurate parameter estimates? Basically we want to understand exactly how much information state samples actually contain about the parameters of various types of systems.

In order to answer these questions, we first examine how parameter estimation relates to well-known concepts like shadowing, hyperbolicity, and structural stability. Chapter 2 discusses how the established theory concerning these concepts relates to the problem of parameter estimation. We also examine what types of systems are guaranteed to have topologically stable sorts of behavior and how this constrains our ability to do parameter estimation.

In chapter 3, we examine one-dimensional maps. Because of the relative simplicity of these systems, they are ideal for investigating how the specific geometry of a system relates to parameter estimation, especially when one is dealing with systems that are not topologically or structurally stable. New quantitative results are obtained concerning how orbits for nearby parameter values shadow each other in certain one-dimensional families of maps.

In chapter 4 we examine non-uniformly hyperbolic systems of dimension greater than one. In such general settings it is difficult to make quantitative statements concerning limits to parameter estimation. However, we extend ideas from the analysis of one-dimensional systems to suggest mechanisms that determine the shadowing behavior of orbits. These mechanisms result from an examination of the stable and unstable manifolds of the systems. Although the conjectures we make are not rigorously proved, they are supported by numerical evidence.

The second major part of the report (comprising chapter 5) describes an effort to use the dynamical systems theory to develop a reasonable algorithm to numerically estimate the parameters of a system given noisy state samples. We discuss why traditional methods of parameter estimation have problems, and some ways to fix these problems.

In chapter 6 we present numerical results demonstrating the effectiveness of the new estimation techniques proposed.

Chapter 7 summarizes the main conclusions of this report, and suggests possible future work.

Chapter 2

Parameter estimation, shadowing, and structural stability

In this chapter we review a variety of established mathematical results and apply these results to an analysis of parameter estimation. In particular, we examine how topological stability results for certain types of systems constrain the feasibility of parameter estimation.

2.1 Preliminaries and definitions

In this section, we introduce some of the basic definitions and tools needed to analyze problems related to parameter estimation. We begin by restating a mathematical description of the problem. We are given the family of discrete mappings, $f_p : M \to M$ where M is a smooth compact manifold and p represents the invariant parameters of the system. For the purposes of this report, we will also assume that p is a scalar so that f_p represents a one-parameter family of maps for $p \in I_p$, where $I_p \subset \mathbb{R}$ is a closed interval of the real line. Note that it will often be convenient to write $f(x,p)$ in place of $f_p(x)$ to denote functional dependence on both x and p. We will assume that this joint function of state and parameters, $f : M \times I_p \to M$, is continuous over its domain.

The data we are given consists of a sequence, $\{y_n\}$, of noisy observations of the state vectors, $\{x_n\}$, where $y_n \in M$, $x_n \in M$, and:

$$\begin{aligned} x_{n+1} &= f_p(x_n) \\ y_n &\in B(x_n, \epsilon) \end{aligned}$$

for all $n \in \mathbb{Z}$ where $\epsilon > 0$ and $B(x_n, \epsilon)$ represents an $\epsilon-$neighborhood of x_n (ie., $y_n \in B(x_n, \epsilon)$ if and only if $d(y_n, x_n) < \epsilon$ for some distance metric d). In other words, the measured data, y_n, consists of the actual state of the system, x_n, plus some noise of magnitude ϵ or less.

Note that if we fix $p_0 \in I_p$, we can generate an orbit, $\{x_n\}$, given an initial condition, x_0. Basically, we would like to know how much information this state orbit contains about the parameters of the system. In other words, within possible measurement error, can we resolve $\{x_n\}$ from orbits of nearby systems in parameter space? In particular, are there parameters near p_0 that have no orbits that closely follow $\{x_n\}$? If so, then we know that such parameters could not possibly produce the state data represented by $\{y_n\}$, and we can thus eliminate these parameters as possible choices for the parameter estimate. Thus, given $p_0 \in I_p$ and a state orbit, $\{x_n\}$, of f_{p_0}, one important question to ask is: For what values of $p \in I_p$ does there exist an orbit, $\{z_n\}$, of f_p such that $d(z_n, x_n) < \epsilon$ for all n?

This relates parameter estimation to the concept of shadowing. Below we describe some definitions for various types of shadowing that will be useful later on:

Definitions: Let $g : M \to M$ be continuous. Suppose $d(g(z_n), z_{n+1}) < \delta$ for all n. Then $\{z_n\}$ is said to be a *δ-pseudo-orbit* of g. We say that a sequence of states, $\{x_n\}$, *ϵ-shadows* another sequence of states, $\{y_n\}$, if $d(x_n, y_n) < \epsilon$ for all n. The map g has the *pseudo-orbit shadowing* property if for any $\epsilon > 0$, there is a $\delta > 0$ such that every δ-pseudo-orbit is ϵ-shadowed by a real orbit of g. The family of maps, $\{f_p | p \in I_p\}$, is said to have the *parameter shadowing* property at $p_0 \in I_p$ if for any $\epsilon > 0$, there exists a $\delta > 0$ such that every orbit of f_{p_0} is ϵ-shadowed by some orbit of f_p for any $p \in B(p_0, \delta)$. Finally, suppose that $g \in X$ where X is some metric space. Suppose further that for any $\epsilon > 0$, there is a neighborhood of g, $U \subset X$, such that if $g' \in U$ then any orbit of g is $\epsilon-$shadowed by an orbit of g'. Then g is said to have a *function shadowing* property in X.

We can see that the various types of shadowing have natural connections to parameter estimation. If two orbits $\epsilon-$shadow each other, then these two orbits will (to first order) be indistinguishable from each other with measurement noise of magnitude ϵ. If f_{p_0} has the parameter shadowing property, then all systems near $p = p_0$ in parameter space have orbits that ϵ-shadow orbits of f_{p_0}. This implies inherent constraints on the attainable accuracy of parameter estimation based on state data, since observable state differences for nearby systems in parameter space are lost in the noise caused by measurement errors.

Thus parameter shadowing is really the property we are most interested in because of its direct relationship with parameter estimation. The concept of function shadowing is simply a generalization of parameter shadowing so that given some function g, we can guarantee that any continuous parameterization of systems containing g must have the parameter shadowing property at g. This situation implies that the state evolution of the system is in some sense stable or insensitive to small perturbations in the system. In the literature, the following language is used to describe this sort of "stability:"

Definitions: Two continuous maps, $f : M \to M$ and $g : M \to M$, are said to be *topologically conjugate* if there exists a homeomorphism, h, such that $gh = hf$. Let $\mathit{Diff}^r(M)$ be the space of C^r diffeomorphisms of M. Then $g \in \mathit{Diff}^r(M)$ is said to be *structurally stable* if for every neighborhood, $U \in \mathit{Diff}^0(M)$, of the identity function, there is a neighborhood, $V \subset \mathit{Diff}^r(M)$, of g such that for each $f \in V$ there exists a homeomorphism, $h_g \in U$, satisfying $f = h_f^{-1} g h_f$. In addition, if there exists a constant $K > 0$ and neighborhood $V' \subset V$ of g such that $\sup_{x \in M} d(h_f(x), x) \leq K \sup_{x \in M} d(f(x), g(x))$, for any $f \in V'$, then g is said to be *absolutely structurally stable*.

Unfortunately, we have introduced a rather large number of definitions. Some of the definitions apply directly to parameter estimation, and others are introduced because they are historically important and are necessary in order to apply results found in the literature. Before we continue, it is important to state clearly how the various properties are related and exactly what they mean for parameter estimation.

2.2 Shadowing and structural stability

We now investigate the relationship between various shadowing properties and structural stability. The goal here is to relate well-known concepts like pseudo-orbit shadowing and structural stability to parameter and function shadowing, so that we can apply results from the literature.

Let us begin with a brief discussion. First of all, given any $p_0 \in I_p$, note that if p is near p_0, then orbits of f_p are pseudo-orbits of f_{p_0}. The pseudo-orbit shadowing property implies that a particular system can shadow all trajectories of nearby systems. That is, any orbit of a nearby system can be shadowed by an orbit of the given system. On the other hand, function shadowing is somewhat the opposite. A system exhibits the function shadowing property if all nearby systems can shadow it. Meanwhile, structural stability implies a one-to-one correspondence between orbits of all systems within a given neighborhood in function space. Thus, if a system is structurally stable, then all nearby systems can shadow each other.

While these three properties are not equivalent in general they are apparently equivalent for certain types of *expansive* maps, where the definition of expansiveness is given below:

Definitions: A homeomorphism $g : M \to M$ is said to be *expansive* if there exists $e(g) > 0$ such that

$$d(g^n(x), g^n(y)) \leq e(g)$$

for $n \in Z$ if and only if $x = y$.[1] $e(g)$ is called the *expansive constant* for g. Also, suppose

[1] Note that in general, if g is a function then we will write g^n to mean the function g composed with

X is a metric space of homeomorphisms. Then a function $g \in X$ is *uniformly expansive* in X if there exists a neighborhood $V \subset X$ of g such that $inf_{f \in V}(e(f)) > 0$.

We now state some properties relating pseudo-orbit shadowing, function shadowing, and structural stability. Many of these results are addressed by Walters in [62]. We refer the reader to [62] and fill in the gaps as necessary in Appendix A.

Theorem 2.2.1 *Let $g : M \to M$ be a structurally stable diffeomorphism. Then g has the function shadowing property.*

Proof: This follows directly from the definitions of structural stability and function shadowing. The conjugating homeomorphism, h, from the definition of structural stability provides a one-to-one connection between shadowing orbits of nearby maps.

Theorem 2.2.2 *(Walters) Let $g : M \to M$ be a structurally stable diffeomorphism of dimension ≥ 2. Then g has the pseudo-orbit shadowing property.*

Proof: This follows directly from Theorem 11 of [62]. The proof is not as simple as the previous theorem, since a pseudo-orbit of g is not necessarily a real orbit of a nearby map. However, Walters shows that given a pseudo-orbit of g, we can pick a (possibly) different pseudo-orbit of g that both shadows the original pseudo-orbit and is in fact a true orbit of a nearby map. Then structural stability can be invoked to to show that there must be a real orbit of g that shadows the original pseudo-orbit.

Theorem 2.2.3 *Let $g : M \to M$ be an expansive diffeomorphism with the pseudo-orbit shadowing property. Suppose there exists a neighborhood, $V \subset \mathrm{Diff}^1(M)$ of g that is uniformly expansive. Then g is structurally stable.*

Proof: This follows from discussions in [62]. See Appendix A.

Theorem 2.2.4 *: Let $g : M \to M$ be an expansive diffeomorphism with the function shadowing property. Suppose there exists a neighborhood, $V \subset \mathrm{Diff}^1(M)$ of g such that V is uniformly expansive. Then g is structurally stable.*

Proof: This is similar to theorem 4 of [62]. See Appendix A.

Summarizing our results relating various forms of shadowing and structural stability, we find that structural stability is the strongest condition considered. Structural stability of a diffeomorphism of greater than one dimension implies both the pseudo-orbit

itself n times. We assume that g^0 is the identity function.

shadowing and parameter shadowing properties for continuous families of mappings. Thus we can use the literature on structural stability to show that certain families of maps must have parameter shadowing properties, making it difficult to accurately estimate parameters given state data. As we shall see, however, most systems we are likely to encounter in physical applications are not structurally stable.

Also, the pseudo-orbit shadowing property, parameter shadowing property, and structurally stability are equivalent for expansive diffeomorphisms $g : M \to M$ of dimension greater than one if there exists a neighborhood of g in $\mathit{Diff}^{\,1}(M)$ that is uniformly expansive. However, again we shall see that most physical systems do not have this expansiveness property. Note also that these results do not apply to the maps of the interval which we consider in the next chapter.

2.3 Absolute structural stability and parameter estimation

There is one more useful property we have not yet addressed. That is the concept of absolute structural stability.

Lemma 2.3.1 *Suppose that $f_p \in \mathrm{Diff}^{\,1}(M)$ for $p \in I_p \subset \mathbb{R}$, and let $f(x,p) = f_p(x)$ for any $x \in M$. Suppose that $f : M \times I_p \to M$ is C^1 and that f_{p_0} is an absolutely structurally stable diffeomorphism for some $p_0 \in I_p$. Then there exist $\epsilon_0 > 0$ and $K > 0$ such that for every positive $\epsilon < \epsilon_0$, any orbit of f_{p_0} can be $\epsilon-$shadowed by an orbit of f_p if $p \in B(p_0, K\epsilon)$.*

Proof: This follows fairly directly from the definition of absolute structural stability. The conjugating homeomorphism provides the connection between shadowing orbits. See Appendix A for a complete explanation.

Thus if an absolutely structurally stable mapping, g, is a member of a continuous parameterization of mappings, then nearby maps in parameter space can ϵ-shadow any orbit of g. Furthermore, from above we see that the range of parameters that can shadow orbits of g varies at most linearly with ϵ for sufficiently small ϵ so that decreasing the measurement error will not result in any dramatic improvements in estimation accuracy. In these systems, it is clear that dynamics does not contribute a great deal to our ability to distinguish between the behavior of nearby systems. In the next section, we shall see that so-called uniformly hyperbolic systems can exhibit this absolute structural stability property, making them poor systems for accurate parameter estimation.

2.4 Uniformly hyperbolic systems

Let us now turn turn our attention to identifying what types of systems exhibit the various shadowing and structural stability properties described in the previous section. Stability is intimately associated with hyperbolicity, so we begin by examining uniformly hyperbolic systems.

Uniformly hyperbolic systems are interesting as the archetypes for complex behavior in nonlinear systems. Because of the definite structure available in such systems, it is generally easier to prove results in this case than for more general situations. Unfortunately, from a practical viewpoint, very few physical systems actually exhibit the properties of uniform hyperbolicity. Nevertheless, understanding hyperbolicity is important as a first step to figuring out what is happening in more general situations.

Our goal in this section is to state some stability results for hyperbolic systems, and to motivate the connections between hyperbolicity, stability, and parameter estimation. Most of the results in this section are well-known and have been written about in numerous sources. The material provided here outlines some of the properties of hyperbolic systems that pertain to our treatment of parameter estimation. The brief discussions use informal arguments in an attempt to motivate ideas rather than provide proofs. References to more rigorous proofs are given. For an overview of some of the material in this section, a few good sources include: Shub [55], Nitecki [43], Palis and de Melo [50], or Newhouse [42].

We first need to know what it means to be hyperbolic:

Definitions:

(1) Given $g : M \to M$, Λ is a (uniformly) *hyperbolic* set of g if there exists a continuous invariant splitting of the tangent bundle, $T_xM = E^s_x \oplus E^u_x$ for all $x \in \Lambda$ and constants $C > 0$ and $\lambda > 1$ such that:
(a)$|Dg^n v| \leq C\lambda^{-n}|v|$ if $v \in E^s_x, n \geq 0$
(b)$|Dg^{-n} v| \leq C\lambda^{-n}|v|$ if $v \in E^u_x, n \geq 0$

(2) A diffeomorphism $g : M \to M$ is said to be *Anosov* if M is uniformly hyperbolic.

One important property for understanding the behavior of hyperbolic systems are the existence of smooth uniformly contracting and expanding manifolds.

Definition: We define the local stable, $W^s_\epsilon(x, g)$, and unstable, $W^u_\epsilon(x, g)$, sets of $g : M \to M$ as follows:

$$
\begin{aligned}
W^s_\epsilon(x, g) &= \{y \in M : d(g^n(x), g^n(y)) \leq \epsilon \text{ for all } n \geq 0 \} \\
W^u_\epsilon(x, g) &= \{y \in M : d(g^{-n}(x), g^{-n}(y)) \leq \epsilon \text{ for all } n \geq 0 \}
\end{aligned}
$$

We define the global stable, $W^s(x, g)$, and unstable, $W^u(x, g)$, sets of $g : M \to M$ as follows:

$$\begin{aligned} W^s(x,g) &= \{y \in M : d(g^n(x), g^n(y)) \to 0 \text{ as } n \to \infty\} \\ W^u(x,g) &= \{y \in M : d(g^{-n}(x), g^{-n}(y)) \to 0 \text{ as } n \to \infty\}. \end{aligned}$$

The following result shows that these sets have definite structure. Based on this result, we replace the word "set" with the word "manifold" in the definitions above, so, for example, $W^s(x, g)$ and $W^u(x, g)$ are the *stable* and *unstable manifolds* of g at x.

Theorem 2.4.1 *(Stable/unstable manifold theorem for hyperbolic sets): Let $g : M \to M$ be a C^r diffeomorphism ($r \geq 1$), and let $\Lambda \subset M$ be a compact invariant hyperbolic set under g. Then for sufficiently small $\epsilon > 0$ the following properties hold for $x \in \Lambda$:*

(1) $W^s_\epsilon(x, g)$ and $W^u_\epsilon(x, g)$ are local C^r disks for any $x \in \Lambda$. $W^s_\epsilon(x, g)$ is tangent to E^s_x at x and $W^u_\epsilon(x, g)$ is tangent to E^u_x at x.

(2) There exist constants $C > 0$ and $\lambda > 1$ such that:

$$\begin{aligned} d(g^n(x), g^n(y)) &< C\lambda^{-n} \text{ for all } n \geq 0 \text{ if } y \in W^s_\epsilon(x) \\ d(g^{-n}(x), g^{-n}(y)) &< C\lambda^{-n} \text{ for all } n \geq 0 \text{ if } y \in W^u_\epsilon(x). \end{aligned}$$

(3) $W^u_\epsilon(x)$ and $W^u_\epsilon(x)$ vary continuously with x.

(4) We can choose an adaptive metric such that $C = 1$ in (2).

Proof: See Nitecki [43] or Shub [55].

Note that from (2) above, we can see that our definitions for the global stable and unstable manifolds are natural extensions of the local manifolds. In particular, $W^s_\epsilon(x, g) \subset W^s(x, g)$, $W^u_\epsilon(x, g) \subset W^u(x)$, and:

$$\begin{aligned} W^s(x,g) &= \bigcup_{n \geq 0} g^{-n}(W^s_\epsilon(g^n x)) \\ W^u(x,g) &= \bigcup_{n \geq 0} g^{n}(W^s_\epsilon(g^{-n} x)). \end{aligned}$$

Thus C^r stable and unstable manifolds vary continuously, and intersect transversally on hyperbolic sets, meaning that the angle of intersection between the stable and unstable manifolds is bounded away from zero on Λ. These manifolds create a foliation of uniformly contracting and expanding sets that provides for a definite structure of the space. We will now argue that uniformly hyperbolic systems obey shadowing properties and are structurally stable.

Lemma 2.4.1 *(Shadowing Lemma): Let $g : M \to M$ be a C^r diffeomorphism ($r \geq 1$), and let $\Lambda \subset M$ be a compact invariant hyperbolic set under g. Then there exists a neighborhood, $U \subset M$, of Λ such that g has the pseudo-orbit shadowing property on U. That is, given $\epsilon > 0$, there exists $\delta > 0$ such that if $\{z_n\}$ is a δ-pseudo-orbit of g, with $z_n \in U$ for all n, then $\{z_n\}$ is $\epsilon-$shadowed by a real orbit, $\{x_n\}$, of g such that $x_n \in \Lambda$ for all integer n.*

Proof: Proofs for this result can be found in [7] and [55]. Here we sketch an informal argument similar to the one given by Conley [16] and Ornstein and Weiss [47] for the case where g is Anosov (ie, $\Lambda = M$ is hyperbolic).

Let $\{z_n\}$ be a δ-pseudo-orbit of g and let $B_n = B(z_n, \epsilon)$. For the pseudo-orbit shadowing property to be true, there must be a real orbit, $\{x_n\}$, of g such that $x_n \in B_n$ for all integer n. Thus it is sufficient to show that for any $\epsilon > 0$ there is a $\delta > 0$ such that given any δ-pseudo-orbit of g, $\{z_n\}$, there exists $x_0 \in \Lambda$ satisfying:

$$x_0 \in \bigcap_{n \in Z} g^{-n}(B(z_n, \epsilon)). \tag{2.1}$$

Since the stable and unstable manifolds intersect transversally (at angles uniformly bounded away from zero), for any $p \in \Lambda$, we can use the structure of the manifolds around p to define a local coordinate system for uniformly large neighborhoods, of $p \in \Lambda$.[2] We can think of this as locally mapping the stable and unstable manifolds onto a patch of $\mathbb{R}^n$ such that stable and unstable manifolds lie parallel to the axes of a Cartesian grid (see figure 2.1). Also we can choose an adapted metric on Λ (specified in part (4) of the stable manifold theorem), for each $p \in \Lambda$ so that g has uniform local contraction/expansion rates. Using this metric on the transformed coordinates, we have a nice, neat model of local dynamical behavior, as we shall see below. From now on we deal exclusively with transformed local coordinates centered around z_n and the adapted metric. Note that the discussion below and the pictures reflect the two-dimensional case (the idea is similar in higher dimensions).

Now for all n pick squares, $S(z_n, \epsilon) = S_n$, of uniformly bounded size centered at z_n with $S(z_n, \epsilon) \subset B(z_n, \epsilon)$ such that the sides of S_n are parallel to the axes of the transformed coordinate system around z_n. The sides of the S_n squares are fibered by stable and unstable manifolds, so when we apply g to S_n, the square is stretched into a rectangle, expanding along the unstable direction, contracting in the stable direction. Meanwhile, the opposite is true for g^{-1}. Note that if we can show that there exists some $x_0 \in \Lambda$ and $\epsilon > 0$ such that:

$$x_0 \in \bigcap_{n \in Z} g^{-n}(S(z_n, \epsilon))$$

[2] The local coordinates we refer to here are known as *canonical coordinates.* For a more rigorous explanation of these coordinates refer to Smale [59] or Nitecki [43].

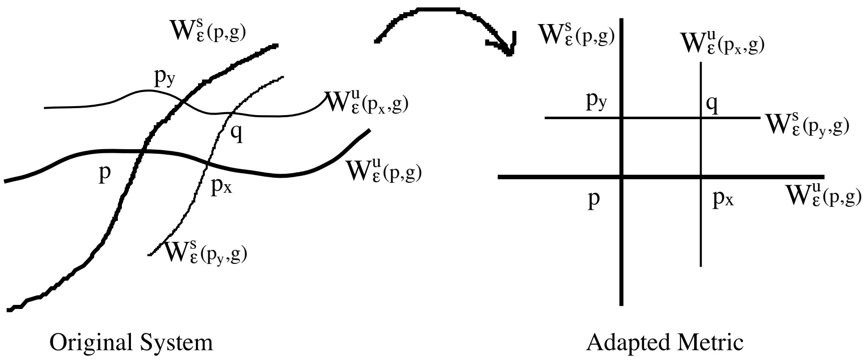

Figure 2.1: First we use the structure of the manifolds of the hyperbolic system to define a local coordinate system with nice geometric properties, so that the manifolds are orthogonal and and expand/contract uniformly under a single application of g.

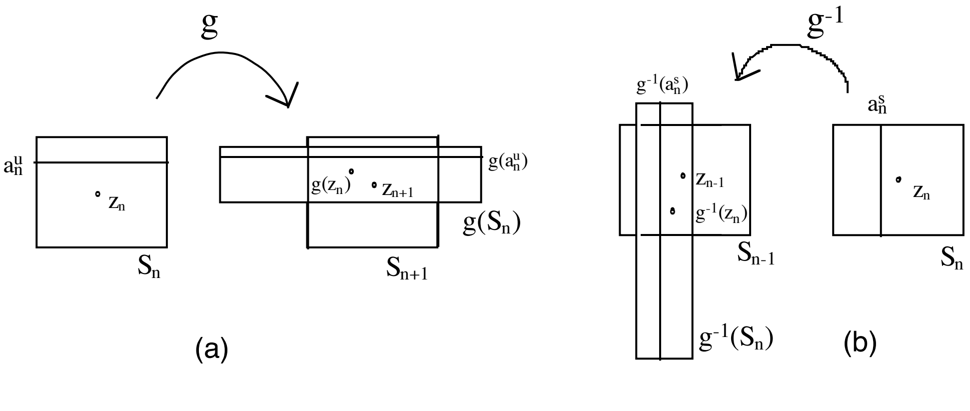

Figure 2.2: For any $\epsilon > 0$ we can choose $\delta > 0$ so that for any $n \in \mathbb{Z}$, (a) any line segment, a^u_n, along the unstable direction in S_n gets mapped by g so that it intersects S_{n+1}, and (b) any line segment, a^s_n, along the stable direction in S_n gets mapped by g^{-1} so that it intersects S_{n-1}.

for any sequence, $\{z_n\}$, that is δ-pseudo-orbit of g, then the shadowing property must be true. This is our goal.

Let $n \in \mathbb{Z}$ and let a_n^u be any line segment extending the length of a side of $S(z_n, \epsilon)$ parallel to the unstable direction inside $S(z_n, \epsilon)$. Set $a_{n+1}^u = g(a_n^u) \cap S(z_{n+1}, \epsilon)$. Then, for any $\epsilon > 0$, we can choose a suitably small $\delta_1 > 0$, such that for any n, a_{n+1}^u must be nonempty if $\{z_n\}$, is a δ_1-pseudo orbit, of g (see figure 2.2). In figure 2.2 we see that $\delta_1 > 0$ represents the possible offset between the centers of the rectangle, $g(S_n)$, and the square, S_{n+1}. As ϵ get smaller, the size of the rectangle and square gets smaller, but we can still choose a suitably small $\delta_1 > 0$ so that $g(a_n^u)$ intersects S_{n+1}. Furthermore we can do exactly the same thing in the opposite direction. That is, let a_n^s be any line segment extending along the stable direction of $S(z_n, \epsilon)$, set $a_{n-1}^s = g^{-1}(a_n^s) \cap S(z_{n-1}, \epsilon)$, and choose $\delta_2 > 0$ suitably small so that a_{n-1}^s must be nonempty for any n if $\{z_n\}$, is a δ_2-pseudo orbit, of g.

Given any $\epsilon > 0$ set $\delta = \min\{\delta_1, \delta_2\}$. Then, for any $n > 0$, let $a_n^s(n)$ be a segment in $S_n = S(z_n, \epsilon)$ parallel to the stable direction. Set $a_{k-1}^s(n) = g^{-1}(a_k^s(n)) \cap S_{k-1}$ for any $k \leq n$. From our previous arguments we know that as long as $\{z_n\}$ is a $\delta-$pseudo orbit of g, then $a_{k-1}^s(n)$ must be a (nonempty) line in the stable direction within S_{k-1} if $a_k^s(n)$ is a line in the stable direction of S_k. Consequently, by induction, $a_0^s(n)$ must be a line in the stable direction of S_0 for any $n > 0$. Furthermore note that $a_k^s(n) \subset S_k$ for any $k \in \{0, 1, \ldots, n\}$. Doing a similar thing for $n < 0$, working with g instead of g^{-1}, and starting with a segment $a_n^u(n)$ parallel to the unstable direction of S_n, we see that for any $n < 0$ there exists a series of line segments, $a_k^u(n) \subset S_k$, for each $k \in \{n, n+1, \ldots, -1, 0\}$ oriented in the unstable direction. Clearly $a_0^s(-n)$ and $a_0^u(n)$ must intersect for any $n > 0$. Now consider the limit of this process as $n \to \infty$. It is easy to show that the intersection point

$$x_0 = (\lim_{n\to\infty} a_0^s(n)) \bigcap (\lim_{n\to-\infty} a_0^u(n))$$

must exist and must in fact be the x_0 we seek satisfying (2.1). This initial condition can then be used to generate a suitable shadowing orbit, $\{x_n\}$.

Theorem 2.4.2 *Anosov diffeomorphisms are structurally stable.*

Proof: Proofs for this result can be found in [4] and [37].

It is also possible to prove this result based on the shadowing lemma. The basic idea is to show that any Anosov diffeomorphism, $g : M \to M$, is uniformly expansive, and then to apply theorem 2.2.3 to get structural stability. Walters does this in [62]. We outline the arguments.

The fact that g is expansive is not too difficult to show. If this were not true, then there must exist $x \neq y$ such that $d(g^n(x), g^n(y)) \leq \epsilon$ all integer n. But satisfying this condition for both $n \geq 0$ and $n \leq 0$ would imply that $y \in W^s_\epsilon(x,g)$ and $y \in W^u_\epsilon(x,g)$, respectively. This cannot happen unless $x = y$. The contradiction shows that the Anosov diffeomorphism, g, must be expansive with expansive constant, $e(g) \geq \epsilon$, where $\epsilon > 0$ is as specified in the stable manifold theorem.

The next step is to observe that there exists a neighborhood, U, of g in $Diff^1(M)$ such that any $f \in U$ is Anosov. Then since the stable and unstable manifolds $W^s_\epsilon(x,f)$ and $W^u_\epsilon(x,f)$ vary continuously with respect to $f \in U$ ([28]),[3] we can show that there exists a neighborhood, $U' \subset U$, of g such that $f \in U'$ is uniformly expansive. Since g has the pseudo-orbit shadowing property, we can apply theorem 2.2.3 to conclude that Anosov diffeomorphisms must be structurally stable. This completes our explanation of theorem 2.4.2.

Theorem 2.4.2, however, is not the most general statement we can make. We need a few more definitions, however, before we can proceed to final result in theorem 2.4.3.

Definitions:

(1) A point x is *nonwandering* if for every neighborhood, U, of x, there exists arbitrarily large n such that $f^n(U) \cap U$ is nonempty.

(2) A diffeomorphism $f : M \to M$ satisfies *Axiom A* if:
(a) the nonwandering set, $\Omega(f) \subset M$, is hyperbolic.
(b) the periodic points of f are dense in $\Omega(f)$.

(3) We say that f satisfies the *strong transversality property* if for every $x \in M$, $E^s \oplus E^u = TM$.

Theorem 2.4.3 *(Franks) If $f : M \to M$ is C^2 then f is absolutely structurally stable if and only if f satisfies Axiom A and the strong transversality property.*

Proof: See Franks [21].

Intuitively, this result seems to be similar to our discussion of Anosov systems, except that hyperbolicity is not available everywhere. However, there has been a great deal of research into questions concerning structural stability, especially whether structurally stable $f \in Diff^1(M)$ implies that f satisfies Axiom A and the strong transversality property. The reader may refer to [55] for discussions and references to this work.

[3] Instead of hiding the details in this statement about stable and unstable manifolds, [62] gives a more direct argument (but one that requires math background which I have tried to avoid in the text). Let $B(M,M)$ be the Banach manifold of all maps from M to M and let $\Phi_f : B(M,M) \to B(M,M)$ so that $\Phi_f(h) = fhg^{-1}$. If $f = g$, $\Phi_g(h)$ has a hyperbolic fixed point near the identity function, id (where by hyperbolic we mean that the spectrum of the tangent map, $T_h\Phi$, is disjoint from the unit circle). Thus for any $f \in U$, $\Phi_f(h)$ has a hyperbolic fixed point near, id, and, since g is expansive, this shows uniform expansiveness for $f \in U$.

For our purposes, however, we now summarize the implications of theorem 2.4.3 to parameter estimation:

Corollary 2.4.1 *Suppose that* $f_p \in \mathrm{Diff}^1(M)$ *for* $p \in I_p \subset \mathbb{R}$*, and let* $f(x,p) = f_p(x)$ *for any* $x \in M$*. Suppose also that* $f : M \times I_p \to M$ *is* C^1 *and that for some* $p_0 \in I_p$*,* f_{p_0} *is a* C^2 *Axiom A diffeomorphism with the strong transversality property. Then there exists* $\epsilon_0 > 0$ *and* $K > 0$ *such that for every positive* $\epsilon < \epsilon_0$*, any orbit of* f_{p_0} *can be* $\epsilon-$*shadowed by an orbit of* f_p *if* $p \in B(p_0, K\epsilon)$*.*

In other words, C^2 Axiom A diffeomorphisms with the strong transversality satisfy a function shadowing property. They are stable in such a way that their dynamics does not magnify differences in parameter values. Chaotic behavior clearly does not lead to improved parameter estimates in this case. However, as noted earlier, most known physical systems do not satisfy the rather stringent conditions of uniform hyperbolicity. In the next two chapters we will investigate results for some systems that are not uniformly hyperbolic, beginning with the simplest possible case: dynamics in one dimension.

Chapter 3

Maps of the interval

In the last chapter we examined systems that are uniformly hyperbolic. In this case, orbits of nearby systems have the same topological properties and shadow each other for arbitrarily long periods of time. We would now consider what happens for other types of systems. To start out with, we will investigate one-dimensional maps, specifically, maps of the interval. One-dimensional maps are useful because they are the simplest systems to analyze; yet as we shall see, even in one dimension there is a great variety of possible behavior, especially if one is interested in geometric relationships between the shadowing orbits of nearby systems. Such relationships are important in assessing the feasibility of parameter estimation, since they determine whether nearby systems can be distinguished from each other in parameter space.

In section 3.1 we begin with a brief overview of what maps of the interval are structurally stable, and in section 3.2 we look at function shadowing properties of these maps. Our purpose here is not to classify maps into various properties. Although it is important to know what types of systems exhibit various shadowing properties, the main goal is to distill out some archetypal mechanisms that may be present in a number of interesting nonlinear systems. Especially of interest are any mechanisms that may help us understand what occurs in higher dimensional problems.

In the process of investigating function shadowing, we will examine how the "folding" behavior around turning points (i.e., relative maxima or minima) of one-dimensional maps governs how orbits shadow each other. This investigation will be extended in section 3.3, where we consider how folding behavior can often lead naturally to asymmetrical shadowing behavior in the parameter space of maps. This, at least, gives us some hint for why we see asymmetrical behavior in a wide variety of numerical experiments. As we will see in chapter 5, this asymmetrical shadowing behavior seems to be crucial in developing methods for estimating parameters, so it is important to try to understand where the behavior comes from.

In order to get definite results, we will restrict our claims to increasingly narrow classes of mappings. In section 3.4 we will apply our results to a specific example, namely the one-parameter family of maps we examined in chapter 1:

$$f_p(x) = px(1-x).$$

Finally, in section 3.6, we conclude with a number of conjectures and suggestions for further research into parameter dependence in one-dimensional maps.

3.1 Structural stability

We first want to examine what types of maps of the interval are structurally stable. These are not the types of maps we are particularly interested in for purposes of parameter estimation, but it is good to identify which maps they are. We briefly state some known results, most of which can be found in de Melo and van Strien [33].[1]

Note that since interesting behavior for maps of the interval occurs only in non-invertible systems, we must slightly revise some of definitions of the previous section in order to account for this. In particular, instead of bi-infinite orbits, we now deal only with forward orbits. These revisions apply, for example, in the definitions for various types of shadowing. Unless we mention a new definition explicitly, the changes are as one would expect.

Let us, however, make the following new definitions, some of which may be a bit different from the analogous terms from chapter 2. In the definitions that follow (and this chapter in general) assume that $I \subset \mathbb{R}$ is a compact interval of the real line.

Definitions: Suppose that $f : I \to I$ is continuous. Then the *turning points* of f are the local extrema of f in the interior I. $C(f)$ is used to designate the set of all turning points of f on I. Let $\mathbb{C}^r(I,I)$ be the set of continuous maps on I such that $f \in \mathbb{C}^r(I,I)$ if the following two conditions hold:

(a) f is C^r (for $r \geq 0$)

(b) $f(I) \subseteq I$.

If in addition, we have that

(c) $f(Bd(I)) \subseteq Bd(I)$ (where $Bd(I)$ denotes the boundary of I),

then we say that $f \in \mathbb{C}^r(I,I)$.

For either $f, g \in \mathbb{C}^r(I,I)$ or $f, g \in \mathbb{C}^r(I,I)$, then let $d(f,g) = \sup_{x \in I} |f(x) - g(x)|$.

Definitions:

(1) $f \in \mathbb{C}^r(I,I)$ is said to be C^r *structurally stable* if there exists a neighborhood U of

[1] [33] is the best source of material I have seen for results involving one-dimensional dynamics.

f in $\mathbb{C}^r(I,I)$ such that for every $g \in U$, there exists a homeomorphism $h_g : I \to I$ such that $gh_g = h_g f$.

(2) Let $f : I \to I$. The ω-limit set of a point, $x \in I$, is:

$$w(x) = \{y \in I : \text{there exists a subsequence } \{n_i\} \text{ such that } f^{n_i}(x) \to y \text{ for some } x \in I\}$$

B is said to be the *basin of a hyperbolic periodic attractor* if $B = \{x \in I : p \in w(x)\}$ where p is a periodic point of f with period n and $|Df^n(p)| < 1$.

(3) $f \in \mathbb{C}^r(I,I)$ is said to satisfy *Axiom A* if
(a) f has a finite number of hyperbolic periodic attractors
(b) Every $x \in I$ is either a member of a (uniformly) hyperbolic set or is in the basin of a hyperbolic periodic attractor.

The following theorem is the one-dimensional analog of theorem 2.4.3.

Theorem 3.1.1 *Suppose that $f \in \mathbb{C}^r(I,I)$ ($r \geq 2$) satisfies Axiom A and the following conditions:*

(1) If $c \in I$ and $Df(c) = 0$, then $c \in C(f)$.

(2) $f^n(C(f)) \cap C(f) = \emptyset$ for all $n > 0$.

Then f is C^2 structurally stable.

Proof: See for example, theorem III.2.5 in [33].

Axiom A maps are apparently prevalent in one-dimensional systems. For example, the following is believed to be true:

Conjecture 3.1.1 *The set of parameters for which $f_p = px(1-x)$ satisfies Axiom A forms a dense set in* $[0,4]$.

Proof: de Melo and van Strien [33] report that Swiatek has recently proved this result in [61].

Assuming that this result is true, we can paint an interesting picture for the parameter space of $f_p = px(1-x)$. Apparently there are a dense set of parameter values for which $f_p = px(1-x)$ has a hyperbolic periodic attractor. The set of parameter values satisfying this property must be consist of a union of open sets, since we know that these systems are structurally stable.

On the other hand, this does not mean that all or almost all of the parameter space of $f_p = px(1-x)$ is taken up by structurally stable systems. In fact, as we shall see in section 3.4, a positive measure of the parameter space is actually taken up by systems that are not structurally stable. These are the parameter values that we will be most interested in.

3.2 Function shadowing

We now consider function and parameter shadowing. In section 2.2 we saw that for uniformly expansive diffeomorphisms, structural stability and function shadowing are equivalent. For more general systems, structural stability still implies function shadowing, however, the converse is not necessarily true. As we shall see, there are many cases where the connections between shadowing orbits of nearby systems cannot be described by a simple homeomorphism. The structure of these connections can in fact be quite complicated.

3.2.1 A function shadowing theorem

There have been several recent results concerning shadowing properties of one-dimensional maps. Among these include papers by Coven, Kan, and Yorke [17], Nusse and Yorke [39], and Chen [12]. This section extends the shadowing results of these papers in order to examine the possibility of parameter and function shadowing for parameterized families of maps of the interval.

Specifically, we will deal with two types of maps: piecewise monotone mappings and uniformly piecewise-linear mappings of a compact interval, $I \subset \mathbb{R}$ onto itself:

Definitions: A continuous map $f : I \to I$ is said to be *piecewise monotone* if f has finitely many turning points. f is said to be a *uniformly piecewise-linear* mappings if it can be written in the form:

$$f(x) = \alpha_i \pm sx \text{ for } x_i \in [c_{i-1}, c_i] \tag{3.1}$$

where $s > 1$, $c_0 < c_1 < \ldots < c_q$ and $q > 0$ is an integer. (We assume $s > 1$ because otherwise there will not be any interesting behavior).

Note that for this section, it is useful to define neighborhoods, $B(x, \epsilon)$, so that they do not extend beyond the confines of I. In other words, let $B(x, \epsilon) = (x - \epsilon, x + \epsilon) \cap I$. With this in mind, we use the following definitions to describe some relevant properties of piecewise monotone maps.

Definition: A piecewise monotone map, $f : I \to I$, is said to be *transitive* if for any two open sets $U, V \subset I$, there exists an $n > 0$ such that $f^n(U) \cap V \neq \emptyset$.

Definitions: Let $f : I \to I$ be piecewise monotone. Then f satisfies the *linking property* if for every $c \in C(f)$ and any $\epsilon > 0$ there is a point $z \in I$ and integer $n > 0$ such that $z \in B(c, \epsilon)$, $f^n(z) \in C(f)$, and $|f^i(c) - f^i(z)| < \epsilon$ for every $i \in \{1, 2, \ldots, n\}$. Suppose, in addition, that we can always choose a $z \neq c$ such that the above condition is satisfied. Then f is said to satisfy the *strong-linking* condition.

We are now ready to state the main result of this section.

Theorem 3.2.1 : *Transitive piecewise monotone maps satisfy the function shadowing property in $\mathbb{C}^0(I, I)$ if and only if they satisfy the strong linking property.*

Proof: The proof may be found in appendix B.

In particular, this theorem implies the following parameter shadowing result. Let $I_p \subset \mathbb{R}$ be a closed interval of the real line. Suppose that $\{f_p : I \to I | p \in I_p\}$ is a continuously parameterized family of one-dimensional maps, and let f_{p_0} be a transitive piecewise monotone mapping with the strong linking property. Then f_p must have the parameter shadowing property at $p = p_0$. Note that f_{p_0} is certainly not structurally stable in $\mathbb{C}^0(I, I)$.[2] The connections between the shadowing orbits are not continuous and one-to-one in general. In the next section we shall further examine what these connections are likely to look like.

Now, however, we would like to present some motivation for why theorem 3.2.1 makes sense. The key to examining the shadowing properties of transitive piecewise monotone maps is to understand the dynamics near the turning points. In regions away from the turning points, these maps look locally hyperbolic, so finite pieces of orbits in these regions shadow each other rather easily. The transitivity condition guarantees hyperbolicity away from the turning points, since any transitive piecewise monotone maps is topologically conjugate to a uniformly piecewise linear map.

Close to the turning points, however, things are more interesting. Suppose, for example, that we are given a family of piecewise monotone maps $f_p : I \to I$, and suppose that we would like to find parameter shadowing orbits for orbits of f_{p_0} that pass near a turning point, c, of f_{p_0}. Consider a neighborhood, $U \subset I$ around the turning point c. Regions of state space near c are folded on top of each other by f_{p_0} (see figure 3.1(a)). This can create problems for parameter shadowing. Consider what the images of U look like under repeated applications of f_{p_0} compared to what they might look like for two other parameter values (p_- and p_+) close to p_0 (see figure 3.1(b)). Under the different parameter values, the forward images of U become offset from each other, since orbits for parameter values near p_0 look like pseudo-orbits of f_{p_0}.

[2]In fact, no map is structurally stable in $\mathbb{C}^0(I, I)$. This is clear, since any $\mathbb{C}^0(I, I)$ neighborhood of $f \in \mathbb{C}^0(I, I)$ contains maps with arbitrary numbers of turning points. Since turning points are preserved by topological conjugacy, f cannot be structurally stable in $\mathbb{C}^0(I, I)$.

The forward images of U for different parameter values tend to consistently either *lag* or *lead* each other, a phenomenon which has interesting consequences for parameter shadowing. For example, in figure 3.1(b), since $f_{p_-}^{k_1}(U)$ lags $f_{p_0}^{k_1}(U)$, it appears that f_{p_-} has a difficult time shadowing the orbit of f_{p_0} emanating from the turning point, c. On the other hand, from the same figure, there is no reason to expect that there are any orbits of f_{p_0} which are not shadowed by suitable orbits of f_{p_+}.

However, this is not the end of the story. If the linking condition is satisfied, then the turning points are recurrent and neighborhoods of turning points keep returning to turning points to get refolded on top of themselves. This allows the orbits of lagging parameter values to catch up as regions get folded back (see figure 3.1(c)). In this case, we see that the forward image of U under f_{p_0} gets folded back into the the corresponding forward image of U under f_{p_-}, thus allowing orbits of f_{p-} to effectively shadow orbits of f_{p_0}.

On the other hand we see that there is an asymmetry in the shadowing behavior of parameter values depending on whether the folded regions around turning point lag or lead one another under the action of different parameter values. The parameter values that lag seem to have a more difficult time shadowing other orbits than the ones that lead. Making this statement more precise is the subject of the next section. Theorem 3.2.1 merely states that if the strong linking condition is satisfied, then regions near turning points are refolded back upon one another in such a way that the parameter shadowing property is satisfied.

3.2.2 An example: the tent map

In [12], Chen proves the following theorem:

Theorem 3.2.2 *The pseudo-orbit shadowing property and the linking property are equivalent for transitive piecewise monotone maps.*

One interesting thing to note is the difference between function shadowing and pseudo-orbit shadowing. For instance, what happens when a transitive map exhibits the linking property but does not satisfy the strong-linking property? We already know that such maps must exhibit the pseudo-orbit shadowing property but must not satisfy the function shadowing property on $\mathbb{C}^0(I, I)$. It is worth a brief look at why this occurs.

As an illustrative example, consider the family of tent maps, $f_p : [0,1] \to [0,1]$, where:

$$f_p(x) \quad = \quad \begin{cases} px \text{ if } x \leq \frac{1}{2} \\ p(1-x) \text{ if } x > \frac{1}{2} \end{cases}$$

for $p \in [0,2]$. Pick $p_0 \in (\sqrt{2}, 2)$ such that $f_{p_0}^5(\frac{1}{2}) = \frac{1}{2}$. It is not difficult to show that such a p_0 exists. Numerically we find that one such value for p_0 occurs near $p_0 \approx 1.5128763969$.

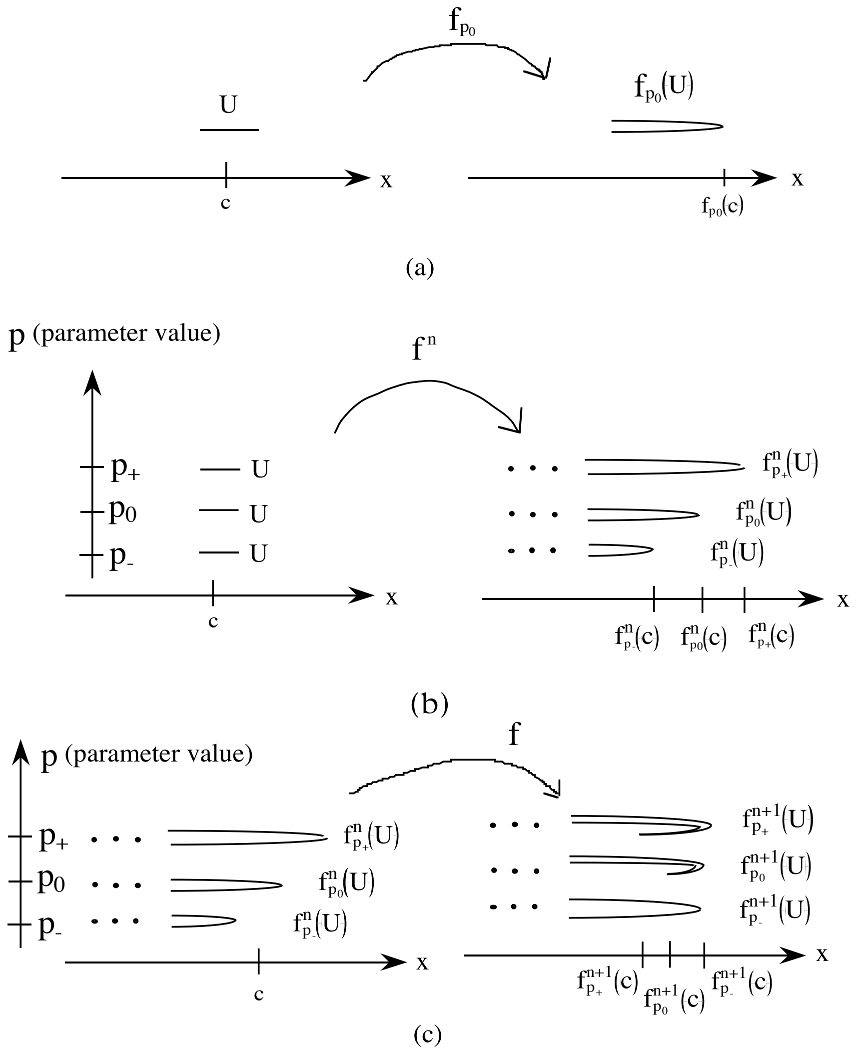

Figure 3.1: Figure 3.1(a) illustrates how neighborhoods near a turning point get folded. (b) shows what might happen for three different parameter values, $p_- < p_0 < p_+$. The images of neighborhoods near the critical point tend to get offset each from other so that the neighborhoods for certain parameters (eg., p_+) may begin to lead while other parameters (eg., p_-) lag behind. Lagging parameters have difficulty shadowing leading parameters. (c) shows how neighborhoods can get refolded on each other as a result of a subsequent encounter with a turning point, allowing lagging parameters to "catch up," so that they are able to shadow parameter values that normally lead.

We can see that f_{p_0} is transitive on the interval $I(p_0) = [f^2_{p_0}(c), f_{p_0}(c)]$ where in this case, $c = \frac{1}{2}$. Given any interval, $U \subset I(p_0)$, since $p_0 > \sqrt{2}$, if $c \notin U$ then $|f_{p_0}(U)| > \sqrt{2}|U|$ and if $c \in U$ then $|f_{p_0}(U)| > \frac{\sqrt{2}}{2}|U|$, where $|U|$ denotes the length of the interval U. Thus either $|f^4_{p_0}(U)| > 2|U|$ or $f^4_{p_0}(U) = I(p_0)$, and for any $U \subset I(p_0)$ there exists a $k \geq 0$ such that $f^k_{p_0}(U) = I(p_0)$. Consequently, f must be transitive on I. Note that even though $I(p)$ is not invariant with respect to p, theorem 3.2.1 still applies, since we could easily rescale the coordinates to eliminate this problem.

Now let p_0 be near 1.5128763969 so that $f^5_{p_0}(c) = c = \frac{1}{2}$. We would like to investigate the shadowing properties of the orbit, $\{f^k_{p_0}(c)\}_{k=0}^{\infty}$. Let $f(x,p) = f_p(x)$. Two important pieces of information are the following:

$$D_p f^5(c, p_0) = \frac{\partial f^5}{\partial p}(c, p_0) \approx -1.2715534 \tag{3.2}$$

$$\sigma_5(c, p_0) = -1 \tag{3.3}$$

where we define:

$$\sigma_i(c,p) = \begin{cases} \text{1 if } c \text{ is a relative maximum of } f^i_p \\ \text{-1 if } c \text{ is a relative minimum of } f^i_p \end{cases}$$

As we shall see in the next section, statistics like (3.2) and (3.3) are important references in evaluating the shadowing behavior for families of maps. For this example, let us consider a combined state and parameter space and examine how a small square in this space around $(x,p) = (c, p_0)$ gets iterated by the map f. We see that because $f^5_{p_0}$ has a relative minimum at $c = \frac{1}{2}$ and because $D_p f^5(c, p_0)$ is negative, parameter values higher than p_0 tend to lead while parameter values less than p_0 tend to lag behind in the manner described earlier in this section. Since the turning point of f_{p_0} at c is periodic with period 5, this type of lead/lag behavior continues for arbitrarily many iterates.

We want to know if nearby maps, f_p, for p near p_0 have orbits that shadow $\{f^k_{p_0}(c)\}_{k=0}^{\infty}$. Consider how the lead/lag behavior affects possible shadowing orbits. Because $c = \frac{1}{2}$ is periodic, it is possible to verify that the quantity, $[\sigma_n(c, p_0^-) D_p f^n(c, p_0^-)]$, grows exponentially as n gets large (where p_0^- indicates that we evaluate the derivative for p arbitrarily close to, but less than p_0). Thus for maps with parameter values $p < p_0$, all possible shadowing orbits diverge away from $\{f^k_{p_0}(c)\}_{k=0}^{\infty}$ at a rate that depends exponentially on the number of iterates. Consequently there exists a $\delta > 0$ such that if $p \in (p_0 - \delta, p_0)$, then no orbit of f_p $\epsilon-$shadows $\{f^k_{p_0}(c)\}_{k=0}^{\infty}$ for any $\epsilon > 0$ sufficiently small. On the other hand the orbit $\{f^k_{p_0}(c)\}_{k=0}^{\infty}$ can be shadowed by f_p for parameter values $p \geq p_0$. In fact, because everything is linear, it is not difficult to show that there must exist a constant $K > 0$ such that that for any $\epsilon > 0$, there is an orbit of f_p that $\epsilon-$shadows $\{f^k_{p_0}(c)\}_{k=0}^{\infty}$ if $p \in [p_0, p_0 + K\epsilon]$.

In summary, we see that the orbit, $\{f^k_{p_0}(c)\}_{k=0}^{\infty}$, cannot be shadowed by parameter values $p < p_0$, but can be shadowed for parameter values $p \geq p_0$. f_{p_0} satisfies the linking but not the strong linking property. Thus f_{p_0} satisfies the pseudo-orbit shadowing property, and any orbit of f_p for p near p_0 can be shadowed by an orbit of f_{p_0}. On the other hand, f_{p_0} does not satisfy function or parameter shadowing properties, since not all nearby systems (for example, f_p for $p < p_0$) have orbits that shadow orbits of f_{p_0}. Also, note how the lead and lag behavior in parameter space results naturally in asymmetrical shadowing properties in parameter space. We will look at this more closely in the next section.

As a final note and preview for the next section, consider briefly how the above example might generalize to other situations. The tent map example may be considered exceptional for two primary reasons: (1) the tent map is uniformly hyperbolic everywhere except for at the turning point, and (2) the turning point of f_{p_0} is periodic. We are generally interested in more generic situations involving parameterized families of piecewise monotone maps, especially maps with positive Lyapunov exponents. Apparently a number of likely scenarios also result in lead/lag behavior in parameter space, producing asymmetries in shadowing behavior similar to that observed in the tent map example. However, this behavior generally gets distorted by local geometry. Also things become more complicated because of folding caused by close returns to turning points. In particular for maps with positive Lyapunov exponents, shadowing orbits for lagging parameter values tend to diverge away at exponential rates, just like in the tent map example, but this only occurs for a certain number of iterates until a close return or linking with a turning point occurs. In such cases, function shadowing properties may exist, but the geometry of the shadowing orbits still reflects the asymmetrical lead/lag behavior. This behavior certainly affects any attempts at parameter estimation.

3.3 Asymmetrical shadowing

In the previous two sections we were primarily interested in topologically-oriented results about whether orbits of nearby one-dimensional systems shadow each other or not. However, topological results really do not provide enough information for us to draw any strong conclusions about the feasibility of estimation problems. Whether orbits shadow each other or not, in general we would also like to know the answers to more specific questions, for example: what is the expected rate of convergence for a parameter estimate, and how does the level of noise or measurement error affect the possible accuracy of a parameter estimate?

In this section we address a more analytical treatment of the subject of shadowing and parameter dependence in one-dimensional maps. The problem with this, of course, is that there is an extremely rich variety of possible behavior in parameterized families of mappings, and it is difficult to say anything concrete without limiting the statements to relatively small classes of maps. Thus some compromises have to be made. However, we approach our investigation with some specific goals in mind. In particular we are interested in definite bounds on how fast the closest shadowing trajectories in nearby systems diverge from each other and some explanation concerning how the observed asymmetrical shadowing behavior gets established in the parameter space. We will concentrate on smooth maps of the interval, especially the quadratic map, $f_p(x) = px(1-x)$.

3.3.1 Lagging parameters

In this subsection, we argue that asymmetries are likely to occur in parameter space. In particular, given a smooth piecewise monotone map with a positive Lyapunov exponent, shadowing orbits for nearby lagging maps tend to diverge away from orbits of the original system at an exponential rate before being folded back by close encounters with turning points.

Preliminaries

We will primarily restrict ourselves to maps with the following properties:

(C0) $g : I \to I$, is piecewise monotone.

(C1) g is C^2 on I.

(C2) Let $C(g)$ be the finite set such that $c \in C(g)$ if and only if g has a local extremum at $c \in I$. Then $g''(c) \neq 0$ if $c \in C(g)$ and $g'(x) \neq 0$ for all $x \in I \setminus C(g)$.

We are also interested in maps that have positive Lyapunov exponents. In particular, we will examine maps satisfying a set of closely related properties known as the Collet-Eckmann conditions, (CE1) and (CE2). We will say that a map g satisfies (CE1) or (CE2), if there exist constants $K_E > 0$ and $\lambda_E > 1$ such that for some $c \in C(g)$:

(CE1) $|Dg^n(g(c))| \geq K_E \lambda_E^n$,

(CE2) $|Dg^n(z)| \geq K_E \lambda_E^n$ if $g^n(z) = c$.

respectively for any $n > 0$.

We also consider one-parameter families of mappings, $f_p : I_x \to I_x$, parameterized by $p \in I_p$, where $I_x \subset \mathbb{R}$ and $I_p \subset \mathbb{R}$ are closed intervals of the real line. Let $f(x,p) = f_p(x)$ where $f : I_x \times I_p \to I_x$. We are primarily interested in one-parameter families of maps with the following characteristics:

(D0) For each $p \in I_p$, $f_p : I_x \to I_x$ satisfies (C0) and (C1). We also require that $C(f_p)$ remains invariant with respect to p for all $p \in I_p$.

(D1) $f : I_x \times I_p \rightarrow I_x$ is C^2 for all $(x,p) \in I_x \times I_p$.

Note that the following notation will be used to express derivatives of $f(x,p)$ with respect to x and p.

$$D_x f(x,p) = \frac{\partial f}{\partial x}(x,p) \tag{3.4}$$

$$D_p f(x,p) = \frac{\partial f}{\partial p}(x,p). \tag{3.5}$$

The Collet-Eckmann conditions specify that derivatives with respect to the state, x, grows exponentially. Similarly we will also be interested in families of maps where derivatives with respect to the parameter, p, also grow exponentially. In other words, we require that there exist constants $K_p > 0$, $\lambda_p > 1$, and $N > 0$ such that for some $p_0 \in I_p$, and $c \in C(f_{p_0})$:

(CP1) $|D_p f^n(c, p_0)| > K_p \lambda_p^n$

for all $n \geq N$. From now on, given a parameterized family of maps, $\{f_p | p \in I_p\}$, we will say that f_{p_0} satisfies (CP1) if the above condition holds.

This may seem to be a rather strong constraint, but in practice it often follows whenever (CE1) holds. We can see this by expanding with the chain rule:

$$D_p f^n(c,p_0) = D_x f(f^{n-1}(c,p_0),p_0) D_p f^{n-1}(c,p_0) + D_p f(f^{n-1}(c,p_0),p_0) \tag{3.6}$$

to obtain the formula for $D_p f^n(x, p_0)$:

$$D_p f^n(x,p_0) = D_p f(f^{n-1}(c,p_0),p_0) + \sum_{i=0}^{n-2} [D_p f(f^i(c,p_0),p_0) \prod_{j=i+1}^{n-1} D_x f(f^j(c,p_0),p_0)].$$

Thus, if $|D_x f^n(f(c,p_0),p_0)|$ grows exponentially, we expect $|D_p f^n(x,p_0)|$ to also grow exponentially unless the parameter dependence is degenerate in some way (eg, if $f(x,p)$ is independent of p).

Now for any $c \in C(f_{p_0})$, define $\sigma_n(c,p)$ recursively as follows:

$$\sigma_{n+1}(c,p) = sgn\{D_x f(f^n(c,p),p)\} \sigma_n(c,p) \tag{3.7}$$

where

$$\sigma_1(c,p) = \begin{cases} \text{1 if } c \text{ is a relative maximum of } f_p \\ \text{-1 if } c \text{ is a relative minimum of } f_p \end{cases}$$

Basically $\sigma_n(c,p) = 1$ if f_p^n has a relative maximum at c and $\sigma_n(c,p) = -1$ if f_p^n has a relative minimum at c. We can use this notion to distinguish a one direction in parameter space from the other.

Definition: Let $\{f_p : I_x \to I_x | p \in I_p\}$ be a one-parameter family of mappings satisfying (D0) and (D1). Suppose that there exists $p_0 \in I_p$ such that f_{p_0} satisfies (CE1) and (CP1) for some $c \in C(f_{p_0})$. Then we say that the turning point, c, of f_{p_0} *favors higher parameters* if there exists $N' > 0$ such that

$$sgn\{D_p f^n(c, p_0)\} \quad = \quad \sigma_n(c, p) \tag{3.8}$$

for all $n \geq N'$. Similarly, the turning point, c, of f_{p_0} *favors lower parameters* if

$$sgn\{D_p f^n(c, p_0)\} \quad = \quad -\sigma_n(c, p) \tag{3.9}$$

for all $n \geq N'$.

The first thing to notice about these two definitions is that they are exhaustive if (CP1) is satisfied. That is, if (CP1) is satisfied for some $p_0 \in I_p$ and $c \in C(f_{p_0})$, then the turning point, c, of f_{p_0} either favors higher parameters or favors lower parameters. We can see this from (3.6). Since $|D_p f(x, p_0)|$ is bounded for $x \in I_x$, if $|D_p f^n(x, p_0)|$ grows large enough then its sign is dominated by the signs of $D_x f(f^{n-1}(c, p_0), p_0)$ and $D_p f^{n-1}(c, p_0)$, so that either (3.8) or (3.9) must be satisfied.

Finally, if $p_0 \in I_p$ and $c \in C(f_{p_0})$, then for any $\epsilon \geq 0$, define $n_e(c, \epsilon, p_0)$ to be the smallest integer $n \geq 1$ such that $|f^n(c, p_0) - c_*| \leq \epsilon$ for any $c_* \in C(f_{p_0})$. We say that $n_e(c, \epsilon, p_0) = \infty$ if no such $n \geq 1$ exists.

Main result

We are now ready to state main results of this subsection.

Theorem 3.3.1 *Let $\{f_p : I_x \to I_x | p \in I_p\}$ be a one-parameter family of mappings satisfying (D0) and (D1). Suppose that (CP1) is satisfied for some $p_0 \in I_p$ and $c \in C(f_{p_0})$. Suppose further that f_{p_0} satisfies (CE1) at c, and that the turning point, c, favors higher parameters under f_{p_0}. Then there exists $\delta p > 0$, $\lambda > 1$, $K' > 0$, and $K \geq 1$, such that if $p \in (p_0 - \delta p, p_0)$, then for any $\epsilon > 0$, the orbit $\{f_{p_0}^n(c)\}_{n=0}^{\infty}$ is not $\epsilon-$shadowed by any orbit of f_p if $|p - p_0| > K' \epsilon \lambda^{-n_e(c, K\epsilon, p_0)}$.*

The analogous result also holds if f_{p_0} favors lower parameters.

Proof: The proof of this result can be found in appendix C.

The proof is actually relatively straightforward, although the details of the analysis becomes a bit tedious. The basic idea is that away from the turning points, everything is hyperbolic, and we can uniformly bound derivatives with respect to state and parameters to grow at an exponential rate. In particular, the lagging behavior for lower parameters is preserved and becomes exponentially more pronounced with increasing numbers of iterates. Shadowing orbits for parameters $p < p_0$ diverge away exponentially fast if higher parameters are favored. However, this only works for orbits that don't return closely to the turning points where derivatives are small.

3.3.2 Leading parameters

Motivation

We have shown in the previous section that if $f : I_x \times I_p \to I_x$ is a one parameter family of maps of the interval and if there exists $N > 0$ such that

$$D_p f^n(c, p_0) > \sigma_n(c, p_0) K \lambda^n \tag{3.10}$$

for all $n > N$, then for $p < p_0$, orbits of f_p tend to diverge at an exponential rate away from orbits of f_{p_0} that pass near the turning point, c. Such orbits of f_{p_0} can only be shadowed by orbits of f_p for $p < p_0$ if the orbits of f_{p_0} are folded back upon themselves by a subsequent encounter with the turning point.

On the other hand, we would like to find a condition like (3.10) under which orbits of f_p for $p \geq p_0$, can shadow any orbit of f_{p_0} indefinitely without relying on folding. This type of phenomenon is indicated by numerical experiments on a variety of systems. Unfortunately however, the derivative condition in (3.10) is local, so we have little control over the long term behavior of orbits. Thus, we must replace this condition with something that acts over an interval in parameter space.

For instance, we are interested in addressing systems like the family of quadratic maps:

$$f(x, p) = px(1 - x). \tag{3.11}$$

It is known that the family of quadratic maps in (3.11) satisfies a property known as the monotonicity of *kneading invariants* in the parameter space of f_p. This condition is sufficient to make one direction in parameter space preferred over the other. We show in this subsection that monotonicity of kneading invariant along with (CE1) is sufficient to guarantee strong shadowing effects for parameters that lead, at least in the case of unimodal (one turning point) maps with negative Schwarzian derivative, a class of maps that include (3.11). Maps with negative Schwarzian derivative have been the focal point of considerable research over the last several years, since they represent some of the simplest smooth maps which have interesting dynamical properties. We take advantage of analytical tools developed recently in order to analyze the relevant shadowing properties.

Definitions and statement of results

Definition: Suppose that $g : I \to I$ is C^3 and $I \subset \mathbb{R}$. Then the *Schwarzian derivative*, Sg, of g is given by the following:

$$Sg(x) = \frac{g'''(x)}{g'(x)} - \frac{3}{2}\left(\frac{g''(x)}{g'(x)}\right)^2.$$

where $g'(x), g''(x), g'''(x)$ here indicate the first, second, and third derivatives of x.

In this section we will primarily restrict ourselves to mappings with the following properties:

(A0) $g : I \to I$, is $C^3(I)$ where $I = [0, 1]$, with $g(0) = 0$ and $g(1) = 0$.

(A1) g has one local maximum at $x = c$; g is strictly increasing on $[0, c]$ and strictly decreasing on $[c, 1]$;

(A2) $g''(c) < 0$, $|g'(0)| > 1$.

(A3) The Schwarzian derivative of g is negative, $Sg(x) < 0$, over all $x \in I$ (we allow $Sg(x) = -\infty$).

Again we will be investigating one-parameter families of mappings, $f : I_x \times I_p \to I_x$, where p is the parameter and $I_x, I_p \subset \mathbb{R}$ are closed intervals. Let $f_p(x) = f(x, p)$ where $f_p : I_x \to I_x$. We are primarily be interested in one-parameter families of maps with the following characteristics:

(B0) For each $p \in I_p$, $f_p : I_x \to I_x$ satisfies (A0), (A1), (A2), and (A3) where $I_x = [0, 1]$. For each p, we also require that f_p has a turning point at c, where c is constant with respect to p.

(B1) $f : I_x \times I_p \to I_x$ is C^2 for all $(x, p) \in I_x \times I_p$.

Another concept we shall need is that of the *kneading invariant.* Kneading invariants and many associated topics are discussed in Milnor and Thurston [34].

Definition: If $g : I \to I$ is a piecewise monotone map with exactly one turning point at c, then the *kneading invariant*, $D(g, t)$, of g is defined as follows:

$$D(g, t) = 1 + \theta_1(g)t + \theta_2(g)t + \ldots + \theta_n(g)t^n + \ldots$$

where

$$\begin{aligned} \theta_n(g) &= \epsilon_1(g)\epsilon_2(g)\ldots\epsilon_n(g) \\ \epsilon_n(g) &= \lim_{x \to c^+} sgn(Dg(g^n(x))) \end{aligned}$$

for $n \geq 1$. If c is a relative maximum of g, then one interpretation of $\theta_n(g)$ is that it represents whether g^{n+1} has a relative maximum ($\theta_n(g) = +1$) or minimum ($\theta_n(g) = -1$) at c.

We can also order these kneading invariants in the following way. We will say that $|D(g, t)| < |D(h, t)|$ if $\theta_i(g) = \theta_i(h)$, for $1 \leq i < n$, but $\theta_n(g) < \theta_n(h)$. A kneading invariant, $D(f_p, t)$, is said to be monotonically decreasing with respect to p if $p_1 > p_0$ implies $|D(f_{p_1}, t)| \leq |D(f_{p_0}, t)|$.

We are now ready to state the main result of this subsection:

Theorem 3.3.2 *Let $\{f_p : I_x \to I_x | p \in I_p\}$ be a one-parameter family of mappings satisfying (B0) and (B1). Suppose that $p_0 \in I_p$ such that f_{p_0} satisfies (CE1). Also, suppose that the kneading invariant, $D(f_p, t)$, is monotonically decreasing with respect to p in some neighborhood of $p = p_0$. Then there exists $\delta p > 0$ and $C > 0$ such that for every $x_0 \in I_x$ there is a set, $W(x_0) \subset I_x \times I_p$, satisfying the following conditions:*

(1) $W(x_0) = \{(\alpha_{x_0}(t), \beta_{x_0}(t)) | t \in [0, 1]\}$ *where* $\alpha_{x_0} : [0, 1] \to I_x$ *and* $\beta_{x_0} : [0, 1] \to I_p$ *are continuous and* $\beta_{x_0}(t)$ *is monotonically increasing with respect to* t *with* $\beta_{x_0}(0) = p_0$ *and* $\beta_{x_0}(1) = p_0 + \delta p$.

(2) *For any* $x_0 \in I_x$, *if* $(x, p) \in W(x_0)$ *then* $|f^n(x, p) - f^n(x_0, p_0)| < C(p - p_0)^{\frac{1}{3}}$ *for all* $n \geq 0$.

Proof: See appendix D

Corollary 3.3.1 *Let $\{f_p : I_x \to I_x | p \in I_p\}$ be a one-parameter family of mappings satisfying (B0) and (B1). Suppose that $p_0 \in I_p$ such that f_{p_0} satisfies (CE1). Also, suppose that the kneading invariant, $D(f_p, t)$, is monotonically decreasing with respect to p in some neighborhood of $p = p_0$. Then there exists $\delta p > 0$ and $C > 0$ such that if $p \in [p_0, p_0 + \delta p]$, then for any $\epsilon > 0$, every orbit of f_{p_0} is ϵ-shadowed by an orbit of f_p if $|p - p_0| < C\epsilon^3$.*

Proof: This is an immediate consequence of theorem 3.3.2.

Overview of proof

We now outline some of the ideas behind the proof of theorem 3.3.2. The proof depends on an examination of the structure of the preimages of the turning point, $x = c$, in the combined space of state and parameters ($I_x \times I_p$ space). The basic idea is to find connected shadowing sets in state-parameter space. These sets have the property that points in the set shadow each other under arbitrarily many applications of f. Certain geometrical properties of these sets can be determined by squeezing the sets between structures of preimage points. In order to discuss the approach further, we first need to introduce some notation.

We consider the set of preimages, $P(n) \subset I_x \times I_p$ satisfying:

$$P(n) = \{(x, p) | f^i(x, p) = c \text{ for some } 0 \leq i \leq n\}.$$

It is also useful to have a way of specifying a particular section of path-connected preimages, $R(n, x_0, p_0) \subset P(n)$, extending from a single point, $(x_0, p_0) \in P(n)$. Let us define $R(n, x_0, p_0)$ so that $(x', p') \in R(n, x_0, p_0)$ if and only if $(x', p') \in P(n)$ and there exists a continuous function, $g : I_p \to I_x$, such that $g(p_0) = x_0$, $g(p') = x'$, and

$$\{(x, p) | x = g(p), p \in [p_0; p']\} \subset P(n),$$

where $[p_0; p']$ may denote either $[p_0, p']$ or $[p', p_0]$, whichever is appropriate.

The first step is to investigate the basic structure of $P(n)$. We show that $P(n)$ contains no regions or interior points and that $P(n)$ cannot contain any isolated points or curve segments. Instead, each point in $P(n)$ must be part of a continuous curve that stretches for the length of the parameter space, I_p. In fact, if $(x_0, y_0) \in P(n)$, then $R(n, x_0, p_0) \cap (I_x \times \{\sup I_p\}) \neq \emptyset$ and $R(n, x_0, p_0) \cap (I_x \times \{\inf I_p\}) \neq \emptyset$.

The next step is to demonstrate that if the kneading invariant of f_p, $D(f_p, t)$, is monotonically decreasing (or increasing), then $P(n)$ has a special topology. It must take on a tree-like structure so that as we travel along one direction in parameter space, branches of $P(n)$ must either always merge or always split away from each other. For example if $D(f_p, t)$ is monotonically decreasing, then branches of $P(n)$ can only split away from each other as we increase the parameter p. In other words, $R(n, y_-, p_0)$ and $R(n, y_+, p_0)$ do not intersect each other in the space, $I_x \times \{p\}$, for for $p \geq p_0$ if $y_+ \neq y_-$ and $y_+, y_- \in I_x$.

Now suppose we want to examine the points that shadow (x_0, p_0) under the action of f given any $x_0 \in I_x$. We first develop bounds on derivatives for differentiable sections of $R(n, x, p_0)$. We then use knowledge about the behavior of $R(n, x, p_0)$ to bound the behavior of the shadowing points. We demonstrate that for maps, f_p, with kneading invariants that decrease monotonically in parameter space, there exist constants $C > 0$ and $\delta p > 0$ such that if $x_0 \in I_x$ and

$$U(p) = \{x | \, |x - x_0| < C(p - p_0)^{\frac{1}{3}}\} \tag{3.12}$$

for any $p \in I_p$, then for any $p' \in [p_0, p_0 + \delta p]$, there exists $x'_+ \in U(p')$ such that $(x'_+, p') \in R(n_+, y_+, p_0)$ for some $y_+ > x_0$ and $n_+ > 0$ assuming that $f^{n_+}(y_+, p_0) = c$. Likewise there exists $x'_+ \in U(p')$ such that $(x'_-, p') \in R(n_-, y_-, p_0)$ for some $y_- < x_0$ and $n_- > 0$ where $f^{n_-}(y_-, p_0) = c$.

However, setting $n = \max\{n_+, n_-\}$, since $R(n, y_-, p_0)$ and $R(n, y_+, p_0)$ do not intersect each other for $p \geq p_0$ and $y_- \neq y_+$, then we also know that for any $y_- < y_+$, there is a region in $I_x \times I_p$ space bounded by $R(n, y_-, p_0)$, $R(n, y_+, p_0)$, and $p \geq p_0$. Take the limit of this region as $y_- \to x_0^-$, $y_+ \to x_0^+$, and $n \to \infty$. Call the resulting region $S(x_0)$. We observe that $S(x_0)$ is a connected set that is invariant under f and is nonempty for every parameter value $p \in I_p$ such that $p \geq p_0$ (by invariant we mean that $f(S(x_0)) = S(f(x_0, p_0))$. Thus, since $S(x_0)$ is bounded by (3.12), there exists a set of points, $S(x_0)$, in combined state and parameter space that shadow any trajectory,

$\{f_{p_0}^n(x_0)\}_{n=0}^{\infty}$ of f_{p_0}. Finally we observe that there exists a subset of $S(x_0)$ that can be represented by the form given for $W(x_0)$.

3.4 Example: quadratic map

In this section we examine how the results of section 3.3 apply to the quadratic map, $f_p : [0,1] \to [0,1]$, where:

$$f_p(x) = px(1-x) \tag{3.13}$$

and $p \in [0,4]$. For the rest of this section, f_p will refer to the map given in (3.13), and $f(x,p) = f_p(x)$ for any $(x,p) \in I_x \times I_p$ where $I_x = [0,1]$ and $I_p = [0,4]$.

We have already seen in conjecture 3.1.1, that there appears to be dense set of parameters in I_p for which f_p is structurally stable and has a hyperbolic periodic attractor. However, by the following result, we find that there is also a large set of parameters for which f_p satisfies the Collet-Eckmann conditions and is not structurally stable:

Theorem 3.4.1 *Let E be the set of parameter values, p, such that (CE1) is satisfied for the family of quadratic maps, f_p, given in (3.13). Then E is a set of positive Lebesgue measure. Specifically, E has a density point at $p = 4$ so that:*

$$\lim_{\epsilon \to 0} \frac{\lambda(E \cap [4-\epsilon, 4])}{\epsilon} = 1. \tag{3.14}$$

where $\lambda(S)$ represents the Lebesgue measure of the set S.

Proof: The first proof of this result was given in [5]. The reader should also consult the proof given in [33].[3]

Apparently, if we pick a parameter, p_0, at random from I_p (with uniform distribution on I_p) there is a positive probability that f_{p_0} will satisfy (CE1). We might note that numerical evidence suggests that the set of parameters, p, resulting in maps, f_p, which satisfy (CE1) are not just concentrated in a small neighborhood of $p = 4$.

In any case, applying the results of the last section, we see that for a positive measure of parameter values, there is a definite asymmetry with respect to shadowing results in parameter space. The following theorem illustrates this fact.

[3] These two references actually deal with the family of maps, $g_a(x) = 1 - ax^2$, where a is the parameter. However, the maps g_a and f_p are topologically conjugate if $a = p^2 - 2p$. The conjugating homeomorphism in this case is simply a linear function. Thus the results in the references immediately apply to the family of quadratic maps, $f_p : I_x \to I_x$ for $p \in I_p$.

Theorem 3.4.2 *Let $I_p = [0,4]$, $I_x = [0,1]$, and $f_p : I_x \to I_x$ be the family of quadratic maps such that $f_p(x) = px(1-x)$ for $p \in I_p$. Then there exist constants $\delta > 0$, $C > 0$, $K > 0$, and set $E(\gamma) \subset I_p$ with positive Lebesgue measure for every $\gamma > 1$ such that:*

(1) If $\gamma > 1$ and $p_0 \in E(\gamma)$, then f_{p_0} satisfies (CE1).

(2) If f_{p_0} satisfies (CE1), then for any $\epsilon > 0$ sufficiently small, any orbit of f_{p_0} can be $\epsilon-$shadowed by an orbit of f_p for $p \in [p_0, p_0 + C\epsilon^3]$.

(3) If $\gamma > 1$ and $p_0 \in E(\gamma)$, then for any $\epsilon > 0$, almost no orbits of f_{p_0} can be $\epsilon-$shadowed by any orbit of f_p for $p \in (p_0 - \delta, p_0 - (K\epsilon)^\gamma)$. That is, the set of possible initial conditions, $x_0 \in I_x$, such that the orbit $\{f^i_{p_0}(x_0)\}_{i=0}^\infty$ can be $\epsilon-$shadowed by some orbit of f_p comprises at most a set of Lebesgue measure zero on I_x if $p \in (p_0 - \delta, p_0 - (K\epsilon)^\gamma)$.

Proof of theorem 3.4.2: The full proof for this result can be found in appendix E.

Before we take a look at an overview of the proof for theorem 3.4.2, it is useful to make a few remarks. First of all, one might wonder whether the asymmetrical situation in theorem 3.4.2 is really generic for all $p_0 \in I_p$ such that f_{p_0} satisfies (CE1). For example, are there other parameter values in I_p for which it is easier to shadow lower parameter values than it is to shadow higher parameter values? Numerical evidence indicates that most if not all $p \in I_p$ exhibit asymmetrical shadowing properties if f_p has positive Lyapunov exponents. Furthermore, it seems that these parameter values favor the same specific direction in parameter space. In fact it is easy to show analytically that condition (2) of theorem 3.4.2 actually holds for all $p_0 \in I_p$ for which f_{p_0} satisfies (CE1). In other words, for f_{p_0} satisfying (CE1), there exists $C > 0$ such that for any $\epsilon > 0$ sufficiently small, f_{p_0} can be $\epsilon-$shadowed by an orbit of f_p if $p \in [p_0, p_0 + C\epsilon^3]$.

We now outline the strategy for the proof of theorem 3.4.2. For parts (1) and (3) we basically want to combine theorem 3.3.1 and theorem 3.4.1 in the appropriate way. There are four major steps. We first bound the return time of the orbit of the turning point, $c = \frac{1}{2}$, to neighborhoods of c. Next we show that f_p satisfies (CP1) and favors higher parameters on a positive measure of parameter values. This allows us to apply theorem 3.3.1. Finally we show that almost every orbit of these maps approach arbitrarily close to c so that if the orbit, $\{f^i_{p_0}(c)\}_{i=0}^\infty$, cannot be shadowed then almost all other orbits of f_{p_0} cannot be shadowed either.

We bound the return time of the orbit of the turning point, c, to neighborhoods of c by examining the proof of theorem 3.4.1. Specifically, as part of the proof of theorem 3.4.1, Benedicks and Carleson [5] show that for any $\alpha > 0$, there is a set of positive measure in parameter space, $S(\alpha) \in I_p$, such that if $p_0 \in S(\alpha)$ then f_{p_0} satisfies (CE1) and the condition:

$$|f^i_{p_0}(c) - c| > e^{-\alpha i} \tag{3.15}$$

for all $i \in \{0, 1, 2, \dots\}$. The set, $S(\alpha)$, has a density point at parameter value $p = 4$.

Next we show that f_p satisfies (CP1) and favors higher parameters on a subset of $S(\alpha)$ of positive measure. This is basically done by looking at what happens for $p = 4$ and extrapolating that result for parameters in a small interval in parameter space around $p = 4$. The result only works for those values of p for which f_p satisfies (CE1). However, since $p = 4$ is a density point of $S(\alpha)$, for any $\alpha > 0$, there is a set, $S_*(\alpha)$, contained in a neighborhood $p = 4$ with a density at $p = 4$ for which $p_0 \in S_*(\alpha)$ implies f_{p_0} satisfies (CE1) and (3.15), and f_p favors higher parameters and satisfies (CP1) at $p = p_0$.

Then by applying theorem 3.3.1 we see that there exist constants $\delta > 0$, $K_0 > 0$ and $K_1 > 0$ such that for any $\alpha > 0$, if $p_0 \in S_*(\alpha)$ then the orbit, $\{f^i_{p_0}(c)\}_{i=0}^{\infty}$, cannot be shadowed by any orbit of f_p for $p \in (p_0 - \delta, p_0 - K_0\epsilon\lambda^{-n_e(c,K_1\epsilon,p_0)})$ (recall that $n_e(c, \epsilon, p_0)$ is defined to be the smallest integer $n \geq 1$ such that $|f^n(c, p_0) - c| \leq \epsilon$.) By controlling $\alpha > 0$ in (3.15) we can effectively control $n_e(c, \epsilon, p_0)$ to be whatever we want. Thus for any $\gamma > 0$ we can choose a set $E(\gamma) \subset I_p$ with a density point at $p = 4$ such that if $p_0 \in E(\gamma)$ then f_{p_0} satisfies (CE1) and no orbits of f_p $\epsilon-$shadow the orbit, $\{f^i_{p_0}(c)\}_{i=0}^{\infty}$, for any $p \in (p_0 - \delta, p_0 - K_0(K_1\epsilon)^{\gamma})$. But since $\gamma > 1$, if we set constant $K = \max\{K_0K_1, K_1\} > 0$ we see that $p_0 - K_0(K_1\epsilon)^{\gamma} > p_0 - (K\epsilon)^{\gamma}$ for any $\epsilon > 0$. Thus, no orbits of f_p may $\epsilon-$shadow $\{f^i_{p_0}(c)\}_{i=0}^{\infty}$, if $p \in (p_0 - \delta, p_0 - (K\epsilon)^{\gamma})$.

Finally it is known that if f_{p_0} satisfies (CE1) then almost every orbit of f_{p_0} approaches arbitrarily close to c. Thus for almost all $x_0 \in I_x$, the orbit, $\{f^i_{p_0}(x_0)\}_{i=0}^{\infty}$, cannot be shadowed by an orbit of f_p if the orbit, $\{f^i_{p_0}(c)\}_{i=0}^{\infty}$, cannot be shadowed by any orbit of f_p. Consequently, we see that for any $\gamma > 1$ if $p_0 \in E(\gamma)$ then f_{p_0} satisfies (CE1) and almost no orbits of f_{p_0} can be shadowed by any orbit of f_p if $p \in (p_0 - \delta, p_0 - (K\epsilon)^{\gamma})$. This would prove parts (1) and (3) of the theorem.

Part (2) of theorem 3.4.2 is a direct result of corollary 3.3.1 and the following result, due to Milnor and Thurston [34]:

Lemma 3.4.1 *The kneading invariant, $D(f_p, t)$, is monotonically decreasing with respect to p for all $p \in I_p$.*

Thus if f_{p_0} satisfies (CE1) for some $p_0 \in E(\gamma)$, there exists a constant $C > 0$ such that any orbit of f_{p_0} can be $\epsilon-$shadowed by an orbit of f_p if $p \in [p_0, p_0 + C\epsilon^3]$. This proves part (2) of the theorem.

3.5 Remarks on convergence of parameter estimates

In order to determine the feasibility of parameter estimation applications, it is important to have some idea about how many state samples are likely to be needed in order to attain a certain accuracy in the parameter estimate. Ergodic theory comes into play here, since we would like to consider the behavior of typical orbits. In particular, suppose that a data stream is generated from an initial condition that is chosen at random after the system has settled into its equilibrium behavior. We would like to estimate the rate at which a parameter estimate is likely to converge with increasing numbers of measurements from the data stream. In this section, we outline ideas on how to approach this question, and make certain conjectures about convergence results. These conjectures closely match numerical results attained from actual parameter estimation techniques as shown in chapter 6 of this report.

We have already seen that the accuracy of a parameter estimate for a piecewise monotone map depends on how close the orbit being sampled comes to the turning points of the map. When an orbit comes close to a turning point, nearby regions in state space are subject to a folding effect that enables us to distinguish small differences in parameters based on state data. With a given level of measurement noise, ϵ, there often exists a lower limit on the parameter estimation accuracy resulting from folding and refolding effects near turning points (see theorem 3.2.1). This bound is related to the amount of time it takes for an orbit near a turning point to return within ϵ distance of a turning point. For most numerical purposes, however, this lower limit is often too small to be of practical importance. Thus, it is important to consider the approximate rate at which a parameter estimate is likely to converge, before the system reaches the lower limit in the accuracy of the parameter estimate.

Assuming that a family of piecewise monotone maps, $\{f_p | p \in I_p \subset \mathbb{R}\}$ has the same number of turning points for all $p \in I_p$, this turns out to be equivalent to asking the following question: Given a typical orbit, $\{x_n\}_{n=1}^N$, of f_{p_0} (with $p_0 \in I_p$), as N increases, for what parameter values, p, do there exist shadowing orbits, $\{y_n(p)\}_{n=1}^N$, of f_p, such that $y_n(p)$ and x_n lie on the same monotone branch of f_{p_0} for each $n \in \{1, 2 \ldots, N\}$. In other words, if $c_1 < c_2 < \ldots < c_m$ are the turning points of f_p for all $p \in I_p$, then for any $n \in \{1, 2 \ldots, N\}$, we require that $x_n \in [c_i, c_{i+1}]$ implies $y_n(p) \in [c_i, c_{i+1}]$. This makes sense because the lower limit in the accuracy of the parameter estimate results from the fact that orbits can shadow each other by evolving on different monotone branches, so that state space regions around an orbit for a map with leading parameters get refolded more than regions around shadowing orbits for maps with lagging parameters. Henceforth, given the family, f_p, of piecewise monotone maps described above, we will say that a sequence of points, $\{y_n\}_{n=1}^N$, *ϵ-monotone-shadows* an orbit $\{x_n\}_{n=1}^N$, of f_{p_0} if y_n and x_n lie on the same monotone branch of f_{p_0} for each $n \in \{1, 2 \ldots, N\}$ and if $|y_n - x_n| < \epsilon$ for each $n \in \{1, 2 \ldots, N\}$.

Using these ideas, we make the following conjectures:

Conjecture 1: *Consider the family of tent maps, $\{g_p : I_x \to I_x | p \in I_p\}$, where $I_x = [0, 1]$,*

$$g_p(x) \;=\; \begin{cases} px \;\; \mathit{if}\; x \le \frac{1}{2} \\ p(1-x) \;\; \mathit{if}\; x > \frac{1}{2} \end{cases}$$

and $I_p = (\frac{3}{2}, 2]$. Given $\epsilon > 0$ and $p_0 \in I_p$, for almost all $x_0 \in I_x$ there is a constant $K > 0$ such that for each positive integer N, there exists a $p \in (p_0 - K\frac{1}{N}, p_0]$ such that the orbit $\{g_{p_0}^n(x_0)\}_{n=0}^N$ is not ϵ-monotone-shadowed by any orbit of g_p.

It turns out that numerical results indicate that the error in the estimate of the parameters of the tent map tends to converge at a rate proportional to $\frac{1}{N}$ where N is the number of observations. Similarly we have:

Conjecture 2: *Consider the family of quadratic maps, $\{f_p : I_x \to I_x | p \in I'_p\}$, where $I_x = [0, 1]$,*

$$f_p(x) = px(1 - x) \tag{3.16}$$

and $I'_p = (2, 4]$. Then there exists a set $E \subset I'_p$ of positive Lebesgue measure such that if $p_0 \in E$, then given $\epsilon > 0$, for almost all $x_0 \in I_x$, there is a constant $K > 0$ such that for each positive integer N, there exists a $p \in (p_0 - K\frac{1}{N^2}, p_0]$ such that the orbit $\{g_{p_0}^n(x_0)\}_{n=0}^N$ is not ϵ-monotone-shadowed by any orbit of f_p.

Furthermore, we expect that the error in the parameter estimate of the quadratic map should converge at a rate proportional to $\frac{1}{N^2}$, where N is the number of observations processed. In chapter 6, we will see that this appears to agree with numerical results.

The rest of this section will be devoted to motivating these two conjectures. In order to estimate the convergence rate of the parameter estimate, we first need an estimate of how fast an orbit is likely to approach a turning point. It turns out that the maps we are interested in are *ergodic* so that the long term average behavior of almost all orbits of the maps can be described by the appropriate *invariant measure* of the map. Thus, in order to estimate how fast most orbits approach a turning point of map, it is helpful to examine the invariant measures of the map.

μ is said to be an invariant measure of the map $h : I_x \to I_x$ if $\mu(h^{-1}(A)) = \mu(A)$ for any open set $A \subset I_x$. Every ergodic map, $h : I_x \to I_x$, has an associated invariant measure, μ, such that for any continuous function $\phi : I_x \to \mathbb{R}$, the relation,

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} \phi(f^n(x_0)) = \int_{x\in I} \phi(x)\mu(dx),$$

holds for μ-almost all $x_0 \in I_x$. Thus, one might say that the "time-average" equals the "space-average" of an ergodic map. The *density*, $\frac{d\mu}{d\lambda}$, of the measure μ satisfies the property that $\int_{x\in A} \frac{d\mu}{d\lambda}(x)\mu(dx) = \mu(A)$ for any open $A \subset I_x$.[4]

Conjecture 1:

Let us now outline the motivation behind Conjecture 1. The tent map, g_p, is ergodic if $p \in I_p$. The density, $\frac{d\mu_p}{d\lambda}$, of the associated invariant measure μ_p of g_p is simply a

[4] For more information regarding invariant measures and ergodic theory of maps of the interval please refer to chapter V in [33].

constant over the region, $[g_p^2(c), g_p(c)]$. We expect that for μ_p-almost all initial conditions $x_0 \in [0,1]$ there exists a $K > 0$ such that:

$$\min_{0 \le i \le n} |g_p^i(x_0) - c| < K \frac{1}{n}$$

is satisfied for all $n \ge 0$ if $p \in I_p$.

Keeping this observation in mind, let $c = \frac{1}{2}$ be the critical point of the tent map. One can show that g_p favors higher parameters for all $p \in I_p$. In other words, using the same notation as in (3.7) and (3.8), we know that

$$sgn\{D_p g^n(c,p)\} \quad = \quad \sigma_n(c,p) \tag{3.17}$$

for all $n \ge 1$.[5] It is also not difficult to show that there exist constants $K_1 > 0$ and $K_2 > 0$ such that

$$K_1 p^n < |D_p g^n(c,p)| < K_2 p^n. \tag{3.18}$$

for all $n \ge 1$ if $p \in I_p$.

Now given $p_0 \in I_p$ and an initial condition $x_0 \in [0,1]$, consider the finite orbit $\{g_{p_0}^n(x_0)\}_{n=0}^N$. We would like to determine if there is an orbit of g_p that $\epsilon-$monotone-shadows this orbit for $p < p_0$. To first order, this is basically determined by the magnitude of

$$\Delta_N = \min_{0 \le n \le N} |g_{p_0}^n(x_0) - c| \tag{3.19}$$

because regions of state space near the critical point c get folded, producing the leading and lagging behavior which in turn leads to asymmetrical shadowing in parameter space. Since the tent map favors higher parameters, g_p cannot $\epsilon-$monotone-shadow the orbit $\{g_{p_0}^n(c + \Delta_N)\}_{n=1}^m$ for $p < p_0$ if:

$$\sigma_n(c,p)[g_{p_0}^n(c + \Delta_N) - g_p^n(c)] > \epsilon. \tag{3.20}$$

for any $n \le m$. Suppose that the inequality in (3.20) is false for all $n \le m-1$. Then, from (3.17) and (3.18),

$$\sigma_m(c,p)[g_{p_0}^m(c + \Delta_N) - g_p^m(c)] > K_1 p^m (p_0 - p) - p_0^m \Delta_N.$$

Now suppose that

$$m = \log_{p_0} \frac{\epsilon}{\Delta_N}. \tag{3.21}$$

[5]Actually g_p favors higher parameters for all $p \in (1,2]$. We confine our discussion here to $p \in I_p = (\frac{3}{2}, 2]$ for convenience since (3.17) may only hold for $n \ge N_0$ for some $N_0 > 1$ if $p \in (1, \frac{3}{2}]$. However, we suspect that Conjecture 2 also holds for any $p \in (1,2]$.

We find that:

$$\sigma_m(c,p)[g_{p_0}^m(c+\Delta_N) - g_p^m(c)] > \epsilon[(\frac{p_0 - p}{\Delta_N})(\frac{\epsilon}{\Delta_N})^{\frac{\log p}{\log p_0}-1} - 1]$$

So for Δ_N sufficiently small and $p \in I_p$, we can see that the inequality in (3.20) holds for the value m given in (3.21) if $p_0 - p > 3\Delta_N$.

Thus, given sufficiently small Δ_N as defined in (3.19), there exists a $p \in (p - 3\Delta_N, p_0]$ such that no orbit of g_p can $\epsilon-$monotone-shadow the orbit $\{x_n\}_{n=0}^{N+m}$. Recall, however, that Δ_N should decrease at rate at least proportional to $\frac{1}{N}$. Applying this fact, we get a result similar to Conjecture 1.

Conjecture 2:

Now let us consider Conjecture 2. The basic idea here is similar to Conjecture 1. However,Conjecture 2 presents some additional complications. First the invariant measures for the quadratic map are more complicated, and cannot be written in closed form. Second, there is no uniform expansion available in state or parameter space, so that it is not a simple matter to bound the quantity $f_{p_0}^n(c+\Delta_N) - f_p^n(c)$ for small Δ_N and for p near p_0.

The invariant measures of the quadratic map have been the subject of vigorous research over the past several years. Nowicki and van Strien show in [46] that for the maps given in (3.16) if f_{p_0} satisfies (CE1) for some $p_0 \in (1,2]$, then f_{p_0} has an ergodic invariant measure μ such that for any measurable set $A \subset [0,1]$ there exists a constant $K > 0$ such that $\mu(A) \leq K|A|^{\frac{1}{2}}$ (where $|A|$ is the Lebesgue of the set A).

Now consider the interval $A_\epsilon = (c - \frac{1}{2}\epsilon, c + \frac{1}{2}\epsilon)$. Note that there exists $K' > 0$ such that for any $\epsilon > 0$, $|f(A_\epsilon)| < K'\epsilon^2$. Thus, from Nowicki and van Strien's result, we know that there exists $K_1 > 0$ such that for any $\epsilon > 0$:

$$\mu(A_\epsilon) = \mu(f_{p_0}(A_\epsilon)) < K|f_{p_0}(A_\epsilon)|^{\frac{1}{2}} < K_1\epsilon$$

Furthermore, it is fairly easy to show that there also exists $K_2 > 0$ such that for any $\epsilon > 0$ $\mu(A_\epsilon) > K_2\epsilon$. Thus, since $K_2\epsilon < \mu(A_\epsilon) < K_1\epsilon$ for any ϵ, we expect that for almost all initial conditions $x_0 \in [0,1]$, the quantity,

$$\Delta_N(x_0) = \min_{0 \leq n \leq N} |f_{p_0}^n(x_0) - c|. \tag{3.22}$$

will decay at a rate proportional to $\frac{1}{N}$.

As in Conjecture 1, given $p_0 \in I_p$, $x_0 \in I_x$, and the finite orbit, $\{f_{p_0}^n(x_0)\}_{n=0}^N$, of f_{p_0}, we would like to determine if there is an orbit of f_p that $\epsilon-$monotone-shadows this orbit for some $p < p_0$. As before, the important statistic to know is $\Delta_N(x_0)$ (we will henceforth assume that x_0 is fixed and refer simply to Δ_N).

As in Conjecture 1, f_p cannot ϵ−shadow the orbit $\{f_{p_0}^n(c+\Delta_N)\}_{n=1}^m$ for $p < p_0$ if:

$$\sigma_i(c,p)[f_{p_0}^i(c+\Delta_N) - f_p^i(c)] > \epsilon. \tag{3.23}$$

for any $i \leq m$. This corresponds to what happens when an orbit, $\{f_{p_0}^n(c+\Delta_N)\}_{n=0}^i$, of f_{p_0} leads the orbit of the critical point of the map f_p by more than ϵ so there is no orbit of f_p that can effectively shadow that orbit. The other way that f_p can fail to ϵ−monotone-shadow the orbit $\{f_{p_0}^n(c+\Delta_N)\}_{n=0}^i$, is if $f_p^i(c)$ lags behind $f_{p_0}^i(c+\Delta_N)$ (by less than ϵ), but $f_p^i(c)$ and $f_{p_0}^i(c+\Delta_N)$ are on different monotone branches (ie, the critical point, $c = \frac{1}{2}$ is between $f_p^i(c)$ and $f_{p_0}^i(c+\Delta_N)$).

Thus, the prove the conjecture it is sufficient to show that given $\epsilon > 0$ sufficiently small, there exists a constant $K > 0$ such that that for each $\Delta_N > 0$ there exists a $p > p_0 - K\Delta_N^2$ and $i < -Clog\Delta_N^2$ such that one of the following is satisfied:

(1) $\sigma_i(c,p)[f_{p_0}^i(c+\Delta_N) - f_p^i(c)] > \epsilon$

(2) $\sigma_i(c,p)[f_{p_0}^i(c+\Delta_N) - f_p^i(c)] > 0$ and $sgn\{c - f_{p_0}^i(c+\Delta_N)\} = -sgn\{c - f_p^i(c)\}$.

The problem is getting a estimate for $f_{p_0}^i(c+\Delta_N) - f_p^i(c)$. Recall that near $p = 4$ there is a set $E \subset (2,4]$ of positive Lebesgue measure such that for each $p_0 \in E$, f_{p_0} satisfies (CE1), (CP1), and favors higher parameters. Thus if $p_0 \in E$, there exists a $K_0 > 0$ and $N_0 > 0$ such that

$$\frac{1}{K_0} < \frac{D_p f^i(c,p_0)}{D_x f^{i-1}(f(c,p_0),p_0)} < K_0. \tag{3.24}$$

for all $i \geq N_0$. So, if $p_0 \in E$, for $p < p_0$ and each $i > N_0$ we have that:

$$\begin{aligned}
&\sigma_i(c,p)[f_{p_0}^i(c+\Delta_N) - f_p^i(c)] \\
&\quad = \sigma_i(c,p)[(f_{p_0}(c) - f_p^i(c)) - (f_{p_0}(c) - f_{p_0}^i(c+\Delta_N)] \\
&\quad > \sigma_i(c,p)[(D_p f^i(c,p_0)(p_0-p) + O((p_0-p)^2) \\
&\qquad\quad -(K' D_x f^{i-1}(f(c,p_0),p_0)\Delta_N^2 + O(\Delta_N^3))] \\
&\quad > |D_x f^{i-1} f(c,p_0),p_0)|[(p_0-p) - K_0 K'\Delta_N^2 + O(\Delta_N^3) + O((p_0-p)^2)]
\end{aligned} \tag{3.25}$$

For each $i > N_0$, the left hand side of (3.23) tends to grow as $(p_0 - p) - K_0 K_1 \Delta_N^2$, at least for small Δ_N^2 and $p_0 - p$. Recall that $D_x f^{i-1}(f(c,p_0),p_0)$ tends to grow exponentially with i and Δ_N tends to decay proportional to $\frac{1}{N}$. Thus, given $\epsilon > 0$ one might expect that there exists $K > 0$ and $C > 0$ such that either condition (1) or (2) are satisfied for some $p > p_0 - K\Delta_N^2$ and $i < -Clog\Delta_N^2$.

This, however, is a somewhat rough calculation, and in order to demonstrate that either conditions (1) or (2) are satisfied, we need to bound the higher order terms in (3.25). This involves getting a uniform estimate of the relationship between $D_p f^i(c, p_0 - \delta p)$ and $D_x f^{i-1}(f(c+\delta x, p_0), p_0)$ for small values of δp and δx as i increases. This does not to be a trivial task and is something that should be looked into more carefully in the future.

3.6 Conclusions, remarks, and future work

The primary goal of this chapter was to examine how shadowing works in one-dimensional maps in order to evaluate the feasibility of parameter estimation on simple chaotic systems. We have been particularly interested in investigating how nonlinear folding affects parameter shadowing and how this might help explain numerical results which show asymmetrical behavior in the parameter space of one-dimensional maps. More specifically, for a parameterized family of maps, f_p, it is apparently the case that an orbit for a particular parameter value, $p = p_0$, is often shadowed much more readily by maps with slightly higher parameter values than by maps with slightly lower parameter values (or vice versa). This phenomenon has important effects on the possibilities for parameter estimation. For example, if we are given noisy observations of the orbit described above and asked what the parameter value was of the map that produced that data, then we would immediately be able to eliminate most values less than p_0 as possible candidates for the actual parameter value. On the other hand, it may be much more difficult to distinguish p_0 from parameter values slightly larger than p_0.

For piecewise monotone maps with positive Lyapunov exponents, we demonstrated that the folding behavior around a turning point generally leads to asymmetrical behavior, unless the parameter dependence is degenerate in some way. In particular, images of neighborhoods of a turning point under f_p tend to separate exponentially fast for perturbations in p. This results in a sort of lead-lag phenomenon as the images for different parameter values separate, causing images for some parameter values to overlap each other more than others. Near the turning point, orbits for parameter values that lag behind cannot shadow orbits for the parameter values that lead unless another folding occurs because of a subsequent approach to a turning point.

For the case of unimodal families of maps with negative Schwarzian derivative, the result is sharper. Apparently, if the parameter dependence is not degenerate, and if a map, f_{p_0}, has positive Lyapunov exponents for some parameter value, p_0, then for any $\epsilon > 0$ sufficiently small, there exists $C > 0$ so that for one direction in parameter space (either $p \geq p_0$ or $p \leq p_0$), all orbits of f_{p_0} can be $\epsilon-$shadowed by an orbit of f_p if $|p - p_0| < C\epsilon^3$. Meanwhile, in the other direction in parameter space, there exist constants $\delta > 0$ and $K > 0$ so that for any $\gamma > 1$ there is a positive Lebesgue measure of parameter values such that if $|p-p_0| < \delta$, then almost no orbits of f_{p_0} can be $\epsilon-$shadowed by any orbit of f_p if $|p - p_0| > (K\epsilon)^\gamma$. This clearly illustrates some sort of preference of direction in parameter space.

One might also note that this result demonstrates that all orbits of certain chaotic (nonperiodic) systems can be shadowed by orbits of systems dominated by hyperbolic periodic attractors (consider, for example, the quadratic map, $f_p(x) = px(1-x)$). Shadowing results have sometimes been cited to justify the use of computers in analyzing dynamical systems, since if one numerically iterates an orbit and finds that it is chaotic, then similar real orbits must exist in that system (or nearby systems). This is true, but one should also be careful, because the real orbits that shadow a numerically generated trajectory are often purely pathological (ie, such orbits are often not qualitatively similar to typical orbits of the system).

In any case, many questions related to this material still remain unanswered. It seems to be quite difficult to come up with crisp general results when it comes to a general topic like parameter dependence in families of maps. For instance, I do not know of a simple way of characterizing exactly when parameter shadowing favors one direction over the other in parameter space for piecewise monotone maps. For unimodal maps, it appears that perhaps a useful connection to topological entropy may be made. If topological entropy is monotonic, and if there is a change in the topological entropy of map f_p with respect to p at $p = p_0$ then certain asymmetrical shadowing results seem likely for orbits of f_{p_0}. However, topological entropy does not appear to be an ideal indicator for asymmetrical shadowing, since it is global in nature. On the other hand, if a piecewise monotone map has multiple turning points, it is possible for some turning points to favor higher parameters while other turning points favor lower parameters. Such examples are interesting, from a parameter estimation point of view, because that means that one may be able to effectively squeeze parameter estimates within a narrow band of uncertainty as the orbit being sampled passes close to turning points which favor different directions in parameter space.

Chapter 4

General nonuniformly hyperbolic systems

In this chapter we examine shadowing behavior for general one-parameter families of C^2 diffeomorphisms, $f_p : M \to M$ for $p \in \mathbb{R}$ where M is a smooth compact manifold. We want to consider why orbits shadow each other (or fail to shadow each other) in maps that are nonuniformly hyperbolic. This is important to investigate so that we can properly evaluate the feasibility of parameter estimation in a wide class of chaotic systems.

The exposition in this chapter will not be rigorous. Most of the arguments will be qualitative in nature. Our goal here is to motivate some possible mechanisms that might help explain results from numerical experiments. In particular we will attempt to draw analogies to our work in chapter 3 to help explain what may be happening in multi-dimensional systems.

4.1 Preliminaries

Let us first outline some basic concepts.

We start by introducing the notion of Lyapunov exponents. Let $f : M \to M$ be a C^2 diffeomorphism. Suppose that M is a compact $q-$dimensional manifold and that for some $x \in M$ there exist subspaces, $R^q = E_x^1 \supset E_x^2 \supset \ldots$ in the tangent space of f at x such that:

$$\lambda_x^i = \lim_{n \to \infty} \frac{1}{n} log|Df^n(x)u| \text{ if } u \in E_x^i \setminus E_x^{i-1}.$$

for some numbers $\lambda_x^1 > \lambda_x^2 > \ldots$. Then the λ_x^i's are the *Lyapunov exponents* of the orbit, $\{f^i(x)\}$. Oseledec's Multiplicative Ergodic Theorem ([48]) demonstrates that for any f-invariant probability measure, μ, these Lyapunov exponents exist for μ-almost all $x \in M$.

If there are no λ_x^i's equal to zero, then there exist local stable manifolds at x tangent to the linear subspace, E_x^i if $\lambda_x^i < 0$. There also exists an analogous unstable manifold. In other words, for almost any $x \in M$ there exists an $\epsilon > 0$ such that:

$$\begin{aligned} W_\epsilon^s(x, f) &= \{y \in M : d(f^n(x), f^n(y)) \leq \epsilon \text{ for all } n \geq 0 \ \} \\ W_\epsilon^u(x, f) &= \{y \in M : d(f^{-n}(x), f^{-n}(y)) \leq \epsilon \text{ for all } n \geq 0 \ \} \end{aligned}$$

These manifolds are locally as differentiable as f. This result is based on Pesin [52] and Ruelle [54]. The difference between these manifolds and manifolds for the uniformly hyperbolic case is that these manifolds do not have to exist everywhere, the angles between the manifolds can approach zero, and the neighborhoods, ϵ, can be arbitrarily small for different $x \in M$.

We can also define global stable and unstable manifolds as follows:

$$\begin{aligned} W^s(x, f) &= \{y \in M : d(f^n(x), f^n(y)) \to 0 \text{ as } n \to \infty\} \\ W^u(x, f) &= \{y \in M : d(f^{-n}(x), f^{-n}(y)) \to 0 \text{ as } n \to \infty\}. \end{aligned}$$

Note that these manifolds are invariant in the sense that $f(W^s(x, f)) = W^s(f(x), f)$. Although locally differentiable, the manifolds can have extremely complicated structure in general.

4.2 Discussion

We now return to the investigation of shadowing orbits.

There have been some attempts to examine the linear theory regarding nonuniformly hyperbolic maps in order to make statements about shadowing behavior (see for example [24]). However, since the nonexistence of shadowing orbits fundamentally results from degeneracy in the linear theory, it is also be useful to consider what happens in terms of the structure of nearby manifolds.

For almost every x, f looks locally hyperbolic. However, in nonhyperbolic systems if we iterate the orbit $\{f^i(x)\}$, we will eventually approach some sort of degeneracy.

For example, one possible scenario is that for some point $a \in \{f^i(x)\}$, $W^s(a, f)$ and $W^u(a, f)$ are nearly tangent and intersect each other at some nearby point, y. As illustrated in figure 4.1, this structure implies a certain scenario for the evolution of the manifolds as we map forward with f or backward with f^{-1}. We will argue that this situation is in some sense a multidimensional analog for the folding behavior we observed in one dimension.

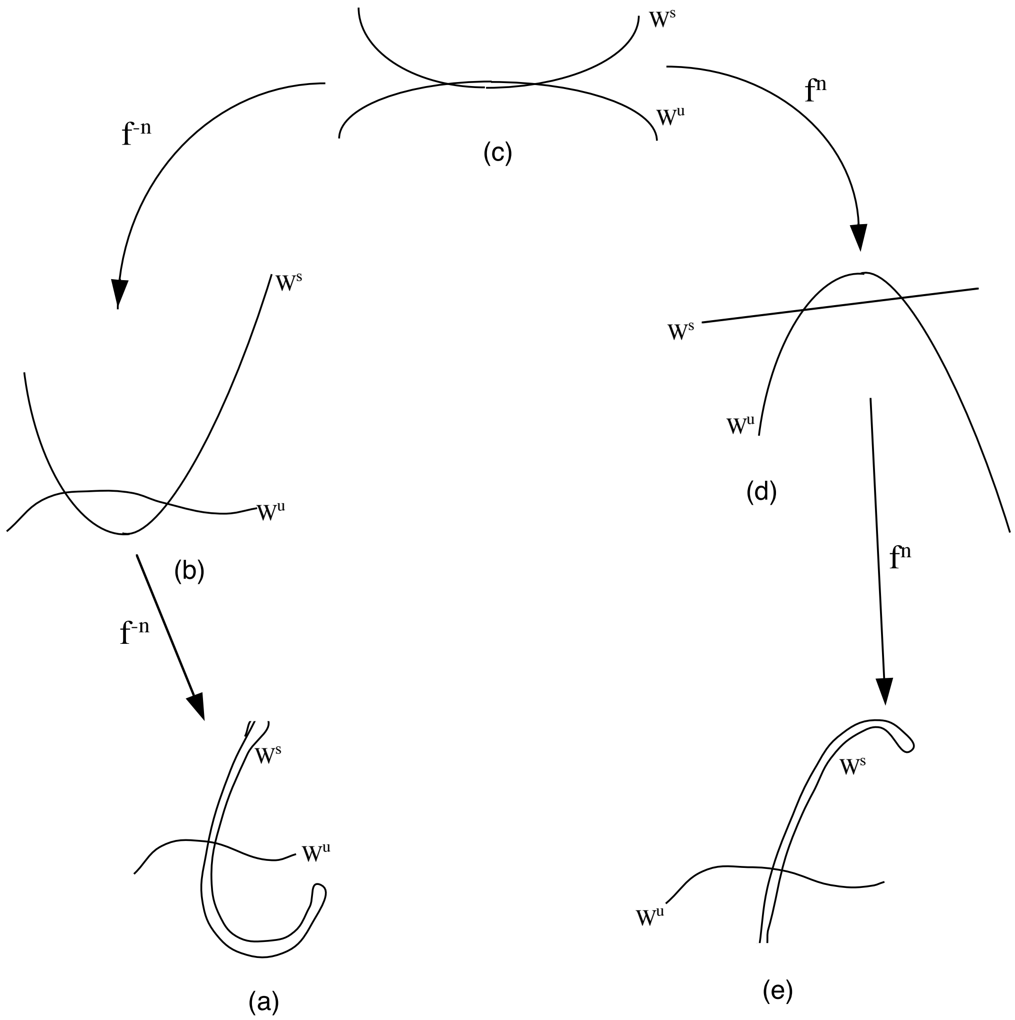

Figure 4.1: Possible situation near a homoclinic tangency. Note how a fold in the unstable manifold is created as we map ahead by f^n, and a fold in the stable manifold is created as we map back by f^{-n}.

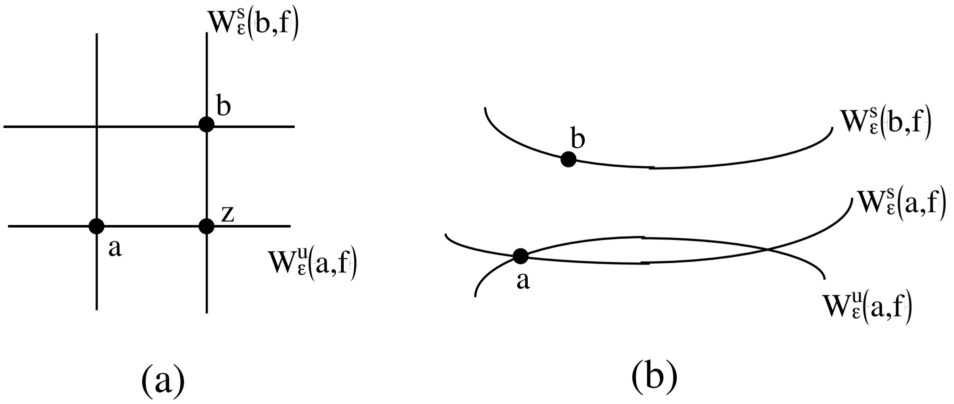

Figure 4.2: An illustrative example of how homoclinic tangencies can cause problems for shadowing.

For one thing, the homoclinic intersection of manifolds can prevent or at least hamper shadowing. We illustrate this in figure 4.2. Consider for example two nearby points a and b such that $d(a,b,) < \delta$ and let $\{c_n\}$ be a δ−pseudo-orbit of f with the following form:

$$c_n = \begin{cases} f^n(a) \text{ if } n < 0 \\ f^n(b) \text{ if } n \geq 0 \end{cases}$$

In a uniformly hyperbolic scenario as shown in figure 4.2(a), we can easily pick a suitable orbit to shadow $\{c_n\}$, namely $\{f^i(z)\}$ where $z = W^u_\epsilon(a,f) \cap W^s_\epsilon(b,f)$. However if a homoclinic intersection is nearby as in figure 4.2(b), we see that there is no obvious way to pick a shadowing orbit, since there may be no point z satisfying $z = W^u_\epsilon(a,f) \cap W^s_\epsilon(b,f)$. Note that the difficulty in finding a shadowing orbit seems to depends on how close a is to the homoclinic tangency, and the geometry of the manifolds nearby.

Homoclinic tangencies could also cause asymmetrical shadowing in parameter space. Numerical experiments with maps that favor higher parameters seem to show the following scenario: As we map a state space region near a homoclinic tangency ahead by f_{p_0} repeatedly, a tongue, or fold of the unstable manifold develops as the manifold expands. If we examine the corresponding situation in a map with a slightly higher parameter value, we find that the corresponding fold in the unstable manifold for the higher parameter system overlaps the fold in the unstable manifold of the original system. In this case we expect that the original system would have difficulty shadowing a trajectory close to the apex of the fold in the higher parameter system. This situation is depicted in figure 4.3. A similar argument works for f^{-1}. Numerical results seem to indicate that for many families of systems at least, there is an ordering in parameter space such that

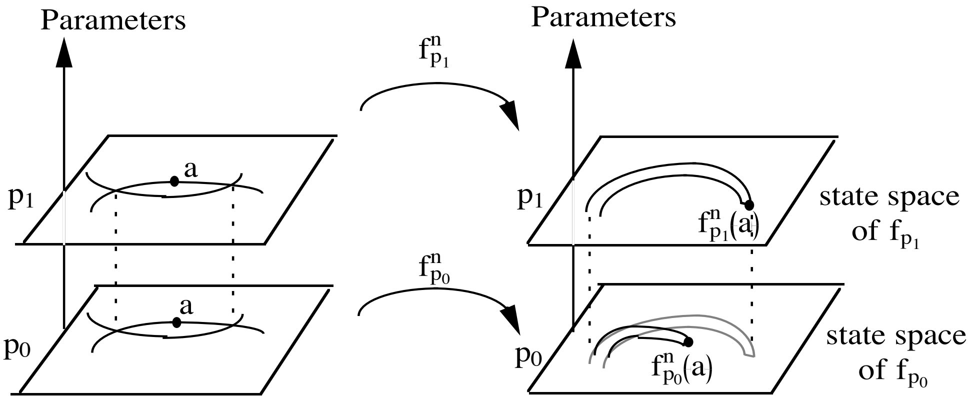

Figure 4.3: Why higher dimensional maps might exhibit asymmetrical shadowing in parameter space.

as we increase (or decrease) the parameter value, the systems get progressively more "flexible" in sense that systems that are more flexible can always shadow orbits of systems that are less flexible. Numerical evidence for this type of shadowing behavior can be found in chapter 6.

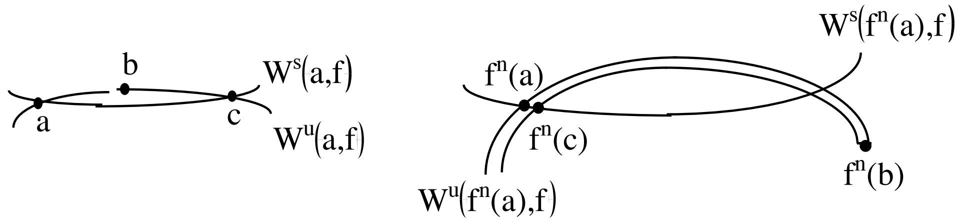

Figure 4.4: Refolding after a subsequent encounter with a homoclinic tangency.

Also recall that with maps of the interval, a folded region can get refolded upon a subsequent encounter with a turning point. A similar thing can also happen in higher dimensions. Consider figure 4.4 for example. Here we see that the folded tongue of the unstable manifold gets refolded back on itself, possibly allowing lagging orbits to catch up so that shadowing is possible. This suggests that there may be interesting shadowing results of the sort described in chapter 3 for one dimension. The situation here, however, is more complicated since in one dimension there were only a finite number of sources of folding, namely the turning point, while here there are likely to be an infinite number of sources for the folding.

Chapter 5

Parameter estimation algorithms

5.1 Introduction

In this chapter we present new algorithms for estimating the parameters of chaotic systems. In particular we will be interested in investigating estimation algorithms for nonuniformly hyperbolic dynamical systems, because these systems include most of the chaotic systems likely to be encountered in physical applications. From our discussion in chapters 3 and 4, we know that there are three basic effects that are important to consider when designing a parameter estimation algorithm for nonuniformly hyperbolic dynamical systems: (1) most data points contribute very little to our knowledge of the parameters of the system, while a relatively few data points may be extremely sensitive to parameters, (2) the sensitive sections of orbits reflect nearby folding behavior which must be accurately modeled in order to extract information about the parameters, and (3) the folding behavior often results in asymmetrical shadowing behavior in the parameter space of the system, so we can generally eliminate only parameters slightly less than or slightly greater than the actual parameter value. The goal is to develop an efficient algorithm that takes all three of these effects into account.

Our basic strategy will be to take advantage of property (1) above by using a linear filtering technique to scan through most of the data and attempt to locate parts of the trajectory where folding occurs. In sections of the trajectory where folding does occur, we will examine the data closely using a type of Monte-Carlo analysis which we have designed to circumvent the numerical pitfalls that accompany work with chaotic systems.

We begin this chapter by surveying some traditional filtering techniques and examining some basic approaches for parameter estimation problems (section 5.3). Those readers who are familiar with traditional estimation theory may wish to skim these sections. We go on in section 5.4 to examine how and why traditional algorithms fail in high-precision estimation of chaotic systems. We then propose a new algorithm for estimating the parameters of a chaotic system in one dimension (section 5.5). This algorithm is generalized in section 5.6 to deal with systems in higher dimensions.

Numerical results of these algorithms describing the performance of these techniques are presented in chapter 6.

5.2 The estimation problem

Let us begin by restating the problem.[1] Let:

$$\begin{aligned} x_{n+1} &= f_p(x_n) && (5.1) \\ \text{and} \qquad y_n &= x_n + v_n && (5.2) \end{aligned}$$

where x_n is the state of the system, y_n are observations, v_n represents noise, f evolves the state, $p \in I_p \subset \mathbb{R}$ is the scalar parameter we are trying to estimate, and I_p is a closed interval of the real line.

It will also be useful to write the system in (5.1) and (5.2) in terms of $u_n = (x_n, p)$, a combined vector of state and parameters:

$$\begin{aligned} u_{n+1} &= g(u_n) && (5.3) \\ y_n &= H_n u_n + v_n && (5.4) \end{aligned}$$

where the map, g, satisfies $g(x,p) = (f_p(x), p)$, and:

$$H_n = \begin{bmatrix} I_q & 0 \\ 0 & 1 \end{bmatrix} \qquad (5.5)$$

where I_q is a $q \times q$ identity matrix if the state, x, has dimension q.

We now make a few remarks about notation. In general, throughout this chapter, the letters x, p, u will correspond to state, parameter, and state-parameter vectors. Set $x^n = (x_0, x_1, \ldots, x_n)$, $y^n = (y_0, y_1, \ldots, y_n)$, and $u^n = (u_0, u_1, \ldots, u_n)$.

The symbol "ˆ" above a vector will be used to denote an estimate. For example. the estimate of the parameter p based on the observations in y^n will be denoted $\hat{p}_n$. We will also use the notation, $\hat{u}_{n|k}$, to denote an estimate of u_n based on observations, y^k. Similarly, the symbol "˜" will be used to denote an error quantity. For example we might write that $\tilde{u}_n = u_n - \hat{u}_{n|n}$.

[1] Note that the setup in (5.1) and (5.2) is somewhat less general than standard formulations of filtering problems. For example one could add an extra term, w_n, to represent the system noise so that $x_{n+1} = f_p(x_n) + w_n$, or one could add an extra function, $h_n(x)$, so that $y_n = h_n(x_n) + v_n$, to reflect the fact that the observations might represent a more general function of the state. However, we have elected to keep problem as simple as possible in order to concentrate on how chaos affects estimation, and to be consistent with the presentation in chapters 2-4.

5.3 Traditional approaches

We now examine some basic methods for approaching parameter estimation. In sections 5.3.1 and 5.3.2 we mainly concentrate on providing the motivation behind linear techniques like the Kalman filter. This treatment is extended in section 5.3.3, where nonlinear techniques are discussed in more detail. The material in this section is well-known in the engineering community, but we explain it here because it provides the basis for new algorithms we develop later to deal with chaotic systems.

There are a variety of ways to approach parameter estimation problems. Engineers have developed a whole host of *ad hoc* tricks that may be applied in different situations. The basic idea, however, is relatively simple. Given observations, $\{y_k\}_{k=0}^n$, and a model for f_p, we would like to pick our parameter estimate, $p = \hat{p}_n$, so that there exists an orbit, $\{\hat{x}_k(p)\}_{n=0}^n$, of f_p that makes the residuals,

$$\epsilon_k(p) = y_k - \hat{x}_k(p)$$

as small as possible for $k \in \{0, 1, \ldots, n\}$. In order to choose the best possible estimate, $\hat{p}_n$, we need some criteria for evaluating how small these residuals are.

From here, there are a number of different ways to approach the problem of how to choose the optimizing criteria to make use of all the known information. In fact, the recursive Kalman filter itself has many different possible interpretations. Many of the different approaches to parameter estimation provide interesting insight into the estimation problem itself. Our objective here will be to motivate some of the different ideas on how to look at parameter estimation, without getting immersed in specific derivations. The reader may consult [3], [29], or [23] for more detailed and/or formal treatments of this subject.

5.3.1 Nonrecursive estimation

Least squares estimation

One of the simplest ideas about how to estimate parameters is to choose the estimate $\hat{p}_n$ so that $p = \hat{p}_n$ minimizes the quantity:

$$S'_n(p) = \inf_{\{\hat{x}_{i|n}(p)\}_{i=0}^n \in Z(p)} \{\sum_{i=0}^n (y_i - \hat{x}_{i|n}(p))^T (R'_i)^{-1} (y_i - \hat{x}_{i|n}(p))\} \tag{5.6}$$

where $Z(p)$ is the set of all orbits of f_p and $(R'_i)^{-1}$ are symmetric positive-definite matrices that weight the relative importance of various measurements. This sort of idea, known as *least squares* estimation, dates back to Gauss [22].

The formulation in (5.6) is not really useful for estimating parameters in practice, since there is no direct way of choosing $\hat{p}_n$ to minimize (5.6). Things become more concrete, however, if we assume the function g in (5.4) is linear in both state and parameters.[2] In this case we can write:

$$y^n = G_n u_0 + v^n \tag{5.7}$$

where G_n is a constant matrix that effectively represents the dynamics of the system. Our goal is to get a good estimate for $u_0 = (x_0, p)$ based on the observations in y^n. In this case, least squares estimation amounts to minimizing

$$S_n(u_0) = (y^n - G_n(u_0))^T R_n^{-1}(y^n - G_n(u^n)) \tag{5.8}$$

with respect to u_0 where R_n^{-1} are positive-definite weighting matrices. Our estimate for u_0 based on y^n, $\hat{u}_{0|n} = (\hat{x}_{0|n}, \hat{p}_n)$, is the value of u_0 that minimizes $S_n(u_0)$. We can find the appropriate minimum of $S_n(u_0)$ by taking the derivative of S_n with respect to u_0. If we do this we find that thus value of u_0 that minimizes $S_n(u_0)$ is:

$$\hat{u}_{0|n} = (G_n^T R_n^{-1} G_n)^{-1} G_n^T R_n^{-1} y^n \tag{5.9}$$

where G_n^T denotes the transpose of G_n.

Stochastic framework

Another way to approach the problem is to think of u_n, y_n, and v_n as random variables. We shall assume that the v_n's are independent random variables with zero mean. The idea is to choose a parameter estimate, $\hat{p}_n$, based on y^n so that the residuals, $\epsilon_i(p) = y_i - \hat{x}_i(p)$, are as close to zero as possible in some statistical sense for $i \in \{0, 1, \ldots, n\}$.

We can write the probability density function[3] for u_n given y^k according to Bayes rule:

$$P(u_n|y^k) = \frac{P(y^k|u_n)P(u_n)}{P(y^k)} \tag{5.10}$$

These density functions describe everything we might know about the states and parameters of the system. Later we will examine more closely how tracking such probability densities in full can provide information about how to choose parameter estimates, especially in cases involving nonlinear or chaotic systems. To start with, however, we concentrate on examining conventional filters which look only at first and second order moments of these densities.

[2]Note that this assumption is extremely restrictive in practice, since even if the system is linear with respect to state, it is generally nonlinear with respect to combined states and parameters. The purpose of this example, however, is to simply motivate linear ideas. We address nonlinearity in the next section.

[3]Contrary to common convention, our choice of the letter p for the parameter necessitates using a capital P to denote probability density functions. Thus $P(u_n|y^k)$ represents the density for for u_n given the value of y^k.

Minimum variance

Given the density function, $P(u_0|y^n)$, one approach is to pick the estimate, $\hat{u}_{0|n}$, to minimize the variance,

$$E[(u_0 - \hat{u}_{0|n})^T(u_0 - \hat{u}_{0|n})] \tag{5.11}$$

where $E[x] = \int xP(x)dx$ denotes the expected value of x. This criterion is called the minimum variance condition. It turns out that this estimator has particularly nice properties. For instance, it is not hard to show (e.g., [57]) that the $\hat{u}_{0|n}$ that minimizes (5.11) also satisfies:

$$\hat{u}_{0|n} = E[u_0|y^n].$$

for any density function, $P(u_0|y^n)$.

Now suppose that g is linear in state and parameters so that (5.7) is satisfied. Let us attempt to find the so called *optimal linear estimator*:

$$\hat{u}_{0|n} = A_n y^n + b_n$$

where the constant matrix, A_n, and constant vector, b_n, are chosen to minimize the variance condition in (5.11). Assuming that the estimator is unbiased (i.e., $E(u_0 - \hat{u}(n|0)) = 0$) then:

$$b_n = E(u_0) - A_n E(y^n).$$

Minimizing $E[(u_0 - \hat{u}_{0|n})^T(u^n - \hat{u}_{0|n})]$ we find ([57]) that

$$A_n = (Q^{-1} + G_n^T R_n^{-1} G)^{-1} G^T R^{-1} \tag{5.12}$$

where $Q = E[u_0 u_0^T]$ is the *covariance matrix* of u_0 and $R_n = E[v^n(v^n)^T]$ is the covariance matrix of v^n. Thus we have:

$$\hat{u}_{0|n} = E(u_0) + A_n(y^n - E[y^n]) \tag{5.13}$$

where A_n is as given in (5.12). Comparing this result with (5.9) we see that the $\hat{u}_{0|n}$ above, which we derived as the linear estimator with minimum variance, actually looks a lot like the estimator from the deterministic least squares approach except for the addition of *a priori* information about u_0 (in the form of $E(u_0)$ and the covariance Q). With the minimum variance approach, the weighting factor R_n also has a definite interpretation as the covariance of the measurement noise.

Furthermore, if we assume that u_n and v_n are Gaussian random variables,[4] and attempt to optimize the estimator $\hat{u}_{0|n}$ for minimum variance, we again find (see [30]) that $\hat{u}_{0|n}$ has the form given in (5.12) and (5.13).

[4] A random variable $v \in R^q$ has Gaussian distribution if

$$P(v) = \frac{1}{(2\pi)^{\frac{q}{2}}} e^{-\frac{1}{2}(v-E(v))^T \Sigma_v^{-1}(v-E(v))}$$

where $E[v]$ is the expected value of v and $\Sigma_v = E[vv^T]$ is the covariance matrix of v.

Thus, in summary, we see that the optimal estimator, $\hat{u}_{0|n}$ as given in (5.12) and (5.13) has a number of different interpretations. If the system, g, is linear then the estimator can be thought of as resulting from a deterministic least squares approach. If u_n and v_n are thought of as random variables, then $\hat{u}_{0|n} = E[u_0|y^n]$, and if we assume that u_n and v_n are Gaussian then the $\hat{u}_{0|n}$ given in (5.13) satisfies the minimum variance condition. Alternatively, if we drop the Gaussian assumption and search of the best linear estimator that minimizes the variance condition, we find that $\hat{u}_{0|n}$ as given in (5.12) and (5.13) is the optimal linear estimator. All these interpretations motivate us to use the estimator given in (5.12) and (5.13).

5.3.2 The Kalman filter

We now have the form of an optimal filter for linear systems. However, the filter has problems computationally. It would be nice if there were a way so that new data could be taken into account easily without having to recompute everything. This is accomplished with the recursive Kalman filter.

The Kalman filter is mathematically equivalent to the linear estimator described in (5.12) and (5.13), except that it has some important computational advantages. The basic premise of the Kalman filter is that the state of the filter can be kept with two statistics, $\hat{u}_{n|n}$ and $\Sigma_{n|n}$, where $\Sigma_{n|n}$ is the covariance matrix, $E[(u_n - \hat{u}_{n|n})(u_n - \hat{u}_{n|n})^T]$. Once we have these two particular statistics, it will be possible, for example, to determine the next state of the filter, $\hat{u}_{n+1|n+1}$ and $\Sigma_{n+1|n+1}$, directly given a new piece of data, y_{n+1}, the filter's present state, $\hat{u}_{n|n}$, $\Sigma_{n|n}$, and knowledge of the map g.

Specifically, suppose we are given the linear system:

$$\begin{aligned} u_{n+1} &= \Phi_n u_n \\ y_n &= H_n u_n + v_n. \end{aligned}$$

where v_n are independent random variables with zero mean and covariance R_n. The recursive Kalman filter can be written in two parts:

Prediction:

$$\hat{u}_{n+1|n} = \Phi_n \hat{u}_{n|n} \tag{5.14}$$
$$\Sigma_{n+1|n} = \Phi_n \Sigma_{n|n} \Phi_n^T + R_{n+1} \tag{5.15}$$

Combination:

$$\hat{u}_{n+1|n+1} = \hat{u}_{n+1|n} + K_{n+1}(y_{n+1} - H_{n+1}\hat{u}_{n+1|n}) \tag{5.16}$$
$$\Sigma_{n+1|n+1} = (I - K_{n+1}H_{n+1})\Sigma_{n+1|n} \tag{5.17}$$

where the Kalman gain, K_{n+1}, is given by:

$$K_{n+1} = \Sigma_{n+1|n} H_{n+1}^T [H_{n+1} \Sigma_{n+1|n} H_{n+1}^T + R_{n+1}]^{-1}. \quad (5.18)$$

Motivation and derivation

The Kalman filter can be motivated in the following way.[5] Consider the metric space, X, of random variables where inner products and norms are defined by:

$$\begin{aligned} \langle x, y \rangle &= E[xy^T] \\ \text{and } \|x\| &= \langle x, x \rangle \end{aligned}$$

if $x, y \in X$. Let $Y_n = span\{y_0, y_1, \ldots, y_n\}$ be the space of a all linear combinations of $\{y_0, y_1, \ldots, y_n\}$. To satisfy the minimum variance condition, we would like to pick $\hat{u}_{n|n} \in Y_n$ to minimize:

$$E[\tilde{u}_n^T \tilde{u}_n] = \|\tilde{u}_n\|.$$

where $\tilde{u}_n = u_n - \hat{u}_{n|n}$. This formulation gives a definite geometric interpretation for the minimization problem and helps to show intuitively what the appropriate $\hat{u}_{n|n}$ is. In order to minimize the distance between u_n and $\hat{u}_{n|n} \in Y_n$, it makes sense to pick $\hat{u}_{n|n}$ so that $\tilde{u}_n$ is *orthogonal* to Y_n. That is, we require:

$$\langle \tilde{u}_n, y \rangle = 0 \quad (5.19)$$

for any $y \in Y_n$. It is not hard to show that this condition is in fact sufficient to minimize $E[\tilde{u}_n^T \tilde{u}_n]$ (see e.g., [3]). From a statistical standpoint, this result also makes sense since it says that the error of the estimate, $\tilde{u}_n$, should be uncorrelated with the measurements. In some sense, the estimate uses all the information contained in the measurements.

We can now derive the equations of Kalman filter. The prediction equations are relatively straightforward:

$$\begin{aligned} \hat{u}_{n+1|n} &= E[u_{n+1}|y^n] = \Phi_n \hat{u}_{n|n} \\ \Sigma_{n+1|n} &= E[(u_{n+1|n} - u_{n\hat{+}1|n})(u_{n+1|n} - u_{n\hat{+}1|n})^T] = \Phi_n \Sigma_{n|n} \Phi_n^T + R_{n+1}. \end{aligned}$$

For the estimator $\hat{u}_{n+1|n+1}$ to be unbiased, $\hat{u}_{n+1|n+1}$ must have the form given in (5.16). Now let us now verify that the formula for K_{n+1} in (5.18) makes the Kalman filter an optimal linear estimator. To do this, we must show that K_{n+1} minimizes the variance, $E[\tilde{u}_{n+1}^T \tilde{u}_{n+1}]$, where $\tilde{u}_{n+1} = u_{n+1} - \hat{u}_{n+1|n+1}$. Since $\hat{u}_{n+1|n+1} \in Y_{n+1}$ we know from (5.19) that a sufficient condition for $E[\tilde{u}_{n+1}^T \tilde{u}_{n+1}]$, to be minimized is that:

$$E[\tilde{u}_{n+1}^T \hat{u}_{n+1|n+1}] = Trace E[\tilde{u}_{n+1} \hat{u}_{n+1|n+1}^T] = 0. \quad (5.20)$$

[5] Much of the explanation here follows the exposition in Siapas [56].

Let us investigate the consequences of this condition. First we have:

$$\begin{aligned}\tilde{u}_{n+1} &= \Phi_n u_n - [\hat{u}_{n+1|n} + K_{n+1}(y_{n+1} - H_{n+1}\hat{u}_{n+1|n})] \\ &= \Phi_n u_n - \Phi_n \hat{u}_{n|n} - K_{n+1}[H_{n+1}u_{n+1} + v_{n+1}] + K_{n+1}H_{n+1}\Phi_n\hat{u}_{n|n} \\ &= (I - K_{n+1}H_{n+1})\Phi_n\tilde{u}_n - K_{n+1}v_{n+1}\end{aligned}$$

So,

$$\begin{aligned}E[\tilde{u}_{n+1}\hat{u}^T_{n+1|n+1}] &= E[\{(I - K_{n+1}H_{n+1})\Phi_n\tilde{u}_n - K_{n+1}v_{n+1}\} \\ &\qquad \{\hat{u}_{n+1|n} + K_{n+1}(y_{n+1} - H_{n+1}\hat{u}_{n+1|n})\}^T] \\ &= E[\{(I - K_{n+1}H_{n+1})\Phi_n\tilde{u}_n - K_{k+1}v_{n+1}\} \\ &\qquad \{\Phi_n\hat{u}_{n|n} + K_{n+1}\Phi_n\tilde{u}_n + K_{n+1}v_{n+1}\}^T]\end{aligned} \tag{5.21}$$

Since we require that $E[\tilde{u}^T_n\hat{u}_{n|n}] = Trace\{E[\tilde{u}_n\hat{u}^T_n]\} = 0$, from (5.21) we get that:

$$\begin{aligned}&Trace\{E[\tilde{u}_{n+1}\hat{u}_{n+1|n+1}]^T\} \\ &= Trace\{(I - K_{n+1}H_{n+1})\Phi_n E[\tilde{u}_n\tilde{u}^T_n]\Phi^T_n H^T_{n+1}K^T_{n+1} - K_{n+1}E[v_{n+1}v^T_{n+1}]K^T_{n+1}\} \\ &= Trace\{\Phi_n\Sigma_{n|n}\Phi^T_n H^T_{n+1}K^T_{n+1} - K_{n+1}H_{n+1}\Phi_n\Sigma_{n|n}\Phi^T_n H^T_{n+1}K^T_{n+1} - K_{n+1}R_{n+1}K^T_{n+1}\} \\ &= Trace\{[\Sigma_{n+1|n}H^T_{n+1} - K_{n+1}(H_{n+1}\Sigma_{n+1|n}H^T_{n+1} + R_{n+1})]K^T_{n+1}\}.\end{aligned}$$

Thus, choosing $K_{n+1} = \Sigma_{n+1|n}H^T_{n+1}[H_{n+1}\Sigma_{n+1|n}H^T_{n+1} + R_{n+1}]^{-1}$ as in (5.18) makes $Trace\{E[\tilde{u}^T_{n+1}\hat{u}_{n+1|n+1}]\} = 0$ and therefore minimizes $E[\tilde{u}^T_{n+1}\tilde{u}_{n+1}]$.

The equation for $\Sigma_{n+1|n+1}$ in (5.17) can then be derived by simply evaluating $\Sigma_{n+1|n+1} = E[\tilde{u}^T_{n+1}\tilde{u}_{n+1}]$.

5.3.3 Nonlinear estimation

Probability densities

The filters we looked at in the previous section are optimal linear estimators in the sense that a minimum variance or least squares condition is satisfied. Estimators like the Kalman filter are only optimal, however, if the system is linear and the corresponding probability densities are Gaussian. Let us now, however, consider how one might approach estimation problems when these rather stringent condition are relaxed.

Let us begin by recalling the density function in (5.10):

$$P(u_n|y^k) = \frac{P(y^k|u_n)P(u_n)}{P(y^k)} \tag{5.22}$$

where $u_n = (x_n, p)$ is the joint vector of state and parameters and $y^k = (y_0, y_1, \ldots, y_k)$ represents a vector of observations. This density function represents everything we know about a state given the observations specified. Techniques that use this density directly to estimate the parameters of a system are known as *Bayesian estimation* algorithms. For example, one might simply attempt to pick an estimate, $\hat{u}_{n|k}$, so that $P(u_n|y^k)$ is maximized at $u_n = \hat{u}_{n|k}$. This is known as a maximum *a posteriori* (MAP) estimate.

If the system, g, is linear in (5.4) and all the *a priori* information and measurement noises are Gaussian, then the MAP estimator gives the same answers as the Kalman filter (e.g., see [23]). We can see this by considering how the appropriate conditional probability densities get transformed by the dynamics of a linear system and combined with new data, as in the prediction and combination steps of the Kalman filter. For example, suppose that the density, $P(u_n|y^n)$ is Gaussian for some value of n (see figure 5.1). The density, $P(u_{n+1}|y^n)$, can be then determined from $P(u_n|y^n)$ by simply

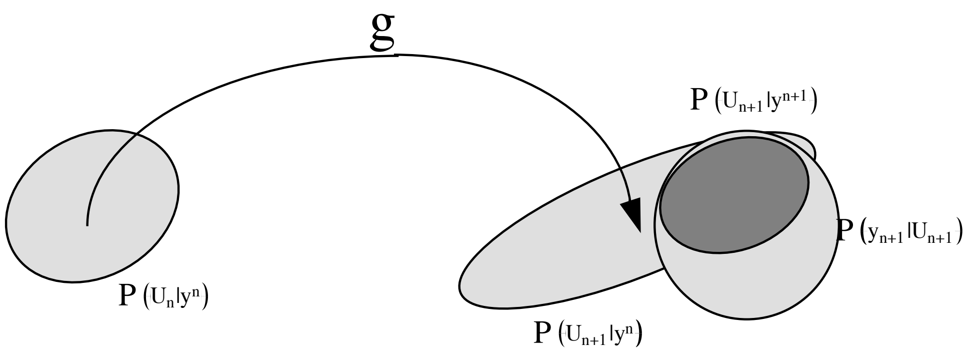

Figure 5.1: Mapping probability densities using g and combining them with new information. This is a probabilistic view of what a recursive estimator like the Kalman filter does. Note that Gaussian densities have equal probability density surfaces that form ellipsoids. In two dimensions we draw the densities as ellipses.

mapping $P(u_n|y^n)$ using the system dynamics, g. More precisely we have that:

$$P(u_{n+1}|y^n) = \sum_{z \in U(u_{n+1})} [P(z|y^n)|Dg(z)|^{-1}] \tag{5.23}$$

where $U(u_{n+1}) = \{z | z = g^{-1}(u_{n+1})\}$ and $|Dg(z)|$ is the determinant of the Jacobian of g evaluated at z. It is not hard to show that if g is linear and $P(u_n|y^n)$ is Gaussian then $P(u_{n+1}|y^n)$ is also Gaussian. Also by Bayes rule, $(P(A,B) = P(A|B)P(B) = P(B|A)P(A))$ we have that:

$$P(u_{n+1}, y_{n+1}|y^n) = P(u_{n+1}|y^{n+1})P(y_{n+1}|y^n) = P(y_{n+1}|u_{n+1}, y^n)P(u_{n+1}|y^n)$$

where $P(y_{n+1}|y^n) = \int P(y_{n+1}|u_{n+1})P(u_{n+1}|z^n)du_{n+1}$. Thus we find that combining information from a new measurement, y_{n+1}, results in the density:

$$P(u_{n+1}|y^{n+1}) = \frac{P(y_{n+1}|u_{n+1})P(u_{n+1}|y^n)}{P(y_{n+1}|y^n)}. \tag{5.24}$$

Since the denominator is independent of u_{n+1}, it is simply a normalizing factor and is therefore not important for our considerations. Also note that since $P(y_{n+1}|u_{n+1})$ and $P(u_{n+1}|y^n)$ are Gaussian, $P(u_{n+1}|y^{n+1})$ must also be Gaussian. Thus, by induction if all the data is Gaussian distributed, then $P(u_k|y^k)$ must be Gaussian for any k. Also, the MAP estimate and minimum variance estimate for u_{n+1} are both the same, namely $\hat{u}_{n+1|n+1} = E[u_{n+1}|y^{n+1}]$.

Now consider what happens if the system is nonlinear. The appropriate densities still describe all we know about the states and parameters. In particular, the equations in (5.23) and (5.24) are still valid descriptions of how to map ahead and combine densities. However, in general there are no constraints on the form of these densities. As a practical matter, the problem becomes how can we deal with these arbitrary probability densities? How can one represent approximations of the densities in a computationally tractable form while still retaining enough information to generate useful estimates? There have been a number of efforts in this area:

Extended Kalman filter

The most basic and widely used trick is to simply linearize the system around the best estimate of the trajectory and then use the Kalman filter. The idea is that if the covariances of the relevant probability densities are small enough, then the system acts approximately linearly on the densities, so linear filtering may adequately describe the situation. For the system,

$$\begin{aligned} u_{n+1} &= g(u_n) && (5.25) \\ y_{n+1} &= H_n u_n + v_n, && (5.26) \end{aligned}$$

as in (5.3), (5.4), and (5.5), the extended Kalman filter is given by the following equations, mirroring the Kalman filter in (5.14)-(5.18):

Prediction:

$$\begin{aligned} \hat{u}_{n+1|n} &= g(\hat{u}_{n|n}) && (5.27) \\ \Sigma_{n+1|n} &= Dg(\hat{u}_{n|n})\Sigma_{n|n}Dg(\hat{u}_{n|n})^T && (5.28) \end{aligned}$$

Combination:

$$\begin{aligned} \hat{u}_{n+1|n+1} &= \hat{u}_{n+1|n} + K_{n+1}(y_{n+1} - H_{n+1}\hat{u}_{n+1|n}) && (5.29) \\ \Sigma_{n+1|n+1} &= (I - K_{n+1}H_{n+1})\Sigma_{n+1|n} && (5.30) \end{aligned}$$

where the Kalman gain, K_{n+1}, is given by:

$$K_{n+1} = \Sigma_{n+1|n} H_{n+1}^T [H_{n+1} \Sigma_{n+1|n} H_{n+1}^T + R_{n+1}]^{-1}. \tag{5.31}$$

Other work in nonlinear estimation

A number of other efforts to do estimation on nonlinear systems have concentrated on developing a better description of the probability densities. For example, in [23] methods are presented that attempt to take into account second order behavior from the dynamics. However, the method still relies on a basically Gaussian assumption of the error distributions, since it computes and propagates only the mean and covariance matrices of densities, adjusting the computations to account for errors due to nonlinearity. Taking into account higher order effects in the densities is in fact a difficult proposition because there is no obvious representation for these densities. Gaussian densities are invariant under linear transformations, and are especially easy to deal with when it comes to combining data from new measurements. However, similar higher order representations do not exist.

Other methods do attempt to get a better representation of the error densities. For example in [2], a method is proposed whereby the densities are represented as a sum of Gaussians. For example, one might write:

$$P(u) = \sum_i \alpha_i N(u; m_i, \Sigma_i)$$

where the α_i's represent scalar constants and $N(u; m_i, \Sigma_i)$ evaluates the Gaussian density function with mean m_i and covariance matrix Σ_i at u.[6] If each of the Gaussians in the sum are localized in state-parameter space (have small covariances) then we might be able to use linear filters to evolve and combine each density in the sum in order to generate a representation of the entire density.

5.4 Applying traditional techniques to chaotic systems

In this section we examine why traditional techniques have a difficult time performing high accuracy parameter estimation on chaotic systems. This investigation will illuminate some of the general difficulties one encounters when dealing with chaotic systems, and will provide some useful ground rules for designing new parameter estimation algorithms.

Let us attempt, for example, to naively apply an estimator like the extended Kalman filter in (5.27)-(5.31) to a chaotic system and see what problems emerge.

[6] In other words, $N(u; m_i, \Sigma_i) = \frac{1}{(2\pi)^{\frac{q}{2}}} e^{-\frac{1}{2}(u-m_i)^T \Sigma_v^{-1}(u-m_i)}$ if q is the dimension of u.

The first problem one is likely to encounter is numerical in nature, and has a relatively well-known solution. It turns out that the formulation in (5.27)-(5.31) is not numerically sound. The problems are especially bad, however, in chaotic systems because covariance matrices become ill-conditioned quickly as densities are stretched exponentially along unstable manifolds and contracted exponentially along stable manifolds. Similar sorts of problems, albeit less severe, have been encountered and dealt with by conventional filtering theory. One solution is to represent the covariance matrix $\Sigma_{n|n}$ as the product of two matrices:

$$\Sigma_{n|n} = S_{n|n} S_{n|n}^T, \tag{5.32}$$

and propagate the matrices $S_{n|n}$ instead of $\Sigma_{n|n}$. These estimation techniques, known as *square root algorithms*, are mathematically the same as the Kalman filter, but have the advantage that they are less sensitive to ill-conditioned covariance matrices. Using square root algorithms, for instance, the resulting covariance matrices are assured to remain positive definite. Since the decomposition in (5.32) is not unique, there are a number of possible implementations for such algorithms. The reader is referred to Kaminski [31] and related papers for detailed implementation descriptions.[7]

Other problems result from the nonlinearity of the system. Some of these problems can be observed in general nonlinear systems, while others seem to be unique to chaotic systems. First of all, using a linearized parameter estimation technique on any nonlinear system can cause trouble, even if the system is not chaotic. Often errors due to nonlinearity cause the filter to become too confident in its estimates, which prevents the filter from updating its information correctly based on new data and eventually locks the filter into a parameter estimate with larger error than expected. This phenomenon is known as divergence.[8] It is not hard to see why divergence can become a problem with estimators like the Kalman filter. For example, in the linear Kalman filter, note that the the estimation error covariance matrix, $\Sigma_{n|n}$, can actually be precomputed without knowledge of the data. In other words there is no feedback between the actual performance of the filter and the filter's estimate of its own accuracy. In the extended Kalman filter there is also virtually no feedback between the observed residuals, $y_n - H_n \hat{u}_n$, and the computed covariance matrix, $\Sigma_{n|n}$.

The divergence problem is considerably worse in nonuniformly hyperbolic systems than it is in other nonlinear applications. This is because folding, a highly nonlinear phenomenon, is crucial to parameter estimation. While linearized strategies may do reasonably well following most chaotic trajectories if the uncertainty variances are small, linearized techniques invariably have great trouble with the sections of trajectories that are most sensitive to parameter perturbations. Figure 5.2 gives a schematic of what happens when folding occurs. The linearized probability densities in that case become

[7] In this report, whenever we refer to numerical results using square root filtering techniques, the implementation we use is the one given in [31] labeled "Square Root Covariance II."

[8] See for example, Ljung [41] for discussion of some related work.

poor approximations to the real densities. Note that the composite densities look extremely long and thin because the densities have gotten stretched and contracted along unstable and stable manifolds.

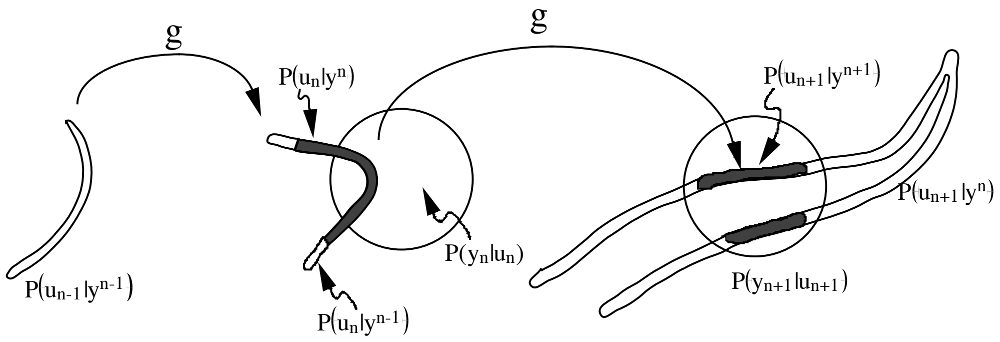

Figure 5.2: In this picture we show a typical example of what can happen to probability densities in chaotic systems. Because of the effects of local folding, linear filters like the Kalman filter sometimes have difficulty tracking nonuniformly hyperbolic dynamical systems.

In chapter 6, we show some examples of the performance of the square root extended Kalman filter on various maps. The filter generally performs reasonably well at first but eventually diverges as the trajectory it is tracking passes close to a folding area. As we observed earlier, once the extended Kalman filter becomes too confident about its estimate, it generally cannot recover. While various *ad hoc* techniques can make small improvements to this problem, none of the standard techniques I encountered did an adequate job of handling the folding. For example, consider the case of the Gaussian sum filter, which is basically the only method that one might expect to have a chance at modeling the folding behavior. Note that the densities in the Gaussian sum have to be re-decomposed into constituent Gaussians every few iterations because of spreading, as expansion along unstable manifolds quickly pushes most of the constituent densities out into regions of near zero probability. In addition, the position of the apex of the fold, which is crucial to estimating the correct parameters, is quite difficult to get a handle on without including many terms in the representation of the density.

5.5 An algorithm in one dimension

In the previous section we saw that traditional techniques do not seem to do a reasonable job modeling the effects of folding on parameter estimation. Since there seems to be no simple way of adequately representing a probability density as it gets folded, we resort to a Monte Carlo representation of densities near folded regions, meaning that the appropriate densities are sampled at many different points in state and parameter space and this data is used as a representation for the density itself. The eventual hope is that we will only have to examine a fraction of the data using computationally-intensive techniques like Monte Carlo, since we know that only a few sections of data are really sensitive to parameter values.

Though the ideas are simple, the actual implementation of such parameter estimation techniques is not as easy one might think because of numerical problems associated with chaotic systems. In this section we examine the basics of how to apply Monte Carlo-type analysis to chaotic systems by looking at an algorithm for one-dimensional noninvertible systems. An algorithm for higher dimensional invertible systems will be considered in section 5.6.

5.5.1 Motivation

Let us consider the following question. Suppose we are given a family of maps of the interval, $f_p : I_x \to I_x$, for $p \in I_p$ and noisy measurement data, $\{y_n\}$, such that:

$$\begin{aligned} x_{n+1} &= f_{p_0}(x_n) \\ \text{and} \qquad y_n &= x_n + v_n, \end{aligned}$$

where $x_n \in I_x$ for all n, $I_x \subset \mathbb{R}$, and $p_0 \in I_p \subset \mathbb{R}$ such that f_{p_0} is chaotic. Suppose also that the v_n's are zero mean Gaussian independent variables with covariance matrix, R_n, and that we have some *a priori* knowledge about the value of p_0. Given this information, we would like to use the state samples, $\{y_n\}$, to get a better estimate of p_0. Let us assume for the moment that we have plenty of computing power and time. What sort of method is likely to extract the most possible information about the parameters of the system given the state data?

The first thing one might try is to simply start picking parameter values, p, near p_0 and initial conditions, x, near y_0, and attempt to iterate orbits of the form $\{f_p^i(x)\}_{i=0}^n$ to see if they come close to $\{y_i\}_{i=0}^n$. If no orbit of f_p follows $\{y_i\}_{i=0}^n$ then we know that $p_0 \neq p$. As we increase n, many orbits of the form $\{f_p^i(x)\}_{i=0}^n$ diverge from $\{y_i\}_{i=0}^n$, and we can gradually discard more and more values of p as candidates for the actual parameter value, p_0.

5.5.2 Overview

In order to implement this idea, we first need some criteria for measuring how close orbits of f_p follow $\{y_i\}$ and some rules for how to use this information to decide whether the parameter value, p, should remain a candidate for our estimate of p_0. Basically, we want

p to be eliminated if the best shadowing orbit, $\{f_p^i(x)\}$, of f_p is far enough away from $\{y_i\}$ that it is highly unlikely that sampling $\{f_p^i(x)\}$ could have resulted in $\{y_i\}$, given the expected measurement noise. As discussed earlier, one way to do this is to think of x_n, y_n, and p_0 as random variables and to consider a probability density function of the form, $P(x_0, p_0|y^n)$. Our goal will be to numerically sample such probability densities and use the results to extract information about the parameters. This is accomplished in stages, since we can only reliably compute orbits for a limited number of iterates at once. Information from various stages can then be combined to construct the composite density, $P(x_0, p_0|y^n)$, for increasing values of n.

So, for example, let us examine how to analyze the kth stage of observations, consisting of the data, $\{y_i\}_{N_k}^{N_{k+1}}$, where N_{k+1} is chosen to be as far away from N_k as possible without greatly affecting the numerical computation of orbits shadowing $\{y_i\}_{N_k}^{N_{k+1}}$. Let $y[a,b] = (y_a, y_{a+1}, \ldots, y_b)$, be a vector of state data. We begin by picking values of p near p_0. For each of these parameter samples, p, we pick a number of initial conditions, x, and iterate out orbits of the form $\{f_p^i(x)\}_{i=N_k}^n$ for $n \geq N_k$ to evaluate $P(x_{N_k}|p_0, y[N_k, n])$ for increasing values of n.[9]

For each $n \geq N_k$ we want to keep track of the set of initial conditions $x_0 \in I_p$ such that $P(x_{N_k}|p_0, y[N_k, n])$ is above a threshold value. If $P(x_{N_k}|p_0, y[N_k, n])$ is below the threshold for some value of x_{N_k}, we discard the orbit $\{f_p^i(x_{N_k})\}_{i=0}^n$ because it is too far from $\{y_i\}_{N_k}^n$ and attempt to repopulate a region, $U_k(p,n) \subset I_x$, in state space with more initial conditions, where $U_k(p,n)$ is constrained so that $x \in U_k(p,n)$ implies that $P(x_{N_k}|p_0, y[N_k, n])$ is above the threshold. Some care must be taken in figuring out how to choose $U_k(p,n)$ so that new initial conditions can be generated effectively. Without care, these regions develop Cantor-set-like structure that is difficult to deal with.

After collecting information from various stages, we then recursively combine the information from consecutive stages (similar to probabilistically combining densities in the Kalman filter) in order to determine the appropriate overall statistics for concatenated orbits over multiple stages. After combining information, at the end of each stage we also take a look at the composite densities for the various parameter samples, p. Values of p whose densities are too low are thrown out, since this means that f_p has no orbits which closely shadow $\{y_i\}_{i=0}^{N_{k+1}}$. The surviving parameter set, i.e., the set in parameter space still being considered for the parameter estimate, must then be repopulated with new parameter samples. The statistics of the new parameter samples may be determined through a combination of interpolation with nearby parameter samples and recomputation of the statistics of nearby stages. Because of the asymmetrical behavior in

[9]Note that $P(x_{N_k}|p_0, y[N_k, n])$ is sufficient to determine $P(x_{N_k}, p_0|y^n)$ for any particular value of p, since

$$P(x_{N_k}, p_0|y^n) = P(x_{N_k}|p_0, y^n)P(p_0)$$

where $P(p_0)$ is a normalizing factor quantifying *a priori* information about the parameters.

shadowing discussed in chapters 3 and 4, we find that $P(x_0, p_0|y^{N_k})$ generally has an extremely asymmetrical structure with respect to p. Specifically, the density $P(x_0, p_0|y^{N_k})$ generally drops off extremely rapidly for parameters at either the higher or lower end of the surviving parameter range (see numerical results in section 6.1). This allows us to get an extremely accurate parameter estimate for p_0 by simply choosing our estimate, $\hat{p}_{N_{k+1}}$, to be the extremum of the surviving parameter range where the density drops off rapidly.

A block diagram summarizing the main steps in algorithm is shown in figure 5.3.

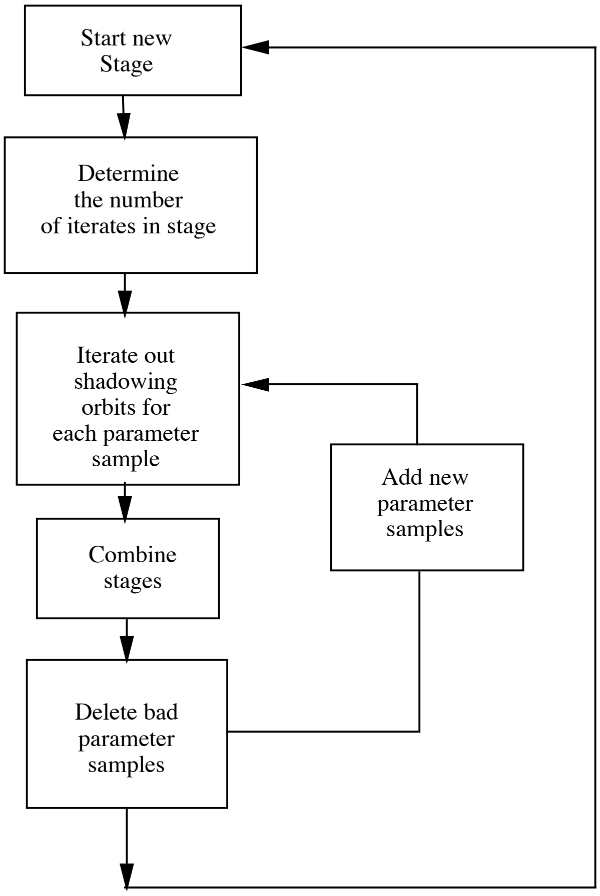

Figure 5.3: This block diagram illustrates the main steps in the proposed estimation algorithm for one-dimensional systems. The algorithm breaks up the data in sections called "stages." The diagram above shows the basic steps the algorithm takes in analyzing each stage of data.

5.5.3 Implementation

Below we explain various aspects of the algorithm in more depth. Note that unless otherwise indicated, x_n, y_n, and p_0 refer to random variables in the discussion below.

Evaluating probability densities

The first thing we must address is how to compute the values of relevant densities. From (5.24) we have that:

$$P(x_0, p_0|y^n) = \frac{P(y_n|x_0, p_0)P(x_0, p_0|y^{n-1})}{P(y_n|y^{n-1})}. \tag{5.33}$$

Expanding the right hand side of this equation recursively we have:

$$P(x_0, p_0|y^n) = K_1 P(x_0, p_0) \prod_{i=0}^{n} N(y_i; f_{p_0}^i(x_0), R_i) \tag{5.34}$$

where K_1 is some constant and $P(x_0, p_0)$ is the probability density representing *a priori* knowledge about the values of x_0 and p_0, while $N(f_{p_0}^i(x_0); y_i, R_i)$ is the value of a Gaussian density with mean $f_{p_0}^i(x_0)$ and covariance matrix R_i evaluated at y_i. In the limit where no *a priori* knowledge about x_0 is available, the weighting factor, $P(x_0, p_0)$, reduces to $P(p_0)$, reflecting *a priori* information about the parameters. Then, taking the natural log of (5.34) we get that:

$$\log[P(x_0, p_0|y^n)] = K_2 + \log[P(p_0)] - \frac{1}{2}\sum_{i=0}^{n}(f_p^i(x_0) - y_i)^T R_i^{-1}(f_{p_0}^i(x_0) - y_i). \tag{5.35}$$

where K_2 is a constant. Note that except for the extra term corresponding to the *a priori* distribution for p_0, maximizing (5.35) is essentially the same as minimizing a least squares criterion. Also note that for any particular value of p_0 we have from (5.35) that:

$$\begin{aligned} \log[P(x_0|p_0, y^n)] &= \log[P(x_0, p_0|y^n)] - \log[P(p_0)] \\ &= K_2 - \frac{1}{2}\sum_{i=0}^{n}(f_p^i(x_0) - y_i)^T R_i^{-1}(f_p^i(x_0) - y_i). \end{aligned} \tag{5.36}$$

Representing and dividing state regions

Given a parameter sample, p_0, and stage, k, we need to specify how to choose sample trajectories, $\{f_{p_0}^i(x_{N_k})\}_{i=0}^{n-N_k}$, to shadow $\{y_i\}_{i=N_k}^{n}$ for $n \in \{N_k, N_k+1, \ldots, N_{k+1}\}$. For each $n \in \{N_k, N_k+1, \ldots, N_{k+1}\}$ we want to keep track of the set of interesting initial conditions, $U_k(p_0, n) \subset I_x$, from which to choose states, x_{N_k}, to evaluate the density, $P(x_{N_k}|p_0, y[N_k, n])$. We require that if $x_{N_k} \in U_k(p_0, n)$, then x_{N_k} must satisfy the following thresholding condition:

$$\log[P(x_{N_k}|p_0, y[N_k, n])] \geq \sup_{x_{N_k} \in I_x} \{\log[P(x_{N_k}|p_0, y[N_k, n])]\} - \sigma^2 \tag{5.37}$$

for some constant, $\sigma > 0$ so that the orbit, $\{f_{p_0}^i(x_{N_k})\}_{i=0}^{n-N_k}$, follows sufficiently close to $\{y_i\}_{i=N_k}^n$. σ can be interpreted to be a measure of the maximum number of standard deviations x_{N_k} is allowed to be from the best shadowing orbit of the map, f_{p_0}. This interpretation arises since if $P(x_{N_k}|p_0, y^n)$ were Gaussian, condition (5.37) would be satisfied by all states, x_{N_k}, within σ standard deviations of the mean, $\hat{x}_{N_k}(p_0, n) = \int_{x_{N_k} \in I_x} x_{N_k} P(x_{N_k}|p_0, y[N_k, n])dx$.[10] To be reasonably sure we don't accidentally eliminate important shadowing orbits of f_{p_0} close to $\{y_i\}$, we might choose, for example, for σ to be between 8 and 12.

Given a parameter sample, p_0, let $V_k(p_0, n) \subset I_x$ represent the set of all $x_{N_k} \in I_x$ satisfying (5.37). Recall that $U_k(p_0, n)$ represents the set of points from which we will choose new sample initial conditions, x_{N_k}. We know that we want $U_k(p_0, n) \subset V_k(p_0, n)$, but problems arise if we always attempt to saturate the set $V_k(p_0, n)$ with sample trajectories. For low values of n, $V_k(p_0, n)$ is an interval. In this case, let $U_k(p_0, n) = V_k(p_0, n)$ and we can simply choose initial conditions, x_{N_k}, at random inside $V_k(p_0, n)$ to generate samples of $P(x_{N_k}|p_0, y[N_k, n])$. As n gets larger, $V_k(p_0, n)$ tends to shrink as $f_{p_0}^{n-N_k}$ expands regions in state space and more trajectory samples get discarded from consideration for failing to satisfy (5.37). However, as long as $V_k(p_0, n)$ is an interval, continue to set $U_k(p_0, n) = V_k(p_0, n)$, since it is not hard to keep track of $V_k(p_0, n)$ to repopulate the region with new trajectory samples.

A problem occurs, however, because of the folding around turning points. If the region, $f_{p_0}^m(V_k(p_0, m))$, contains a turning point for some integer $m > 0$, then as n grows larger than m, $V_k(p_0, n)$ may split into two distinct intervals, $V_k^+(p_0, n)$ and $V_k^-(p_0, n)$. Folding causes the two separate regions to get mapped into each other by $f_{p_0}^{m+1}$ (i.e., $f_{p_0}^{m+1}(V_k^+(p_0, n)) = f_{p_0}^{m+1}(V_k^-(p_0, n))$). In addition, the new intervals, $V_k^+(p_0, n)$ and $V_k^-(p_0, n)$, can also be split apart into other separate intervals by similar means as n increases. In principle, this sort of phenomenon can happen arbitrarily many times, turning $V_k(p_0, n)$ into a collection of thin, disjoint intervals. This makes it difficult to keep up with a characterization of $V_k(p_0, n)$, and makes it difficult to know how to choose new initial conditions, $x_{N_k} \in V_k(n, p)$, to replace trajectory samples that have been eliminated.

Instead of attempting to keep up with all the separate areas of $V_k(p_0, n)$, and trying to repopulate all these areas with new state samples, we let $U_k(p_0, n) \subset V_k(p_0, n)$ be the single connected interval of $V_k(p_0, n)$ where $P(x_{N_k}|p_0, y[N_k, n])$ is a maximum.[11] We

[10] One might think that this Gaussian assumption may be a bad one and that in general we might, for instance, want to make sure that we kept a set, Q, of initial states such that $Pr(x_{N_k} \in Q|p_0) > 1 - \alpha$ for $\alpha > 0$ small, where $Pr(X)$ is the probability of event X. However, in practice, the condition (5.37) is simpler to evaluate and works well for all the problems encountered. The choice of thresholding value is not critically important as long as it is not so high that close shadowing orbits are thrown away from consideration.

[11] Strictly speaking we actually want to maximize $P(x_{N_{k-1}}|p_0, y[N_{k-1}, N_k])P(x_{N_k}|p_0, y[N_k, n])$, (see the section on how to combine data). In practice this almost always amounts to maximizing

know that the separate areas of $V_k(p_0, n)$ eventually get mapped into each other, so there is no way that one of the separate areas of $V_k(p_0, n)$ can end up shadowing $\{y_i\}$ if no states in $U_k(p_0, n)$ can shadow $\{y_i\}$. Since we are primarily interested in the best shadowing orbit of f_{p_0}, keeping up with orbits with initial conditions in $U_k(p_0, n)$ is adequate.

Finally, note also that it is sometimes obvious that the parameter sample, p_0, cannot possibly be the correct parameter value. This happens if no orbit of f_{p_0} comes anywhere close to shadowing $\{y_i\}$. In this case we can immediately discard parameter sample, p_0, from consideration.

Deciding what parameters to keep

We need to evaluate how good a parameter sample is, so we know which parameter samples to keep and which parameters to eliminate as a possible choice for the parameter estimate. After the completion of stage k, we evaluate a parameter sample, p_0, according to the following criterion:

$$L_{k+1}(p_0) \quad = \quad \sup_{x_{N_k} \in I_x} \{\log[P(x_{N_k}, p_0 | y^{N_{k+1}})]\} \tag{5.38}$$

which is what one would expect if we were interested in obtaining a MAP estimate. Let $\mathcal{P}_k$ be the set of parameter samples valid at the start of the kth stage. We will eliminate a parameter sample, p_0, after the kth stage if it satisfies the following formula:

$$L_{k+1}(p_0) < \sup_{p' \in \mathcal{P}_k} \{L_{k+1}(p')\} - \sigma^2.$$

where $\sigma > 0$ is some measure of the number of standard deviations p is allowed to be from the most likely parameter value.

Choosing the number of iterates per stage

The necessity of breaking up orbits into stages is apparent, since orbits can be reliably computed only for a limited number of iterates. We now explain how to determine the number of iterates in each stage. Let $\hat{p}_{MAP}(k)$, be the MAP estimate for p_0, at the beginning of stage k (ie $p = \hat{p}_{MAP}(k)$ is the parameter sample that maximizes $L_k(p)$ for any $p \in \mathcal{P}_k$). We want to choose N_{k+1} to be as large as possible provided we are still able to reliably compute orbits of the form $\{f^i_{p_0}(x_{N_k})\}_{i=0}^{N_{k+1}-N_k}$ to shadow $\{y_i\}_{i=N_k}^{N_{k+1}}$.

Suppose that $x_{N_k} \in U_k(p_0, n)$. A reasonable measure of the number of iterates we can reliably compute for an orbit like $\{f^i_{p_0}(x_{N_k})\}_{i=0}^{n-N_k}$ is given by the size of $U_k(p_0, n)$. If $U_k(p_0, n)$ is small, this implies that small changes or errors in initial state get magnified to magnitudes on the order of the measurement noise. Since we need to compute states to accuracies better than the measurement noise, it makes sense to pick N_{k+1} so that $U_k(p_0, N_{k+1})$ is a few orders of magnitude above the precision of the computer.

$P(x_{N_k} | p_0, y[N_k, n])$ because $U_k(p_0, n)$ is generally much smaller than $f^{N_k - N_{k-1}}(U_{k-1}(p_0, N_k))$.

One complication that can arise, is that the sequence of states, $\{y_{N_k}, y_{N_k+1}, \dots\}$, might correspond to an especially parameter-sensitive stretch of points, so that there may be no orbit of $f_{\hat{p}_{MAP}(k)}$ that shadows the data, $\{y_i\}_{i=N_k}^{n}$. In this case, we cannot use the size of $U_k(\hat{p}_{MAP}(k), n)$ to determine N_{k+1}. Instead of using $\hat{p}_{MAP}(k)$ pick the next best parameter sample in $\mathcal{P}_k$, $\hat{p}'(k)$, where $\hat{p}'(k)$ maximizes $L_{N_k}(p)$ for any $p \in \mathcal{P}_k$, besides $\hat{p}_{MAP}(k)$. We then try to play the same procedure with $\hat{p}'$ that we described for $\hat{p}_{MAP}(k)$. Similarly, if $f_{\hat{p}'}$ cannot shadow the data choose another parameter value from $\mathcal{P}_k$, and so forth. Eventually some parameter value in $\mathcal{P}_k$ must work, or else either: (1) there are not enough parameter samples, or (2) p_0 is not in the parameter space region specified upon entrance to the kth stage. This can be especially be a problem at the beginning of the estimation process when the parameters are not known well, and parameter samples are more sparse in parameter space. The solution is to choose parameters intelligently, choosing varying numbers of parameter samples in different regions of parameter space and in different situations (for example, to initialize the estimation routine).

Combining data from stages

As in the Kalman filter, we want to build a recursive algorithm so that data summarizing information for stages 1 through $k-1$ can be combined with information from stage k to produce results which summarize all knowledge about stages 1 through k. Specifically, suppose that $y[N_k, N_{k+1}] = (y_{N_k}, y_{N_k+1}, \dots, y_{N_{k+1}})$ represents the state samples of the kth stage. We propose to compute $L_{k+1}(p_0)$ using information given in $L_k(p_0)$, $P(x_{N_{k-1}}|p_0, y[N_{k-1}, N_k])$, and $P(x_{N_k}, p_0|y[N_k, N_{k+1}])$. Then all information about stages 1 through k can be represented by $L_{k+1}(p_0)$ and $P(x_{N_k}|p_0, y[N_k, N_{k+1}])$.

From (5.38) we see that $L_k(p_0)$ depends only on $P(x_{N_{k-1}}, p_0|y^{N_k})$ evaluated on the orbit that best shadows the first N_k state samples. In other words if $\{\hat{x}_{i|N_k}\}_{i=0}^{N_k}$ is the best shadowing orbit based on the first N_k state samples, then from (5.38) and (5.35):

$$\begin{aligned} L_k(p_0) &= \log[P(x_{N_{k-1}} = \hat{x}_{N_{k-1}|N_k}, p_0|y^{N_k})] \\ &= K_2 + \log[P(p_0)] - \frac{1}{2}\sum_{i=0}^{N_k}(\hat{x}_{i|N_k} - y_i)^T R_i^{-1}(\hat{x}_{i|N_k} - y_i). \end{aligned} \tag{5.39}$$

One key thing to notice is that $U_{k-1}(p_0, N_k)$ and $U_k(p_0, N_{k+1})$ should be very small compared to the measurement noise, R_i, for any i. This is a reasonable assumption as long as none of the measurements have relative accuracies on the order of the machine precision. Therefore we can approximate $\hat{x}_{i|N_{k+1}}$ with $\hat{x}_{i|N_k}$ for $i \in \{0, 1, \dots, N_{k-1}\}$ in (5.39) and if we let:

$$A_k(p_0) = \log[P(p_0)] - \frac{1}{2}\sum_{i=0}^{N_{k-1}}(\hat{x}_{i|N_{k+1}} - y_i)^T R_i^{-1}(\hat{x}_{i|N_{k+1}} - y_i) \tag{5.40}$$

Then from (5.36), (5.39), and (5.40):

$$L_k(p_0) \approx A_k(p_0) + \sup_{x_{N_{k-1}} \in I_x} \{\log[P(x_{N_{k-1}}|p_0, y[N_{k-1}, N_k])]\} \tag{5.41}$$

and also:

$$
\begin{aligned}
L_{k+1}(p_0) \quad \approx \quad & A_k(p_0) - \frac{1}{2}\sum_{i=N_{k-1}}^{N_k} (\hat{x}_{i|N_{k+1}} - y_i)^T R_i^{-1}(\hat{x}_{i|N_{k+1}} - y_i) \\
& + \sup_{x_{N_k}\in I_x} \{\log[P(x_{N_k}|p_0, y[N_k, N_{k+1}])]\}. \qquad (5.42)
\end{aligned}
$$

We can now evaluate (5.42) given the appropriate representations of $L_k(p)$, $P(x_{N_{k-1}}|p_0, y[N_{k-1}, N_k])$, and $P(x_{N_k}|p_0, y[N_k, N_{k+1}])$. The term on the right hand side of (5.42) involving $\sup_{x_{N_k}\in I_x}$ can be approximated from our representation of the density $P(x_{N_k}|p_0, y[N_k, N_{k+1}])$ by simply taking the maximum density value over all the trajectory samples. Likewise $A_k(p_0)$ can be evaluated from (5.41) in a similar manner given $L_k(p_0)$. The trajectory $\{\hat{x}_{i|N_{k+1}}\}_{i=N_{k-1}}^{N_k}$ can be approximated by looking for trajectory sample $x' \in U_{k-1}(p_0, N_k)$ in the representation for $P(x_{N_{k-1}}|p_0, y[N_{k-1}, N_k])$ that makes $f_{p_0}^{N_k - N_{k-1}}(x')$ as close to $U_k(p_0, N_{k+1})$ as possible. Then let $\hat{x}_{i|N_{k+1}} = f_{p_0}^{i-N_{k-1}}(x')$ for $i \in \{N_{k-1}, \ldots, N_k\}$.

Note that this assumes that $U_k(p_0, N_{k+1}) \subset f_{p_0}^{N_k - N_{k-1}}(U_{k-1}(p_0, N_k))$. If this is not true then no orbit of f_{p_0} adequately shadows $\{y_i\}_{i=0}^{N_{k+1}}$, and we can throw out the parameter sample p_0.

Choosing new parameter samples and evaluating associated densities

Once a parameter sample is deleted because it does not satisfy (5.37), a new parameter sample must be chosen along with the appropriate statistics and densities. We want choose new parameters after stage k so that they adequately describe $L_{k+1}(p)$ over the surviving parameter range. In other words we attempt to choose new parameters to fill in gaps in parameter space where nearby parameter samples, p_1 and p_2, for example, have very different values of $L_{k+1}(p_1)$ and $L_{k+1}(p_2)$.

Once we choose the new parameter sample, p_*, we need to evaluate the relevant statistics, namely $L_{k+1}(p_*)$ and $P(x_{N_k}|p_0 = p_*, y[N_k, N_{k+1}])$. We could, of course, do this by going back through all of data $\{y_i\}_{i=0}^{N_{k+1}}$ and sampling the appropriate densities. This, however, would be quite time-consuming, and would likely not reveal much more information about the parameters than we could get by much simpler means, assuming that enough parameter samples are used. Instead, we interpolate $A_k(p_*)$ given $A_k(p)$ for all valid parameter samples, $p \in \mathcal{P}_k$. We then compute $P(x_{N_{k-1}}|p_0, y[N_{k-1}, N_k])$ and $P(x_{N_k}|p_0, y[N_k, N_{k+1}])$ by iterating trajectory samples. We can then evaluate $L_{k+1}(p_*)$ according to (5.42).

Efficiency concerns

This algorithm is not designed to be especially efficient. Rather, it is intended to try to extract as much information about the parameters of a one-dimensional map as reasonably possible. For a discussion of some performance issues, see the next section where we apply the algorithm to the family of quadratic maps.

One way to increase the efficiency of this algorithm would be to attempt to locate the sections of the data orbit that are sensitive to parameters, and perform the appropriate analysis only on these observations. For maps of the interval this corresponds to locating sections of orbit that pass near turning points. The problem, however, is not as obvious in higher dimensions. Rather than address this issue in a one-dimensional setting, in section 5.6 we will look at how this might be done in higher dimensional systems using linear analyses.

5.6 Algorithms for higher dimensional systems

In this section we develop an algorithm to estimate the parameters of general nonuniformly hyperbolic systems. Suppose we are given a family of maps, $f_p : M \to M$, for $p \in I_p$ and noisy measurement data, $\{y_n\}$, where:

$$\begin{aligned} x_{n+1} &= f_{p0}(x_n) \\ \text{and} \qquad y_n &= x_n + v_n \end{aligned}$$

where $x_n \in M$ for all n, M is some metric space, and $p_0 \in I_p \subset \mathbb{R}$ such that f_{p_0} is nonuniformly hyperbolic. Suppose also that the v_n's are zero mean Gaussian independent random variables with covariance matrix, R_n, and that we have some *a priori* knowledge about the value of p_0. Our goal in this section is to develop an algorithm to estimate p_0 given $\{y_n\}$.

Like the algorithm for one-dimensional systems discussed in the last section, the estimation technique presented here is based on an analysis of probability densities using a Monte-Carlo-like approach. The idea, however, is to avoid the heavy computational burden typical of Monte Carlo methods by selectively choosing which pieces of data to fully analyze. Since most of the state data in a nonuniformly hyperbolic systems apparently do not contribute much information about the parameters of the system, the objective is to quickly bypass the vast majority of data, but still construct extremely accurate parameter estimates by performing intensive analyses on the small sections of data that really matter.

5.6.1 Overview

The parameter estimation algorithm has two primary components. The first component sifts through the data to locate orbit sections that might be sensitive. The second component performs an analysis on the parameter-sensitive data sections to determine the parameter estimate.

The data is first scanned using a linear estimator like the square root extended Kalman filter. As described in chapter 4, linear analyses can indicate the presence of degeneracy in the hyperbolic structure of a system. In the case of a recursive linear filter, degeneracies corresponding to parameter-sensitive stretches of data are indicated by a sharp drop in the covariance matrix of the estimate. We simply run the data through the appropriate filter, look for a drop in covariance estimate over a small number of iterates, and note the appropriate sections of data for further analysis.

The second component of the estimation technique consists of Monte-Carlo-based technique. The underlying basis for this analysis is similar to what was described in section 5.5 for one-dimensional systems. Basically the estimate is constructed by using information obtained by sampling the appropriate probability densities in state and parameter space. There are, however, a few important differences to point out from the one-dimensional algorithm. First, since the systems are invertible, we iterate the map both forwards and backwards in time[12] in order to obtain information about probability densities. Also the higher dimensionality of the systems causes a few problems with how to represent and choose regions of state space in which to generate samples. Finally instead of concatenating consecutive stages by matching initial and final conditions of sample trajectories, we generate only one stage for each section of sensitive state data. The stages are separated in space and time, so there is no matching of initial and final conditions.

5.6.2 Implementation

In this section we detail some of the basic issues that need to be addressed in order to implement the proposed algorithm.

Top-level scan filter

The data is first scanned by a square root extended Kalman filter. The implementation is straightforward: simply process the data and look for drops in the error covariance matrix. There are two parameters that may be adjusted: (1) a parameter, N, to set the number of iterates (time scale) to look for degeneracies, (2) a parameter, α, to set the threshold that governs whether a section of data is sent to the Monte-Carlo algorithm for further analysis. α is expressed in terms of a ratio of the square roots of the variances of the parameter error.

Evaluating densities

Let $y^n = (y_0, y_1, \ldots, y_n)$. To estimate parameters, we are interested in densities of

[12] For lack of a better term we use "time" to refer to increasing iterations of the discrete map f_p. For example applying f_p to a state will sometimes be called mapping forwards in time and applying $f_{p_0}^{-1}$ will be referred to as mapping backwards in time.

the form $P(x_0, p_0|y^n)$. From (5.36) we have that:

$$\begin{aligned}
&\log[P(x_0, p_0|y^n)] \\
&\qquad = \log P(p_0) + \log[P(x_0|p_0, y^n)] \\
&\qquad = K_2 + \log P(p_0) - \frac{1}{2}\sum_{i=0}^{n}(f_{p_0}^i(x_0) - y_i)^T R_i^{-1}(f_{p_0}^i(x_0) - y_i) \qquad (5.43)
\end{aligned}$$

where K_2 is a constant.

Information about probability densities is obtained by sampling in state and parameter space. For a MAP estimator, we expect that the relative merit of various parameters samples, p_0, would be evaluated according to the formula:

$$\begin{aligned}
L(p_0|y^n) &= \sup_{x_0 \in I_x} \log[P(x_0, p_0|y^n)] \\
&= \log P(p_0) + \sup_{x_0 \in I_x} \log[P(x_0|p_0, y^n)] \\
&= K_2 + \log P(p_0) - \frac{1}{2}\sup_{x_0 \in I_x}\{\sum_{i=0}^{n}(f_{p_0}^i(x_0) - y_i)^T R_i^{-1}(f_{p_0}^i(x_0) - y_i)\}.
\end{aligned}$$

In general, however, we will only consider a few sets of observations in the sequence, $\{y_i\}$. For example, suppose that for any integer, $n > 0$, the linear filter has identified $k(n)$ groups or stages of measurements that may be sensitive to parameters. Then for each $j \in \{1, 2, \ldots, k(n)\}$, define $Y_j = \{y_i | i \in S_j\}$ to be a set of sensitive measurements that have been singled by the linear filter, where the sets, $S_j \subset \mathbb{Z}$, represent the indices that can be used to identify the measurements. From our arguments in chapters 3 and 4 we expect that most of the information about the parameters of the system can be extracted locally by looking at each group of measurements individually. Thus we consider the statistic, $L_{k(n)}(p_0)$, as a replacement for $L(p_0|y^n)$ where:

$$\begin{aligned}
L_{k(n)}(p_0) &= K_2 + \log P(p_0) + \sum_{j=1}^{k(n)} \sup_{x_0 \in Y_j} \log[P(x_0, p_0|Y_j)] \\
&= K_4(k(n)) + \log P(p_0) - \frac{1}{2}\sum_{j=1}^{k(n)}[\sup_{x_0 \in I_x}\{\sum_{i \in S_j}(f_{p_0}^i(x_0) - y_i)^T R_i^{-1}(f_{p_0}^i(x_0) - y_i)\}]
\end{aligned}$$

and $K_4(k(n))$ depends only on $k(n)$.

As in the one-dimensional case, we eliminate parameter samples, p, that fail to satisfy a thresholding condition: $L_{k(n)}(p) \geq \sup_{p' \in \mathcal{P}_{k(n)}}\{L_{k(n)}(p')\} - \sigma^2$ for some $\sigma > 0$ where $\mathcal{P}_{k(n)}$ is the set of parameter samples at stage $k(n)$. In practice, if Y_j for $j \in \{1, 2 \ldots, k(n)\}$ are really the main measurements sampling parameter-sensitive areas of local folding, then $L_{k(n)}(p_0)$ in fact mirrors $L(p_0|y^n)$, at least with respect to eliminating parameter values that are not favored. This is the most important property of $L_{k(n)}(p_0)$

with respect to parameter estimation, since, as in the one-dimensional case, we would like to choose the parameter estimate, $\hat{p}_n$, to reflect the extremum of the surviving parameter range where $L(p_0|y^n)$ drops off rapidly.

Stages

Suppose that the linear filter decides that the data, $\{y_i\}$, might be sensitive near iterate $i = N_k$. Given parameter sample, p_0, we begin to examine the density, $P(x_{N_k}|p_0, y[N_k - n, N_k + n])$, for increasing values of n by generating trajectory samples of the form $\{f^i_{p_0}(x_{N_k})\}^n_{i=-n}$ and evaluating:

$$\log[P(x_{N_k}|p_0, y[N_k - n, N_k + n])] = K - \frac{1}{2}\sum_{i=-n}^{n} (f^i_{p_0}(x_{N_k}) - y_i)^T R_i^{-1} (f^i_{p_0}(x_{N_k}) - y_i)$$

for some constant, K. As in the one-dimensional case, for each n we keep only trajectory samples, x_{N_k}, that satisfy a thresholding condition like:

$$\begin{aligned} &\log[P(x_{N_k}|p_0, y[N_k - n, N_k + n])] \\ &\qquad \geq \sup_{x_{N_k} \in M} \{\log[P(x_{N_k}|p_0, y[N_k - n, N_k + n])]\} - \sigma^2 \end{aligned} \tag{5.44}$$

for some $\sigma > 0$. As n is increased, we replace trajectory samples that have been thrown out for failing to satisfy (5.44) by trying new initial conditions chosen at random from a bounded region in state space which we will denote $B_0(p_0, N_k, n)$. $B_0(p_0, N_k, n) \subset M$ plays a role analogous to $U_k(p_0, N_{k+1})$ in the one-dimensional case, except that it is a multidimensional neighborhood instead of simply an interval.

Representing sample regions

Given a specific parameter sample, p_0, we now discuss how to choose trajectory samples. In particular we examine the proper choice of $B_0(p_0, N_k, n)$ for $n \geq 0$. For any $n \geq 0$, the objective is to choose $B_0(p_0, N_k, n)$ so that it is a reasonably efficient representation of the volume of space occupied by $X_0(p_0, N_k, n)$ where $X_0(p_0, N_k, n) \subset M$ is a bounded region in state space such that $x \in X_0(p_0, N_k, n)$ satisfies (5.44). We want to choose a simple representation for $B_0(p_0, N_k, n)$ so that $B_0(p_0, N_k, n)$ is large enough that $B_0(p_0, N_k, n) \supset X_0(p_0, N_k, n)$, but small enough so that if an initial condition x is chosen at random from $B_0(p_0, N_k, n)$ then there is high probability that $x \in X_0(p_0, N_k, n)$. We get an idea for what $X_0(p_0, N_k, n)$ is by iterating old trajectory samples of the density, $P(x_{N_k}|p_0, y[N_k - (n-1), N_k + (n-1)])$, and deleting the initial conditions that do not satisfy (5.44). Based on these trajectory samples, we choose $B_0(p_0, N_k, n)$ to be a simple parallelepiped enclosing the surviving initial conditions. As new trajectory samples are chosen by picking random initial conditions in $B_0(p_0, N_k, n)$, we get a better idea about the geometry of $X_0(p_0, N_k, n)$ and can in turn choose a more efficient $B_0(p_0, N_k, n)$ to generate additional trajectory samples.

In our implementation of the algorithm, $B_0(p_0, N_k, n)$ is always represented as a box. This method has the advantage that it is extremely simple and also makes it very easy to generate a random initial condition within the region, $B_0(p_0, N_k, n)$. One could also use more sophisticated approximations for $B_0(p_0, N_k, n)$. However, no matter what representation we use for $B_0(p_0, N_k, n)$, we are likely to have trouble after a while choosing new initial conditions and iterating new sample trajectories to satisfy (5.44).

Dividing sample regions

There are two main reasons why the default choice of $B_0(p_0, N_k, n)$ as described above can cause problems. First, just as in the one-dimensional case, high probability density areas in state space can split apart into separate regions. For example, in figure 5.4 we see that regions A and B converge towards each other both forwards and backwards in time (i.e., under the action of both f_p and f_p^{-1}). Both regions include orbits that shadow $\{y_i\}_{i=N_k-n}^{N_k+n}$ for large values of n. Note that this sort of phenomenon is particularly likely to happen near areas of folding, which are the areas we are most interested in investigating. This situation is a problem because if we attempt to choose $B_0(p_0, N_k, n)$ to be a large region enclosing both A and B, then there is low probability that an initial condition chosen at random from $B_0(p_0, N_k, n)$ will satisfy (5.44). The solution to this problem, however, is not too difficult. As in one-dimensional case we simply choose $X_0(p_0, N_k, n)$ to be whichever region, A or B, has the highest density values and concentrate on sampling that region.

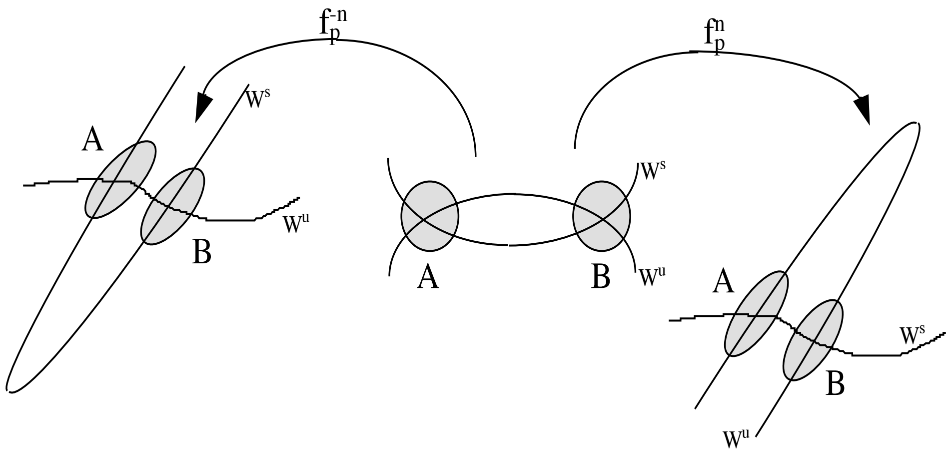

Figure 5.4: Here we illustrate why there can be multiple regions shadowing the same orbit. Near areas of folding, two regions, A and B, can be separate, yet can get asymptotically mapped toward each other both forwards and backwards in time. Note that in the picture, A and B are located at intersections of the same stable and unstable manifolds. This situation must be dealt with when sampling probability densities and searching for optimal shadowing orbits.

Avoiding degenerate sample regions

The other problem is that $X_0(p_0, N_k, n)$ tends to collapse onto a lower dimensional surface as n gets large. This is due to the fact that the map, $f_{p_0}^n$, generally contracts and expands some directions in state space more than others. Our ability to compute orbits like $\{f_{p_0}^i(x)\}_{i=-n}^n$ is related to the largest expansion factor of either $f_{p_0}^n$ or $f_{p_0}^{-n}$ (e.g., the square root of $Df_{p_0}^n(x)^T Df_{p_0}^n(x)$). If $X_0(p_0, N_k, n)$ collapses onto a lower dimensional surface, that means that across the width of the surface of $X_0(p_0, N_k, n)$, tiny differences in initial conditions get magnified to the level of the measurement noise by either $f_{p_0}^n$ or $f_{p_0}^{-n}$. For example, if $f_{p_0}^n$ is responsible for collapsing $X_0(p_0, N_k, n)$ onto a surface with thickness comparable to the machine precision, then we cannot expect to choose trajectory samples of the form $f_{p_0}^i(x)$ for $i > n$ without experiencing debilitating roundoff errors.

Ideally, as n increases, we would like $X_0(p_0, N_k, n)$ to converge toward smaller and smaller ball-shaped regions while maintaining approximately the same thickness in every direction. Besides having better numerical behavior than regions that collapse onto a lower-dimensional surface, it is also much easier to represent such regions and choose initial conditions inside these regions.

There is a degree of freedom that is available and can be used to adjust the shape of the region where initial conditions are sampled. We can simply choose to iterate trajectory samples further backwards in time than forwards in time or vice-versa. In other words, if $f_{p_0}^n$ expands one direction much more than $f_{p_0}^{-n}$ expands any direction in state space then we may iterate orbits of the form $\{f_{p_0}^i(x)\}_{i=-n_a}^{n_b}$ where $n_a > n_b$. The relative sizes of n_a and n_b can then be adjusted to match the rates of convergence of the region where initial conditions are sampled.

In practice it can be a bit tedious to adjust the number of iterates in sample trajectories and attempt to figure out what effect iterating forwards or backwards has on the shape of a particular region in state space. A better way to approach the problem is to examine regions of the form:

$$X_j(p_0, N_k, n) = f_{p_0}^j(X_0(p_0, N_k, n))$$

for $j \in \{-n, -n+1, \ldots, n-1, n\}$. For any particular p_0, N_k, and n, if $X_0(p_0, N_k, n)$ starts to become an inadequate region for choosing new sample trajectories, we simply search for a j so that the region, $X_j(p_0, N_k, n)$, is not degenerate in any direction in state space (This process is described in the next section). We can then pick new initial conditions, $x \in X_j(p_0, N_k, n)$ and iterate orbits of the form $\{f_{p_0}^i(x)\}_{i=-n-j}^{n-j}$ in order to evaluate the proper densities. Note that instead of deleting sample trajectories according to (5.44), new sample trajectories are now thrown out if they fail to satisfy

$$\log[P(x_{N_k-j}|p_0, y[N_k-n, N_k+n])] \geq \sup_{x_{N_k-j}\in M}\{\log[P(x_{N_k-j}|p_0, y[N_k-n, N_k+n])]\} - \sigma^2.$$

This procedure is thus equivalent to sampling trajectories from $X_0(p_0, N_k, n)$, except that it is better numerically.

Evaluating and choosing new sample regions

We now describe how to decide when an initial condition sample region like $X_{j_0}(p_0, N_k, n)$ has become inadequate and how to choose a new $j^* \in \{-n, -n+1, \ldots, n-1, n\}$ so that $X_{j^*}(p_0, N_k, n)$ makes an effective sample region.

Basically, as long as we can pick $B_0(p_0, N_k, n)$ so that most initial conditions, x, chosen from $B_0(p_0, N_k, n)$ satisfy $x \in X_{j_0}(p_0, N_k, n)$, then things are satisfactory, and there is no need to search for a new sample region. However, suppose that it becomes difficult to choose $x \in B_0(p_0, N_k, n)$ so that $x \in X_{j_0}(p_0, N_k, n)$. It might be the case that $X_{j_0}(p_0, N_k, n)$ is collapsing in multiple directions, and we simply cannot increase n without running into numerical problems. If this is not the case, then we first search for whether $X_{j_0}(p_0, N_k, n)$ can be divided into two separate high density regions. If so, then we concentrate on one of these regions. Otherwise we have to search for a new $j^* \in \{-n, -n+1, \ldots, n-1, n\}$ and a new sample region, $X_{j^*}(p_0, N_k, n)$.

This is done in the following manner. We take the trajectory samples marking the region, $X_{j_0}(p_0, N_k, n)$, and iterate them forwards and backwards in time looking at samples of

$$X_j(p_0, N_k, n) = f_p^{j-j_0}(X_{j_0}(p_0, N_k, n))$$

for $j \in \{-n + j_0, -n + j_0 + 1, \ldots, n + j_0\}$. We would like to pick j^* to be a value for j such that $X_j(p_0, N_k, n)$ is not degenerate, so that it is easy to pick $B_0(p_0, N_k, n)$ such that $x \in B_0(p_0, N_k, n)$ implies $x \in X_j(p_0, N_k, n)$ with high probability.

We would also like to pick j^* so that $X_{j^*}(p_0, N_k, n)$ is a well balanced region and is not degenerate in any direction. The first thing to check is to simply generate the box, $B_j(p_0, N_k, n)$, enclosing $X_j(p_0, N_k, n)$ for each j and make sure that none of its side lengths are degenerate. This condition is not adequate, however, since one could end up with a j^* in which $X_{j^*}(p_0, N_k, n)$ is actually long and thin but curls back on itself so that its bounding box, $B_j(p_0, N_k, n)$, is not long and thin. In order to check for this case, one thing to do is to partition the box, $B_j(p_0, N_k, n)$, into a number of subregions and check to see how many of these subregions are actually occupied by the trajectory samples demarking $X_j(p_0, N_k, n)$. If very few subregions are occupied then we have to reject j as a possible choice for j^*. An adequate choice for j^* can then be made using this constraint along with information about the ratio of the side lengths of $B_j(p_0, N_k, n)$.

Chapter 6

Numerical results

In this chapter we present results from various numerical experiments. In particular, we demonstrate the effectiveness of the algorithms proposed in chapter 5 for estimating the parameters of chaotic systems.

The algorithms are applied to four different systems. The first system, the quadratic map, is the same one-dimensional system that was examined in chapter 3 of this report. The second system we look at is the Henon map, a dissipative two-dimensional mapping with a strange attractor. The third system is the standard map, an area-preserving map that exhibits chaotic behavior. Finally in contrast to the first three systems, which are all nonuniformly hyperbolic, we also take a brief look at the Lozi map, one of the few nonpathological examples of a chaotic map exhibiting uniformly hyperbolic behavior.

We find that with the exception of the Lozi map, the other maps in this chapter all exhibit asymmetrical shadowing behavior on the parameter space of the map. Furthermore, this asymmetrical behavior always seems to favor one direction in parameter space regardless of locality in state space.

Note that many of the basic comments and explanations applicable to all the systems are included in section 6.1 on the quadratic map, where the issues are first encountered.

6.1 Quadratic map

In this section we describe numerical experiments on the quadratic map:

$$f_p(x) = px(1-x) \tag{6.1}$$

where $x \in [0,1]$ and $p \in [0,4]$. For values of p between 3.57 and 4.00, numerical experiments suggest that there are a large number of parameter values where (6.1) exhibits chaotic behavior. In particular we will concentrate on parameters near $p_0 = 3.9$. For $p_0 = 3.9$, numerical results indicate that f_{p_0} has a Lyapunov exponent of about 0.49.

Let us begin by presenting a summary of our results for one particular orbit of the quadratic map, the orbit with initial condition $x_0 = 0.4$. These results are summarized in figure 6.1. Our discussion in this section will seek to answer the following questions: (1) what each of the lines in figure 6.1 mean, (2) why each of the data sets graphed has the behavior shown, and (3) what we expect the asymptotic behavior for each of the traces might be if the simulations were continued for higher numbers of data points.

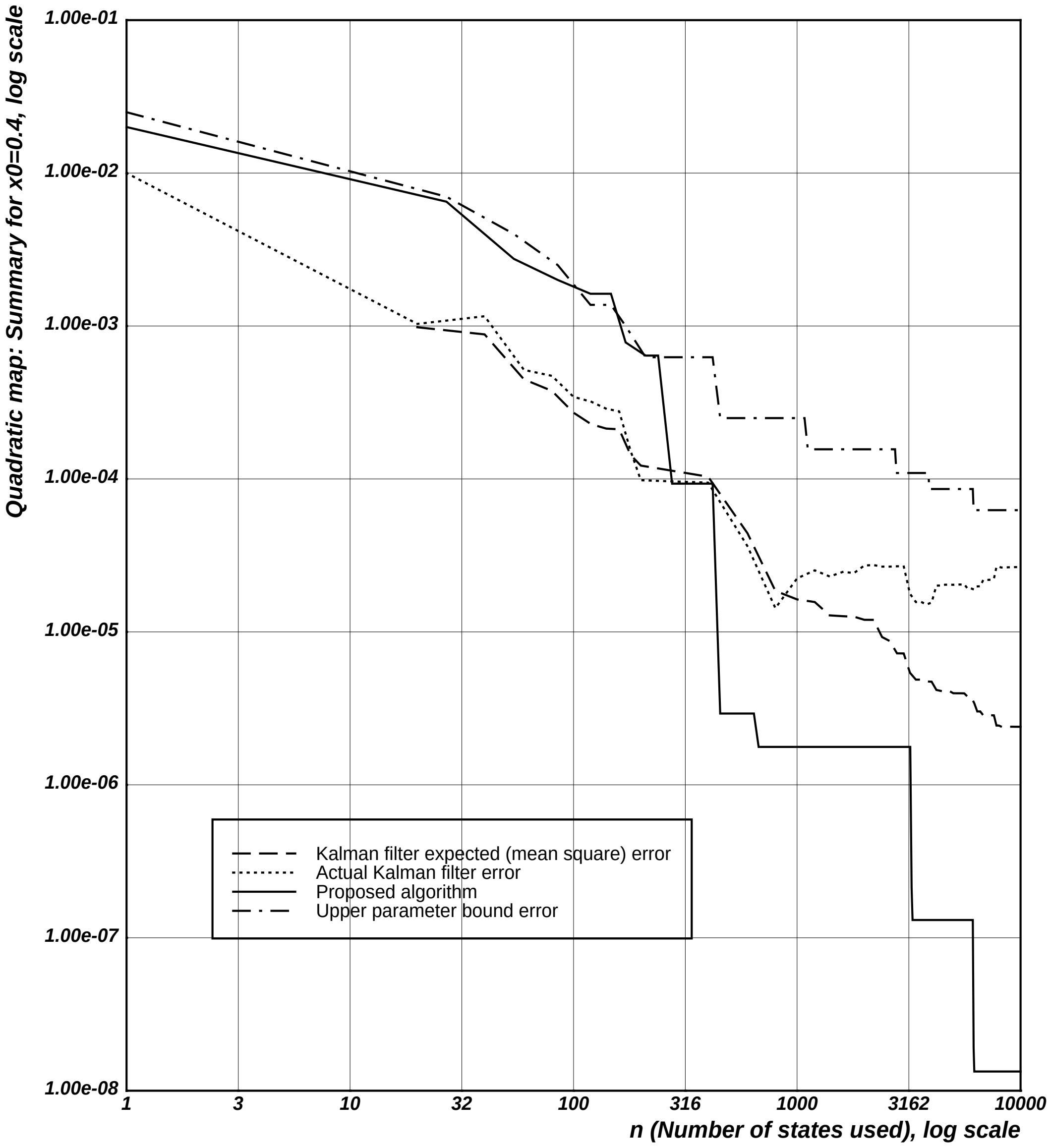

Figure 6.1: This graph summarizes results related to estimating the parameter p in the quadratic map for data generated using the initial condition $x_0 = 0.4$.

6.1.1 Setting up the experiment

In order to test parameter estimation algorithms numerically, we first pick a parameter value, p_0 and generate a sequence of data points $\{y_i\}_{i=0}^n$, to represent noisy measurements of f_{p_0}. This is done by choosing an initial condition, x_0, and numerically iterating the orbit $\{x_i = f_{p_0}^i(x_0)\}_{i=0}^n$. The noisy measurements, $\{y_i\}_{i=0}^n$, are then simulated by setting $y_i = x_i + v_i$ where the v_i's are randomly generated values for $i \in \{0, 1, \ldots, n\}$. For the experiments in this section, the v_i's are chosen to simulate independent identically distributed Gaussian random variables with standard deviation 0.001.

We then use the simulated data, $\{y_i\}_{i=0}^n$, as input to the parameter estimation algorithm to see whether the algorithm can figure out what parameter value was used to generate the data in the first place. In general the parameter estimation algorithm may also use *a priori* information like an initial parameter estimate along with some measure of how good that estimate is. In this chapter we generally choose the initial parameter estimate to be a random value within .025 of p_0.

6.1.2 Kalman filter

Let us now examine what happens when we apply the square root extended Kalman filter to the quadratic map. We investigate the Kalman filter for data generated from four different initial conditions: $x_0 = \{0.1, 0.2, 0.3, 0.4\}$.

Figure 6.2 illustrates perhaps the most important feature of the simulations, namely that the Kalman filter eventually "diverges." Each trace in figure 6.2 represents the average of ten different runs using ten different sets of numerically generated data from each initial condition. On the $y-$axis we plot the ratio of the actual error of the parameter estimate versus the estimated mean square error obtained from the covariance matrix of the filter. If the filter is working, we generally expect this ratio to be close to 1. Note also that the filter seems to start fine, but then the error jumps to many "standard deviations" of the expected error and never returns to the normal operating range.

In fairness, plotting an average can be somewhat misleading because the average might be skewed by outliers and runs that fail massively. There are in fact significant differences from run to run. However, numerous experiments with the Kalman filter suggest that divergence pretty much always occurs if one allows the filter to run long enough. In addition, none of the standard techniques for addressing divergence difficulties seem to be able to adequately solve the problem (eg, exponential forgetting of data). It seems that one is stuck with either letting the filter diverge, or somehow decreasing confidence in the covariance matrix so much that accurate estimates cannot be attained.

In figure 6.3 we plot the actual error of the Kalman filter versus number of state

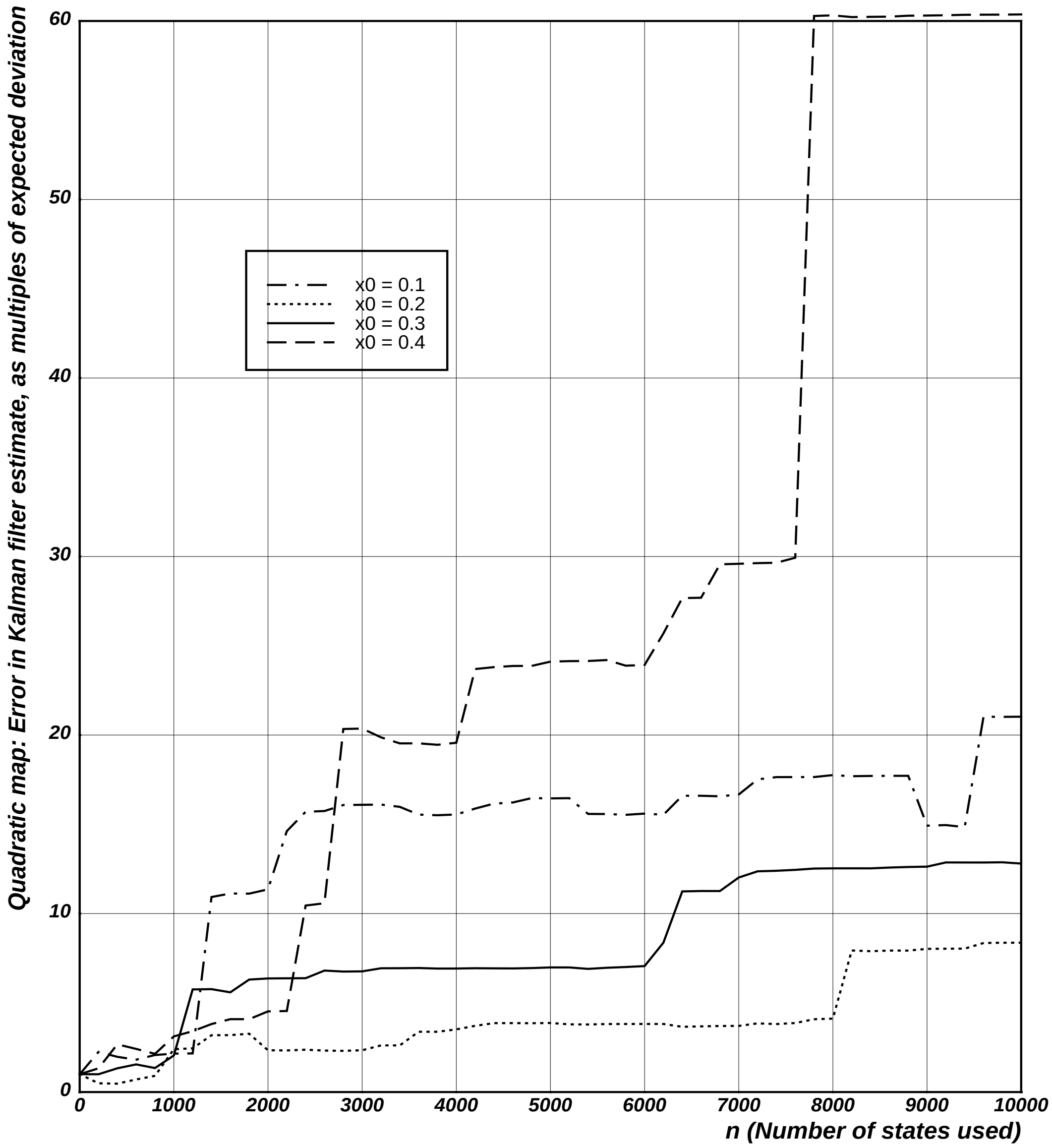

Figure 6.2: This figure shows results for applying the square root extended Kalman filter to estimating the parameters of the quadratic map with $p = 3.9$. Each trace represents the average ratio of the actual parameter estimate error to the estimated mean square error as calculated by the Kalman filter over 10 different trials. The different traces represent experiments based on orbits with different initial conditions. Note how the error jumps up to levels on the order of 10 or higher, indicating divergence.

samples used on a log-log scale. Again the errors plotted are the average of the errors of ten different runs. We see that the error makes progress for a little while but then divergence occurs. The Kalman filter rarely makes any real progress after divergence occurs, not even exhibiting the $\frac{1}{\sqrt{n}}$ improvement characteristic of purely stochastic convergence (ie, the filter is not getting any information from the dynamics), since the over-confident covariance matrix prohibits the parameter estimate from moving much unless the state data drifts many deviations away from what the filter expects. [1]

6.1.3 Analysis of proposed algorithm

We now examine the performance of the algorithm presented in section 5.5. The results in this section reflect an implementation of the algorithm based on 9 samples in parameter space and 50 samples in state space (250 when representations for different stages are being combined). Each stage is iterated until the state sample region is of length 1×10^{-9} or less. We use $\sigma = 8$ so that the sample spaces in state and parameters are 8 deviations wide.

One of the most striking things about the results of the algorithm is the asymmetry of the merit function, $L(p)$, in parameter space. As shown in figure 6.4, the parameter merit function typically shows a very sharp dropoff on the low end of the parameter space. Based on this asymmetry we choose the parameter estimate to be the parameter value at which the sharp dropoff in $L(p)$ occurs.

In figure 6.5 we see the performance of the algorithm on data based on the initial conditions, $x_0 \in \{0.1, 0.2, 0.3, 0.4\}$. Each trace in the figure represents one run of the algorithm. Rerunning the algorithm multiple times on data based on the same initial condition produces similar results, except that the scanning linear filter sometimes defers a few more or less points to the Monte Carlo estimator for analysis.

Note how the error in the estimate tends to converge in sudden large jumps over small numbers of iterates, while staying approximately constant in between these jumps. The large decreases in error level occur when the data orbit makes a close approach to the turning point, causing a stretch of state samples to become sensitive to parameters. This is not simply a product of discretization in the algorithm, since the Monte Carlo estimator sometimes makes no gains at all, while other times great gains are made, and a large number of parameter samples are deleted on the lower end of the parameter sample range.

One might wonder how this graph would look like if we were to extend it for arbitrarily many iterates. Consider the theory presented in chapter 3. First of all, it is likely

[1] Interestingly, this actually does occur, apparently near areas of folding, since the filter models the folding phenomena so poorly. Occasionally this can even cause the filter to get back in sync, moving the parameter estimate just the right amount to lower the error. This seems to be quite rare, however.

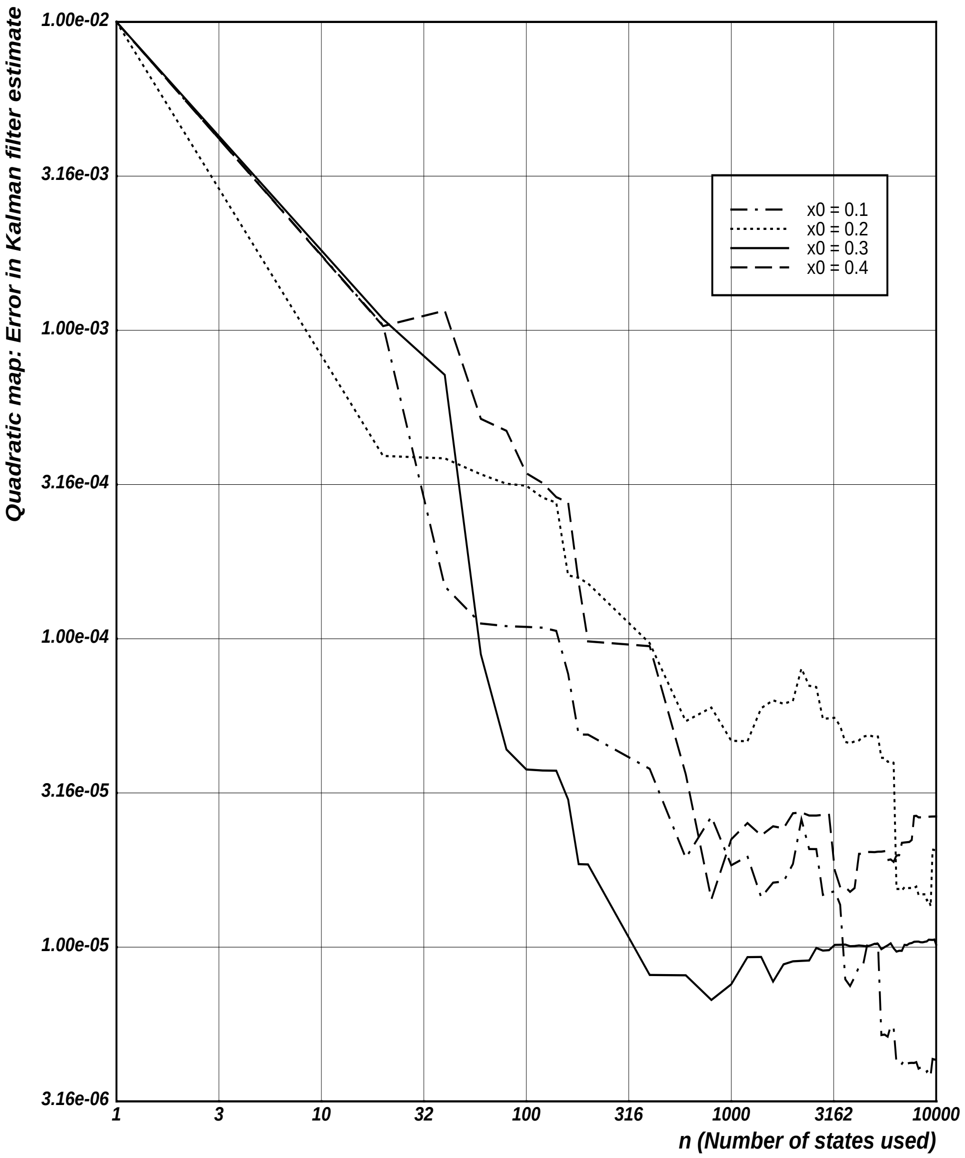

Figure 6.3: Graph of the average error in the parameter estimate as computed by square root extended Kalman filter applied to the quadratic map with parameter value $p = 3.9$. Data represents average error over 10 runs.

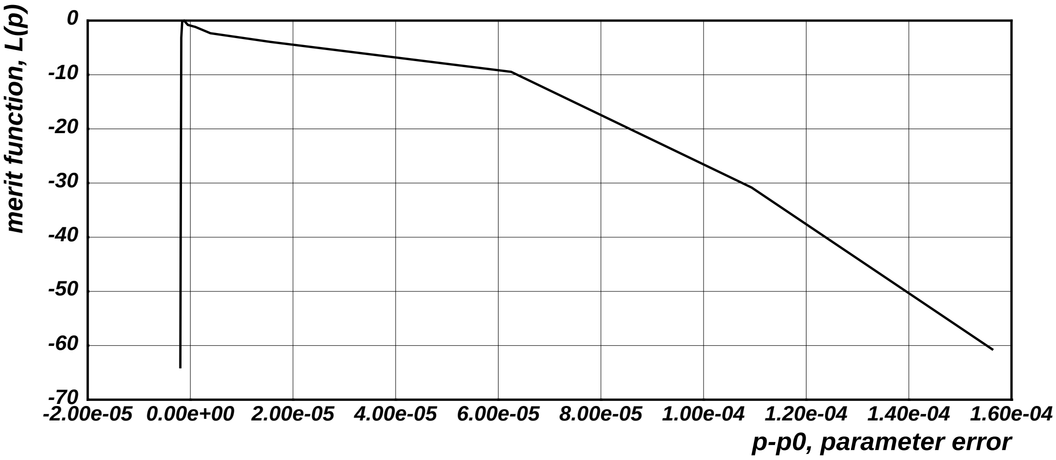

Figure 6.4: Asymmetry in the parameter space of the quadratic map: Here we graph the parameter merit function $L(p)$ after processing 2500 iterates of an orbit with initial condition $x_0 = 0.4$. The merit function is normalized so that $L(p) = 0$ at the maximum. Since $\sigma = 8$, a parameter sample, p, is deleted if $L(p) < -64$. This sort of asymmetrical merit function is typical of all orbits encountered in the quadratic map, Henon map, and standard map.

that f_{p_0} satisfies the linking condition, and therefore exhibits a parameter shadowing property. This means there is essentially an end to the progress that can be made in the estimate based on dynamical information, after which stochastic convergence would be the rule. However, there is evidence that the level of accuracy at which this effect becomes important is probably many, many orders of magnitude smaller from the level we are dealing with. [2]

This leads us to ask: assuming that we do not see the effects of parameter shadowing, how does the parameter estimation accuracy converge with respect to n, the number of state samples processed by the algorithm? As conjectured in section 3.5, we believe that the accuracy converges at a rate proportional to $\frac{1}{n^2}$. A line with a slope of -2 is drawn in figure 6.5 to suggest the conjectured asymptotic behavior. Note that the conjecture seems plausible from the picture, although more data would be needed to really make the evidence convincing.

In figure 6.6 we show the error in the upper bound of the parameter range being considered by the algorithm. While the lower bound of this range is used as the parameter estimate, the upper bound has significantly different behavior. After an initial period, the convergence of the upper bound is governed purely by stochastic means (ie, without any help from the dynamics). This is predicted by Theorem 3.4.2. Thus we expect that

[2] It is difficult to calculate this directly, since it requires knowing the exact number of iterates it takes an orbit from the turning point to return near the turning point. However, rough calculations suggest that for most parameters around $p_0 = 3.9$ we expect that parameter shadowing would not be seen until parameter deviations are less than 1×10^{-50} for noise levels of 0.001.

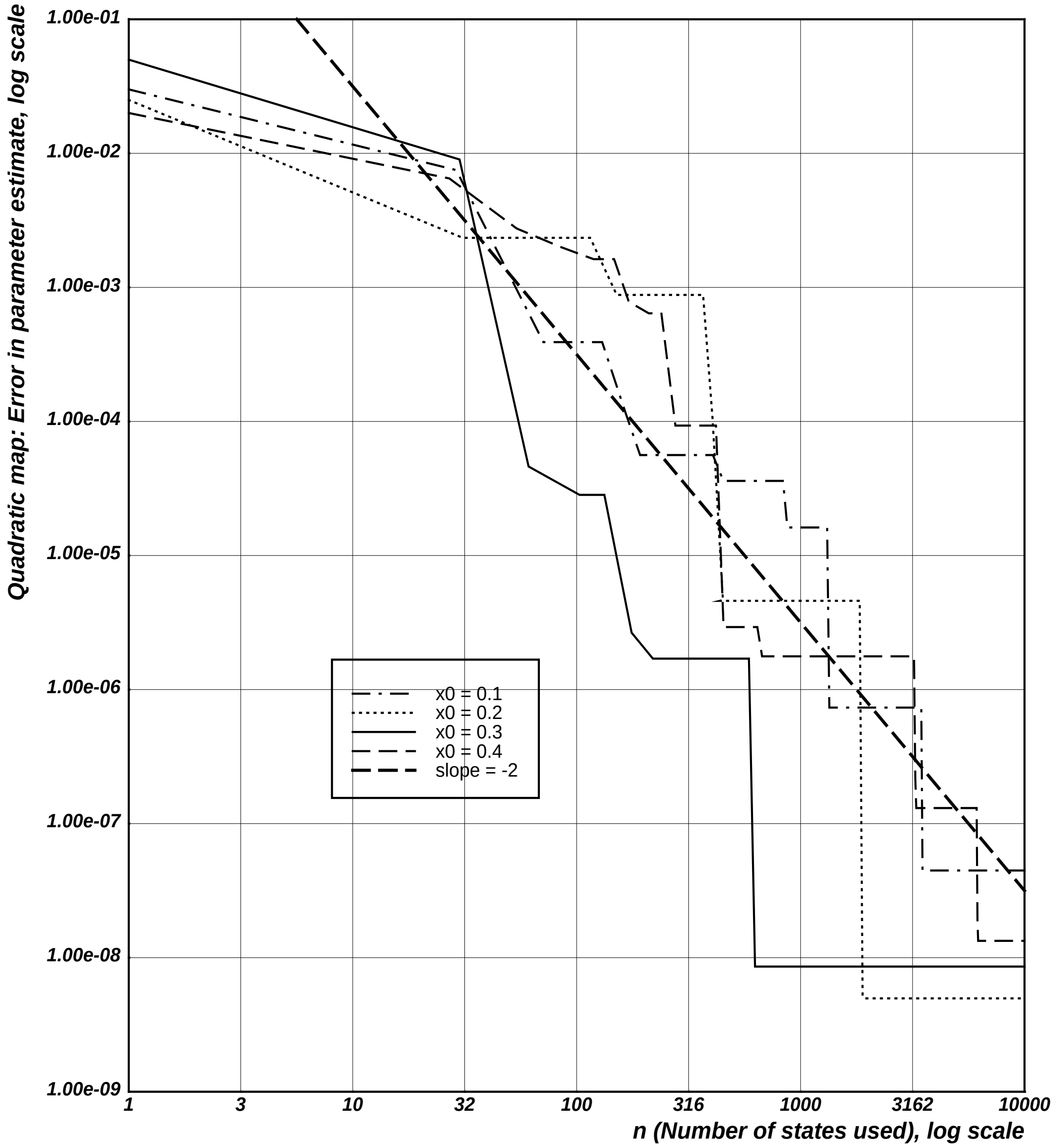

Figure 6.5: Graph of the actual error in the parameter estimate of the proposed algorithm when applied to data from the quadratic map with $p = 3.9$. A line of slope -2 is drawn on the graph to indicate the conjectured asymptotic rate of convergence for the estimate.

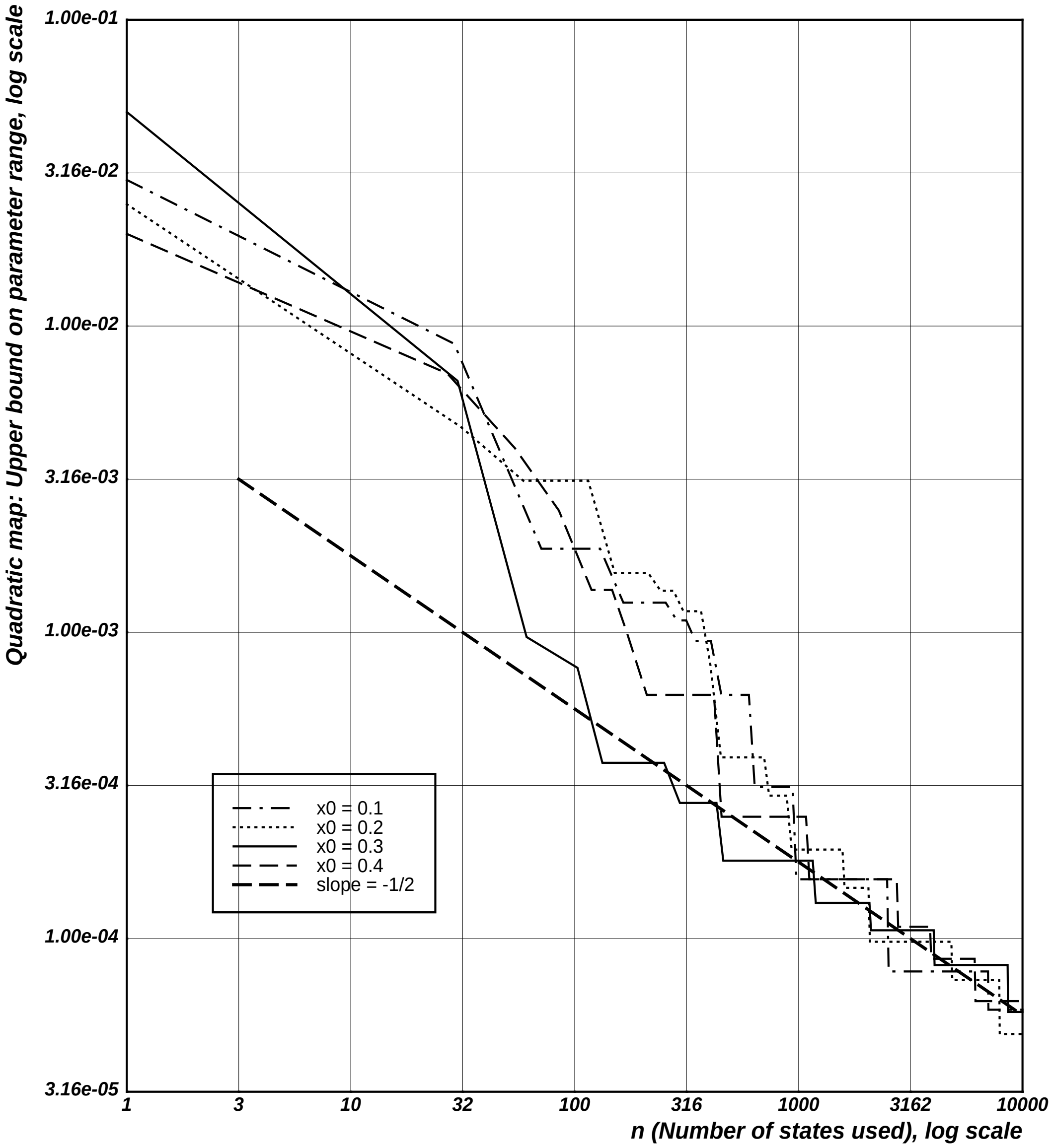

Figure 6.6: Graph of the error in the upper bound of the parameter range being considered by the proposed algorithm for the quadratic map with $p = 3.9$. A line with a slope of $-\frac{1}{2}$ is drawn to indicate the expected asymptotic convergence of the error.

the convergence will be on the order of $\frac{1}{\sqrt{n}}$, as suggested by the line with a slope of $-\frac{1}{2}$ as shown in the figure. The small jumps in the graphs for figure 6.6 are simply the result of the discrete nature of how parameter space is sampled.

6.1.4 Measurement noise

One other important question to ask is, what happens if we change the level of measurement noise? The short answer is that the parameter estimate results presented here are surprisingly insensitive to measurement noise. If we ignore the parameter shadowing effects caused by close returns to the turning point (which we have already argued are negligible for our experiments), then shadowing of any finite orbit is really an all or nothing property in parameter space. Consider a stretch of state orbit with initial condition x_0 close to the turning point. Then for a parameter value in the unfavored direction, either the parameter value can shadow that stretch of orbit (presumably with initial condition closer to the turning point than x_0), or the parameter value cannot shadow the orbit, in which case it loses track of the original orbit exponentially fast. Asymptotically, the *measurement noise actually makes no difference in the parameter estimate* other than through parameter shadowing effects caused by linking. Thus, once the measurement noise is lower than a certain level, the actual measurement noise makes very little difference in the accuracy of parameter estimates.

Measurement noise does have a large affect on figure 6.6, the upper parameter bound, and the possibility of parameter shadowing caused by linking. If the measurement noise is large, then there is likely to be more parameter shadowing effects caused by linking. On the other hand, if the measurement noise is really small, then the asymmetrical effect in parameter space will in fact get drowned out for quite a while (until the sampled orbit comes extremely close to the turning point). In most reasonable cases however, the asymmetry in parameter space is likely to be quite important if we want to get accurate parameter estimates for reasonably large data sets.

6.2 Henon map

We now discuss numerical experiments with the Henon map:

$$x_{n+1} = y_n + 1 - ax_n^2 \tag{6.2}$$

$$y_{n+1} = bx_n \tag{6.3}$$

where the state $(x_n, y_n) \in \mathbb{R}^2$ and the parameter values, a and b, are invariant. For parameter values $a = 1.4$ and $b = 0.3$, numerical evidence indicates the existence of a chaotic *attractor* as shown in figure 6.7. See Henon [27] for a more detailed description of the basic properties of Henon map.

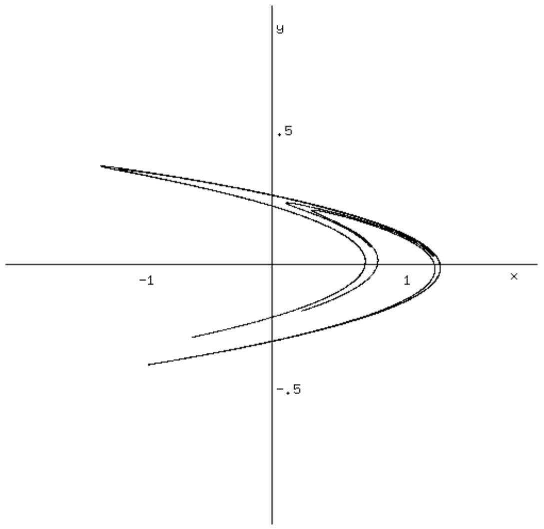

Figure 6.7: The Henon attractor for $a = 1.4$, $b = 0.3$.

For the purposes of testing out parameter estimation algorithms, we fix $b = 0.3$ and attempt to estimate the parameter, a. State data is chosen from an orbit on the attractor of the Henon map. Noisy measurement data is generated using a state orbit and adding Gaussian noise with standard deviation 0.001 to each state value.

Applying the square root extended Kalman filter to an orbit on the attractor results in figure 6.8. Observe that the filter diverges after about 15,000 iterates and does not recover. Note that the figure represents data for only one run. However, the results in figure 6.8 are representative for other sequences of data that we have tried. Although the performance of the Kalman filter is quite sensitive to noise, the key point is that divergence inevitably occurs, sooner or later, and the performance of the filter is generally unreliable.

Note in figure 6.8 that the expected mean square error of the Kalman filter tends to change suddenly in jumps. In most cases these jumps probably correspond to sections of orbits that are especially sensitive to parameters because of folding in state space. The Kalman filter has a tough time handling the folding and typically divergence occurs during one of these jumps in the mean square error. This phenomenon is also apparent in figure 6.12. Note also that even after divergence, the parameter estimate sometimes changes by many standard deviations, indicating that the state space error residual must have been many deviations off. This again reflects the fact that the Kalman filter does not model folding well.

We now apply the algorithm described in section 5.6. We choose to examine the top-level scan filter every 20 iterates or so looking for covariance matrix drops of around a factor of .7 or less. The algorithm is relatively insensitive to changes in these parameters so their choice is not particularly critical.

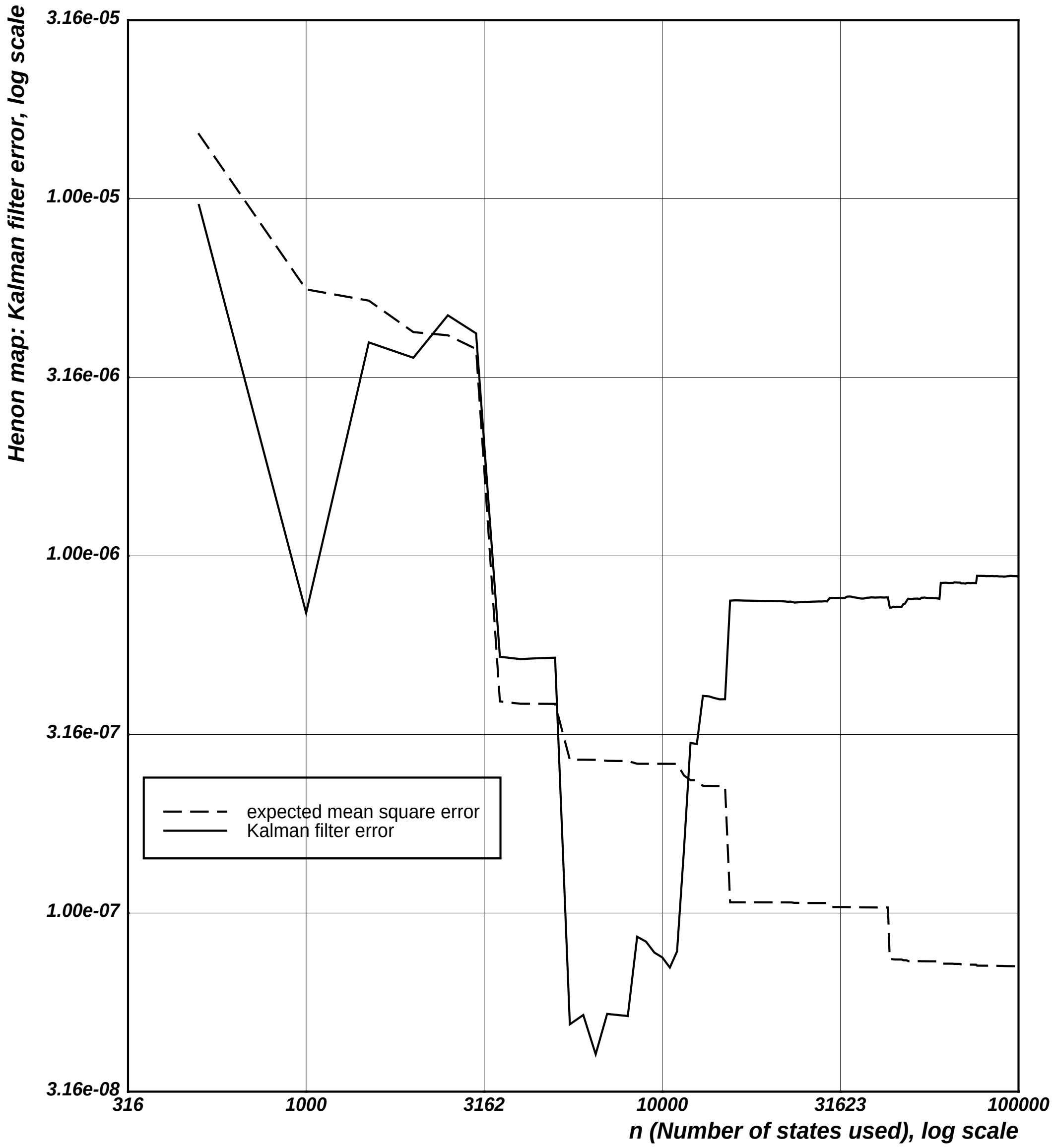

Figure 6.8: This graph depicts the performance of the Kalman filter in estimating parameter a for one sequence of noisy state data from the Henon map for $a = 1.4$ and $b = 0.3$. The data was generated using the initial condition, $(x_0, y_0) = (.633135448, 18940634)$, which is very close to the attractor.

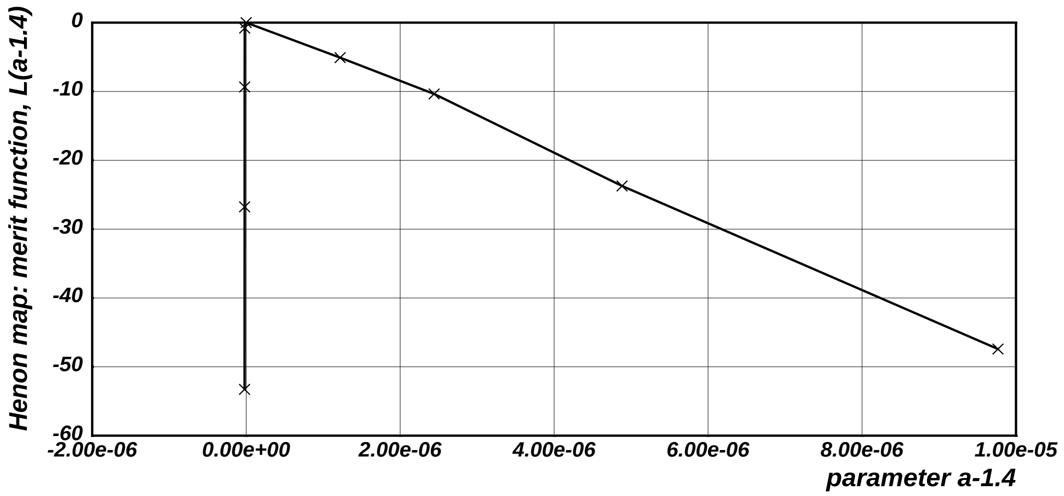

Figure 6.9: Asymmetry in the parameter space of the Henon map (with $a = 1.4$, $b = 0.3$): Here we graph the parameter merit function $L(a)$ after 200000 iterates of an orbit with initial condition on the attractor near $x_0 = (.423, .208)$. Note that this merit function is actually based on only the most sensitive 931 data points, since the linear filter threw out over 199,000 points.

As in the quadratic map, we find that the parameter merit function, $L(a)$, is asymmetrical in parameter space. Specifically, $L(a)$ always has a sharp dropoff in its lower bound, indicating that the Henon map favors higher parameters for parameter a (see figure 6.9). This property seems to be true for any orbit on the attractor. It also seems to be true for all the parameter values of the Henon that have been tried. We thus take advantage of the asymmetry in parameter space in order to estimate the parameters of the system.

Figure 6.10 shows the estimation effort for data generated from several different initial conditions on the attractor. The tick marks on the traces of the graph denote places where the top level scan filter deferred to the Monte-Carlo analysis. Note that as with the quadratic map, improvements in the estimate seem to be made suddenly. Because relatively few numbers of points are analyzed by the Monte-Carlo technique, and because the state samples scanned by the Kalman filter do not contribute to the parameter estimate, almost all the gain in parameter estimate must have been made because of the dynamics.

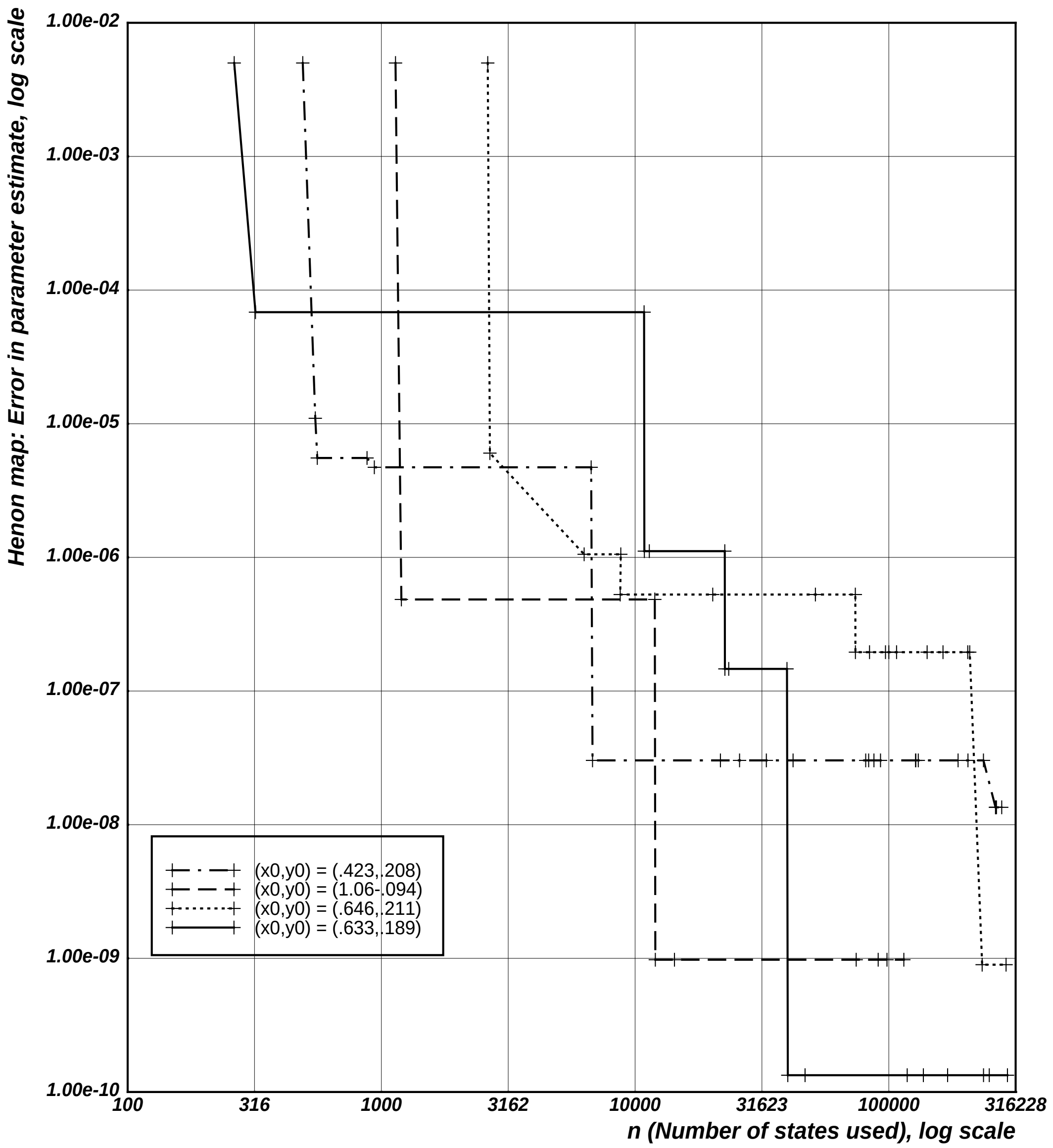

Figure 6.10: Graph of the actual error of the parameter estimate for a using the proposed algorithm on the Henon map (with $a = 1.4$ and $b = 0.3$). This graph contains results for four different sets of data corresponding to four different initial conditions, all chosen on the attractor of the system. The tick marks on each trace denote places where the top level Kalman filter deferred to a Monte-Carlo-based approach for additional analysis.

6.3 Standard map

We now discuss numerical experiments with the standard map:

$$x_{n+1} = (x_n + y_n + K \sin x) \mod 2\pi \quad (6.4)$$
$$y_{n+1} = (y_n + K \sin x) \mod 2\pi \quad (6.5)$$

where K is the parameter of the system and the state, $(x_n, y_n) \in T^2$, lives on the 2-torus, T^2. The standard map is a Hamiltonian (area-preserving) system, and thus does not have any attractors. Instead, for example, for $K = 1$, there is apparently a mixture of invariant tori and seas of chaos where non-periodic orbits wander around. This is illustrated in figure 6.11. See Chirikov [13] for more discussion on the properties of the standard map.

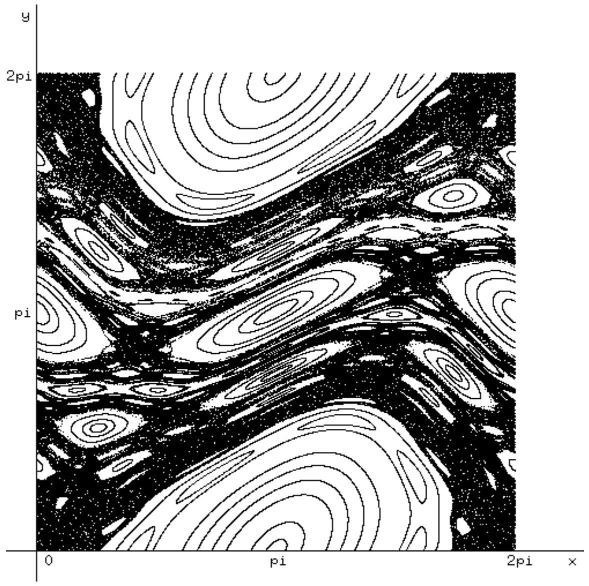

Figure 6.11: This picture shows various orbits of the standard map near $K = 1$. Note that since the space is a torus, the sides of the square are actually overlapping. This picture shows a number of different orbits. Some orbits fill out dark zones of chaotic behavior, while others remain on circular tori.

In order to test the parameter estimation technique, we picked $K = 1$ and generated data based on orbits chosen to be in a chaotic region. To each state, we added random Gaussian measurement noise with standard deviation 0.001 to produce the data set. The results of applying the square root extended Kalman filter are shown in figure 6.12. As in the quadratic map and Henon map, we see that the Kalman filter diverges.

In figure 6.14 we show the result of applying the algorithm in section 5.6 to the standard map. In particular we investigate data for five different initial conditions in

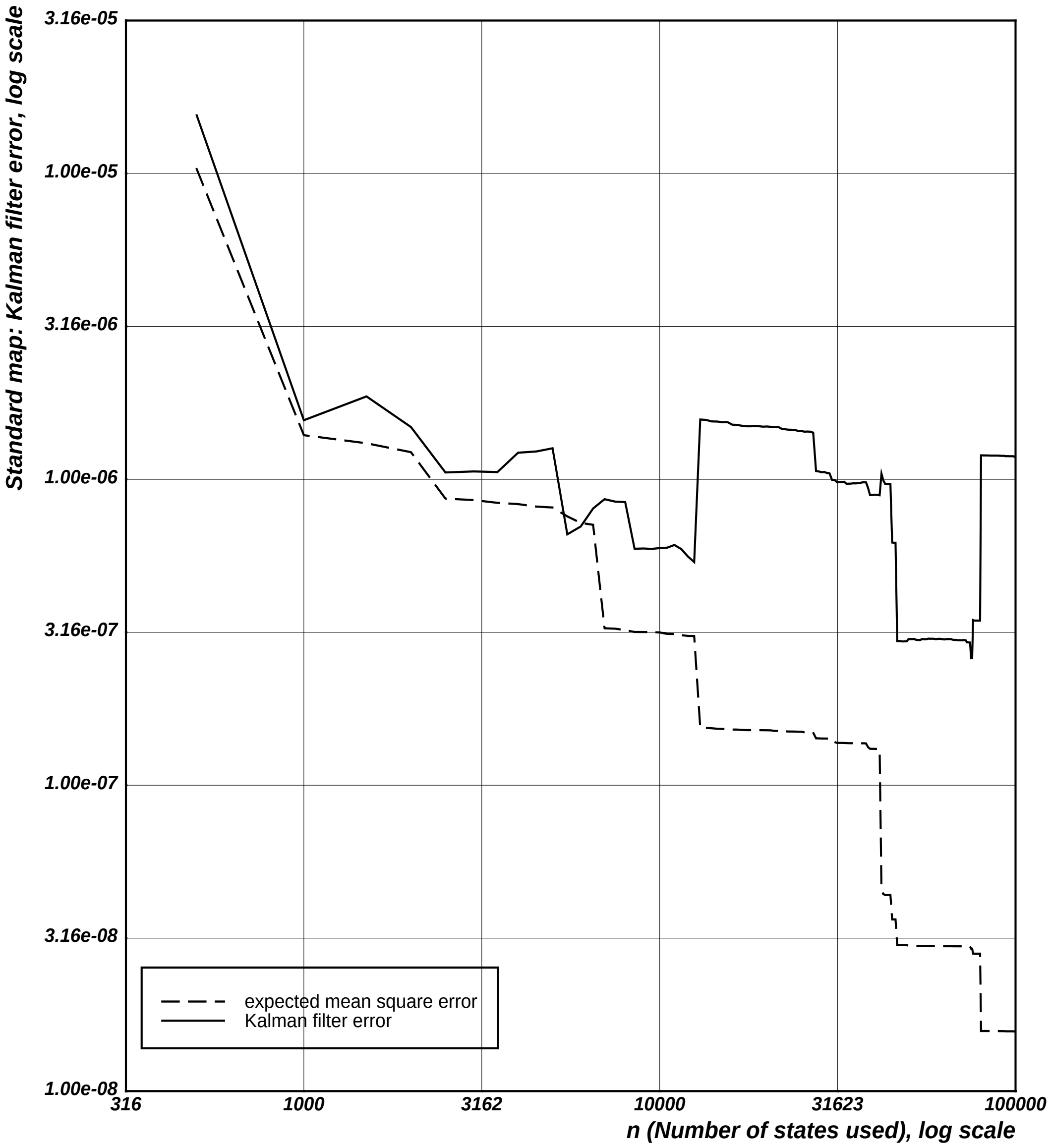

Figure 6.12: This graph depicts the performance of the square root extended Kalman filter for estimating parameter K using one sequence of noisy state data from the standard map with $K = 1$. The data was generated using the initial condition, $(x_0, y_0) = (0.05, 0.05)$. This initial condition results in a trajectory that wanders around in a chaotic zone.

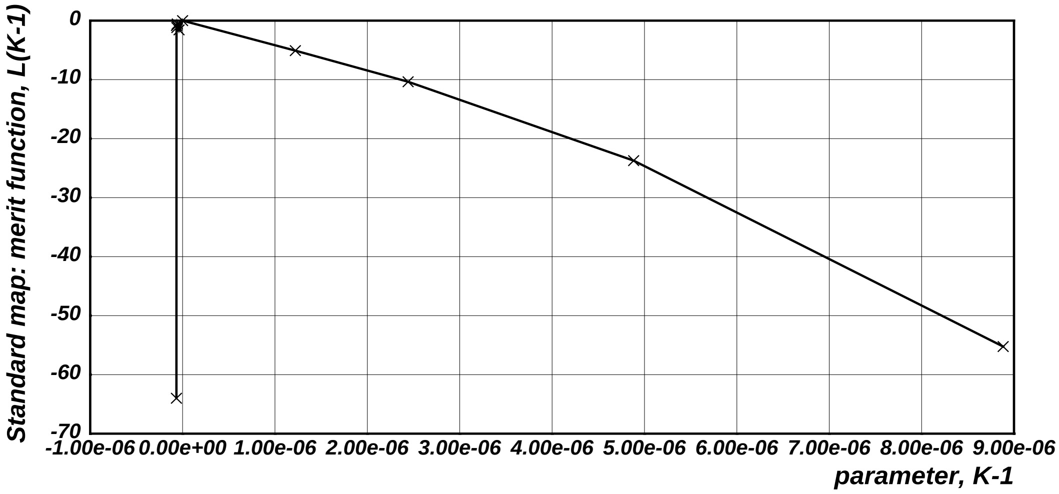

Figure 6.13: Asymmetry in the parameter space of the standard map (with $K = 1$): Here we graph the parameter merit function $L(K)$ after 250000 iterates of an orbit with initial condition $x_0 = (.423, .208)$.

the chaotic zone. In figure 6.13 we see the effects of asymmetric shadowing in the standard map. The algorithm used in these trials is exactly the same as the one used for the experiments with the Henon map (not even the tunable parameters of the algorithm were changed). This indicates that the algorithm is relatively flexible and does not have to be tuned precisely to generate reasonable results.

6.4 Lozi map

We now discuss numerical experiments with the Lozi map:

$$x_{n+1} = y_n + 1 - a|x_n| \tag{6.6}$$

$$y_{n+1} = bx_n \tag{6.7}$$

where the state $(x_n, y_n) \in \mathbb{R}^2$ and the parameter values, a and b, are invariant. The Lozi map may be thought of as a piecewise linear version of the Henon map. Unlike the Henon map, however, the Lozi map is uniformly hyperbolic where the appropriate derivatives exist ([36]). For parameter values $a = 1.7$ and $b = 0.5$, the Lozi map has a hyperbolic attractor ([36]) as shown in figure 6.15.

For the purposes of testing out parameter estimation algorithms, we fix $b = 0.5$ and attempt to estimate a. State data is chosen from an orbit on the attractor of the Lozi map.

In figure 6.16 we show the result of applying a square root extended Kalman filter to the Lozi map. Unlike with the quadratic, Henon, and standard maps, the Kalman

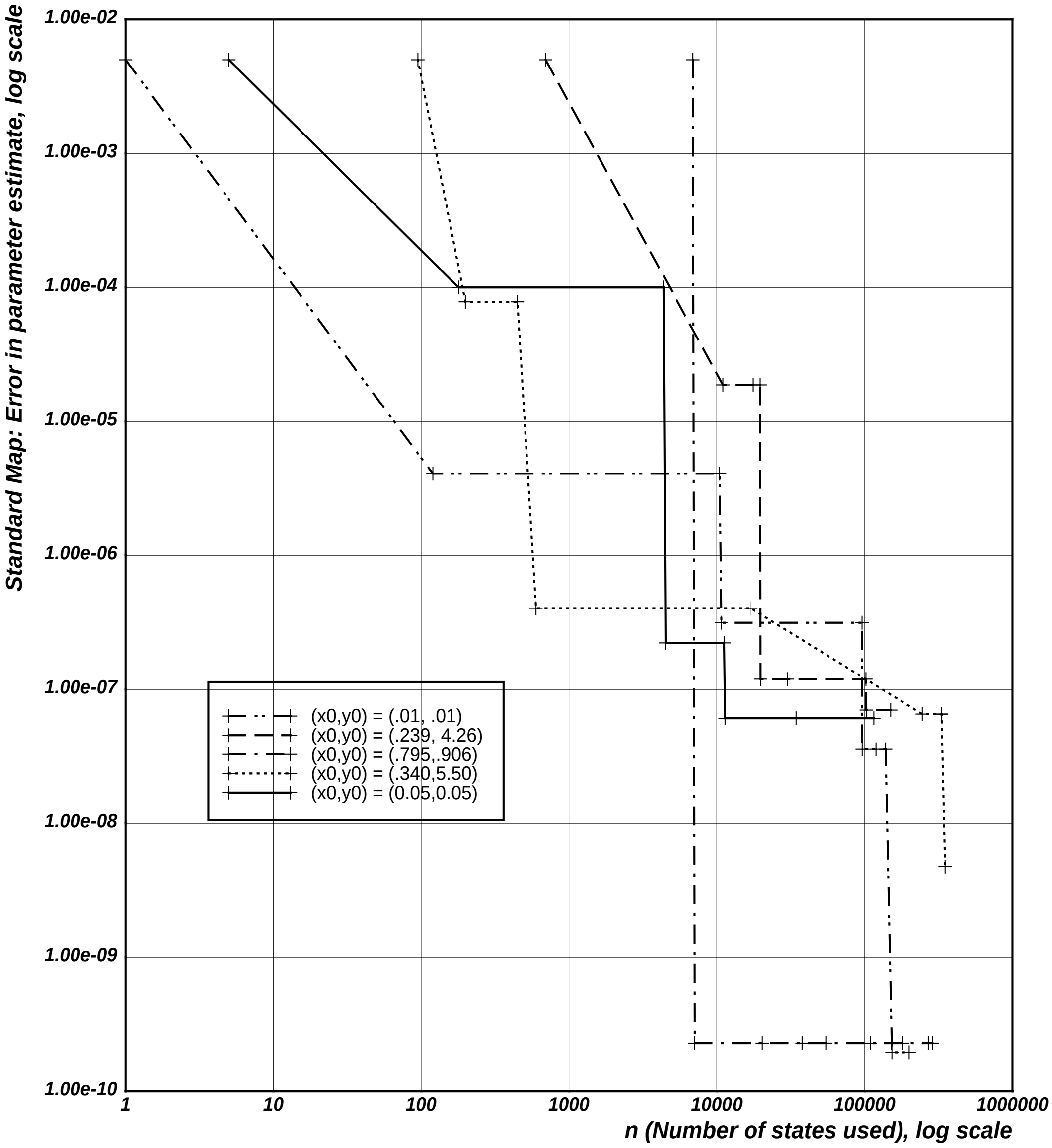

Figure 6.14: This graph depicts the performance of the proposed algorithm for estimating parameter K using one sequence of noisy state data from the standard map with $K = 1$.

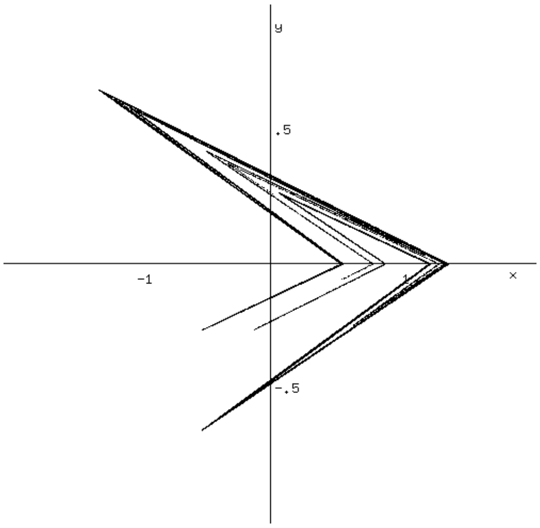

Figure 6.15: The Lozi attractor for $a = 1.7$, $b = 0.5$.

filter applied to the Lozi map shows no signs of divergence, at least within 100,000 iterates. Note that the convergence of the expected mean square parameter estimation error falls almost exactly at the $\frac{1}{\sqrt{n}}$ rate indicated by pure stochastic convergence. Thus, the dynamics makes no asymptotic contribution to the parameter estimate, as one would expect with a uniformly hyperbolic system.

We cannot really apply the algorithm from section 5.6 to the Lozi map because there are basically no sensitive orbit sections to investigate. The whole data set would pass right through the top level scanning filter without further review. However, even if we did force the Monte-Carlo algorithm to consider all the data points, we should again find purely stochastic convergence.

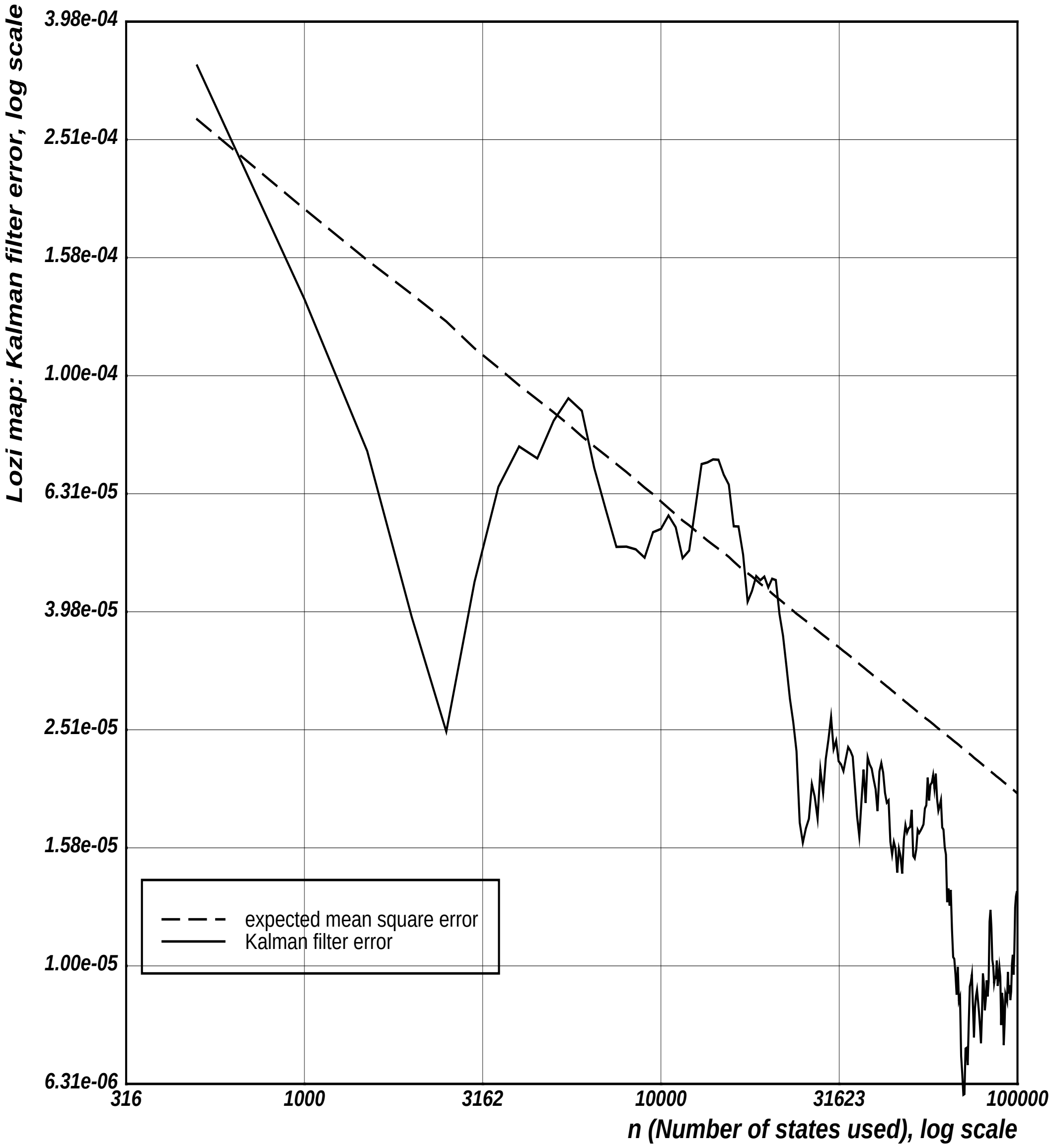

Figure 6.16: This graph plots the performance of a square root extended Kalman filter in estimating the parameter, a, in the uniformly hyperbolic Lozi map. The data here represents the average over five runs based on data with different measurement noises bit generated using the parameters $a = 1.7$, $b = 0.5$, and the same initial condition on the attractor, near $(x_0, y_0) = (-.407, .430)$. Note the lack of divergence, and the fact that convergence is purely stochastic.

Chapter 7

Conclusions and future work

7.1 Conclusions

This report examines how to estimate the parameters of a chaotic system given observations of the state behavior of the system. This problem is interesting in light of recent efforts to use chaotic systems for control and signal processing applications, and because of the possibilities for using parameter estimation in chaotic systems to develop extremely sensitive measurement techniques. In order to evaluate the possible application of parameter estimation techniques to chaotic systems, we approached this report with two main goals in mind: (1) to examine the extent to which it is theoretically possible to estimate the parameters of a chaotic system, and (2) to develop an algorithm to do the parameter estimation. Significant progress was made on both objectives.

7.1.1 Theoretical considerations

In order to examine the theoretical possibilities of parameter estimation, we first broke chaotic systems down into two categories: structurally stable systems and systems that are not structurally stable. Structurally stable systems are probably not that interesting for measurement applications, since small perturbations in the parameters of these systems do not result in qualitatively different state orbits. Consequently, we cannot extract asymptotic information about the parameters by observing the dynamics of structurally stable systems.

The situation, however, is significantly different for systems that are not structurally stable. It turns out that the accuracy of parameter estimates is closely related to how orbits shadow each other for systems with slightly different parameter values. Thus, investigating the possibilities for parameter estimation required us to examine shadowing orbits. We discovered two interesting properties of shadowing orbits for parameterized families of nonuniformly hyperbolic systems. First, we found that there is often an asymmetrical shadowing behavior in the parameter space of these systems. That is, for one-parameter families of systems, it is typically much easier for systems with slightly higher parameter values to shadow orbits of systems with slightly lower parameter values (or vice versa). To illustrate this property in at least one case, we proved a specific shadowing result showing there truly is a preferred direction in parameter space for certain maps of the interval with negative Schwarzian derivative satisfying a Collet-Eckmann-like condition for state and parameter space derivatives.

In addition, we also found that given a typical orbit of a nonuniformly hyperbolic system, most iterates of the orbit look locally hyperbolic, so that only a few rare stretches of the orbit are sensitive to parameters and exhibit the asymmetrical shadowing behavior in parameter space. These sensitive stretches of orbit seem to correspond to local nonhyperbolic folding behavior in state space.

7.1.2 Parameter estimation algorithms

In designing the new parameter estimation algorithm, we took advantage of the two theoretical observations described above. First, since most of the state data is apparently insensitive to parameter changes, we chose a fast top-level filter to scan through the data before concentrating on data that might be especially sensitive. The observation about asymmetrical shadowing behavior in parameter space is also extremely important, since it means that we have only to investigate the sharp boundary in parameter space between parameters that do and do not shadow the data in order to estimate what the true parameters are.

The resulting algorithm is shown to perform significantly better than standard parameter estimation algorithms like the extended Kalman filter. The extended Kalman filter typically diverges for most problems involving parameter estimation of chaotic systems. That is, the filter's covariance matrix becomes too confident about the estimation error, effectively fixing the parameter estimate to an incorrect value without accepting new information from additional data points. This occurs because most of the information about the parameters of the system can be derived from observations that experience local folding in state space, a phenomenon that is inherently difficult to model with the local linearization techniques used by the Kalman filter.

Our algorithm, on the other hand, does not have the divergence problem of the extended Kalman filter. In several numerical experiments we demonstrated that the algorithm described in this report achieved accuracies at least 3 to 4 orders of magnitude better than the extended Kalman filter before the experiment was stopped. Presumably, we should be able to get even better accuracies with the proposed algorithm simply by using more data points. Meanwhile, the divergence problem places a fairly strict bound on the accuracy of the extended Kalman filter.

Furthermore, it appears that the estimation accuracy of the proposed algorithm converges at a rate of $\frac{1}{n^2}$ for certain systems (where n is the number of state samples processed). This is interesting because it is significantly better than the $\frac{1}{\sqrt{n}}$ stochastic convergence one might typically expect from most nonchaotic or structurally stable systems. This indicates that the chaotic dynamics of a system can indeed help parameter estimation to some extent, and opens the door to some interesting possible applications like high precision measurement.

7.2 Future work

7.2.1 Theory

Many questions still remain unanswered. First of all, I would like to know how to really characterize the ability of a system to shadow other systems. Is there a simple set of properties of a parameterized family of mappings that guarantee the asymmetry in parameter space shadowing behavior for a large class of mappings? How widespread is this asymmetrical behavior in parameter space shadowing? It seems likely that the situation is "generic" in some sense, but how can we make this statement more concrete?

Shadowing is particularly not well understood in higher dimensional systems. It might be helpful to further investigate the invariant manifolds of nonuniformly hyperbolic systems in order to better understand shadowing results. In particular, it would be interesting to investigate more quantitative results concerning the folding behavior observed in this report and to specify how this phenomenon affects shadowing behavior in general.

There is also work to be done in figuring out exactly what the rate of convergence is likely to be for parameter estimation algorithms, in particular when those algorithms are applied to multi-dimensional nonuniformly hyperbolic systems. This is important if we would like to choose a system to optimize for parameter sensitivity. The conjectures of section 3.5 seem to be a good place to start.

7.2.2 Parameter estimation algorithms

There are a number of ways in which the parameter estimation algorithm could probably be improved. For instance, the biggest problem now seems to be in the behavior of the top-level scanning Kalman filter. Is there a better way of detecting where the parameter-sensitive stretches of data occur? Perhaps a better solution would be to use some sort of fixed-lag smoother so that data is taken from both forwards and backwards in time in order to smooth out local stretches of parameter-sensitive data.

Also, is there a nicer way of representing the state-parameter space probability densities? It is clear that linear representations like those in the extended Kalman filter cannot do the job. I have tried a number of other representation forms without success, and eventually resorted to a Monte-Carlo based method. Perhaps a more efficient but still effective representation form for the densities can be found.

7.2.3 Applications

Most importantly, there are still questions about how to apply parameter estimation in chaotic time series to problems like high precision measurement, control, or other possible applications. This report shows that many chaotic systems exhibit some special properties that would aid someone who is interested in knowing the parameters of a system based on state data. Now that we have a better theoretical base for understanding what factors affect parameter estimation in chaotic systems, it should be easier to understand how and when to apply the resulting algorithmic tools.

As for the possibility of high precision measurement applications, this idea certainly merits additional research in light of the results in this report. The main problem here would be to find a suitable application where the quantity to be measured is physically interesting and the chaotic system involved satisfies all the right properties. For instance, this technique would ideally be applied to a system that is well-modeled by a relatively simple set of equations. The problem would be to find a suitable setup that would make the application worthwhile, and/or to increase the sophistication of the parameter estimation algorithms to handle a larger set of experimental situations.

Appendix A

Proofs from Chapter 2

This appendix contains notes on three proofs from Chapter 2. Note that in the first two theorems (sections A.1 and A.2), we reverse the names of the functions f and g from the corresponding theorems in the text of this report. This is done to conform with the notation used in Walters' paper, [62]. The notation in the appendix is the same as in Walters, while the notation in the text is switched.

A.1 Proof of Theorem 2.2.3

Theorem 2.2.3: *(Walters) Let $f : M \to M$ be an expansive diffeomorphism with the pseudo-orbit shadowing property. Suppose there exists a neighborhood, $V \subset \mathrm{Diff}^1(M)$ of f that is uniformly expansive. Then f is structurally stable.*

Proof: This is based on theorem 4 and 5 and the remark on page 237 in [62]. In theorem 4, Walters states that an expansive homeomorphism with the pseudo-orbit shadowing property is "topologically stable." However, Walters' definition of topological stability is weaker than our definition of structural stability. In particular, for topological stability of f, Walters requires that there exist a neighborhood, $U \subset \mathit{Diff}^1(M)$, of f such that for each $g \in U$, there is a continuous map $h : M \to M$ such that $hg = fh$. For structural stability, this h must be a homeomorphism. We can get the injectiveness of h from the uniform expansiveness of nearby maps (apply theorem 5 of [62]). We can get the surjectiveness of h from the compactness of M based on an argument from algebraic topology (see Lemma 3.11 in [38], page 36). Since M is compact, and h is injective and surjective, h must be a homeomorphism.

A.2 Proof of Theorem 2.2.4

Theorem 2.2.4: *Let $f : M \to M$ be an expansive diffeomorphism with the function shadowing property. Suppose there exists a neighborhood, $V \subset \text{Diff}^1(M)$ of f such that V is uniformly expansive. Then f is structurally stable.*

Proof: The proof given here is similar to theorem 4 of [62] except that the effective roles of f and g are reversed (where g denotes maps near f in $Diff^1(M)$). Instead of knowing that all orbits of nearby systems can be shadowed by real orbits of f (pseudo-orbit shadowing), here we are given that all orbits of f can be shadowed by real orbits of any nearby system (function shadowing).

We shall prove that there is a neighborhood $U \subset V$ of f in $Diff^1(M)$ such that for any $g \in U$, there exists a continuous h such that $hf = gh$ (note that the h we use here is the inverse of the one in theorem 2.2.3). From this result we can use the arguments outlined for theorem 2.2.3 to show that h is a homeomorphism because of the uniform expansiveness of f and the compactness of M.

First we need to show the existence of a function $h : M \to M$ such that $hf = gh$. From the function shadowing property, given any $\epsilon > 0$, there exists a neighborhood, $U_\epsilon \subset V$ of f such that any orbit of f is $\epsilon-$shadowed by an orbit of $g \in U_\epsilon$.

Now suppose that $\epsilon < \frac{1}{2}\inf_{g\in V} e(g)$. In this case, we claim that there is exactly one orbit of g that $\epsilon-$shadows any particular orbit of f. If this were not true then two different orbits of g, $\{x_n\}$ and $\{y_n\}$, must shadow the same orbit of f. But because of the expansiveness of g there must exist an integer, N, such that $d(x_N, y_N)) > 2\epsilon$, so that $\{x_n\}$ and $\{y_n\}$ clearly cannot ϵ-shadow the same orbit of f. Thus we can see that there must be a function h which maps each orbit of f to a shadowing orbit of g.

Consequently, for any $\epsilon > 0$, there exists a neighborhood U_ϵ such that for any $g \in U_\epsilon$, we can define a function h such that $hf = gh$ and:

$$sup_{x\in M} d(h(x), x) < \epsilon. \tag{A.1}$$

We now need to show that this h is also continuous. To do this we first need the following lemma from [62]:

Lemma A.2.1 *(Lemma 2 in [62]) Let f be expansive with expansive constant $e(f) > 0$. Given any $\delta > 0$, there exists $N \geq 1$ such that $d(f^n(x), f^n(y)) \leq e(f)$ for $|n| < N$ implies $d(x, y) < \delta$.*

Proof of Lemma: Given $\delta > 0$, suppose that the lemma is not true so that no such N can be chosen. Then there are exists a sequence of points, $\{x_i\}_{i=1}^{\infty}$ and $\{y_i\}_{i=1}^{\infty}$ (not orbits), such that for any $N \geq 1$, $d(x_N, y_N) \geq \delta$ and $d(f^n(x_N), f^n(y_N)) \leq e(f)$ for all $|n| < N$. There exists a subsequence of points $\{x_{n_i}\}_{i=0}^{\infty}$ and $\{y_{n_i}\}_{i=0}^{\infty}$ such that $x_{n_i} \to x$

and $y_{n_i} \to y$ as $i \to \infty$ such that $d(x,y) \geq \delta$. By continuity of f this implies that $d(f^n(x), f^n(y)) \leq e(f)$ for all n, which is a direct contradiction of the expansiveness of f. This completes the proof of lemma A.2.1.

Returning to the proof of theorem 2.2.4, we now want to show the continuity of h. In other words, given any $\alpha > 0$ we need to show there exists a $\delta > 0$ such that $d(x,y) < \delta$ implies $d(h(x), h(y)) < \alpha$.

Our strategy is as follows: Since g is expansive, from lemma A.2.1 we know that for any $\alpha > 0$ we can choose N_α such that if $d(g^n(h(x)), g^n(h(y))) \leq e(g)$ for $|n| < N_\alpha$ then $d(h(x), h(y)) < \alpha$. Thus suppose that for any $\alpha > 0$ there exists $\delta > 0$ such that $d(x,y) < \delta$ implies $d(g^n(h(x)), g^n(h(y))) \leq e(g)$ for all $|n| < N_\alpha$. Then $d(h(x), h(y)) < \alpha$, and h must be continuous. This is what we shall show.

Given $\alpha > 0$, pick $\delta > 0$ such that $d(f^n(x), f^n(y)) < \delta$ if $|n| < N_\alpha$. Set $e(V) = sup_{g \in V} e(g)$ and fix $\epsilon = \frac{1}{3}e(V)$. From equation (A.1) we know that given this $\epsilon > 0$, there exists a neighborhood, $U_\epsilon \subset V$, of f in $Diff^1(M)$ such that for any $g \in U_\epsilon$, there exists h such that $hf = gh$ and $sup_{x \in M} d(h(x), x) < \epsilon$. Thus for any $g \in U_\epsilon$ and corresponding $h : M \to M$, if $d(x,y) < \epsilon$ then we have:

$$\begin{aligned} d(g^n(h(x)), g^n(h(y))) &= d(h(f^n(x)), h(f^n(y))) \\ &\leq d(h(f^n(x)), f^n(x)) + d(f^n(x), f^n(y)) + d(f^n(y), h(f^n(y))) \\ &\leq \epsilon + \frac{1}{3}e(V) + \epsilon \\ &\leq e(V) \leq e(g) \text{ for all } |n| < N_\alpha \end{aligned}$$

From the argument in the previous paragraph, this shows that h must be continuous which completes the proof of theorem 2.2.4.

A.3 Proof of Lemma 2.3.1

Lemma 2.3.1: *Suppose that $f_p \in \mathrm{Diff}^1(M)$ for $p \in I_p \subset \mathbb{R}$, and let $f(x,p) = f_p(x)$ for any $x \in M$. Suppose also that f is C^1 and that f_{p_0} is an absolutely structurally stable diffeomorphism for some $p_0 \in I_p$. Then there exists $\epsilon_0 > 0$ and $K > 0$ such that for every positive $\epsilon < \epsilon_0$, any orbit of f_{p_0} can be $\epsilon-$shadowed by an orbit of f_p for $p \in B(p_0, K\epsilon)$.*

Proof: This follows from the definition of absolute structural stability. From that definition, we know that there exists $\epsilon_0 > 0$, $K_1 > 0$, and conjugating homeomorphisms, $h_p : M \to M$, such that if $p \in B(p_0, \epsilon_0)$, then:

$$\sup_{x \in M} d(h_p^{-1}(x), x) \leq K_1 \sup_{x \in M} d(f_{p_0}(x), f_p(x))).$$

where $f_{p_0} = h_p f_p h_p^{-1}$. Given an orbit, $\{x_n\}$, of f_{p_0} we claim that h_p^{-1} maps x_n onto a suitable shadowing orbit, $z_n(p)$ of f_p for each $n \in \mathbb{Z}$. Also, since f is C^1 for $(x,p) \in M \times I_p$, there exists a constant, $K_2 > 0$, such that $sup_{x\in M} d(f_{p_0}(x), f_p(x)) \leq K_2|p - p_0|$ for any $p \in I_p$. Thus, setting $z_n(p) = h_p^{-1}(x_n)$, for all n we see that:

$$\begin{aligned}
\sup_{n\in\mathbb{Z}} d(z_n(p), x_n) &\leq \sup_{x\in M} d(h_p^{-1}(x), x) \\
&\leq K_1 \sup_{x\in M} d(f_{p_0}(x), f_p(x)) \\
&\leq K_1 K_2 |p - p_0|
\end{aligned}$$

for all integer n. Now setting $K = \frac{1}{2K_1K_2}$, we have the desired result that $sup_{n\in Z} d(z_n(p), x_n) < \epsilon$ if $p \in B(p_0, K\epsilon)$, for all n and any positive $\epsilon < \epsilon_0$. This completes the proof of lemma 2.3.1.

Appendix B

Proof of theorem 3.2.1

In this appendix, we present the proof for theorem 3.2.1.

B.1 Preliminaries

We first repeat the related definitions which are the same as those found in chapter 3. Throughout this appendix we shall assume that $I \subset \mathbb{R}$ represents a compact interval of the real line.

Definitions: Suppose that $f : I \to I$ is continuous. Then the *turning points* of f are the local extrema of f in the interior I. $C(f)$ is used to designate the set of all turning points of f on I. $\mathbb{C}^r(I, I)$ is the set of continuous maps on I such that $f \in \mathbb{C}^r(I, I)$ if:
(a) f is C^r (for $r \geq 0$)
(b) $f(I) \subseteq I$, and
(c) $f(Bd(I)) \subseteq Bd(I)$ (where $Bd(I)$ denotes the boundary of I).
If $f \in \mathbb{C}^r(I, I)$ and $g \in \mathbb{C}^r(I, I)$, let $d(f, g) = \sup_{x \in I} |f(x) - g(x)|$.

Definitions: A continuous map $f : I \to I$ is said to be *piecewise monotone* if f have finitely many turning points. f is said to be a *uniformly piecewise-linear* mappings if it can be written in the form:

$$f(x) = \alpha_i \pm sx \text{ for } x_i \in [c_{i-1}, c_i] \tag{B.1}$$

where $s > 1$, $c_0 < c_1 < \ldots < c_q$ and $q > 0$ is an integer. (We assume $s > 1$ because otherwise there will not be any interesting behavior).

Note that for this section, it is useful to define neighborhoods, $B(x, \epsilon)$, so that they do not extend beyond the confines of I. In other words, let $B(x, \epsilon) = (x - \epsilon, x + \epsilon) \cap I$. With this in mind, we use the following definitions to describe some relevant properties of piecewise monotone maps.

Definition: A piecewise monotone map, $f : I \to I$, is said to be *transitive* if for any two open sets $U, V \subset I$, there exists an $n > 0$ such that $f^n(U) \cap V \neq \emptyset$.

Definitions: Let $f : I \to I$ be piecewise monotone. Then f satisfies the *linking property* if for every $c \in C(f)$ and any $\epsilon > 0$ there is a point $z \in I$ such that $z \in B(c, \epsilon)$, $f^n(z) \in C(f)$ for some integer $n > 0$, and $|f^i(c) - f^i(z)| < \epsilon$ for every $i \in \{1, 2, \ldots, n\}$. Suppose, in addition, that we can always pick $z \neq c$ such that the above condition is satisfied. Then f is said to satisfy the *strong-linking* condition.

We are now ready to state the objective of this appendix:

Theorem 3.2.1 *Transitive piecewise monotone maps satisfy the function shadowing property in $\mathbb{C}^0(I, I)$ if and only if the satisfy the strong linking property.*

We note Liang Chen [12] proves a similar result, namely that the pseudo-orbit shadowing property is equivalent to the linking property for maps topologically conjugate to uniformly piecewise linear mappings. Some parts of the proof we describe below are also similar to the work of Coven, Kan, and Yorke [17] for tent maps (uniformly piecewise linear maps with one turning point). The main difference is that they prove a pseudo-orbit shadowing property while we are interested in parameter and function shadowing.

B.2 Proof

This section will be devoted to the proof of theorem 3.2.1 and related results. The basic strategy of the proof will be as follows. First we relate piecewise monotone mappings to piecewise linear mappings through a topological conjugacy (lemmas B.2.1 and B.2.2). This provides for uniform hyperbolicity away from the turning points. Second we capture the effects of "folding" near turning points and show how this leads to function shadowing (lemmas B.2.4,B.2.5,B.2.6). Finally in lemma B.2.7 we show that the local folding effects of lemmas B.2.4, B.2.5, or B.2.6 are satisfied for the maps we are interested in.

Lemma B.2.1 *: Let $f : I \to I$ be a transitive piecewise-monotone mapping. Then f is topologically conjugate to uniformly piecewise-linear mapping.*

Proof: See Parry [51] and Coven and Mulvey [18].

The following lemma is necessary for the application of the topological conjugacy result.

Lemma B.2.2 *Let $f : I \to I$ and $g : I \to I$ be two topologically conjugate continuous maps. If f has the linking or strong linking property then g must have these properties* *also. If f satisfies has the function shadowing property on $\mathbb{C}^0(I, I)$, then g must also satisfy the function shadowing property on $\mathbb{C}^0(I, I)$.*

Proof: Since f and g are conjugate, the orbits of f and g are connected through a homeomorphism, h, such that $g = h^{-1}fh$. Because h is continuous and one-to-one, the of turning points of f and g must be preserved by the topological conjugacy. Thus if f has the linking or strong linking properties, then g must have these properties also.

Now suppose that f has the function shadowing property on $\mathbb{C}^0(I, I)$. We want to show that g also has this function shadowing property which means that for any $\epsilon > 0$, there exists a neighborhood, V, of g in $\mathbb{C}^0(I, I)$ such that if $g_* \in V$ then any orbit of g is $\epsilon-$shadowed by an orbit of g_*.

Since h is continuous, and I is compact, we know that given $\epsilon > 0$ there exists $\delta > 0$ such that $|x - y| < \delta$ implies $|h(x) - h(y)| < \epsilon$ if $x, y \in I$. Given this $\delta > 0$, since f has the function shadowing property, there is a neighborhood $U \subset \mathbb{C}^0(I, I)$ of f such that if $f_* \in U$, then any orbit of f can be δ-shadowed by an orbit of f_*. Let $V = h^{-1}Uh$. Since $g = h^{-1}fh$, V must contain a neighborhood of g in $\mathbb{C}^0(I, I)$. We now must show if $g_* \in V$, then any orbit of g can be $\epsilon-$shadowed by an orbit of g_*.

Suppose we are given an orbit, $\{x_n\}$, of g and any $g_* \in V$. Let $\{w_n\}$ be the corresponding orbit of f such that $w_n = h^{-1}(x_n)$. Set $f_* = h^{-1}(g_*)$. Since $f_* \in U$, there exists an orbit, $\{y_n\}$, of f_* that $\delta-$shadows $\{w_n\}$. Then if $z_n = h(y_n)$, $\{z_n\}$ must be an orbit of g_* that $\epsilon-$shadows $\{x_n\}$, since $|h(x) - h(y)| < \epsilon$ if $|x - y| < \delta$. This proves the lemma.

Thus, combining lemmas B.2.1 and B.2.2, we see that the problem of proving the function shadowing property for transitive piecewise-monotone maps with the strong linking property reduces to proving the function shadowing property for uniformly piecewise linear maps with the strong-linking property.

We now introduce one more result that will be useful later on:

Lemma B.2.3 *Let $f : I \to I$. Suppose f^n satisfies the function shadowing property on $\mathbb{C}^0(I, I)$ for some integer $n > 0$. Then f has the function shadowing property on $\mathbb{C}^0(I, I)$.*

Proof: Given any $\epsilon > 0$ we need to show that there exists a neighborhood, U of f in $\mathbb{C}^0(I, I)$ such that if $g \in U$, then any orbit of f is $\epsilon-$shadowed by an orbit of g. Since f is continuous and I is compact, there exists a $\delta > 0$ such that if $|x - y| < \delta$, then

$$|f^i(x) - f^i(y)| < \frac{1}{2}\epsilon \tag{B.2}$$

for any $i \in \{0, 1, \ldots, n\}$ and $x, y \in I$. We also know that there exists a neighborhood, V_1 of f in $\mathbb{C}^0(I, I)$ such that if $g \in V_1$:

$$|f^i(x) - g^i(x)| < \frac{1}{2}\epsilon \tag{B.3}$$

for all $x \in I$ and $i \in \{0, 1, \ldots, n\}$.

Combining (B.2) and (B.3) and using the triangle inequality we see that for any $\epsilon > 0$ there exists a $\delta > 0$ and a neighborhood, V_1, of f in $\mathbb{C}^0(I, I)$ such that if $g \in V_1$ and $|x - y| < \delta$, then:

$$|f^i(x) - g^i(y)| < \epsilon. \tag{B.4}$$

for all $i \in \{0, 1, \ldots, n\}$ if $x, y \in I$. Given $\epsilon > 0$, fix $\delta > 0$ and $V_1 \in \mathbb{C}^0(I, I)$ to satisfy (B.4).

Using this $\delta > 0$, since f^n has the function shadowing property, we know there exists a neighborhood, V_2, of f^n in $\mathbb{C}^0(I, I)$ such that if $g^n \in V_2$, then any orbit of f^n is $\delta-$shadowed by an orbit g^n. Given this neighborhood, V_2, of f^n, we can always pick a neighborhood, $V_3 \subset \mathbb{C}^0(I, I)$ of f such that $g \in V_3$ implies that $g^n \in V_2$. This is apparent, since for any $\alpha > 0$ there exists a neighborhood V_3 of f in $\mathbb{C}^0(I, I)$ such that

$$d(f^n, g^n) = \sup_{x \in I} |f^n(x) - g^n(x)| < \alpha.$$

if $g \in U$. Thus, for any $\epsilon > 0$, if $g \in V_3$, then any orbit of f^n is δ-shadowed by an orbit of g^n.

Now set $U = V_1 \cap V_3$. Note that U must be a contain neighborhood of f in $\mathbb{C}^0(I, I)$. If we fix $g \in U$, we find that given any orbit, $\{x_i\}_{i=0}^{\infty}$, of f, there is an orbit, $\{y_i\}_{i=0}^{\infty}$, of g such that $y_i \in B(x_i, \delta)$ if $i = kn$ for any $k \in \{0, 1, \ldots\}$. Thus, from (B.4), we know that $y_i \in B(x_i, \epsilon)$ for all $i \geq 0$. Consequently, given any $\epsilon > 0$, there exists a neighborhood U of f in $\mathbb{C}^0(I, I)$ such that if $g \in U$, then any orbit of f can be $\epsilon-$shadowed by an orbit of g. This is what we set out to prove.

We now examine the mechanism underlying shadowing in one-dimensional maps. In the next three lemmas we look at how local "folding" can lead to shadowing.

Lemma B.2.4 *Given $f \in \mathbb{C}^0(I, I)$, suppose that for any $\epsilon > 0$ sufficiently small there exists a neighborhood, U, of f in $\mathbb{C}^0(I, I)$ such that if $g \in U$,*

$$g(B(x, \epsilon)) \supseteq (B(f(x), \epsilon)) \tag{B.5}$$

for all $x \in I$. Then f has the function shadowing property in $\mathbb{C}^0(I, I)$.

Proof: Let $\{x_n\}$ be an orbit of f and suppose that (B.5) is satisfied. Then if $g \in U$, for any $y_1 \in I$ with $y_1 \in B(x_1, \epsilon)$ we can choose a $y_0 \in I$ so that $y_0 \in B(x_0, \epsilon)$ and $y_1 = g(y_0)$. Similarly for any $y_2 \in I$ with $y_2 \in B(x_2, \epsilon)$, we can pick y_1 and y_0 within ϵ distance of x_1 and x_0, respectively. Extending this argument for arbitrarily many iterates we see that (B.5) implies that there exists an orbit, $\{y_i\}$, of g so that $y_i \in B(x_i, \epsilon)$ for all integer $i \geq 0$. Thus, given any $\epsilon > 0$ sufficiently small, there exists a neighborhood, U, of f in $\mathbb{C}^0(I, I)$ such that if $g \in U$, then any orbit orbit of f can be $\epsilon-$shadowed by an orbit of g.

Lemma B.2.5 *Let $f \in \mathbb{C}^0(I, I)$. Suppose that for any $\epsilon > 0$ sufficiently small, there exists $N > 0$ and a neighborhood, U, of f in $\mathbb{C}^0(I, I)$ such that for any $g \in U$, there exists a function $n : I \to \mathbb{Z}^+$ so that for each $x \in I$:*

$$\{g^{n(x)}(y) : \ |f^i(x) - g^i(y)| < \epsilon, 0 \le i \le n(x)\} \supseteq (B[f^{n(x)}(x), \epsilon]) \tag{B.6}$$

where $1 \le n(x) < N$ for all $x \in I$. Then f has the function shadowing property in $\mathbb{C}^0(I, I)$.

Proof: The idea is very similar to lemma B.2.4. Let $\{x_n\}$ be an orbit of f. In lemma B.2.4, given sufficiently small $\epsilon > 0$ and $g \in U$, we could always choose $y_0 \in B(x_0, \epsilon)$ given a $y_1 \in B(x_1, \epsilon)$ so that $y_1 = g(y_0)$. A similar thing applies here except that we have to consider the iterates in groups. Suppose that the premise of lemma B.2.5 is satisfied. Given sufficiently small $\epsilon > 0$, fix $g \in U$. Then, for any $y_{n(x_0)} \in B(x_{n(x_0)}, \epsilon)$, there exists a finite orbit $Y_0 = \{y_i\}_{i=0}^{n(x_0)}$ of g such that $|x_i - y_i| < \epsilon$, for $i \in \{0, 1, \ldots, n(x_0)\}$. Similarly, we can play the same trick starting with $y_{n(x_0)}$ for the next $n(x_{n(x_0)})$ group of iterates constructing another finite orbit, $Y_1 = \{y_i\}_{i=n(x_0)}^{n(x_0)+n(x_{n(x_0)})}$, of g. Since we are free choose Y_0 from any $y_{n(x_0)} \in B(x_{n(x_0)}, \epsilon)$, it is clear that given any Y_1 we can pick a Y_0 belonging to the same infinite forward orbit of g, thereby allowing us to concatenate Y_0 and Y_1 to construct a single finite orbit of g, $\{y_i\}_{i=0}^{n(x_0)+n(x_{n(x_0)})}$ that $\epsilon-$shadows $\{x_i\}_{i=0}^{n(x_0)+n(x_{n(x_0)})}$. This process can be repeated indefinitely for arbitrarily many groups of iterates, gluing together each group of iterates as we go. Thus the function shadowing property holds.

Lemma B.2.6 *Let $f \in \mathbb{C}^0(I, I)$. Suppose that for any $\epsilon > 0$ sufficiently small, there exists $N > 0$ and a neighborhood, U, of f in $\mathbb{C}^0(I, I)$ such that for any $g \in U$, there exists a function $n : I \to \mathbb{Z}^+$ so that for each $x \in I$:*

$$\begin{aligned}\{g^{n(x)+1}(y) : \ |x - y| < \epsilon, |f^i(x) - g^i(y)| < 8\epsilon, 1 \le i \le n(x)\} \\ \supseteq g[B(f^{n(x)}(x), \epsilon)]\end{aligned} \tag{B.7}$$

where $1 \le n(x) < N$ for all $x \in I$. Then f has the function shadowing property in $\mathbb{C}^0(I, I)$.

Proof: (compare with lemma 2.4 of [17]). We shall show that given sufficiently small $\epsilon > 0$ and any $g \in U$, if (B.7) is satisfied, then for any orbit, $\{x_i\}_{i=0}^{\infty}$ of f, there exists an orbit, $\{y_i\}_{i=0}^{\infty}$, of g such that $|x_i - y_i| < 8\epsilon$ for all integer $i \ge 0$. By condition (B.7), given any $y^0_{n(x_0)+1} \in g(B[x_{n(x_0)}, \epsilon])$ we can choose a finite orbit, $Y_0 = \{y^0_i\}_{i=0}^{n(x_0)}$, of g that 8ϵ-shadows $\{x_i\}_{i=0}^{n(x_0)}$ and satisfies $g(y^0_{n(x)}) = y^0_{n(x)+1}$. Similarly, using the same trick with the next $n(x_{n(x_0)})$ iterates, we can construct a finite orbit, $Y_1 = \{y^1_i\}_{i=n(x_0)}^{n(x_0)+n(x_{n(x_0)})}$, of g that 8ϵ-shadows $\{x_i\}_{i=n(x_0)}^{n(x_0)+n(x_{n(x_0)})}$ and satisfies $y^1_{n(x)} \in B(x_{n(x_0)}, \epsilon)$.

Also, notice that given Y_1 we can always choose a Y_0 so that $g(y^0_{n(x_0)}) = y^1_{n(x_0)+1}$. This is because we know that $y^1_{n(x_0)} \in B[x_{n(x_0)}, \epsilon]$ and because we are free to choose any $y^0_{n(x_0)+1} \in g(B[x_{n(x_0)}, \epsilon])$ to construct Y_0. Consequently we can concatenate Y_0 and Y_1 to form an orbit that 8ϵ-shadows $\{x_i\}_{i=0}^{n(x_0)+n(x_{n(x_0)})}$. We can continue this construction by concatenating more groups of $n(x_i)$ iterates for increasingly large i. Thus given (B.7) it is apparent that we can choose an orbit, $\{y_i\}_{i=0}^{\infty}$, of g that 8ϵ-shadows any orbit of f if $g \in U$. This proves the lemma.

Now we must show that lemma B.2.6 is satisfied for any uniformly piecewise-linear map. Note that condition (B.6) in lemma B.2.5 in fact implies (B.7) in lemma B.2.6, so it is sufficient to show that either (B.6) or (B.7) is true for any particular $x \in I$. This is done in lemma B.2.7 below. We can then combine lemma B.2.7 with lemma B.2.3 to prove theorem 3.2.1.

First, however, we introduce the following notation, in order to state our results more concisely.

Definition: Given a map, $f \in \mathbb{C}^0(I, I)$, define:

$$\begin{aligned} D_k(x, g, \epsilon) &= \{g^k(y) : y \in I, |f^i(x) - g^i(y)| < \epsilon \text{ for } i \in \{0, 1, \ldots, k\}\}. \\ E_k(x, g, \epsilon) &= \{g^k(y) : y \in I, |x - y| < \epsilon, \text{ and } |f^i(x) - g^i(y)| < 8\epsilon \text{ for } i \in \{1, 2, \ldots, k\}\}. \end{aligned}$$

for any $x \in I$, $k \in Z^+$, and $\epsilon > 0$ where $g \in \mathbb{C}^0(I, I)$ is a C^0 perturbation of f. Although $D_k(x, g, \epsilon)$ and $E_k(x, g, \epsilon)$ also depend on f we leave out this dependence because f will always refer to the uniformly piecewise linear map specified in the statement of lemma B.2.7 below.

Lemma B.2.7 *: Let $f : I \to I$ be a uniformly piecewise linear map with slope $s > 9$. Suppose that f satisfies the strong linking property. Then for any $\epsilon > 0$ there exists $N > 0$ and a neighborhood, U, of f in $\mathbb{C}^0(I, I)$ such that for any $g \in U$ at least one of the following two properties hold for each $x \in I$:*

(I) $D_{n(x)}(x, g, \epsilon) \supseteq B[f^{n(x)}(x), \epsilon]$

(II) $g(E_{n(x)}(x, g, \epsilon)) \supseteq g(B[f^{n(x)}(x), \epsilon])$

where $n : I \to \mathbb{Z}^+$ and $1 \le n(x) < N$ for all $x \in I$.

Proof of lemma B.2.7: Let $C(f) = \{c_1, c_2, \ldots, c_q\}$ where $c_1 < c_2 < \ldots < c_q$. Assume that $\epsilon > 0$ is small enough such that

$$|c_k - c_i| > 16\epsilon$$

for any $k \neq i$.

We now utilize the strong linking property. For each $j \in \{1, 2, \ldots, q\}$ and $k \in \mathbb{Z}^+$ define $w_k(j, \epsilon) \subset I$ such that:

$$w_k(j,\epsilon) = \{g^k(y) : y \in I, |f^i(c_j) - f^i(y)| < \frac{5}{2}\epsilon \text{ for } i \in \{0, 1, \ldots, k\}\} \tag{B.8}$$

Given $\epsilon > 0$, for each $j \in \{1, 2, \ldots, q\}$ let m_j be the minimum k such that

$$w_k(j, \epsilon) \bigcap C(f) \neq \emptyset. \tag{B.9}$$

The strong linking property implies that such m_j's exist and are finite for each $j \in \{1, 2, \ldots, q\}$ and for any $\epsilon > 0$. From (B.8) and (B.9) we can also see that for each $j \in \{1, 2, \ldots, q\}$, there exists some $r(j) \in \{1, 2, \ldots, q\}$ such that

$$c_{r(j)} \in w_k(j, \epsilon).$$

Now set:

$$\delta_x = \frac{1}{10} \min_{j \in \{1,2,\ldots,q\}} |f^{m_j}(c_j) - c_{r(j)}| \tag{B.10}$$

and note that from (B.8) and (B.9):

$$|f^{m_j}(c_j) - c_{r(j)}| < \frac{5}{2}\epsilon \tag{B.11}$$

for any $j \in \{1, 2, \ldots, q\}$. Thus it is evident that:

$$\delta_x < \frac{1}{4}\epsilon. \tag{B.12}$$

Because of the strong linking property, we know that $\delta_x > 0$.

Also, set $M = max_{j \in \{1,2,\ldots,q\}} m_j$, define $\Delta_x(g) : \mathbb{C}^0(I, I) \to \mathbb{R}$ such that:

$$\Delta_x(g) = \max_{i \in \{1,2,\ldots,M\}} \sup_{x \in I} |f^i(x) - g^i(x)|, \tag{B.13}$$

and choose U to be a neighborhood of f in $\mathbb{C}^0(I, I)$ such that $\Delta_x(g) < \delta_x$ for any $g \in U$. Thus for any $g \in U$, any $x \in I$, and any $i \in \{1, 2, \ldots, M\}$:

$$|f^i(x) - g^i(x)| < \frac{1}{4}\epsilon. \tag{B.14}$$

Now, let $(a; b]$ indicate either the interval, $(a, b]$, or the interval, $[b, a)$, whichever is appropriate. Then, since $s > 9$, for any $\epsilon > 0$ we assert that:

$$D_i(c_j, f, \epsilon) = (f^i(c_j) - \sigma_i(c_j)\epsilon \ ; \ f^i(c_j)] \tag{B.15}$$

for each $j \in \{1, 2, \ldots, q\}$ and every $i \in \{1, 2, \ldots, m_j\}$ where:

$$\sigma_i(c) = \begin{cases} +1 & \text{if } f^i \text{ has a a relative maximum at } c \in C(f) \\ -1 & \text{if } f^i \text{ has a a relative minimum at } c \in C(f). \end{cases}$$

Note that (B.9) guarantees that that $D_i(c_j, f, \epsilon) \cap C(f) = \emptyset$ for any $i \in \{1, 2, \ldots, m_j - 1\}$. Thus, since $s > 9$, (B.15) can be shown by a simple induction on i.

We now proceed to the main part of the proof for lemma B.2.7:

Given any $g \in U$ we must show that for each $x \in I$ either condition (I) or (II) holds in the statement of the lemma for some $n(x) < N$. We now break up the problem into two separate cases. Given some $\epsilon > 0$ first suppose that x is more than ϵ distance away from any turning point. In other words suppose that $|x - c_j| \geq \epsilon$ for all $j \in \{1, 2, \ldots, q\}$. Then we can set $n(x) = 1$ and it is easy to verify that condition (I) of the lemma holds:

$$\begin{aligned} D_1(x, g, \epsilon) &= g(B(x, \epsilon)) \bigcap B(f(x), \epsilon) \\ &= B(g(x), \epsilon) \end{aligned}$$

since $s > 9$ and $|f(x) - g(x)| < \frac{\epsilon}{4}$ for all $x \in I$.

The other possibility is that x is within ϵ distance of one of the turning points, in other words that $x \in V$ where:

$$V = \{x \in I : |x - c_j| < \epsilon \text{ for } j \in \{1, 2, \ldots, q\}\}.$$

Below we show that for all $g \in U$, if $x \in V$ does not satisfy condition (I) then x satisfies condition (II) of the lemma. This would complete the proof of lemma B.2.7.

Suppose that $|x - c_j| < \epsilon$ for some $j \in \{1, 2, \ldots, q\}$ and suppose that x does not satisfy condition (I) for any $n(x) \in \{1, 2, \ldots, m_j\}$. In qualitative terms, since f is expansive by a factor of $s > 9$ everywhere except at the turning points, the only way for x not to satisfy condition (I) is if x is close enough to c_j so that $D_i(x, g, \epsilon)$ represents a "folded" line segment for every $i \in \{1, 2, \ldots, m_j\}$.

More precisely, for each $i \in \{1, 2, \ldots, m_j$ if we let

$$J_i(x, g, \epsilon) = \{y \in I : |f^k(x) - g^k(y)| < \epsilon \text{ for } k \in \{0, 1, \ldots, i\}\}.$$

so that $D_i(x, g, \epsilon) = g^i(J_i(x, g, \epsilon))$, then following claim is true.

Claim: Given $g \in U$, suppose that $x \in B(c_j, \epsilon)$ does not satisfy condition (I) of lemma B.2.7 for any $n(x) \in \{1, 2, \ldots, m_j\}$. Then for each $j \in \{1, 2, \ldots, q\}$ we claim that the following three statements are true:

(1) For any $i \in \{1, 2, \ldots, m_j\}$, if we define $y_i(j) \in J_i(x, g, \epsilon)$ such that:

$$g^i(y_i(j)) = \begin{cases} \sup_{z \in J_i(x,g,\epsilon)} g^i(z) & \text{if } \sigma_i(c_j) = +1 \\ \inf_{z \in J_i(x,g,\epsilon)} g^i(z) & \text{if } \sigma_i(c_j) = -1 \end{cases} \tag{B.16}$$

then

$$D_i(x,g,\epsilon) = (f^i(x) - \sigma_i(c_j)\epsilon \ ; \ g^i(y_i(j))] \tag{B.17}$$

and $g^i(y_i(j)) \in (f^i(x) - \epsilon, f^i(x) + \epsilon)$.
(2) For any $i \in \{1, 2, \ldots, m_j - 1\}$, $D_i(x, f, \epsilon) \cap C(f) = \emptyset$.
(3) For any $i \in \{1, 2, \ldots, m_j\}$, $y_i(j) \in J_i(x, f, \epsilon)$.

Proof of claim: We prove parts (1) and (2) of this claim by induction on i.

First we demonstrate that if conditions (1) and (2) above are true for each $i \in \{1, 2, \ldots, k\}$ where $k \in \{1, 2, \ldots, m_j - 1\}$, then condition (1) is true for $i = k + 1$. Thus we assume that $D_k(x, g, \epsilon)$ has the form given in (B.17), if $x \in B(x, \epsilon)$, so that:

$$D_k(x,g,\epsilon) \supset (f^k(x) - \sigma_k(c_j)\epsilon \ ; \ g^k(x)].$$

Since $|f^k(x) - g^k(x)| < \frac{1}{4}\epsilon$, this means:

$$D_k(x,g,\epsilon) \supset (f^k(x) - \sigma_k(c_j)\epsilon \ ; \ f^k(x) - \frac{1}{4}\sigma_k(c_j)\epsilon].$$

In particular $(f^k(x) - \frac{1}{2}\sigma_k(c_j))\epsilon \in D_k(x, g, \epsilon)$. Since $D_k(x, f, \epsilon) \supset (f^k(x) - \sigma_k(c_j)\epsilon \ ; \ f^i(x)]$ and $D_k(x, f, \epsilon) \cap C(f) = \emptyset$ (assuming that (2) is true for $i = k$) we know that $[C(f) \cap (f^k(x) - \frac{1}{2}\sigma_k(c_j)\epsilon \ ; \ f^i(x))] = \emptyset$. Thus, since $s > 9$:

$$g(f^k(x) - \frac{1}{2}\sigma_k(c_j)\epsilon) \in (f^k(x) - \frac{1}{2}s\sigma_{k+1}(c_j)\epsilon - \delta_x \ ; \ f^k(x) - \frac{1}{2}s\sigma_{k+1}(c_j)\epsilon + \delta_x)$$

Now suppose that c_j is a relative maximum of the map f^{k+1} so that $\sigma_{k+1}(c_j) = +1$ (the case where $\sigma_{k+1}(c_j) = -1$ is analogous). Then we find that:

$$g(f^k(x) - \frac{1}{2}\sigma_k(c_j)\epsilon) < f^k(x) - \epsilon$$

where $g(f^k(x) - \frac{1}{2}\sigma_k(c_j)\epsilon) \in g(D_k(x, g, \epsilon))$. Thus, since $D_k(x, g, \epsilon)$ and hence $g(D_k(x, g, \epsilon))$ are connected sets, this means that since

$$D_{k+1}(x,g,\epsilon) \quad = \quad g(D_k(x,g,\epsilon)) \cap B(f^{k+1}(x),\epsilon)$$

we know that $f^k(x) - \epsilon$ must be the lower endpoint of $D_{k+1}(x, g, \epsilon)$. Also we know that

$$D_{k+1}(x,g,\epsilon) \subset (f^{k+1}(x) - \epsilon \ ; \ f^{k+1}(x) + \epsilon)$$

because otherwise condition (I) is satisfied for $n(x) = k + 1$. Consequently by the definition of $y_k(j)$ in (B.16), we see that:

$$D_{k+1}(x,g,\epsilon) = (f^{k+1}(x) - (c_j)\epsilon \ ; \ g^k(y_{k+1}(j))].$$

where $g^k(y_{k+1}(j)) \in (f^{k+1}(x) - \epsilon \ ; \ f^{k+1}(x) + \epsilon)$ if $\sigma_{k+1}(c_j) = +1$. Combing this with the corresponding result for $\sigma_{k+1}(c_j) = -1$ proves that condition (1) is true for $i = k+1$ given that (1) and (2) are true for $i = k$.

Next we show that if (1) and (2) are true for each $i \in \{1, 2, \ldots, k\}$ where $k \in \{1, 2, \ldots, m_j - 2\}$, then (2) is true for $i = k+1$. Suppose on the contrary that (2) is not true for $k = i+1$ so that $D_{k+1}(x, f, \epsilon) \cap C(f) \neq \emptyset$. Since $D_{k+1}(x, f, \epsilon) \subset B(f^{k+1}(x), \epsilon)$ we know that:

$$f^{k+1}(x) \in B(c, \epsilon) \tag{B.18}$$

for some $c \in C(f)$. From (B.8) and (B.9) we also know that:

$$f^i(c_j) \notin (c \ ; \ c + \frac{5}{2}\sigma_i(c_j)\epsilon) \tag{B.19}$$

for any $c \in C(f)$ if $i \in \{1, 2, \ldots, m_j - 2\}$.

We now address two cases. First suppose that there exists some $t \in \{1, 2, \ldots, k\}$ and $c \in C(f)$ such that:

$$c \in (f^t(x) \ ; \ f^t(c_j)) \tag{B.20}$$

Let t be the minimum value for which (B.20) holds for any $c \in C(f)$. Since t is minimal we know that f^t must be monotone on $(x; c_j)$ so that:

$$\sigma_t(c_j)(f^t(c_j) - f^t(x)) \geq 0.$$

Combining this result with (B.20) and (B.19) we find that:

$$\sigma_t(c_j)(f^t(c_j) - f^t(x)) > \frac{5}{2}\epsilon. \tag{B.21}$$

Now suppose there exists no $i \in \{1, 2, \ldots, k\}$, such that:

$$c \in (f^i(x) \ ; \ f^i(c_j))$$

for any $c \in C(f)$. Note that since we assume (2) is true for $i \leq k$, this means there exists no $i \in \{1, 2, \ldots, k\}$, such that:

$$c \in (f^i(x) \ ; \ f^i(c_j)) \cup D_i(x, f, \epsilon).$$

for any $c \in C(f)$. Then for any $i \in \{1, 2, \ldots, k+1\}$, we know that f^i is monotone on $(x; c_j) \cup J_i(x, f, \epsilon)$. Thus, for any $z \in D_i(x, f, \epsilon)$ we have:

$$\sigma_i(c_j)(f^i(c_j) - z) \geq 0$$

and from (B.18) and (B.19):

$$\sigma_{k+1}(c_j)(f^{k+1}(c_j) - f^{k+1}(x)) > \frac{3}{2}\epsilon. \tag{B.22}$$

From (B.21) and (B.22) we have shown that if (2) is satisfied for any $i \in \{1, 2, \ldots, k\}$ then there exists $t \leq k+1$ such that:

$$\sigma_t(c_j)(f^t(c_j) - f^t(x)) > \frac{3}{2}\epsilon.$$

This implies that:

$$\sigma_t(c_j)(g^t(c_j) - f^t(x)) > \epsilon$$

so $c_j \notin J_t(x, g, \epsilon)$. Thus there exists some $\ell \in \{0, 1, \ldots, t-1\}$ such that $c_j \in J_i(x, g, \epsilon)$ for any i satisfying $1 \leq i \leq \ell$ but $c_j \notin J_{\ell+1}(x, g, \epsilon)$. Since $D_i(x, g, \epsilon) \cap C(f) = \emptyset$ for any $i \in \{1, 2, \ldots, \ell\}$ we know that:

$$\sigma_{\ell+1}(c_j)(f^{\ell+1}(c_j) - f^{\ell+1}(x)) \geq 0.$$

Consequently, since $c_j \notin J_{\ell+1}(x, g, \epsilon)$, it is apparent that:

$$\sigma_{\ell+1}(c_j)(g^{\ell+1}(c_j) - f^{\ell+1}(x)) > \epsilon.$$

Thus, since $D_\ell(x, g, \epsilon)$ is connected, and since $g^{\ell+1}(c_j) \in g(D_\ell(x, g, \epsilon)$, we know that $f^{\ell+1}(x) + \sigma_{\ell+1}(c_j)\epsilon$ must be an endpoint of $D_{\ell+1}(x, g, \epsilon) = g(D_\ell(x, g, \epsilon) \cap B(f^\ell(x), \epsilon)$ where $\ell + 1 \leq t \leq k+1$. This contradicts (1) for $i = \ell + 1 \leq k+1$. But we have already shown that if (1) and (2) are satisfied for $i \in \{1, 2, \ldots, k\}$, then (1) is satisfied for $i = k+1$. Thus if (1) and (2) are satisfied for $i \in \{1, 2, \ldots, k\}$, then (2) is also satisfied for $i = k+1$.

We now need to show that (1) is true for $i = 1$. By definition, we can write: $D_1(x, g, \epsilon) = g[(x-\epsilon, x+\epsilon)] \cap B(f(x), \epsilon)$. If condition (I) is not satisfied, then $D_1(x, g, \epsilon) \subset (f(x)-\epsilon, f(x)+\epsilon)$ and at least one endpoint of $D_1(x, g, \epsilon)$ has to correspond either to a maximum or minimum point of g in the interior of $J_1(x, g, \epsilon)$. Since $s > 9$, and since all the turning points of f are separated by at least 16ϵ, we know that the other endpoint of $D_1(x, g, \epsilon)$ must be $f(x) - \sigma_1(c_j)\epsilon$. Thus $D_1(x, g, \epsilon)$ has the form given in (B.17).

Now we show that (2) is true for $i = 1$. Suppose that $D_1(x, g, \epsilon) \cap C(f) \neq \emptyset$. Then $\sigma_1(c_j)(f(x) - c) \leq \epsilon$ for some $c \in C(f)$. If $x \in B(c_j, \epsilon)$ and $m_j > 1$ then $\sigma_1(c_j)(f(c_j) - c) \geq \frac{5}{2}\epsilon$ for any $c \in C(f)$. Thus $\sigma_1(c_j)(f(c_j) - f(x)) \geq \frac{3}{2}\epsilon$ which means that $\sigma_1(c_j)(g(c_j) - f(x)) \geq \epsilon$. This contradicts (1) for $i = 1$ and completes the proof of parts (1) and (2) of the claim.

We now show that condition (3) of the claim holds. Suppose on the contrary that there exists $x \in B(c_j, \epsilon)$ for some $j \in \{1, 2, \ldots, q\}$ such that $y_i(j) \notin J_i(x, f, \epsilon)$ for some $i \in \{1, 2, \ldots, m_j\}$. Then there exists a $k \in \{0, 1, \ldots, i-1\}$ such that $y_{k+1}(j) \notin J_{k+1}(x, f, \epsilon)$ but $y_\ell(j) \in J_k(x, f, \epsilon)$ for any integer ℓ satisfying $1 \leq \ell \leq k$. We know that:

$$f^{k+1}(y_i(j)) \notin (f^{k+1}(x) - \epsilon \ , \ f^{k+1}(x) + \epsilon),$$
$$g^{k+1}(y_i(j)) \in (f^{k+1}(x) - \epsilon \ , \ f^{k+1}(x) + \epsilon).$$

And, since $|f^{k+1}(y_{k+1}(j)) - g^{k+1}(y_{k+1}(j))| < \delta_x$, we find that:

$$\begin{aligned} f^{k+1}(y_i(j)) &\in (f^{k+1}(x) - \epsilon - \delta_x \ , \ f^{k+1}(x) - \epsilon) \\ &\bigcup \ (f^{k+1}(x) + \epsilon \ , \ f^{k+1}(x) + \epsilon + \delta) \qquad \text{(B.23)} \\ g^{k+1}(y_i(j)) &\in (f^{k+1}(x) - \epsilon \ , \ f^{k+1}(x) - \epsilon + \delta_x) \\ &\bigcup \ (f^{k+1}(x) + \epsilon - \delta_x \ , \ f^{k+1}(x) + \epsilon). \qquad \text{(B.24)} \end{aligned}$$

Also, substituting $f = g$ in part (1) of the claim, we can see that:

$$D_i(x, f, \epsilon) = (f^i(x) - \sigma_i(c_j)\epsilon \ ; \ f^i(c_j)] \qquad \text{(B.25)}$$

where $f^i(c_j) \in (f^i(x) - \epsilon \ , \ f^i(x) + \epsilon)$ for any $i \in \{1, 2, \ldots, m_j\}$ provided condition (I) of the lemma is not satisfied. Now suppose $\sigma_{k+1}(c_j) = +1$ (the other case is analogous). Then, since $y_i(j) \in J_k(x, f, \epsilon)$, we know that it cannot be true that $f^{k+1}(y_i(j)) \geq f^{k+1}(x) + \epsilon$, since that would contradict (B.25). Thus we can drop one of the intervals in each the unions in (B.23) and (B.24). In particular we find that:

$$g^{k+1}(y_i(j)) \in (f^{k+1}(x) - \sigma_{k+1}(c_j) \ ; \ f^{k+1}(x) - \sigma_{k+1}(c_j)(\epsilon - \delta_x)). \qquad \text{(B.26)}$$

This implies $i \neq k+1$ since:

$$\begin{aligned} \text{if } \sigma_{k+1}(c_j) = +1\text{:} \quad g^{k+1}(y_{k+1}(j)) &= \sup_{z \in J_{k+1}(x,g,\epsilon)} g^{k+1}(z) \geq f^{k+1}(x) > g^{k+1}(y_i(j)) \\ \text{if } \sigma_{k+1}(c_j) = -1\text{:} \quad g^{k+1}(y_{k+1}(j)) &= \inf_{z \in J_{k+1}(x,g,\epsilon)} g^{k+1}(z) \leq f^{k+1}(x) < g^{k+1}(y_i(j)). \end{aligned}$$

But since $D_{k+1}(x, f, \epsilon) \cap C(f) = \emptyset$ for $k+1 < m_j$ we know from (B.25) that:

$$(f^{k+1}(x) + \sigma_{k+1}(c_j)\epsilon) \ ; \ f^{k+1}(x)) \ \bigcap \ C(f) = \emptyset.$$

Thus from (B.26), since $s > 9$, it is clear that

$$g^{k+2}(y_i(j)) \notin D_{k+2}(x, g, \epsilon).$$

This means that $y_i(j) \notin J_\ell(x, g, \epsilon)$ for any $\ell \geq k+2$, so $i \leq k+1$. But we have already shown that $i \neq k+1$. Therefore $i \leq k$. But this contradicts our assumption that $k \in \{0, 1, \ldots, i-1\}$. This proves condition (3) and completes the proof of the claim.

Returning to the proof of lemma B.2.7 we now assert that:

$$E_{m_j}(x, g, \epsilon) \supseteq (f^{m_j}(x) - 8\sigma_{m_j}(c_j)\epsilon \ , \ g^{m_j}(y_{m_j}(j))]. \qquad \text{(B.27)}$$

if x does not satisfy condition (I) of the lemma for any $n(x) \in \{1, 2, \ldots, m_j\}$. It is clear that $D_i(x, g, \epsilon) \subseteq E_i(x, g, \epsilon)$ for each each $i \in \{1, 2, \ldots, m_j\}$. We also know that $|f(x) - g(x)| < \frac{1}{4}\epsilon$ for all $x \in I$ so that given the form of $D_i(x, g, \epsilon)$ in (B.17) and because of the expansion factor, $s > 9$, we have that:

$$E_{i+1}(x.g, \epsilon) \supseteq g(D_i(x, g, \epsilon)) \cap B(f^{i+1}(x), 8\epsilon).$$

for any $i \in \{1, 2, \ldots, m_j - 1\}$. Setting $i = m_j - 1$, and substituting $D_i(x, g, \epsilon)$ in the equation above using (B.17), we get (B.27).

Now suppose that $\sigma_{m_j}(c_j) = +1$ (the case where $\sigma_{m_j}(c_j) = -1$ is analogous). Then, from (B.10):

$$f^{m_j}(c_j) - c_{r(j)} \geq 10\delta_x. \tag{B.28}$$

Also, if condition (I) is not satisfied for some $x \in B(c_j, \epsilon)$, then since $y_{m_j}(j) \in D_{m_j}(x, f, \epsilon)$ we know that $f^{m_j}(c_j) > f^{m_j}(y_{m_j}(j))$ since $D_{m_j-1}(x, f, \epsilon) \cap C(f) = \emptyset$. Thus, because $|f^{m_j}(x) - g^{m_j}(x)| < \delta_x$:

$$\begin{aligned}
g^{m_j}(y_{m_j}(j)) - f^{m_j}(c_j) &< (f^{m_j}(y_{m_j}(j)) + \delta_x) - f^{m_j}(c_j) \\
&< (f^{m_j}(c_j) + \delta_x) - f^{m_j}(c_j) \\
&< \delta_x && \text{(B.29)} \\
g^{m_j}(y_{m_j(j)}) - f^{m_j}(c_j) &\geq g^{m_j}(c_j) - f^{m_j}(c_j) > -\delta_x. && \text{(B.30)}
\end{aligned}$$

Note that f has either a local maximum or a local minimum at $c_{r(j)}$. For definiteness, assume that f has a local maximum at $c_{r(j)}$ (the other case is again analogous). Then, since $|f(x) - g(x)| < \delta_x$ for all $x \in I$, there exists a local maximum of the map, g, at $y_1(r(j))$ such that:

$$\begin{aligned}
g(y_1(r(j))) &= \sup_{x \in B(c_{r(j)}, 8\epsilon)} g(x) && \text{(B.31)} \\
\text{and} \qquad y_1(r(j)) &\in B(c_{r(j)}, 2\frac{\delta_x}{s}). && \text{(B.32)}
\end{aligned}$$

since the turning points of f are separated by at least 16ϵ distance.

Consequently from (B.28), (B.30), (B.32), and since $s > 9$ we see that:

$$\begin{aligned}
&g^{m_j}(y_{m_j}(j)) - y_1(r(j)) \\
&\quad = [c_{r(j)} + (f^{m_j}(c_j) - c_{r(j)}) + (g^{m_j}(y_{m_j}(j)) - f^{m_j}(c_j))] - [c_{r(j)} + (y_1(r(j)) - c_{r(j)})] \\
&\quad > [c_{r(j)} + 10\delta_x - \delta_x] - [c_{r(j)} + 2\frac{\delta_x}{s})] \\
&\quad > 0. && \text{(B.33)}
\end{aligned}$$

Also, from (B.29), (B.11), and (B.32) and since $s > 9$ and $\delta < \frac{1}{4}\epsilon$:

$$\begin{aligned}
&g^{m_j}(y_{m_j}(j)) - y_1(r(j)) \\
&\qquad = (g^{m_j}(y_{m_j}(j)) - f^{m_j}(c_j)) + (f^{m_j}(c_j) - c_{r(j)}) - (c_{r(j)}) - y_1(r(j))) \\
&\qquad < \delta_x + \frac{5}{2}\epsilon - 2\frac{\delta_x}{s} \\
&\qquad < 3\epsilon
\end{aligned} \tag{B.34}$$

Consequently, from (B.33), (B.34), and (B.27) we see that if $x \in B(c_j, \epsilon)$ does not satisfy condition (I), then

$$y_1(r(j)) \in E_{m_j}(x, g, \epsilon). \tag{B.35}$$

Furthermore, from (B.31) we also know that:

$$g(y_1(r(j))) = \sup_{z \in E_{m_j}(x,g,\epsilon)} g(z). \tag{B.36}$$

If we assume $\sigma_{m_j}(c_j) = +1$, then from (B.27), (B.29), (B.11), (B.32), and since $s > 9$ and $\delta_x < \frac{1}{4}\epsilon$ we have:

$$\begin{aligned}
g^{m_j}(x) &\le g^{m_j}(y_{m_j}(j)) \\
&< f^{m_j}(c_j) + \delta_x \\
&< c_{r(j)} + \frac{5}{2}\epsilon + \delta_x \\
&< y_1(r(j)) + 2\frac{\delta_x}{s} + \frac{5}{2}\epsilon + \delta_x \\
&< y_1(r(j)) + 3\epsilon
\end{aligned} \tag{B.37}$$

Still assuming $\sigma_{m_j}(c_j) = +1$, then from (B.27), (B.36), (B.37), and since $\delta_x < \frac{1}{4}\epsilon$, and $|f(x) - g(x)| < \delta_x$ for all $x \in I$:

$$\begin{aligned}
g(E_{m_j}(x, g, \epsilon)) &\supseteq (g(g^{m_j}(x) - 8\epsilon)\ ,\ g(y_1(r(j))] \\
&\supseteq (g(y_1(r(j)) - 5\epsilon)\ ,\ g(y_1(r(j))] \\
&\supseteq (g(y_1(r(j))) - 5s\epsilon + \delta_x\ ,\ g(y_1(r(j))] \\
&\supseteq (g(y_1(r(j))) - \frac{9}{2}s\epsilon\ ,\ g(y_1(r(j))]
\end{aligned} \tag{B.38}$$

Finally, if $\sigma_{m_j}(c_j) = +1$, then since $c_{r(j)} < f^{m_j}(c_j) < c_{r(j)} + \frac{5}{2}\epsilon$ and $s > 9$, we know from (B.32) that $c_{r(j)} - \frac{1}{2}\epsilon < y_1(r(j))) < c_{r(j)} + 3\epsilon$. Thus:

$$\begin{aligned}
g(B[f^{m_j}(x), \epsilon]) &\subseteq (g(y_1(r(j))) - 4s\epsilon - \delta_x\ ,\ g(y_1(r(j)))] \\
&\subseteq (g(y_1(r(j))) - \frac{9}{2}s\epsilon\ ,\ g(y_1(r(j)))]
\end{aligned} \tag{B.39}$$

Consequently, from (B.38) and (B.39), we have that if $x \in V$ does not satisfy condition (I) of lemma B.2.7 for any $n(x) \in \{1, 2, \ldots, m_j\}$, then:

$$g(E_{m_j}(x, g, \epsilon)) \supseteq g(B[f^{m_j}(x), \epsilon]),$$

satisfying condition II of the lemma. We already saw that condition I of the lemma is satisfied for $n(x) = 1$ if $x \in I \setminus V$. This proves lemma B.2.7.

Proof of theorem 3.2.1:
Strong linking condition → Function shadowing: Note that (B.6) in lemma B.2.5 may be rewritten as:

$$D_{n(x)}(x, g, \epsilon) \supseteq B[f^{n(x)}(x), \epsilon]$$

and (B.7) in lemma B.2.6 may be rewritten as

$$g(E_{n(x)}(x, g, \epsilon)) \supseteq g(B[f^{n(x)}(x), \epsilon])$$

so we can see these two statements are the same as conditions in lemma B.2.7.

For any $x \in I$, condition (I) of lemma B.2.7 implies that condition (II) must also be true, since clearly $E_{n(x)}(x, g, \epsilon) \supseteq D_{n(x)}(x, g, \epsilon)$. Thus, combining lemmas B.2.7 and B.2.6, we see that if $f : I \to I$ is uniformly piecewise linear with $s > 9$ and the strong linking property, then f must satisfy the function shadowing property on $\mathbb{C}^0(I, I)$. Furthermore, using lemma B.2.3, we can drop the requirement that $s > 9$. We can do this since $s > 1$ for any uniformly piecewise linear map f, so there always exists $n > 0$ such that f^n is uniformly piecewise linear and satisfies $s > 9$. Thus, from lemmas B.2.1 and B.2.2, we know that any transitive map $f : I \to I$ with the strong linking property must also satisfy a the function shadowing property on $\mathbb{C}^0(I, I)$.

Function shadowing → Strong linking condition: Suppose that f is a piecewise linear map that does not satisfy the strong linking condition. We shall first show that f does not satisfy the function shadowing property on $\mathbb{C}^0(I, I)$.

If f does not satisfy the strong linking condition, then there is a $c \in C(f)$ and $\epsilon_0 > 0$ such that there exists no $z \in \{B(c, \epsilon) \setminus c\}$ and $n \in \mathbb{Z}^+$ satisfying $f^n(z) \in C(f)$ and $|f^i(c) - f^i(z)| < \epsilon_0$ for every $i \in \{1, 2, \ldots, n\}$. We will show that if $\epsilon \in (0, \frac{1}{2}\epsilon_0)$, then for any $\delta > 0$ there exists a $g \in \mathbb{C}^0(I, I)$ that satisfies $d(f, g) \leq \delta$ but has the property that no orbit of g $\epsilon-$shadows the orbit, $\{f^i(c)\}_{i=0}^{\infty}$, of f.

Now given $\delta > 0$ and $\epsilon < \frac{1}{2}\epsilon_0$, choose g to be any map that satisfies the following properties:

(1) $g \in \mathbb{C}^0(I, I)$

(2) $g(c) = f(c) - \sigma_1(c)\delta$

(3) $g(x) = f(x)$ for any $x \in \{I \setminus B(c, \epsilon_0)\}$.

(4) $sup_{x \in B(c,\epsilon)}[\sigma_1(c)g(x)] = \sigma_1(c)g(c)$

(5) $d(f, g) \leq \delta$

Set $x_i = f^i(c)$ and let $y_i = g^i(c)$ so that $\{y_i\}$ is an orbit of g. Suppose that $k \in \mathbb{Z}^+$ such that $\sigma_i(c)(x_i - y_i) < \epsilon_0$ for all $i \in \{0, 1, \ldots, k\}$. We assert that

$$\sigma_i(c)(x_i - y_i) \geq s^{i-1}\delta \tag{B.40}$$

for any $i \in \{1, 2, \ldots, k+1\}$. It is not hard to show this assertion by induction. For any $i \in \{1, 2, \ldots, k\}$ we have that $C(f) \cap (x_i; y_i) = \emptyset$ and $\sigma_{i+1}(c)(f(y_i) - g(y_i)) \geq 0$. Thus, since $\sigma_{i+1}(c)(f(x_i) - f(y_i)) = s\sigma_i(c)(x_i - y_i)$, we have that

$$\sigma_{i+1}(c)(f(x_i) - g(y_i)) \geq \sigma_{i+1}(c)(f(x_i) - f(y_i)) = s\sigma_i(c)(x_i - y_i) \tag{B.41}$$

so that if (B.40) is true for i, then it also must be true for $i+1$, provided that $i \in \{1, 2, \ldots, k\}$.

But $\{y_i\}_{i=0}^{k+1}$ does not $\epsilon-$shadow $\{x_i\}_{i=0}^{k+1}$. We can see this from (B.40) and from our choice of k, since $\epsilon < \frac{1}{2}\epsilon_0$. Furthermore there is no orbit of g that more closely shadows $\{x_i\}_{i=0}^{k+1}$ than $\{y_i\}_{i=0}^{k+1}$. This is because for any $u \in I$, if $i \in \{1, 2, \ldots, k\}$ and $u \in J_i(c, g, \epsilon)$, then $(g^i(u); x_i) \cap C(f) = \emptyset$ since $\epsilon < \frac{1}{2}\epsilon_0$. Also, using property (4) of our choice of g, we can show that $sup_{z \in J_i(c,g,\epsilon)}[\sigma_i(c)g^i(z)] = \sigma_i(c)g^i(c)$ for any $i \in \{1, 2, \ldots, k+1\}$ by induction on i.

Consequently, if f is a piecewise linear map that does not satisfy the strong linking condition, then it cannot satisfy the function-shadowing in $\mathbb{C}^0(I, I)$. Since the function shadowing property is preserved by topological conjugacy (lemma B.2.2) this implies that a transitive piecewise monotone map cannot exhibit function shadowing in $\mathbb{C}^0(I, I)$ if it does not satisfy the strong linking condition.

This concludes the proof of theorem 3.2.1.

Appendix C

Proof of theorem 3.3.1

This appendix contains the proof for theorem 3.3.1. I have made an effort to make the appendix as self-contained as possible, so that the reader should be able to find most of the relevant definitions and explanations in this appendix. Naturally, this means that the appendix repeats some material found elsewhere in this report.

C.1 Definitions and statement of theorem

We first repeat the related definitions which are the same as those found in chapter 3. Throughout this appendix we shall assume that $I \subset \mathbb{R}$ represents a compact interval of the real line.

Definitions: Suppose that $f : I \to I$ is continuous. Then the *turning points* of f are the local extrema of f in the interior I. $C(f)$ is used to designate the set of all turning points of f on I. Let $\mathbb{C}^r(I, I)$ be the set of continuous maps on I such that $f \in C^r(I, I)$ if the following three conditions hold:
(a) f is C^r (for $r \geq 0$)
(b) $f(I) \subseteq I$.
(c) $f(Bd(I)) \subseteq Bd(I)$ (where $Bd(I)$ denotes the boundary of I),
If $f \in \mathbb{C}^r(I, I)$ and $g \in \mathbb{C}^r(I, I)$, let $d(f, g) = \sup_{x \in I} |f(x) - g(x)|$.

We will primarily restrict ourselves to maps with the following properties:

(C0) $g : I \to I$, is piecewise monotone.

(C1) g is C^2 on I.

(C2) Let $C(g)$ be the finite set such that $c \in C(g)$ if and only if g has a local extremum at $c \in I$. Then $g''(c) \neq 0$ if $c \in C(g)$ and $g'(x) \neq 0$ for all $x \in I \setminus C(g)$.

Under the Collet-Eckmann conditions, there exist constants $K_E > 0$ and $\lambda_E > 1$ such that for some $c \in C(g)$:

(CE1) $|Dg^n(g(c))| > K_E \lambda_E^n$

(CE2) $|Dg^n(z)| > K_E \lambda_E^n$ if $g^n(z) = c$.

for any $n > 0$.

We consider one-parameter families of mappings, $f_p : I_x \to I_x$, parameterized by $p \in I_p$, where $I_x \subset \mathbb{R}$ and $I_p \subset \mathbb{R}$ are closed intervals of the real line. Let $f(x,p) = f_p(x)$ where $f : I_x \times I_p \to I_x$. We are primarily interested in one-parameter families of maps with the following characteristics:

(D0) For each $p \in I_p$, $f_p : I_x \to I_x$ satisfies (C0) and (C1). We also require that $C(f_p)$ remains invariant with respect to p for all $p \in I_p$.

(D1) $f : I_x \times I_p \to I_x$ is C^2 for all $(x,p) \in I_x \times I_p$.

Note that the following notation will be used to express derivatives of $f(x,p)$ with respect to x and p.

$$D_x f(x,p) = \frac{\partial f}{\partial x}(x,p) \tag{C.1}$$

$$D_p f(x,p) = \frac{\partial f}{\partial p}(x,p). \tag{C.2}$$

The Collet-Eckmann conditions specify that derivatives with respect to the state, x, grows exponentially. Similarly we will also be interested in families of maps where derivatives with respect to the parameter, p, also grow exponentially. In other words, we require that there exist constants $K_p > 0$, $\lambda_p > 1$, and $N > 0$ such that for some $p_0 \in I_p$, and $c \in C(f_{p_0})$:

(CP1) $|D_p f^n(c, p_0)| > K_p \lambda_p^n$

for all $n \geq N$. This may seem to be a rather strong constraint, but in practice it often follows whenever (CE1) holds. We can see this by expanding with the chain rule:

$$D_p f^n(c,p_0) = D_x f(f^{n-1}(c,p_0),p_0) D_p f^{n-1}(c,p_0) + D_p f(f^{n-1}(c,p_0),p_0) \tag{C.3}$$

to obtain the formula for $D_p f^n(x,p_0)$:

$$D_p f^n(x,p_0) = D_p f(f^{n-1}(c,p_0),p_0) + \sum_{i=0}^{n-2} [D_p f(f^i(c,p_0),p_0) \prod_{j=i+1}^{n-1} D_x f(f^j(c,p_0),p_0)].$$

Thus, if $|D_x f^n(f(c,p_0),p_0)|$ grows exponentially, we expect $|D_p f^n(x,p_0)|$ to also grow exponentially unless the parameter dependence is degenerate in some way.

Now for any $c \in C(f_{p_0})$ define $\sigma_n(c,p)$ recursively as follows:

$$\sigma_{n+1}(c,p) \quad = \quad sgn\{D_x f(f^n(c,p),p)\}\sigma_n(c,p)$$

where

$$\sigma_1(c,p) \quad = \quad \begin{cases} \text{1 if } c \text{ is a relative maximum of } f_p \\ \text{-1 if } c \text{ is a relative minimum of } f_p \end{cases}$$

Basically $\sigma_n(c,p) = 1$ if f_p^n has a relative maximum at c and $\sigma_n(c,p) = -1$ if f_p^n has a relative minimum at c. We can use this notion to distinguish a particular direction in parameter space.

Definition C.1.1 *Let $\{f_p : I_x \to I_x | p \in I_p\}$ be a one-parameter family of mappings satisfying (D0) and (D1). Suppose that there exists $p_0 \in I_p$ such that f_{p_0} satisfies (CE1) and (CP1) for some $c \in C(f_{p_0})$. Then we say the that turning point c of f_{p_0} favors higher parameters if there exists $N' > 0$ such that*

$$sgn\{D_p f^n(c,p_0)\} \quad = \quad sgn\{\sigma_n(c,p)\} \tag{C.4}$$

for all $n \geq N'$. Similarly, the turning point, c, of f_{p_0} favors lower parameters if

$$sgn\{D_p f^n(c,p_0)\} \quad = \quad -sgn\{\sigma_n(c,p)\} \tag{C.5}$$

for all $n \geq N'$.

The first thing to notice about these two definitions is that they are exhaustive if (CP1) is satisfied. That is, if (CP1) is satisfied for some $p_0 \in I_p$ and $c \in C(f_{p_0})$, then the turning point, c, of f_{p_0} either favors higher parameters or favors lower parameters. We can see this from (C.3). Since $|D_p f(x,p_0)|$ is bounded for $x \in I_x$, if $|D_p f^n(x,p_0)|$ grows large enough then its sign is dominated by the signs of $D_x f(f^{n-1}(c,p_0),p_0)$ and $D_p f^{n-1}(c,p_0)$, so that either (C.4) or (C.5) must be satisfied.

Finally, if $p_0 \in I_p$ and $c \in C(f_{p_0})$, then for any $\epsilon \geq 0$, define $n_e(c,\epsilon,p_0)$ to be the smallest integer $n \geq 1$ such that $|f^n(c,p_0) - c_*| \leq \epsilon$ for any $c_* \in C(f_{p_0})$. We say that $n_e(c,\epsilon,p_0) = \infty$ if no such $n \geq 1$ exists.

We are now ready to state main result of this appendix.

Theorem 3.3.1 *Let $\{f_p : I_x \to I_x | p \in I_p\}$ be a one-parameter family of mappings satisfying (D0) and (D1). Suppose that (CP1) is satisfied for some $p_0 \in I_p$ and $c \in C(f_{p_0})$. Suppose further that f_{p_0} satisfies (CE1) at c, and that the turning point, c, favors higher parameters under f_{p_0}. Then there exists $\delta p > 0$, $\lambda > 1$, $K' > 0$, and $K \geq 1$, such that if $p \in (p_0 - \delta p, p_0)$, then for any $\epsilon > 0$, the orbit $\{f_{p_0}^n(c)\}_{n=0}^{\infty}$ is not $\epsilon-$shadowed by any orbit of f_p if $|p - p_0| > K'\epsilon\lambda^{-n_e(c,K\epsilon,p_0)}$.*

The analogous result also holds if f_{p_0} favors lower parameters.

C.2 Proof

Lemma C.2.1 *Let $\{f_p : I_x \to I_x | p \in I_p\}$ be a one-parameter family of mappings satisfying (D0) and (D1). Then given $p_0 \in I_p$, there exist constants $K_1 > 0$, $K_2 > 0$, and $K_3 > 0$ such that the following properties are satisfied:*

(1) $|D_x f(x_1, p_0) - D_x f(x_2, p_0)| < K_1 |x_1 - x_2|$ *for any* $x_1 \in I_x$ *and* $x_2 \in I_x$.

(2) Let $\delta x > 0$ *to be the maximal value such that* $|x - c_*| < \delta x$ *implies* $|D_x^2 f(x, p_0)| > 0$ *for any* $c_* \in C(f_{p_0})$. *Then* $|Df(x, p_0)| > K_2|x - c|$ *if* $|x - c| < \delta x$ *for some* $c \in C(f_{p_0})$.

(3) Fix $c \in C(f_{p_0})$. *Then,* $|D_x f(x, p) - D_x f(x, p_0)| < K_3 |x - c||p_1 - p_2|$ *for any* $x \in I_x$ *and* $p \in I_p$.

Proof of (1): (1) is true since $f(x, p)$ is C^2 and $I_x \times I_p$ is compact.

Proof of (2): From (C2) we know that it is possible to choose a $\delta x > 0$ as specified. Let $c \in C(f_{p_0})$ and $x \in I_x$. By the mean value theorem:

$$|D_x f(x, p_0)| = |D_x^2 f(y, p_0)||x - c|$$

for some $y \in [c; x]$. Now set:

$$K_2 = \frac{1}{2} \inf_{y \in [c - \frac{1}{2}\delta x, c + \frac{1}{2}\delta x]} |D_x^2 f(y, p_0)|.$$

From our choice of δx, we know $K_2 > 0$. Thus if $|x - c| < \frac{1}{2}\delta x$, we have that:

$$|Df(x, p_0)| > 2K_2|x - c|.$$

But since $|D_x^2 f(y, p_0)| > 0$ if $|x - c| < \delta x$, it is evident that $|Df(y, p_0)| \geq |Df(x + \frac{1}{2}\delta, p_0)|$ for any $y \in (c + \frac{1}{2}\delta x, c + \delta x)$. Similarly $|Df(y, p_0)| \geq |Df(x - \frac{1}{2}\delta, p_0)|$ if $y \in (c - \delta x, c - \frac{1}{2}\delta x)$. Thus:

$$|Df(x, p_0)| > K_2|x - c|$$

for any x satisfying $|x - c| < \delta x$.

Proof of (3): Fix $c \in C(f_{p_0})$ and $p_0 \in I_p$. Then for any $x \in I_x$ and $p \in I_p$, let:

$$q(x, p) = D_x f(x, p) - D_x f(x, p_0).$$

Since f is C^2, q must be C^1. It is clear that:

$$q(c, p) = 0 \tag{C.6}$$

for all $p \in I_p$ and

$$q(x, p_0) = 0 \tag{C.7}$$

for all $x \in I_x$.

From (C.7) and since $q(x,p)$ is C^1, $q(x,p)$ satisfies a Lipschitz condition on $I_x \times I_p$ so that there exists a constant $C > 0$ such that:

$$|q(x,p)| < C|p - p_0|. \tag{C.8}$$

for any $(x,p) \in I_x \times I_p$. Now define

$$r(x,p) = \begin{cases} \frac{q(x,p)}{p-p_0} & \text{if } p \neq 0 \\ D_p q(x, p_0) & \text{if } p = p_0 \end{cases} \tag{C.9}$$

Note that from (C.8), $|r(x,p)| < C|I_p|$ for any $(x,p) \in I_x \times I_p$ such that $p \neq p_0$. Since r is bounded and $q(x,p)$ is C^1, it is fairly easy to check that $r(x,p)$ is C^1 for all $(x,p) \in I_x \times I_p$.

From (C.9) and (C.7), we see that:

$$q(x,p) = r(x,p)(p - p_0) \tag{C.10}$$

for all $(x,p) \in I_x \times I_p$. Also from (C.6) we know $r(c,p) = 0$ for all $p \in I_p$. Thus since $r(x,p)$ is C^1, there exists $K_3 > 0$ such that $|r(x,p)| < K_3|x - c|$ for any $(x,p) \in I_x \times I_p$. Substituting this into (C.10) we find that:

$$|q(x,p)| < K_3|x - c||p - p_0|$$

for any $(x,p) \in I_x \times I_p$. This proves part (3) of the lemma.

Lemma C.2.2 *Let $\{f_p : I_x \to I_x | p \in I_p\}$ be a one-parameter family of mappings satisfying (C0) and (C1). Suppose that f_{p_0} satisfies (CE1) for $p_0 \in I_p$ and some turning point, $c \in C(f_{p_0})$. Suppose that turning point c of f_{p_0} favors higher parameters. Given any $\lambda_0 > \lambda_1 > 1$, there exist constants $K \geq 1$, $\delta p > 0$ and $\epsilon_0 > 0$ such that for any $\epsilon < \epsilon_0$, if $|p - p_0| < \delta p$, $|f^i(c,p) - f^i(c,p_0)| < \epsilon$, and $|f^i(c,p_0) - c_*| > K\epsilon$ for all $c_* \in C(f_{p_0})$ and $1 \leq i \leq n$ then:*

$$\frac{|D_x(f^i(c,p),p)|}{|D_x(f^i(c,p_0),p_0)|} < \frac{\lambda_1}{\lambda_0} \tag{C.11}$$

for all $1 \leq i \leq n$.

Proof: We first describe possible choices for $K \geq 1$, $\delta p > 0$, and $\epsilon_0 > 0$. We then show that these choices in fact satisfy (C.11).

Fix $\delta x > 0$ such that

$$D_x^2 f(x, p_0) \neq 0 \text{ if } |x - c_*| < \delta x$$

for any $c_* \in C(f_{p_0})$. Then let:

$$J_x = \{x \in I_x | \; |x - c_*| \geq \delta x \text{ for any } c_* \in C(f_{p_0})\}.$$

Set $M_x = \inf_{x \in J_x} |D_x f(x, p_0)|$ and define:

$$\Delta(a) = \sup_{x \in I_x} \sup_{p \in [p_0 - a, p_0 + a]} |D_p f(x, p) - D_p(x, p_0)|.$$

Now let $K_1 > 0$, $K_2 > 0$, and $K_3 > 0$ be the constants from lemma C.2.1. Choose:

$$K = \frac{2K_1}{K_2(1 - \frac{\lambda_1}{\lambda_0})}. \tag{C.12}$$

Note that since $K_1 \geq K_2$, we know that $K \geq 1$. Choose $\delta p_1 > 0$ such that:

$$\Delta(\delta p_1) < \frac{M_x}{2}(1 - \frac{\lambda_1}{\lambda_0}). \tag{C.13}$$

Let $\delta p_2 = \frac{K_2}{K_3}(1 - \frac{\lambda_1}{\lambda_0})$ and set

$$\delta p = \min\{\delta p_1, \delta p_2\}. \tag{C.14}$$

Finally, fix

$$\epsilon_0 = \min\{\frac{M_x}{2K_1}(1 - \frac{\lambda_1}{\lambda_0}), \frac{\delta x}{K}\}. \tag{C.15}$$

In order to show (C.11) it is sufficient to show:

$$A(i, p, p_0) \leq 1 - \frac{\lambda_1}{\lambda_0} \tag{C.16}$$

where

$$A(i, p, p_0) = \frac{|D_x f(f^i(c, p), p) - D_x f(f^i(c, p_0), p_0)|}{|D_x f(f^i(c, p_0), p_0)|}. \tag{C.17}$$

For each $1 \leq i \leq n$ we now consider two possibilities:

(1) $|f^i(c, p) - c_*| \geq \delta x$ for some $c_* \in C(f_{p_0})$

(2) $K\epsilon \leq |f^i(c, p_0) - c_*| < \delta x$ for some $c_* \in C(f_{p_0})$.

(Note that we know $K\epsilon < \delta x$ from (C.15).)

From now on we assume that $|p-p_0| < \delta p$, $|f^i(c,p)-f^i(c,p_0)| < \epsilon$, and $|f^i(c,p_0) - c_*| > K\epsilon$ for all $c_* \in C(f_{p_0})$ and $1 \leq i \leq n$. We wish to show that (C.16) is true for both cases (1) and (2) above for each $1 \leq i \leq n$.

In case (1) using (C.13), (C.14),(C.15), (C.17), and lemma C.2.1 we have:

$$\begin{aligned} A(i,p,p_0) &\leq \frac{1}{|D_x f(f^i(c,p_0),p_0)|}(|D_x f(f^i(c,p),p) - D_x f(f^i(c,p_0),p)| \\ &\qquad +|D_x f(f^i(c,p_0),p) - D_x f(f^i(c,p_0),p_0)|) \qquad \text{(C.18)} \\ &\leq \frac{K_1|f^i(c,p)-f^i(c,p_0)|}{M_x} + \Delta(|p-p_0|) \\ &< \frac{K_1\epsilon_0}{M_x} + \frac{M_x}{2}(1-\frac{\lambda_1}{\lambda_0}) \\ &< \frac{K_1}{M_x}\frac{M_x}{2K_1}(1-\frac{\lambda_1}{\lambda_0}) + \frac{1}{2}(1-\frac{\lambda_1}{\lambda_0}) \\ &< 1-\frac{\lambda_1}{\lambda_0} \end{aligned}$$

which proves the lemma for case (1).

In case (2), if $K\epsilon \leq |f^i(c,p_0)-c_*| < \delta x$, for some $c_* \in C(f_{p_0})$ then from lemma C.2.1, (C.18), (C.15), and (C.12):

$$\begin{aligned} A(i,p,p_0) &< \frac{K_1|f^i(c,p)-f^i(c,p_0)| + K_3|f^i(c,p_0)-c_*||p-p_0|}{K_2|f^i(c,p_0)-c_*|} \\ &< \frac{K_1\epsilon}{K_2(K\epsilon)} + \frac{K_3|p-p_0|}{K_2} \\ &< \frac{1}{2}(1-\frac{\lambda_1}{\lambda_0}) + \frac{1}{2}(1-\frac{\lambda_1}{\lambda_0}) \\ &< 1-\frac{\lambda_1}{\lambda_0}. \end{aligned}$$

This proves the lemma.

Lemma C.2.3 *Suppose that there exist constants $C > 0$, $N_0 > 0$ and $\lambda_0 > 1$ such that*

$$|D_p f^i(c,p_0)| > C\lambda_0^i \qquad \text{(C.19)}$$

for all $i \geq N_0$ where $p_0 \in I_p$. Suppose also that there exists $\delta p > 0$ and $\lambda_1 \in (1,\lambda_0)$ such that for some $n \geq N_0$:

$$\frac{|D_x f(f^i(c),p)|}{|D_x f(f^i(c),p_0)|} > \frac{\lambda_1}{\lambda_0} \qquad \text{(C.20)}$$

for all $1 \leq i \leq n$ *if* $|p - p_0| < \delta p$. *Then for any* $\lambda_2 \in (1, \lambda_1)$, *there exists* $N_1 > 0$ *(independent of* n *and* δp*) and* $\delta p_1 > 0$ *(independent of* n*) such that*

$$|D_p f^i(c,p)| > C\lambda_2^i$$

for all $N_1 \leq i \leq n+1$ *if* $|p - p_0| < \delta p_1$.

Proof: Given $\lambda_0 > 1$, fix $1 < \lambda_2 < \lambda_1 < \lambda_0$. Set $M_p = \sup_{x \in I_x} |D_p f(x, p_0)|$ and define:

$$z(i) = (\frac{\lambda_2}{\lambda_0} - 1)C_0\lambda_2^{i+1} - M_p\frac{\lambda_1}{\lambda_2}(\frac{\lambda_2}{\lambda_0})^{i+1} - 2M_p \tag{C.21}$$

It is apparent that $z(i) \to \infty$ as $i \to \infty$. Thus, it is possible to choose $N_2 > 0$ (independent of n and δp) so that $z(i) > K_0|I_p|$ for all $i \geq N_2$ where $K_0 > 0$ is the constant from lemma C.2.1 such that:

$$|D_p f(x,p) - D_p f(x,p_0)| < K_0|p - p_0|$$

for any $x \in I_x$ and $p \in I_p$. Let $N_1 = \max\{N_0, N_2\}$.

We now prove the lemma by induction on i for $N_1 \leq i \leq n$. From (C.19), and since $|D_p f^i(c,p)|$ is continuous with respect to p, there exists $\delta p_2 > 0$ such that

$$|D_p f^{N_1}(c,p)| > C\lambda_1^{N_1} \tag{C.22}$$

if $|p - p_0| < \delta p_2$. Set $\delta p_1 = \min\{\delta p, \delta p_2\}$. Thus, since $\delta p_1 > 0$ is independent of n, to prove the lemma it is sufficient to show that:

$$\frac{|D_p f^i(c,p)|}{|D_p f^i(c,p_0)|} > (\frac{\lambda_2}{\lambda_0})^i \tag{C.23}$$

implies

$$\frac{|D_p f^{i+1}(c,p)|}{|D_p f^{i+1}(c,p_0)|} > (\frac{\lambda_2}{\lambda_0})^{i+1}.$$

for any $|p - p_0| < \delta p_1$ if $N_1 \leq i \leq n$.

Let $E = \frac{|D_p f^{i+1}(c,p)|}{|D_p f^{i+1}(c,p_0)|}$ and let $A = |D_x f(f^i(c,p_0),p_0)D_p^i(c,p_0)|$. Then, expanding by the chain rule:

$$\begin{aligned} E = & \frac{|D_p f^{i+1}(c,p)|}{|D_p f^{i+1}(c,p_0)|} \\ > & \frac{|D_x f(f^i(c,p),p)D_p f^i(c,p)| - |D_p f(f^i(c,p),p)|}{|D_x f(f^i(c,p_0),p_0)D_p f^i(c,p_0)| + |D_p f(f^i(c,p_0),p_0)|} \end{aligned} \tag{C.24}$$

Using (C.20) and (C.23):

$$\begin{aligned}
&|D_x f(f^i(c,p),p) D_p f^i(c,p)| \\
&= \frac{\lambda_1}{\lambda_0}|D_x f(f^i(c,p_0),p_0)|(\frac{\lambda_2}{\lambda_0})^i |D_p f^i(c,p_0)| \\
&= (\frac{\lambda_2}{\lambda_0})^{i+1}\frac{\lambda_1}{\lambda_2}A
\end{aligned} \tag{C.25}$$

Also, we know for lemma C.2.1 that there exists $K_0 > 0$ such that:

$$\begin{aligned}
&|D_p f(f^i(c,p),p)| \\
&\leq |D_p f(f^i(c,p),p) - D_p f(f^i(c,p),p_0)| + |D_p f(f^i(c,p),p_0) - D_p f(f^i(c,p_0),p_0)| \\
&\leq K_0|p-p_0| + 2M_p
\end{aligned} \tag{C.26}$$

Thus, substituting (C.25) and (C.26) into (C.24):

$$\begin{aligned}
E &> \frac{(\frac{\lambda_2}{\lambda_0})^{i+1}\frac{\lambda_1}{\lambda_2}A - (K_0|p-p_0| + 2M_p)}{A+M_p} \\
&> (\frac{\lambda_2}{\lambda_0})^{i+1} + \frac{(\frac{\lambda_2}{\lambda_0})^{i+1}(\frac{\lambda_1}{\lambda_2}-1)A - (K_0|p-p_0|+2M_p) - M_p(\frac{\lambda_2}{\lambda_0})^{i+1}}{A+M_p}.
\end{aligned} \tag{C.27}$$

Since $|D_p f^{i+1}(c,p_0)| < A + M_p$ and from (C.19) we have that

$$A > C_0\lambda_0^{i+1} - M_p \tag{C.28}$$

Substituting (C.28) into (C.27) and from (C.21) we have:

$$\begin{aligned}
E &> (\frac{\lambda_2}{\lambda_0})^{i+1} + \frac{(\frac{\lambda_2}{\lambda_0}-1)C_0\lambda_2^{i+1} - M_p\frac{\lambda_1}{\lambda_2}(\frac{\lambda_2}{\lambda_0})^{i+1} - 2M_p - K_0|p-p_0|}{A+M_p} \\
&> (\frac{\lambda_2}{\lambda_0})^{i+1} + \frac{z(i) - K_0|p-p_0|}{A+M_p}
\end{aligned}$$

Since $z(i) > K_0|p-p_0|$, for $i \geq N_1$, we have that:

$$E > (\frac{\lambda_2}{\lambda_0})^{i+1},$$

if $N_1 \leq i \leq n$ which proves the lemma.

Lemma C.2.4 *Let $\{f_p : I_x \to I_x | p \in I_p\}$ be a one-parameter family of mappings satisfying (C0) and (C1). Suppose that f_{p_0} satisfies (CE1) and (CP1) for $p_0 \in I_p$ and some $c \in C(f_{p_0})$. Then there exist constants $\epsilon_0 > 0$, $K \geq 1$, $N_1 > 0$, $\lambda > 1$, and $\delta p > 0$ such that for any positive $\epsilon < \epsilon_0$, if $p \in B(p_0, \delta p)$ then for any $n < n_e(c, \epsilon, p_0)$ the following two conditions are true:*

(1) If $|f^i(c,p) - f^i(c,p_0)| < \epsilon$ for every $1 \le i \le n$, then

$$|D_p f^j(c,p)| > C\lambda^j$$

for any $N_1 \le j \le n+1$.

(2)

$$\max_{N_1 \le i \le n} |f^i(c,p) - f^i(c,p_0)| \ge \min\{\epsilon, C\lambda^i |p - p_0|\}.$$

Proof: If $f(x,p_0)$ for $c \in C(f_{p_0})$ then there exists $C > 0$, $N_0 > 0$, and $\lambda_0 > 0$ such that:

$$|D_p f^i(c,p_0)| > C\lambda_0^i$$

for all $i \ge N_0$. Choose λ and λ_1 such that $1 < \lambda < \lambda_1 < \lambda_0$. Then from lemma C.2.2 we know that there exists $K \ge 1$, $\delta p_1 > 0$, and $\epsilon_1 > 0$ such that for any $\epsilon < \epsilon_1$, if $p \in B(p_0, \delta p_1)$, $n < n_e(c, K\epsilon, p_0)$, and $|f^i(c,p) - f^i(c,p_0)| < \epsilon$ for $1 \le i \le n$, then:

$$\frac{|D_x(f^i(c,p),p)|}{|D_x(f^i(c,p_0),p_0)|} < \frac{\lambda_1}{\lambda_0}$$

for any $1 \le i \le n$. From lemma C.2.3, this implies that there exists $\epsilon_0 > 0$, $\delta p_2 > 0$, and $N_1 > 0$ such that for any $\epsilon < \epsilon_0$, if $p \in B(p_0, \delta p_2)$ and $|f^i(c,p) - f^i(c,p_0)| < \epsilon$ for $1 \le i \le n$, then:

$$|Df^j(c,p)| > C\lambda^j \tag{C.29}$$

for any j satisfying $N_1 \le j \le n+1$, provided that $n < n_e(c, K\epsilon, p_0)$. This proves part (1) of the lemma. It also implies that

$$|f^i(c,p) - f^i(c,p_0)| \ge C\lambda^i |p - p_0| \tag{C.30}$$

for any $N_1 \le i \le n+1$ if $n < n_e(c, K\epsilon, p_0)$.

Now define:

$$g(p) = \max_{1 \le i \le N_1} |f^i(c,p) - f^i(c,p_0)|$$

for any $p \in I_p$. Since $f(x,p)$ is C^2 and $|D_p f^{N_1}(c,p_0)| > C\lambda_0^{N_1}$, there exists $\delta p_3 > 0$ such that $g(p)$ is monotonically increasing in the interval $[p_0, p_0 + \delta p_3]$ and monotonically decreasing in the interval $[p_0 - \delta p_3, p_0]$. Choose $\delta p = min\{\delta p_2, \delta p_3\}$.

Now fix $\epsilon < \epsilon_0$. For each $n > 0$, define J_n to be the largest connected interval such that $p \in J_n$ implies that $|f^i(c,p) - f^i(c,p_0)| < \epsilon$ for $1 \le i \le n$, $p_0 \in J_n$, and $J_n \subset B(p_0, \delta p)$. In order to prove part (2) of the lemma it is sufficient to show that for any $p \in B(p_0, \delta p)$ if $N_1 \le n \le n_e(c, K\epsilon, p_0)$, then either (a) $p \in J_n$ which implies $|f^i(c,p) - f^i(c,p_0)| \ge C\lambda^i$

for all $N_0 \leq i \leq n$ or (b) $p \notin J_n$ which implies that $|f^i(c,p) - f^i(c,p_0)| \geq \epsilon$ for some $N_1 \leq i \leq n$. Case (a) has already been proved above (see (C.30)). We now prove case (b).

First of all note that by our choice of δp and J_n, if $p \in B(p_0, \delta p)$, then either $p \in J_{N_1}$ or $|f^i(c,p) - f^i(c,p_0)| \geq \epsilon$ for some $1 \leq i \leq N_1$. Now fix $p_1 \in B(p_0, \delta)$ and suppose that $p_1 \notin J_n$, for some n satisfying $N_1 \leq n \leq n_e(c, K\epsilon, p_0)$. Then, since $J_i \supset J_{i+1}$ for all $i \geq N_1$, we know that if there exists $k < n$ such that $p_1 \in J_k \setminus J_{k+1}$ where $N_1 \leq k < n_e(c, K\epsilon, p_0)$. But for any $p \in J_k$ we know (see (C.29)) that $|Df^{k+1}(c,p)| > C\lambda^{k+1}$. Thus $(f^{k+1}(c,p) - f^{k+1}(c,p_0))$ must be monotone with for all $p \in J_k$. Consequently if $p_1 \in J_k \setminus J_{k+1}$ then $|f^{k+1}(c,p_1) - f^{k+1}(c,p_0)| \geq \epsilon$ where $N_1 \leq k < n_e(c, K\epsilon, p_0)$. This proves the lemma.

Lemma C.2.5 *Let $\{f_p : I_x \to I_x | p \in I_p\}$ be a one-parameter family of mappings satisfying (C0) and (C1). Suppose that f_{p_0} satisfies (CE1) for some $p_0 \in I_p$ and $c \in C(f_{p_0})$. For any $p \in I_p$ and $n \geq 0$ define:*

$$V_n(p,\epsilon) = \{x \in I_x | \ |f^i(x,p) - f^i(c,p_0)| \leq \epsilon, \ \textit{for all } 0 \leq i \leq n\}$$

Then there exists $\epsilon_0 > 0$ such that for any positive $\epsilon < \epsilon_0$, and any $1 \leq n \leq n_e(c, \epsilon, p_0)$:

$$\sup_{x \in V_n(p,\epsilon)} \{\sigma_n(c,p_0) f^n(x,p)\} \leq \sigma_n(c,p_0) f^n(c,p). \tag{C.31}$$

Proof: Proof by induction. Suppose that the elements of $C(f_{p_0})$ are $c_1 < c_2 < \ldots < c_m$, for some $m \geq 1$. Assume that

$$\epsilon_0 < \min_{i \in \{1,2,\ldots,m-1\}} |c_{i+1} - c_i|$$

In this case, (C.31) clearly holds for $n = 1$ since $\sigma_1(c,p_0) = 1$ implies that c is relative maximum of f_{p_0} and $\sigma_1(c,p_0) = -1$ implies that c is relative minimum of f_{p_0}. Now assuming that (C.31) holds for some $n = k$ where $1 \leq k < n_e(c, \epsilon, p_0)$, we need to show that (C.31) holds for $n = k+1$.

Since $k < n_\epsilon(\epsilon)$, $|f^k(c,p_0) - c_i| > \epsilon$ for any $i \in \{1,2,\ldots,m\}$. Consequently, since $|f^k(x,p) - f^k(c,p_0)| \leq \epsilon$ for any $x \in V_k(p,\epsilon)$, we see that there exists $i \in \{1,2,\ldots,m-1\}$ such that $c_i < x < c_{i+1}$ for every $x \in V_k(p,\epsilon)$. In other words, all elements of $V_k(p,\epsilon)$ must lie on one monotone branch of f_p and:

$$sgn\{Df(f^k(x,p),p)\} = sgn\{Df(f^k(c,p_0),p_0)\} \tag{C.32}$$

for all $x \in V_k(p,\epsilon)$.

From our specification of $\sigma_k(c,p_0)$ we have that:

$$\sigma_{k+1}(c,p_0) = sgn\{Df(f^k(c,p_0),p_0)\}\sigma_k(c,p_0). \tag{C.33}$$

We can consider four cases: $sgn\{Df(f^k(c,p_0),p_0)\} = \pm 1$ and $\sigma_k(c,p_0) = \pm 1$. Suppose that $\sigma_k(c,p_0) = 1$. By assumption, if $\sigma_k(c,p_0) = 1$, then

$$\sup_{x \in V_n(p,\epsilon)} f^n(x,p) \le f^n(c,p). \tag{C.34}$$

Thus, if $sgn\{Df(f^k(c,p_0),p_0)\} = 1$, then, from (C.33), $\sigma_{k+1}(c,p_0) = 1$. Also, from (C.32), we know that $sgn\{Df(f^k(x,p),p)\} = 1$ for all $x \in V_k(p,\epsilon)$, and we know that all elements of $V_k(p,\epsilon)$ lie on a monotonically increasing branch of f_p. Combining this result with (C.34) implies that:

$$\sup_{x \in V_{k+1}(p,\epsilon)} f^{k+1}(x,p) \le f^{k+1}(c,p).$$

On the other hand, if $sgn\{Df(f^k(c,p_0),p_0)\} = -1$, then $\sigma_{k+1}(c,p_0) = -1$ and

$$\inf_{x \in V_{k+1}(p,\epsilon)} f^{k+1}(x,p) \ge f^{k+1}(c,p).$$

In both cases above we can see that (C.31) is satisfied for $n = k+1$. Similarly we can verify that (C.31) is also satisfied for $n = k+1$ in the two cases where $\sigma_k(c,p_0) = -1$. This proves the lemma.

Proof of theorem 3.3.1:

We are given that f_{p_0} satisfies (CE1) for some $p_0 \in I_p$ and $c \in C(f_{p_0})$. Then, from part (1) of lemma C.2.4, there exist constants $K \ge 1$, $C > 0$, $N_2 > 0$, $\epsilon_0 > 0$, $\delta p > 0$, and $\lambda > 1$ such that for any $\epsilon < \epsilon_0$, if $p \in B(p_0, \delta p)$, and $|f^i(c,p) - f^i(c,p_0)| < \epsilon$ for all i satisfying $1 \le i \le n-1$, then:

$$|D_p f^n(c,p)| > C\lambda^n \tag{C.35}$$

for any n such that $N_2 \le n \le n_e(c, K\epsilon, p_0)$.

Now suppose that there exists $c \in C(f_{p_0})$ that favors higher parameters. Then there exists $N_3 > 0$ such that for any $n \ge N_3$:

$$sgn\{D_p f^n(c,p_0)\} = \sigma_n(c,p_0). \tag{C.36}$$

Set $N_1 = max\{N_2, N_3\}$. From (C.35) and since f is C^2 it is clear that $D_p f^n(c,p)$ can not change signs for any $p \in B(p_0,\delta p)$ if $N_2 \le n \le n_e(c, K\epsilon, p_0)$. Consequently, from (C.36) we have that:

$$sgn\{D_p f^n(c,p)\} = \sigma_n(c,p_0)$$

for any $N_1 \le n \le n_e(c, K\epsilon, p_0)$ if $p \in B(p_0,\delta p)$ and $|f^i(c,p) - f^i(c,p_0)| < \epsilon$ for $1 \le i \le n-1$. In this case:

$$sgn\{f^n(c,p) - f^n(c,p_0)\} = \sigma_n(c,p_0) sgn\{p - p_0\}. \tag{C.37}$$

Now suppose that $p < p_0$. Then from (C.37) if $\sigma_n(c, p_0) = 1$, then $f^n(c,p) \leq f^n(c,p_0)$ and if $\sigma_n(c,p_0) = -1$, then $f^n(c,p) \geq f^n(c,p_0)$ for any $p \in B(p_0, \delta p)$ such that $|f^i(c,p) - f^i(c,p_0)| < \epsilon$ for $1 \leq i \leq n-1$, provided that $N_1 \leq n \leq n_e(c, K\epsilon, p_0)$. Combining this result with lemma C.2.5 we find that:

$$\begin{aligned}\sup_{x \in V_n(p,\epsilon)} f^n(x,p) &\leq f^n(c,p_0) \text{ if } \sigma_n(c,p_0) = 1 \\ \inf_{x \in V_n(p,\epsilon)} f^n(x,p) &\geq f^n(c,p_0) \text{ if } \sigma_n(c,p_0) = -1\end{aligned}$$

which implies that

$$\inf_{x \in V_n(p,\epsilon)} |f^n(x,p) - f^n(c,p_0)| \geq |f^n(c,p) - f^n(c,p_0)| \tag{C.38}$$

for any $p \in [p_0 - \delta p, p_0]$, if $N_1 \leq n \leq n_e(c, K\epsilon, p_0)$ (where $V_n(p,\epsilon)$ is as defined in the statement of lemma C.2.5).

Finally, from lemma C.2.4 we also know that

$$\max_{N_1 \leq i \leq n} |f^i(c,p) - f^i(c,p_0)| \geq \min\{\epsilon, C\lambda^i|p - p_0|\}. \tag{C.39}$$

if $N_1 \leq n \leq n_e(c, K\epsilon, p_0)$ and $p \in B(p_0, \delta p)$. Combining (C.38) and (C.39) we find that:

$$\inf_{x \in V_n(p,\epsilon)} |f^n(x,p) - f^n(c,p_0)| \geq \min\{\epsilon, C\lambda^i|p - p_0|\}. \tag{C.40}$$

if $N_1 \leq n \leq n_e(c, K\epsilon, p_0)$ and $p \in [p_0 - \delta p, p_0]$. Clearly the orbit $\{f^i(c,p_0)\}_{i=0}^{\infty}$ cannot be $\epsilon-$shadowed by an orbit of f_p if

$$\inf_{x \in V_n(p,\epsilon)} |f^n(x,p) - f^n(c,p_0)| > \epsilon \tag{C.41}$$

for any finite value of n. Consequently from (C.40) and (C.41) we see that for any $\epsilon < \epsilon_0$, the orbit, $\{f^i(c,p_0)\}_{i=0}^{\infty}$, cannot be ϵ-shadowed by f_p if

$$|p - p_0| > \frac{1}{C}\epsilon\lambda^{-n_\epsilon(K\epsilon)} \tag{C.42}$$

and $p \in [p_0 - \delta p, p_0]$. Setting $K' = \frac{1}{C}$, this proves the theorem.

Appendix D

Proof of theorem 3.3.2

This appendix contains the proof for theorem 3.3.2. I have made an effort to make the appendix as self-contained as possible, so that the reader should be able to find most of the relevant definitions and explanations in this appendix. Naturally, this means that the appendix repeats some material found elsewhere in this report.

D.1 Definitions and statement of theorem

Definition: Suppose that $g : I \to I$ is C^3 and $I \subset \mathbb{R}$. Then the *Schwarzian derivative*, Sg, of g is given by the following:

$$Sg(x) = \frac{g'''(x)}{g'(x)} - \frac{3}{2}(\frac{g''(x)}{g'(x)})^2.$$

where $g'(x), g''(x), g'''(x)$ here indicate the first, second, and third derivatives of x.

In this section we will primarily restrict ourselves to mappings with the following properties:

(A0) $g : I \to I$, is $C^3(I)$ where $I = [0, 1]$, with $g(0) = 0$ and $g(1) = 0$.

(A1) g has one local maximum at $x = c$; g is strictly increasing on $[0, c]$ and strictly decreasing on $[c, 1]$;

(A2) $g''(c) < 0$, $|g'(0)| > 1$.

(A3) The Schwarzian derivative of g is negative, $Sg(x) < 0$, over all $x \in I$ (we allow $Sg(x) = -\infty$).

Under the Collet-Eckmann conditions, there exist constants $K_E > 0$ and $\lambda_E > 1$ such that for some $c \in C(g)$:

(CE1) $|Dg^n(g(c))| > K_E\lambda_E^n$

(CE2) $|Dg^n(z)| > K_E\lambda_E^n$ if $g^n(z) = c$.

for any $n > 0$.

We will be investigating one-parameter families of mappings, $f : I_x \times I_p \to I_x$, where p is the parameter and $I_x, I_p \subset \mathbb{R}$ are closed intervals. Let $f_p(x) = f(x,p)$ where $f_p : I_x \to I_x$. We are primarily be interested in one-parameter families of maps with the following characteristics:

(B0) For each $p \in I_p$, $f_p : I_x \to I_x$ satisfies (A0), (A1), (A2), and (A3) where $I_x = [0,1]$. For each p, we also require that f_p has a turning point at c, where c is constant with respect to p.

(B1) $f : I_x \times I_p \to I_x$ is C^2 for all $(x,p) \in I_x \times I_p$.

Another concept we shall need is that of the *kneading invariant*. Kneading invariants and many associated topics are discussed in Milnor and Thurston [34].

Definition: If $g : I \to I$ is a piecewise monotone map with exactly one turning point at c, then the *kneading invariant*, $D(g,t)$, of g is defined as follows:

$$D(g,t) = 1 + \theta_1(g)t + \theta_2(g)t + \ldots + \theta_n(g)t^n + \ldots$$

where

$$\begin{aligned} \theta_n(g) &= \epsilon_1(g)\epsilon_2(g)\ldots\epsilon_n(g) \\ \epsilon_n(g) &= \lim_{x\to c^+} sgn(Dg(g^n(x))) \end{aligned}$$

for $n \geq 1$. If c is a relative maximum of g, then one interpretation of $\theta_n(g)$ is that it represents whether g^{n+1} has a relative maximum ($\theta_n(g) = +1$) or minimum ($\theta_n(g) = -1$) at c.

We can also order these kneading invariants in the following way. We will say that $|D(g,t)| < |D(h,t)|$ if $\theta_i(g) = \theta_i(h)$, for $1 \leq i < n$, but $\theta_n(g) < \theta_n(h)$. A kneading invariant, $D(f_p,t)$, is said to be monotonically decreasing with respect to p if $p_1 > p_0$ implies $|D(f_{p_1},t)| \leq |D(f_{p_0},t)|$.

We are now ready to state the main result of this appendix:

Theorem 3.3.2 *Let $\{f_p : I_x \to I_x | p \in I_p\}$ be a one-parameter family of mappings satisfying (B0) and (B1). Suppose that $p_0 \in int(I_p)$* [1] *such that f_{p_0} satisfies (CE1).*

[1] Henceforth, if $A \subset \mathbb{R}$, let int(A) denote the interior of A.

Also, suppose that the kneading invariant, $D(f_p, t)$, is monotonically decreasing with respect to p in some neighborhood of $p = p_0$. Then there exists $\delta p > 0$ and $C > 0$ such that for every $x_0 \in I_x$ there is a set, $W(x_0) \subset I_x \times I_p$, satisfying the following conditions:

(1) $W(x_0) = \{(\alpha_{x_0}(t), \beta_{x_0}(t)) | t \in [0,1]\}$ *where* $\alpha_{x_0} : [0,1] \to I_x$ *and* $\beta_{x_0} : [0,1] \to I_p$ *are continuous and* $\beta_{x_0}(t)$ *is monotonically increasing with respect to* t *with* $\beta_{x_0}(0) = p_0$ *and* $\beta_{x_0}(1) = p_0 + \delta p$.

(2) For any $x_0 \in I_x$, *if* $(x,p) \in W(x_0)$ *then* $|f^n(x,p) - f^n(x_0,p_0)| < C(p-p_0)^{\frac{1}{3}}$ *for all* $n \geq 0$.

D.2 Tools for maps with negative Schwarzian derivative

There has been a significant amount of interest in recent years into one-dimensional maps, particularly maps with negative Schwarzian derivative. Below we state some useful properties and analytical tools that have been developed to analyze these maps. For the most part, the results are only stated here, and references provided to appropriate proofs. We do not attempt to trace the history of the development of these results.

The only results in this section that are new are contained in lemmas D.2.11, D.2.12, and D.2.13.

Lemma D.2.1 *If* g *satisfies (A0), (A1), and (A2) then there exist constants* $K_0 > 0$, *and* $K_1 > 0$ *such that for all* $x \in I$:

(1) $K_0|x-c| < |Dg(x)| < K_1|x-c|$

(2) $\frac{1}{2}K_0|x-c|^2 < |g(x) - g(c)| < \frac{1}{2}K_1|x-c|^2$

Proof: This is clear, since $g''(c) \neq 0$.

Lemma D.2.2 *If* $f(x,p)$ *satisfies (B0) and (B1), then there exist constants* $K_0 > 0$, *and* $K_1 > 0$ *such that for any* $x \in I_x$, $y \in I_x$, $p_0 \in I_p$, *and* $p_1 \in I_p$:

(1) $|D_x f(x,p_0) - D_x f(y,p_0)| < K_0|x-y|$

(2) $|D_x f(x,p_0) - D_x f(x,p_1)| < K_1|p_0 - p_1|$

Proof: This is clear, since $f(x,p)$ is C^2 and $I_x \times I_p$ is compact.

Lemma D.2.3 *(Minimum Principle). Suppose that g has negative Schwarzian derivative. Let $J = [x_0, x_1]$ be an interval on which g is monotone. Then*

$$|Dg(x)| \geq min\{|Df(x_0)|, |Df(x_1)|\}$$

for all $x \in J$.

Proof: See, for example, page 154 of [33].

Definition: Given map $g : I \to I$, we say that x is in the basin of attraction of an orbit, $\{y_i\}_{i=0}^{\infty}$, of g if there exists an $m \geq 0$ such that $lim_{i\to\infty}(g^{i+m}(x) - y_i) = 0$.

Lemma D.2.4 *(Singer) If $g : I \to I$ is C^3 and has negative Schwarzian derivative, then the basin of attraction of any stable periodic orbit contains either a critical point or one of the boundary points of I.*

Proof: See Singer [58].

Definition D.2.1 *We will say that a piecewise monotone map, $g : I \to I$, has a sink if there exists an interval $J \subset I$ such that that g is monotone on J^n and $g^n(J) \subset J$ for some $n > 0$.*

Lemma D.2.5 *If $g : I \to I$ satisfies (A0), (A1), (A2), (A3), and (CE1). Then g has no sinks.*

Proof: It is relatively simple to show that the existence of such a sink implies the existence of a stable periodic point (see for example Collet and Eckmann [14], lemma II.5.1). From Singer's theorem, we know that $g : [0, 1] \to [0, 1]$ does not have a stable periodic orbit unless $x = 0$, $x = c$, or $x = 1$ is in the basin of attraction of that periodic orbit. From (CE1) we know that the critical point does not tend to a stable orbit and from (A2) we know that $x = 0$ and $x = 1$ do not tend to a stable periodic orbit. Thus g has no sinks.

Lemma D.2.6 *(Koebe Inequality). Suppose that $g : I \to I$ has negative Schwarzian derivative. Let $T = [a, b]$ be an interval on which g is a diffeomorphism. Given $x \in T$, let L and R be the components of $T \setminus \{x\}$. If there exists $\tau > 0$ such that:*

$$\frac{|g(L)|}{|g(T)|} \geq \tau \text{ and } \frac{|g(R)|}{|g(T)|} \geq \tau$$

then there exists $K(\tau) > 0$ such that:

$$|Dg(x)| \geq K(\tau) \sup_{z\in T} |Dg(z)|$$

where $K(\tau)$ depends only on τ.

Proof: See, for example, theorem 3.2 in van Strien [60].

Lemma D.2.7 *Let $g : I \to I$ satisfy (A0), (A1), (A2), (A3) and (CE1). Then g satisfies (CE2).*

Proof: See Nowicki [44].

Lemma D.2.8 *Let $g : I \to I$ satisfy (A0), (A1), (A2), (A3) and (CE1). There exists $K > 0$ and $\lambda_1 > 1$ such that for any $n > 0$, if $g^n(x) = c$ then $|x - c| > K\lambda_1^{-n}$.*

Proof: From lemma D.2.1, we know there exists $K_0 > 0$ such that $|Dg(x)| < K_0|x-c|$ for any $x \in I$. Now set $a = \sup_{x\in I} |Dg(x)|$. Then we have:

$$|Dg^n(x)| \leq a^{n-1}K_0|x - c|$$

However, by lemma D.2.7, we also know that g satisfies (CE2), so that $Dg^n(x) > K_E\lambda^n$ for some constants $K_E > 0$ and $\lambda > 1$. Thus $a^{n-1}K_0|x - c| < K_E\lambda^n$ which implies that $|x - c| < \frac{aK_E}{K_0}(\frac{\lambda}{a})^n$. This proves the lemma if we set $K = \frac{aK_E}{K_0}$ and $\lambda_1 = (\frac{\lambda}{a})$.

Lemma D.2.9 *Let $g : I \to I$ satisfy (A0), (A1), (A2), (A3) and (CE1). Let $J_n \subset of I$ be any interval such that g^n is monotone on J_n. Then there exist constants $K > 0$ and $\lambda_2 > 1$ such that for any $n \geq 0$:*

$$|J_n| < K\lambda_2^{-n}$$

Proof: See Nowicki [44].

Lemma D.2.10 *Let $g : I \to I$ satisfy (A0), (A1), (A2), (A3) and (CE1). Suppose that g^n is monotone on $J = [a, b]$ where $J \subset I$ and $g^n(a) = c$ for some $n \geq 0$. Then there exist a constant, $K > 0$, such that for any $n \geq 0$:*

$$\frac{|g^n(J)|}{|J|} \geq K$$

Proof: See lemma 6.2 in Nowicki [45].

Lemma D.2.11 *Suppose that $g : I \to I$ satisfies (A0), (A1), (A2), (A3), and (CE1). Let $x \in I$ such that $|g^i(x) - c| > \epsilon$ for $0 \leq i < n$. Then, for any $\epsilon > 0$ there exist constants $C > 0$ and $\lambda > 1$ (independent of x) such that:*

$$|Dg^i(x)| > C\epsilon^2\lambda^i$$

for $0 \leq i \leq n$.

Proof: For any $i \geq 0$, let $\Delta_i(x)$ be the maximal interval such that $x \in \Delta_i(x)$ and g^i is monotone on $\Delta_i(x)$. The proof of the lemma is based on the following claim:

Claim: Let $x \in I$, and suppose that there exists $b \in \Delta_n(x)$ such that $g^n(b) = c$ for some $n \geq 0$. If $|g^i(x) - c| > \epsilon$ for $0 \leq i \leq n$, then there exist $C_0 > 0$ and $\lambda > 1$ (independent of x) such that:

$$|Dg^{n+1}(x)| > C_0 \epsilon^2 \lambda^{n+1}.$$

We shall now describe the proof of the lemma using this claim, leaving the proof of the claim for later.

Fix $x \in I$ and $i \leq n$. Suppose that $\Delta_i(x) = [a, a']$ and let $x_i = f^i(x)$, $a_i = f^i(a)$, and $a'_i = f^i(a')$. For definiteness, assume that $|x_i - a_i| < |a'_i - x_i|$ (the other case is analogous). Since $\Delta_i(x)$ is maximal, each endpoint of $\Delta_i(x)$ must map either into (1) the critical point, or (2) into the boundary of I. If case (2) is true, there must exists $k < i$ such that $g^k(a) = 0$, or $g^k(a) = 1$ (since $I = [0,1]$ by (A2)). This means either $a = 0$, $a = 1$ or $g^j(a) = c$ for some $j < k$. If $g^j(a) = c$ then case (1) is also satisfied. Otherwise, if $a = 0$ or $a = 1$, then $f^i(\Delta_i(x)) \cap \{c\} \neq \emptyset$, and the lemma may be proved by a direct application of the claim described above.

Otherwise, if case (1) is true, there must exist $k < i$ such that $g^k(a) = c$. By (CE1), we know there exist constants, $K_E > 0$ and $\lambda_E > 1$ (independent of i and k) such that:

$$|Dg^{i-k-1}(g^{k+1}(a))| > K_E \lambda_E^{i-k-1} \tag{D.1}$$

Now set $y \in [a, a']$ so that $y_i = g^i(y) = \frac{1}{2}(a_i + a'_i)$. By the Koebe Inequality, since $|y_k - a_k| < |a'_k - y_k|$, there exists $K_0 = K(\tau = \frac{1}{2}) > 0$ such that:

$$|Dg^{i-k-1}(g^{k+1}(y))| > K_0 |Dg^{i-k-1}(g^{k+1}(a))|$$

Combining this with (D.1) we have:

$$|Dg^{i-k-1}(g^{k+1}(y))| > K_0 K_E \lambda_E^{i-k-1} \tag{D.2}$$

Also, since $|x_i - a_i| < |a'_i - x_i|$, we know $x_i \in [a_i; y_i]$ (where $[a; b]$ means either $[a, b]$ or $[b, a]$ whichever is appropriate). Thus by using the minimum principle with (D.1) and (D.2) we find that there exists $K_1 > 0$ such that:

$$|Dg^{i-k-1}(g^{k+1}(x))| > K_1 \lambda_E^{i-k-1}. \tag{D.3}$$

We are now ready to apply the claim. It is clear that $a \in \Delta_k(x)$. Since $g^k(a) = c$, the claim implies that there exists $C_0 > 0$, and $\lambda_0 > 1$ such that:

$$|Dg^{k+1}(x)| > C_0 \epsilon^2 \lambda_0^{k+1} \tag{D.4}$$

Combining (D.3) and (D.4) we find that there exists $C > 0$, and $\lambda > 1$ such that:

$$|Dg^i(x)| = |Dg^{i-k-1}(g^{k+1}(x))||Dg^{k+1}(x)| > C\epsilon^2\lambda^i.$$

This proves the lemma, except for the proof of the claim, which we describe below.

Proof of Claim: Let $\Delta_n(x) = [a, a']$. If $b = a$ or $b = a'$ then the proof is trivial since g satisfies (CE2) from lemma D.2.7. So suppose that $b \in (a, a')$. For definiteness suppose that $x < b$ so that $x \in [a, b]$ (the other case is analogous). As before, since $\Delta_n(x)$ is maximal, the endpoints of $\Delta_n(x)$ must map either into the critical point, or into the boundary of I. Let us address the critical point case now, and come back to the other case at the end of the proof.

Assume that there exists $k \leq n$ such that $g^k(a) = c$. Let $a_k = g^k(a)$ and $b_k = g^k(b)$ and let $y \in [a, b]$ such that $y_k = g^k(y) = \frac{1}{2}(a_k + b_k)$. By the Koebe Inequality we know that there exists $K_2 = K(\tau = \frac{1}{2})$ such that $|Dg^k(y)| > K_2|Dg^k(a)|$. Also, since g satisfies (CE2), there exists $K_E > 0$ and $\lambda_E > 1$ such that:

$$|Dg^k(a)| > K_E\lambda_E^k. \tag{D.5}$$

Combining the last two statements, we find that

$$|Dg^k(y)| > K_2K_E\lambda_E^k. \tag{D.6}$$

Now let $y' \in [a, b]$ so that $y'_k = g^k(y') = a_k + \frac{1}{2}sgn(b_k - a_k)\epsilon) \in [a_k; b_k]$. Since $x_k = g^k(x) \in [a_k; b_k]$, we know $|x_k - a_k| = |x_k - c| > \epsilon$. Consequently $|b_k - a_k| > \epsilon$ which implies $|y_k - a_k| > \frac{1}{2}\epsilon$. Thus, since $|y'_k - a_k| = \frac{1}{2}\epsilon$, we have $y'_k \in [a_k; y_k]$.

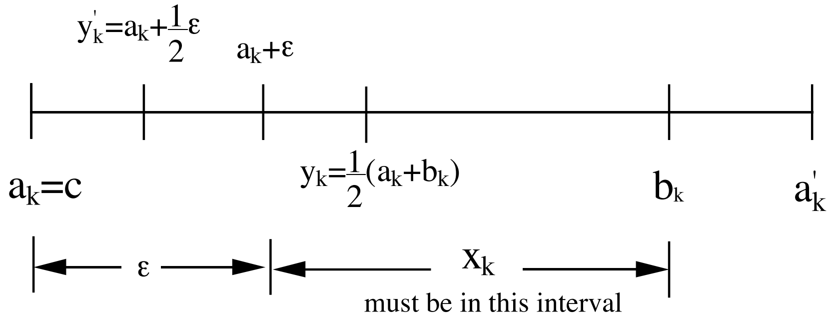

Figure D.1: The interval $g^k(\Delta(x)) = [a_k, a'_k]$ and associated variables are shown. The figure is drawn assuming that $a'_k > a_k$, $b \in (a, a')$, and that $x \in [a, b]$.

Applying the minimum principle to this interval and using (D.5) and (D.6), we find that there exists $K_3 > 0$ such that:

$$|Dg^k(y')| > K_3 \lambda_E^k. \tag{D.7}$$

Also, for any $\epsilon > 0$, we know from lemma D.2.1 that there exists $K_4 > 0$ such that

$$|Dg(y'_k)| > \frac{1}{2} K_4 \epsilon. \tag{D.8}$$

From (D.7) and (D.8) and setting $K_5 = \frac{1}{2} K_3 K_4$, we have:

$$|Dg^{k+1}(y')| > K_5 \epsilon \lambda_E^{k+1}. \tag{D.9}$$

Also, since $g^k(a) = c$, from (CE1) we know that $|Dg^{n-k-1}(g^{k+1}(a))| > K_E \lambda_E^{n-k-1}$. Since $g^n(b) = c$, we know from (CE2) that $|Dg^{n-k-1}(g^{k+1}(b))| > K_E \lambda_E^{n-k-1}$. Thus, by the minimum principle, $|Dg^{n-k-1}(g^{k+1}(y'))| > K_E \lambda_E^{n-k-1}$. Combining this with (D.9) we find:

$$|Dg^n(y')| > K_5 K_E \epsilon \lambda_E^n. \tag{D.10}$$

From (CE2) we also know that

$$|Dg^n(b)| > K_E \lambda_E^n. \tag{D.11}$$

In addition, since $|x_k - a_k| > \epsilon$, we know that $x_k \in [y'_k; b_k]$ so that $x \in [y', b]$. Thus, from the (D.10), (D.11), and the minimum principle, we can conclude that there exists $K_6 > 0$ such that:

$$|Dg^n(x)| > K_6 \epsilon \lambda_E^n.$$

Finally, since $|g^n(x) - c| > \epsilon$, we can use lemma D.2.1 to bound $|Dg(g^n(x))| < K_4 \epsilon$ for $K_4 > 0$. Consequently there exists $C_1 > 0$ such that:

$$|Dg^{n+1}(x)| > C_1 \epsilon^2 \lambda_E^n \tag{D.12}$$

which proves the claim for the case where $g^k(a) = c$ for some $k < n$.

The other possibility is that $g^k(a) \in Bd(I)$ for some $k < n$ where $Bd(I)$ denotes the boundary of I. But this implies that either $a \in Bd(I)$ or possibly that $g^{k-1}(a) = c$. The possibility where $g^{k-1}(a) = c$ has already been covered by the previous case. On the other hand, if $a \in Bd(I)$ then by (A2) there exists $\lambda_0 > 1$ such that $|Dg^n(a)| > \lambda_0^n$. From (CE2) we also know that $|Dg^n(b)| > K_E \lambda_E^n$. Thus, by the minimum principle, there exists $K_7 > 0$ and $\lambda_1 > 0$ such that $|Dg^n(x)| > K_7 \lambda_1^n$ for any $x \in [a, b]$. Then, since $|g^n(x) - c| > \epsilon$ we can use lemma D.2.1 to bound $|Dg(g^n(x))|$ so that there exists $C_2 > 0$ satisfying:

$$|Dg^{n+1}(x)| > C_2 \epsilon \lambda_1^n \tag{D.13}$$

Combining (D.12) and (D.13) shows that we can pick $C > 0$ and $\lambda > 1$ to prove the claim.

Lemma D.2.12 *Let $g : I \to I$ satisfy (A0), (A1), (A2), (A3), and (CE1). Suppose there exists $a \in I$ and $n \geq 0$ such that $g^n(a) = c$. Given any $\alpha > 0$ sufficiently small, either $\min_{0 \leq i < n} |g^i(a) - c| \geq \alpha$ or there exists $b \in I$, $n' \geq 0$, and constants $K > 0$ and $K' > 0$ such that $g^{n'}(b) = c$, $|b - a| < K\alpha$, and $n' < n - K' \log \frac{1}{\alpha}$.*

Proof: Suppose that $\min_{0 \leq i < n} |g^i(a) - c| < \alpha$. Then there exists $m < n$ such that $|g^m(a) - c| < \alpha$ and $|g^i(a) - c| \geq \alpha$ for $0 \leq i < m$.

Since $g^m(y_0)$ approaches close to c, we can bound m away from n using lemma D.2.8:

$$n - m \geq \frac{\log \frac{1}{\alpha}}{\log \lambda_1} \tag{D.14}$$

where $\lambda_1 > 1$ is a constant dependent only on g.

We now consider two possibilities: (1) there exists $b \in I$ such that $g^m(b) = c$ and g^m is monotone on $[a; b]$ or (2) there exists $b \in I$ and $k < m$ such that g^m is monotone on $[a; b]$, $g^k(b) = c$, and $g^m(b) \in [g^m(a); c]$. One of these two cases must be true.

Let $a_i = g^i(a)$ and $b_i = g^i(b)$ for $i > 0$. In the first case, from lemma D.2.10, there exists $K_3 > 0$ such that:

$$|b - a| < \frac{1}{K_3}|b_m - c| < \frac{\alpha}{K_3}. \tag{D.15}$$

Also, from (D.14) we know $m \leq n - \frac{\log \frac{1}{\alpha}}{\log \lambda_1}$. Thus, in this case the lemma is proved if we set $K = \frac{1}{K_3}$, $K' = \frac{1}{\log \lambda_1}$ and $n' = m$.

Now we address the second case. From lemma D.2.1 we know there exists $K_0 > 0$ and $K_1 > 0$ such that $K_0|x - c|^2 \leq |f(x) - f(c)| \leq K_1|x - c|^2$. Thus if we set $K_2 = \frac{K_1}{K_0}$ we see that for any $\delta > 0$ and $\delta^* > K_2\delta$ we have that:

$$g([c \pm \delta; c]) \quad \subset \quad g([c; c \pm \delta^*]) \tag{D.16}$$

where the $\pm$ notation means that the relation holds for all four possible combinations. Also note that since $b_k = c$ and $b_m \in [a_m; c]$ we have:

$$[a_{k+1}; b_{k+1}] \quad = \quad g([a_k; b_k]) = g([a_k; c]) \tag{D.17}$$

$$[a_{m+1}; b_{m+1}] \quad = \quad g([a_m; b_m]) \subset g([a_m; c]). \tag{D.18}$$

We now assert that $|a_k - b_k| < K_2\alpha$. Suppose to the contrary that $|a_k - c| = |a_k - b_k| \geq K_2\alpha > K_2|a_m - c|$. Then, combining this with (D.16), (D.17), and (D.18) implies that:

$$[a_{m+1}; b_{m+1}] \subset [a_{k+1}; b_{k+1}]. \tag{D.19}$$

However, since g satisfies (CE1), it cannot have any sinks (from lemma D.2.5). In particular this means:

$$[a_{m+1}; b_{m+1}] \not\subset [a_{k+1}; b_{k+1}]$$

if $k < m$ since g^{m+1} is monotone on $[a; b]$ if $\alpha > 0$ is sufficiently small. Thus, (D.19) cannot be true so we conclude that:

$$|a_k - b_k| \leq K_2 \alpha.$$

Finally, since $b_k = c$, we can use D.2.10 to show that there exists $K_3 > 0$ such that:

$$|b - a| < \frac{1}{K_3} |a_k - b_k| = \frac{1}{K_3} K_2 \alpha \tag{D.20}$$

Thus combining (D.14) and (D.20) we see that the lemma is satisfied if we set $K = \frac{K_2}{K_3}$, $K' = \frac{1}{\log \lambda_1}$ and $n' = k < m \leq n - \frac{\log \frac{1}{\alpha}}{\log \lambda_1}$.

Thus, combining the results from (D.15) and (D.20), proves the lemma.

Lemma D.2.13 *Suppose $g : I \to I$ satisfies (A0), (A1), (A2), (A3), and (CE1). Then there exists $C > 0$ and $\epsilon_0 > 0$ so that given any positive $\epsilon < \epsilon_0$, and any $x \in I$ such that $x + \epsilon \in I$, then there is a $y \in (x, x + \epsilon)$ such that $N(y, g) < \infty$ and* $\min_{0 \leq i < N(y,g)} |g^i(y) - c| \geq C\epsilon$. *Similarly if $x - \epsilon \in I$, then there exists $y' \in (x - \epsilon, x)$ such that $N(y', g) < \infty$ and* $\min_{0 \leq i < N(y',g)} |g^i(y) - c| \geq C\epsilon$.

Proof: We show the proof for $y \in (x, x + \epsilon)$. The proof for $y' \in (x - \epsilon, x)$ is exactly analagous.

Our plan is to apply lemma D.2.12 as many times as necessary to find an appropriate y to satisfy the lemma. In other words, lemma D.2.12 implies that given any $y_i \in I$ such that $n_i = N(y_i, g) < \infty$ and $\min_{0 \leq i < n_i} |g^i(y_i) - c| \geq \alpha$, then there exists a $y_{i+1} \in I$ such that $|y_{i+1} - y_i| < K\alpha$ and

$$n_{i+1} = N(y_{i+1}, g) < n_i - K' \frac{1}{\alpha} \tag{D.21}$$

for positive constants K and K'. Thus given y_0, we can generate a sequence $\{y_i\}_{i=0}^{i=m}$ in this manner for increasing i until $i = m$ such that

$$\min_{0 \leq i < n_m} |g^i(y_m) - c| \geq \alpha. \tag{D.22}$$

For example, given any $\alpha > 0$, and any $x_0 \in I$ we know from lemma D.2.9 that if $x_0 + \alpha \in I$, then there exists $y_0 \in (x_0, x_0 + \alpha)$ such that $g^{n_0}(y_0) = c$ for some integer satisfying:

$$n_0 \leq \frac{\log \frac{1}{\alpha}}{\log \lambda_2} + 1 \tag{D.23}$$

where $\lambda_2 > 0$ is a constant dependent only on g. If we generate $\{y_i\}_{i=0}^{i=m}$ from the y_0 specified above, then from (D.21) and (D.23) we find that:

$$n_i \leq (\frac{1}{\log \lambda_2} - iK')(\log \frac{1}{\alpha}) + 1 \tag{D.24}$$

for all $0 \leq i \leq m$. Set $M = \frac{1}{K' \log \lambda_2} + 1$. Then for sufficiently small $\alpha > 0$ we find that $m < M$ because otherwise (D.24) would imply that $n_i < 0$ for $i > m$.

So given $x \in I$ and positive $\epsilon < \epsilon_0$ from the statement of the lemma, set $x_0 = x + KM\alpha$ and $\alpha = \frac{1}{2KM+1}\epsilon$. Note that we can choose $\epsilon_0 > 0$ to insure that $\alpha > 0$ is sufficiently small so that the above arguments work. Also, note that since $x_0 + \alpha = x + \frac{KM+1}{2KM+1}\epsilon < x + \epsilon$, if $x + \epsilon \in I$ then $x_0 + \alpha \in I$. From our choice of $y_0 \in (x_0, x_0 + \alpha)$, we also know that since $|y_{i+1} - y_i| < K\alpha$, we have $|y_m - y_0| < Km\alpha$. Consequently $y_m > x + KM\alpha - Km\alpha > x$ and $y_m > x + KM\alpha + \alpha + Km\alpha > x + (2KM+1)\alpha \leq x + \epsilon$. Thus $y_m \in (x, x+\epsilon)$ and from (D.22), we have that $\min_{0 \leq i < n_m} |g^i(y_m) - c| \geq \alpha = C\epsilon$ where $C = \frac{1}{2KM+1}$. Setting $y = y_m$, this proves the lemma.

D.3 Analyzing preimages

In this section we will investigate one-parameter family of mappings, $\{f_p | p \in I_p\}$, that satisfy (B0) and (B1). Our discussion depends on an examination of the preimages of the critical point, $x = c$ in $I_x \times I_p$ space. We first need to introduce some notation in order to describe the relevant concepts.

For the remainder of this section, $\{f_p | p \in I_p\}$ will refer to a given one-parameter family of mappings satisfying (B0) and (B1). We will consider the set of preimages, $P(n) \in I_x \times I_p$ satisfying:

$$P(n) = \{(x,p) | f^i(x,p) = c \text{ for some } 0 \leq i \leq n\}.$$

First of all, it will be useful to have a way of specifying particular "sections" of preimages, $R(n, x_0, p_0)$, extending from a particular point $(x_0, p_0) \in I_x \times I_p$. So let $R(n, x_0, p_0) \subset I_x \times I_p$ denote the set of path-connected elements, consisting of all points $(x', p') \in I_x \times I_p$ such that there exists a continuous function $g : I_p \to I_x$ satisfying $g(p_0) = x_0$, $g(p') = x'$, and

$$\{(x,p) | x = g_{(x_0,p_0)}(p), p \in [p_0; p']\} \subset P(n).$$

where $[p_0; p']$ may denote either $[p_0, p']$ or $[p', p_0]$, whichever is appropriate.

A roadmap of the development in this section is as follows. In lemma D.3.1 we show that $P(n)$ cannot have isolated points or curve segments. Instead, each point in $P(n)$

must be part of a path-connected set of points in $P(n)$ that stretches for the length of the parameter space, I_p. In lemma D.3.2 we demonstrate that if the kneading invariant of f_p, $D(f_p, t)$, is monotonically decreasing (or increasing), then $P(n)$ must have a branching tree-like structure. As we travel along one direction in parameter space, branches of $P(n)$ must either always merge or always split away from each other. For example if $D(f_p, t)$ is monotonically decreasing, then branches of $P(n)$ can only split away from each other as we increase the parameter p. Thus in this case, $R(n, y_-, p_0)$ and $R(n, y_+, p_0)$ cannot intersect each other for $p \geq p_0$ if $y_+ \neq y_-$, and $y_+, y_- \in I_x$.

In lemmas D.3.3,D.3.4, D.3.5, and D.3.6 we develop bounds on the derivatives for differentiable branches of $R(n, x, p_0)$. The basic idea behind lemma D.3.7 is that we can use these bounds to demonstrate that for maps, f_p, with kneading invariants that decrease monotonically in parameter space, there exist constants $C > 0$ and $\delta p > 0$ such that if $x_0 \in I_x$ and

$$U(p) = \{x | \, |x - x_0| < C(p - p_0)^{\frac{1}{3}}\} \tag{D.25}$$

for any $p \in I_p$, then for any $p' \in [p_0, p_0 + \delta p]$, there exists $x'_+ \in U(p')$ such that $(x'_+, p') \in R(n_+, y_+, p_0)$ for some $y_+ > x_0$ and $n_+ > 0$ assuming that $f^{n_+}(y_+, p_0) = c$. Likewise there exists $x'_+ \in U(p')$ such that $(x'_-, p') \in R(n_-, y_-, p_0)$ for some $y_- < x_0$ and $n_- > 0$ where $f^{n_-}(y_-, p_0) = c$.

However, setting $n = \max\{n_+, n_-\}$, since $R(n, y_-, p_0)$ and $R(n, y_+, p_0)$ do not intersect each other for $p \geq p_0$ and $y_- \neq y_+$, we also know that for any $y_- < y_+$, there is a region in $I_x \times I_p$ space bounded by $R(n, y_-, p_0)$, $R(n, y_+, p_0)$, and $p \geq p_0$. Given any $x_0 \in I_x$, take the limit of this region as $y_- \to x_0^-$, $y_+ \to x_0^+$, and $n \to \infty$. Call the resulting region $S(x_0)$. Observe that $S(x_0)$ is a connected set that is invariant under f and is nonempty for every parameter value $p \in I_p$ such that $p \geq p_0$. Thus since $S(x_0)$ is bounded from (D.25), there exists a set of points, $S(x_0)$, in combined state and parameter space that "shadow" any trajectory, $\{f_{p_0}^n(x_0)\}_{n=0}^{\infty}$ of f_{p_0}. Finally we observe that a subset of $S(x_0)$ can be represented by the form given for $W(x_0)$.

We are now ready to examine these arguments more formally.

Lemma D.3.1 *Let $\{f_p : I_x \to I_x | p \in I_p\}$ be a one-parameter family of mappings satisfying (B0) and (B1). Suppose that $x_0 \in I_x$ satisfies $n = N(x_0, f_{p_0}) < \infty$ for some $p_0 \in int(I_p)$. Then the following statements hold true:*

(1) There exists a closed interval $J_p(x_0, p_0) \subset I_p$, and a C^2 function $h_{(x_0,p_0)} : J_p(y, p_0) \to I_x$ such that $p_0 \in int(J_p(x_0, p_0))$, $h_{y,p_0}(p_0) = p_0$, and $f^n(h_{y,p_0}(p), p) = c$ for all $p \in J_p(y, p_0)$. Also, if $J_p(y, p_0) = [a, b]$ then a is either an endpoint of I_p or $f^i(h_{y,p_0}(a), a) = c$ for some $i < n$, and similarly for b.

(2) There exists a continuous function, $g_{(x_0,p_0)} : I_p \to I_x$ such that $g_{(x_0,p_0)}(p_0) = x_0$ and

$$\{(x, p) | x = g_{(x_0,p_0)}(p), p \in I_p\} \subset P(n).$$

Proof: Suppose that $f^{m_0}(x_0, p_0) = c$ for $m_0 \leq n$ and $f^i(x_0, p_0) \neq c$ for $0 \leq i < m_0$. Then define the set $S(x_0, p_0) \subset I_x \times I_p$ to be the maximal path-connected set satisfying the following conditions:

(1) $(x_0, p_0) \in S(x_0, p_0)$

(2) $(x, p) \in S(x_0, y_0)$ if $p \in I_p$ and $f^i(x, p) \neq c$ for every $0 \leq i < m_0$.

Note that $S(x_0, p_0)$ must contain an open neighborhood around (x_0, p_0) because of the continuity of f.

Now let $\overline{S}(x_0, p_0)\}$ be the closure of of $S(x_0, p_0)$, define $Q_{(x_0,p_0)}(p) = \{x|(x, p) \in S(x_0, p_0)\}$, and let

$$J_p(x_0, p_0) = [\inf_{(x,p)\in S(x_0,p_0)} p, \sup_{(x,p)\in S(x_0,p_0)} p] \tag{D.26}$$

We claim that $Q_{(x_0,y_0)}(p) \in I_x$ must consist of a single connected interval for every $p \in J_p(x_0, p_0)$. Otherwise if there existed $x_1 < x_2 < x_3$ such that $x_1 \in Q_{(x_0,p_0)}(p)$, $x_2 \notin Q_{(x_0,p_0)}(p)$, and $x_3 \in Q_{(x_0,p_0)}(p)$ then there would exist $i < m_0$ such that $c \in [f^i(x_0, p); f^i(x_3, p)]$. But since $(x_1, p) \in S(x_0, p_0)$ and $(x_3, p) \in S(x_0, p_0)$ there exists a connected path, $\{(x(t), p(t))|t \in [0, 1]\} \subset S(x_0, p_0)$, joining (x_1, p) and (x_3, p), where where $x(t) : [0, 1] \to I_x$ and $p(t) : [0, 1] \to I_p$ are continuous functions. Along this path, $f^i(x(t), p(t))$ is continuous and $f^i(x(t), p(t)) \neq c$ for any $t \in [0, 1]$. This contradicts the assertion that $c \in [f^i(x_0, p); f^i(x_3, p)]$ and proves the claim that $Q_{(x_0,y_0)}(p)$ must consist of a single interval for all $p \in J_p(x_0, p_0)$.

Returning to the proof of the lemma we find that, since $(x, p) \in S(x_0, p_0)$ implies $f^i(x, p) \neq c$ for every $0 \leq i < m_0$, we know that $f_p^{m_0}(x)$ must be strictly monotonic on $Q_{(x_0,y_0)}(p)$ for each $p \in J_p(x_0, p_0)$. Thus for each $p \in [p_0, p_1)$ there is exactly one $x \in Q_{(x_0,y_0)}(p)$ such that $f^{m_0}(x, p) = c$. Consequently there exists a function $h_{(x_0,p_0)} : I_p \to I_x$ such that $f^{m_0}(h_{(x_0,p_0)}(p), p) = c$ and $h_{(x_0,p_0)}(p) \in Q_{(x_0,y_0)}(p)$ if $p \in J_p(x_0, p_0)$. Furthermore, the function, $h_{(x_0,p_0)}$, must be C^2 for $p \in int(J_p(x_0, p_0))$ since $f(x, p)$ is C^2 and $f_p^{m_0}(x)$ is strictly monotonic in for $x \in Q_{(x_0,y_0)}(p)$. Finally, from our choice of $S(x_0, p_0)$ and $h_{(x_0,p_0)}(p)$, it is clear that $(h_{(x_0,p_0)}(p), p) \in P(n)$ for all $p \in J_p(x_0, p_0)$. This proves property (1) of the lemma.

We now have to construct a continuous $g_{(x_0,p_0)}(p)$ that is valid over the entire range of I_p. Suppose that $J_p(x_0, y_0) = [p_{-1}, p_1]$. Let $g_{(x_0,p_0)}(p_1) = x_1$. From our specification of $S(x_0, p_0)$ it is clear that $f^j(x_1, p_1) = c$ for some $j < m_0$. Thus there exists $m_1 < m_0$ such that $f^{m_1}(x_1, p_1) = c$ and $f^i(x_1, p_1) \neq c$ for $0 \leq i < m_1$. Consequently, we can use the same arguments as before to consider the set $S(x_1, p_1)$, and generate a continuous function, $h_{(x_1,p_1)}(p)$ such that $(h_{(x_1,p_1)}(p), p) \in P(n)$ for all $p \in J_p(x_1, p_1)$ where $J_p(x_1, y_1) \supset [p_1, p_2]$ for some $p_2 > p_1$. This argument can be carried out repeatedly for $m_0 > m_1 > m_2, \ldots$ and so forth. However, since $f^{m_i}(x_i, p_i) = c$, we see that

$\sup(I_p) \in J_p(x_i, p_i)$ for some $i \leq n$. Similarly we can also use the same arguments for $p < p_0$, working in the opposite direction in parameter space in order to successively generate $(h_{(x_{-i},p_{-i})}(p), p) \in P(n)$ for increasing values of i. Consequently, there exists $-n \leq a \leq 0$ and $0 \leq b \leq n$ such that $I_p = \cup_{i=a}^{b} J_p(x_i, p_i)$. Now if we set $h : I_p \to I_x$ to be

$$g_{(x_0,p_0)}(p) = h_{(x_i,p_i)}(p) \text{ if } p \in J_p(x_i, p_i), \tag{D.27}$$

we can see that $g_{(x_0,p_0)}(p)$ is continuous since $h_{(x_i,p_i)}(p)$ is C^2 if $p \in int(J_p(x_i, p_i))$, and $h_{(x_i,p_i)}(p_i) = h_{(x_{i-1},p_{i-1})}(p_i)$ for all $a < i \leq b$. Finally, since $(h_{(x_i,p_i)}(p), p) \in P(n)$ for all $a \leq i \leq b$ we see that $g_{(x_0,p_0)}(p)$ has all the properties guaranteed by the lemma.

Lemma D.3.2 *Let $\{f_p : I_x \to I_x | p \in I_p\}$ be a one-parameter family of mappings satisfying (B0) and (B1). Suppose that there exists $\delta p > 0$ such that the kneading invariant $D(f_p, t)$ is monotonically decreasing for $p \in [p_0, p_0 + \delta p]$. Then*

$$R(n, y_0, p_0) \cap R(n, y_1, p_0) \cap (I_x \times [p_0, p_0 + \delta p]) = \emptyset \tag{D.28}$$

for any $y_0 \neq y_1$ and any $n \geq 0$ such that $y_0 \in I_x$ and $y_1 \in I_x$.

Proof: Suppose that there exists $y_0 \in I_x$ and $y_1 \in I_x$ such that

$$R(n, y_0, p_0) \cap R(n, y_1, p_0) \cap (I_x \times [p_0, p_0 + \delta p]) \neq \emptyset. \tag{D.29}$$

for some $n \geq 0$ where $N(y_0, f_{p_0}) < n$ and $N(y_1, f_{p_0}) < n$. It is sufficient to show that this statement contradicts the condition that $D(f_p, t)$ is monotonically decreasing for $p \in [p_0, p_0 + \delta p]$.

Let $p' > p_0$ be the smallest value such that there exists a pair of points $y_2 \in I_x$ and $y_3 \in I_x$ with $y_2 < y_3$ satisfying:

$$R(n, y_2, p_0) \cap R(n, y_3, p_0) \cap (I_x \times [p_0, p']) \neq \emptyset. \tag{D.30}$$

Assuming that $(D.29)$ is true, we know that $p' < p_0 + \delta p$. Now fix y_2 in the right hand side of (D.30) and let y_3 take on all values such that $y_3 > y_2$ and $y_3 \in I_x$. Let y_4 be the smallest possible value of y_3 that satisfies (D.30) and set $x' \in I_x$ such that $(x', p') \in R(n, y_2, p_0)$ and $(x', p') \in R(n, y_4, p_0)$.

Let G_2 be the set of all continuous functions, $\tilde{g}_2 : I_p \to I_x$, such that $\tilde{g}_2(p') = x'$ and $f(\tilde{g}_2(p), p) \in R(n, y_2, p_0)$ for all $p \in I_p$. By lemma D.3.1, there exist at least one element in G_2. Set

$$g_2(p) = \sup_{\tilde{g}_2 \in G_2} \tilde{g}_2(p). \tag{D.31}$$

Clearly $g_2(x)$ must be also be continuous function that satisfies $g_2(p') = x'$ and $f(g_2(p), p) \in R(n, y_2, p_0)$ for all $p \geq p_0$ if $p \in I_p$. Similarly we can define $g_4(x)$ in analagous way, making

$$g_4(x) = \inf_{\tilde{g}_4 \in G_4} \tilde{g}_4(x) \tag{D.32}$$

where G_4 is the set of all functions $\tilde{g_4} : I_p \to I_x$, satisfying $\tilde{g}_4(p') = x'$ and $f(\tilde{g}_4(p), p) \in R(n, y_4, p_0)$ for all p satisfying $p \in I_p$ and $p \geq p_0$.

Because of our choice of p', we know that $g_2(p) \neq g_4(p)$ if $p \in [p_0, p')$. Now let

$$\begin{aligned} J_2 &= \{(f(g_2(p), p), p) | p \in I_p\} \\ J_4 &= \{(f(g_4(p), p), p) | p \in I_p\}. \end{aligned}$$

And let $M \in I_x \times I_p$ be the interior of the region bounded by $J_2 \cup J_4 \cup (I_x \times \{p_0\})$. From our choice of p' we know that

$$\begin{aligned} J_2 \cap R(n, y, p_0) \cap (I_x \times [p_0, p')) &= \emptyset \\ J_4 \cap R(n, y, p_0) \cap (I_x \times [p_0, p')) &= \emptyset \end{aligned}$$

for any $y \neq y_2$ and $y \neq y_4$. From our choice of y_4 we also know that $(x', p') \notin R(n, y, p_0)$ for any $y \in (y_2, y_4)$. Thus we conclude that no $R(n, y, p_0)$ intersects M for any $y \in I_x$ satisfying $y \neq y_2$, $y \neq y_4$, and $N(y, f_{p_0}) \leq n$. Finally, from our choice of of $g_2(x)$ and $g_4(x)$ it is also apparent that neither $R(n, y_2, p_0)$ nor $R(n, y_4, p_0)$ intersects M. Consequently, we see that:

$$M \cap P(n) = \emptyset. \tag{D.33}$$

Now let

$$M_x(p) = \{x | (x, p) \in \overline{M}\}$$

where $\overline{M}$ denotes the closure of M. From (D.33) we know that f_p^i is strictly monotonic on $M_x(p)$ for any $0 \leq i \leq n$. Note in particular that this implies that there can exist no $0 \leq i \leq n$ such that

$$g_2^i(p) = g_4^i(p) = c \tag{D.34}$$

for any $p \in [p_0, p')$.

Now let $\{a_k\}_{k=0}^{\infty}$ be a monotonically increasing sequence such that $a_0 = p_0$ and $a_k \to p'$ as $k \to \infty$. We know that for any $p \in [p_0, p']$, there exists an $k \leq n$ such that $f^k(g_2(p), p) = c$. Thus consider the sequence $\{b_k\}_{k=0}^{\infty}$ where $b_k = N(g_2(a_k), f_{a_k})$. Since b_k can only take on a finite number of values ($0 \leq b_k \leq n$), we know there exists an infinite subsequence $\{k_i\}_{i=0}^{\infty}$ such that $b_{k_i} = b$ if $i \geq 0$ for some $0 \leq b \leq n$. This implies that $f^b(g_2(a_{k_i}), a_{k_i}) = c$ for all $i \geq 0$. Also, since f is continuous and $a_{k_i} \to p'$ as $i \to \infty$, we can also conclude that

$$f^b(g_2(p'), p') = f^b(x', p') = c. \tag{D.35}$$

We also play the same game with g_4 instead of g_2. Consider the sequence $\{d_i\}_{i=0}^{\infty}$ where $d_i = N(g_4(a_{k_i}), f_{a_{k_i}})$. We know that d_i can only take on a finite number of values, so there exists an infinite subsequence, $\{i_j\}_{j=0}^{\infty}$ and a number $0 \leq d \leq n$ such that $d_{i_j} = d$ for all $j \geq 0$. In this case, $f^d(g_2(a_{k_{i_j}}), a_{k_{i_j}}) = c$ for all $j \geq 0$. Since $a_{k_{i_j}} \to p'$ as $j \to \infty$ this implies that

$$f^d(g_4(p'), p') = f^d(x', p') = c. \tag{D.36}$$

However, from (D.34) we also know that $d_i \neq b_{k_i}$ for all $i \geq 0$. Thus $d \neq b$. For definiteness assume $b < d$. There exists $\delta p_1 > 0$ such that if $p \in [p' - \delta p_1, p']$ then $g_2^i(p) \neq c$ whenever $g_2^i(p') \neq c$ for any i satisfying $b < i < d$. Choose $p^* = a_{k_{i_j}}$ for some $j \geq 0$ large enough such that $p^* > p' - \delta p_1$. Note that by this choice of p^*, we know that $f^b(g_2(p^*), p^*) = c$ and $f^d(g_4(p^*), p^*) = c$.

Now recall the definition of the kneading invariant:

$$D(f_p, t) = 1 + \sum_{i=1}^{\infty} \theta_k(f_p) t^i.$$

where

$$\begin{aligned} \theta_i(f_p) &= \epsilon_1(f_p)\epsilon_2(f_p)\ldots\epsilon_i(f_p) \\ \epsilon_i(f_p) &= \lim_{x \to c^+} sgn(Df(f^i(c,p))) \end{aligned}$$

We claim that

$$|1 + \sum_{i=1}^{d-b-1} \theta_k(f_{p'})t^i| \geq |1 + \sum_{i=1}^{d-b-1} \theta_k(f_{p^*})t^i| \tag{D.37}$$

If this claim is true, the rest of the lemma follows. At this point we shall finish the proof of the lemma before coming back to the proof of the claim.

From (D.35) and (D.36) we know that

$$\theta_{d-b}(f_{p'}) = +1 \tag{D.38}$$

Also, since $g_2(p) \neq g_4(p)$ for $p \in [p_0, p')$, and $f^d(g_4(p^*), p^*) = c$, we know $f^d(g_2(p^*), p^*) = f^{d-b}(c, p^*) \neq c$. Combining this result with the fact that $f^d_{p^*}$ is monotone on $M_x(p^*)$ we see that if $f^{d-b}(c, p^*) > c$ then f^{d-b} has a maximum at $x = c$, which implies that f^{d-b+1} must have a minimum at $x = c$. Otherwise, if $f^{d-b}(c, p^*) < c$ then f^{d-b} has a minimum at $x = c$, and again f^{d-b+1} has a minimum at $x = c$. Thus we conclude that:

$$\theta_{d-b}(f_{p'}) = -1. \tag{D.39}$$

Finally, combining (D.38) with (D.39) with the claim above we find that $|D(f_{p'}, t)| > |D(f_{p^*}, t)|$. But since $p' > p^*$, this contradicts the assumption that the kneading invariant of f_p is monotonically decreasing with respect to p. This proves the theorem, except for the proof of the claim which we give below:

We now prove the claim given in (D.37) by induction on i. Suppose that $\theta_{i-1}(f_{p'}) = \theta_{i-1}(f_{p^*})$. We shall show that $\theta_i(f_{p'}) \geq \theta_i(f_{p^*})$.

Since $f^b(g_2(p'), p') = f^b(g_2(p^*), p^*) = c$, we can see that

$$sgn(Df(f^i(c,p))) = sgn(Df(f^{b+i}(g_2(p),p),p))$$

for either $p = p'$ or $p = p^*$. Since $R(n, y, p_0)$ does not cross the boundary of M for any $y \in I_x$, we can see that either both $f^{b+i}(g_2(p'), p') \geq c$ and $f^{b+i}(g_2(p^*), p^*) \geq c$ or both $f^{b+i}(g_2(p'), p') \leq c$ and $f^{b+i}(g_2(p^*), p^*) \leq c$ since both $(g_2(p'), p')$ and $(g_2(p^*), p^*)$ are on the boundary of M. Furthermore from our choice of p^* and $\delta p_1 > 0$ we know that if $g^i(c, p') \neq c$ then $g^i(c, p^*) \neq c$ for $0 < i \leq b - d$. Consequently we can see that if $g^i(c, p') \neq c$ then

$$\epsilon_i(f_{p'}) = \epsilon_i(f_{p^*}). \tag{D.40}$$

This in turn implies $\theta_i(f_{p'}) = \theta_i(f_{p^*})$ since $\theta_i(f_p) = \epsilon_i(f_p)\theta_{i-1}(f_p)$. On the other hand, if $g^i(c, p') = c$, then $\theta_i(f_{p'}) = +1$ so we automatically know that $\theta_i(f_{p'}) \geq \theta_i(f_{p^*})$.

Finally, note that the $\theta_i(f_{p'}) \geq \theta_i(f_{p^*})$ is satisfied for $i = 1$ since we have $\theta_1(f_{p'}) = \theta_1(f_{p^*})$ from (D.40) if $g(c, p') = c$ and $\theta_1(f_{p'}) \geq \theta_1(f_{p^*})$ if $g(c, p') = c$. This completes the proof of the claim.

Lemma D.3.3 *Let $\{f_p : I_x \to I_x | p \in I_p\}$ be a one-parameter family of mappings satisfying (B0) and (B1). Let $p_0 \in int(I_p)$ and $M_p = sup_{x \in I_x}(D_p f(x, p_0))$. Given $x_0 \in I_x$ such that $n = N(x_0, f_{p_0}) < \infty$, then for each $p \in J(x_0, p_0)$:*

$$|h'_{(x_0,p_0)}(p)| \leq \frac{M_p}{|D_x f(f^{n-1}(h_{(x_0,p_0)}(p), p), p)|} \sum_{i=0}^{n-1} |\frac{1}{D_x f^i(h_{(x_0,p_0)}(p), p)}|$$

Proof: In order to prove the lemma, we first need the following result (which can be found, for example, on page 417 of [33]).

Claim: For any $x \in I_x$ and $n \geq 1$:

$$|D_p f^n(x,p)| \leq M_p \sum_{i=0}^{n-1} |D_x f^{n-1-i}(f^i(x,p),p)| \tag{D.41}$$

Proof of claim: Proof by induction on n. For $n = 1$ the claim is clearly true. By the chain rule, for any $n \geq 1$:

$$D_p f^n(x,p) = D_p f(f^{n-1}(x,p),p) + D_x f(f^{n-1}(x,p),p) D_p f^{n-1}(x,p)$$

Thus we have the following

$$\begin{aligned}
|D_p f^n(x,p)| &\leq M_p + |D_x f(f^{n-1}(x,p),p)||D_p f^{n-1}(x,p)| \\
&\leq M_p + |D_x f(f^{n-1}(x,p),p)| M_p \sum_{i=0}^{n-2} |D_x f^{n-2-i}(f^i(x,p),p)| \\
&\leq M_p + M_p \sum_{i=0}^{n-2} |D_x f^{n-1-i}(f^i(x,p),p)| \\
&\leq M_p \sum_{i=0}^{n-1} |D_x f^{n-1-i}(f^i(x,p),p)|
\end{aligned}$$

This completes the induction argument and proves the claim.

Returning to the proof of the lemma, we know that since $f^n(h_{(x_0,p_0)}(p),p) = c$ for $p \in J(x_0,p_0)$. Consequently

$$\frac{\partial}{\partial p}[f^n(h_{(x_0,y_0)}(p),p)] = 0 \tag{D.42}$$

By the chain rule:

$$\frac{\partial}{\partial p}[f^n(h_{(x_0,y_0)}(p),p)] = (h'_{(x_0,p_0)}(p))(D_x f^n(h_{(x_0,p_0)}(p),p)) + D_p f^n(h_{(x_0,p_0)}(p),p) \tag{D.43}$$

Thus, combining (D.42) and (D.43), we have:

$$|h'_{(x_0,p_0)}(p)| = \frac{|D_p f^n(h_{(x_0,p_0)}(p),p)|}{|D_x f^n(h_{(x_0,p_0)}(p),p)|} \tag{D.44}$$

Let $x_p = h_{(x_0,p_0)}(p)$. Then, combining (D.41) and (D.44) we have:

$$\begin{aligned}
|h'_{(x_0,p_0)}(p)| &\leq \frac{M_p \sum_{i=0}^{n-1} |D_x f^{n-1-i}(f^i(x_p,p),p)|}{|D_x f^n(x_p,p)|} \\
&\leq \frac{M_p}{|D_x f^n(x_p,p)|} \sum_{i=0}^{n-1} \frac{|D_x f^{n-1}(f^i(x_p,p),p)|}{|D_x f^i(f^i(x_p,p),p)|} \\
&\leq \frac{M_p}{|D_x f(f^{n-1}(x_p,p),p)|} \sum_{i=0}^{n-1} \frac{1}{|D_x f^i(f^i(x_p,p),p)|}.
\end{aligned}$$

provided $p \in J(x_0,p_0)$. This proves the lemma.

Lemma D.3.4 *Let $\{f_p : I_x \to I_x | p \in I_p\}$ be a one-parameter family of mappings satisfying (B0) and (B1). Suppose that $p_0 \in int(I_p)$, and f_{p_0} satisfies (CE1). Also, suppose that $x_0 \in I_x$ such that $n = N(x_0, f_{p_0}) < \infty$, and $\min_{0 \leq i < n} |f^i(x_0,p_0) - c| = \alpha_{x_0} > 0$. Then there exist constants $C_1 > 0$ (independent of x_0) such that*

$$|h'_{(x_0,p_0)}(p_0)| \leq C_1 \frac{1}{\alpha_{x_0}^2}$$

Proof: From lemma D.3.3 :

$$|h'_{(x_0,p_0)}(p_0)| \leq \frac{M_p}{|D_x f(f^{n-1}(x_0,p_0),p_0)|}\sum_{i=0}^{n-1}\frac{1}{|D_x f^i(x_0,p_0)|} \tag{D.45}$$

From lemma D.2.7, we also know that f_{p_0} satisfies condition (CE2). Thus, since $f^n(x_0,p_0) = c$, we know there exists $K_E > 0$ such that $|D_x f(f^{n-1}(x_0,p_0),p_0)| > K_E$. Substituting this into (D.45) we have:

$$|h'_{(x_0,p_0)}(p_0)| \leq \frac{M_p}{K_E}\sum_{i=0}^{n-1}\frac{1}{|D_x f^i(x_0,p_0)|} \tag{D.46}$$

From lemma (D.2.11) we know that there exists $C > 0$ and $\lambda > 0$ such that:

$$|Dg^i(x)| > C\alpha_{x_0}^2\lambda^i$$

Then from (D.46),

$$|h'_{(x_0,p_0)}(p_0)| \leq \frac{M_p}{K_E}\sum_{i=0}^{n-1}\frac{1}{C\alpha_{x_0}^2\lambda^i} \leq \frac{M_p}{K_E C\alpha_{x_0}^2}(\frac{1}{1-\lambda^{-1}}) \leq C_1\frac{1}{\alpha_{x_0}^2}$$

if we set $C_1 = \frac{M_p}{K_E C}(\frac{1}{1-\lambda^{-1}})$. This proves the lemma.

Lemma D.3.5 *Let $\{f_p : I_x \to I_x | p \in I_p\}$ be a one-parameter family of mappings satisfying (B0) and (B1). Let $p_0 \in I_p$ and suppose that $x_0 \in I_x$ such that $n = N(x_0, f_{p_0}) < \infty$ and $\min_{0\leq i<n} |f^i(x_0,p_0) - c| = \alpha_{x_0} > 0$. Then for any $0 < \beta < 1$ there exists $0 < C_2 < \frac{1}{2}$ such that if $x_1 \in I_x$ and $p_1 \in I_p$ satisfy:*

(1) $|p_1 - p_0| \leq C_2\alpha_{x_0}$.

(2) $|f^i(x_1,p_1) - f^i(x_0,p_0)| \leq C_2\alpha_{x_0}$ *for* $0 \leq i < n$

then

$$\frac{|D_x f^i(x_1,p_1)|}{|D_x f^i(x_0,p_0)|} \geq \beta^i.$$

for $0 \leq i \leq n$.

Proof: Combining lemmas D.2.1 and D.2.2 with conditions (1) and (2) above we find that there exists $K_0 > 0$, $K_1 > 0$, and $K_2 > 0$ such that:

$$|D_x f(f^i(x_1,p_1),p_1) - D_x f(f^i(x_1,p_1),p_0)| < K_0|p_1 - p_0| < K_0C_2\alpha_{x_0} \tag{D.47}$$

$$|D_x f(f^i(x_1,p_1),p_0) - D_x f(f^i(x_0,p_0),p_0)| < K_1|f^i(x_1,p_1) - f^i(x_0,p_0)| < K_1C_2\alpha_{x_0} \tag{D.48}$$

$$|D_x f(f^i(x_0,p_0),p_0)| < K_2|f^i(x_0,p_0) - c| < K_2\alpha_{x_0} \tag{D.49}$$

for all $0 \leq i < n$.

From (D.47) and (D.48) we have:

$$\begin{aligned} & |D_x f(f^i(x_1,p_1),p_1) - D_x f(f^i(x_0,p_0),p_0)| \\ \leq \;& |D_x f(f^i(x_1,p_1),p_1) - D_x f(f^i(x_1,p_1),p_0)| \\ & + |D_x f(f^i(x_1,p_1),p_0) - D_x f(f^i(x_0,p_0),p_0)| \\ < \;& K_0 C_2 \alpha_{x_0} + K_1 C_2 \alpha_{x_0} = C_2(K_0+K_1)\alpha_{x_0} \end{aligned} \tag{D.50}$$

for all $0 \leq i < n$.

Now set $C_2 = \min\{\frac{1}{2}, \frac{K_2}{K_0+K_1}(1-\beta)\}$. Then from (D.50) and (D.49):

$$\begin{aligned} \frac{|D_x f(f^i(x_1,p),p_1)|}{|D_x f(f^i(x_0,p_0),p_0)|} &\geq 1 - \frac{|D_x f(f^i(x_1,p_1),p_1) - D_x f(f^i(x_0,p_0),p_0)|}{|D_x f(f^i(x_0,p_0),p_0)|} \\ &> 1 - \frac{C_2(K_0+K_1)\alpha_{x_0}}{K_2\alpha_{x_0}} \\ &\geq 1 - (\frac{K_2}{K_0+K_1})(1-\beta)(\frac{K_0+K_1}{K_2}) = \beta \end{aligned}$$

for all $0 \leq i < n$. Thus we have:

$$\frac{|D_x f^i(x_1,p_1)|}{|D_x f^i(x_0,p_0)|} = \prod_{j=0}^{i-1} \frac{|D_x f(f^j(x_1,p_1),p_1)|}{|D_x f(f^j(x_0,p_0),p_0)|} > \beta^i$$

if $0 \leq i \leq n$, which proves the lemma.

Lemma D.3.6 *Let $\{f_p : I_x \to I_x | p \in I_p\}$ be a one-parameter family of mappings satisfying (B0) and (B1). Suppose that $p_0 \in int(I_p)$, and f_{p_0} satisfies (CE1). Let $x_0 \in I_x$ such that $n = N(x_0, f_{p_0}) < \infty$ and $\min_{0 \leq i < n} |f^i(x_0,p_0) - c| = \alpha_{x_0} > 0$. Then there exist $C_3 > 0$ and $C_4 > 0$ (independent of x_0) such that*

$$|h'_{(x_0,p_0)}(p)| < C_3 \frac{1}{\alpha_{x_0}^2}$$

if $p \in V(x_0,p_0)$ where $V(x_0,p_0) = [p_0, p_0+\delta p_1]$, $\delta p_1 = C_4 \alpha_{x_0}^3$, and $h_{(x_0,p_0)} : V(x_0,p_0) \to I_x$ is a C^2 function satisfying $h_{(x_0,p_0)}(p_0) = x_0$ and $f^n(h_{(x_0,p_0)}(p),p) = c$ for all $p \in V(x_0,p_0)$.

From lemma D.3.1 we know that there exists a C^2 function $h_{(x_0,p_0)}(p)$ such that $h_{(x_0,p_0)}(p_0) = x_0$ and $f^n(h_{(x_0,p_0)}(p),p) = c$ if $p \in J(x_0,p_0)$ where $J(x_0,p_0) \subset I_p$ is a interval containing p_0. Also from lemma D.3.1 we know that there exists a continuous function $g_{(x_0,p_0)}(p)$ satisfying $g_{(x_0,p_0)}(p_0) = x_0$ and $f^n(g_{(x_0,p_0)}(p),p) = c$ for all $p \in I_p$.

By lemma D.2.11, there exists $C > 0$ and $\lambda > 0$ such that:

$$D_x f^i(x_0, p_0) > C\alpha_{x_0}^2 \lambda^i. \tag{D.51}$$

for any $0 \le i \le n$.

Now fix $\lambda_1 = \frac{1+\lambda}{2} > 1$ and let $\beta = \frac{\lambda_1}{\lambda} < 1$. Then given $g_{(x_0,p_0)}(p)$, we know from lemma D.3.5 that there exists a constant $0 < C_2 < \frac{1}{2}$ (dependent only on β) such that if $V(x_0, p_0) \subset I_p$ is the maximal interval satisfying the following conditions:

(1) If $p \in V(x_0, p_0)$, then $|p - p_0| \le C_2\alpha_{x_0}$.

(2) If $p \in V(x_0, p_0)$, then $|f^i(g_{(x_0,p_0)}(p), p) - f^i(x_0, p_0)| \le C_2\alpha_{x_0}$ for $0 \le i < n$,

then $p \in V(x_0, p_0)$ implies that:

$$\frac{|D_x f^i(g_{(x_0,p_0)}(p), p)|}{|D_x f^i(x_0, p_0)|} \ge \beta^i \tag{D.52}$$

for any $0 \le i \le n$. Note that by setting $\lambda_1 > 0$, we have also set the constants $0 < \beta < 1$ and $0 < C_2 < \frac{1}{2}$, so these constants are fixed for the discussion that follows.

Note, also, that from condition (2) above it is apparent that $g_{(x_0,p_0)} \ne c$ for any $p \in V(x_0, p_0)$. From lemma D.3.1, this implies that $V(x_0, p_0) \subset J(x_0, p_0)$ so that $g_{(x_0,p_0)}(p) = h_{(x_0,p_0)}(p)$ is C^2 when $p \in V(x_0, p_0)$.

Now consider the sequence $\{y_{-i}\}_{i=0}^n$ where $y_{-i} = f^{n-i}(x_0, p_0)$ so that $y_{-n} = x_0$ and $y_0 = c$. Then, from (D.51), (D.52), and our choice of β, we know that:

$$|D_x f^i(h_{(y_{-i},p_0)}(p), p)| \ge |D_x f^i(y_{-i}, p_0)|\beta^i \ge C\alpha_{x_0}^2\lambda^i\beta^i \ge C\alpha_{x_0}^2\lambda_1^i$$

if $p \in V(y_{-i}, p_0)$ for any $0 < i \le n$. Substituting this into lemma D.3.3 we find that if $p \in V(x_0, p_0)$:

$$|h'_{(y_{-i},p_0)}(p)| \le \frac{M_p}{|D_x f(z(p), p), p)|} \sum_{j=0}^{i} \frac{1}{|D_x f^j(h_{(y_{-i},p_0)}(p), p)|} \tag{D.53}$$

Where $z(p) = f^{n-1}(h_{(x_0,p_0)}(p), p)$. Since f_{p_0} satisfies (CE2) and $f(z(p), p) = c$, we can bound $|Df(z(p_0), p_0)| > K_E$ for some constant $K_E > 0$ independent of x_0. Consequently from condition (2) above and lemma D.2.1 there must exist $K'_E > 0$ (independent of x_0) such that $|Df(z(p), p_0)| > K'_E$ if $p \in V(x_0, p_0)$. Substituting this into (D.53) we have:

$$\begin{aligned} |h'_{(y_{-i},p_0)}(p)| &\le \frac{M_p}{K'_E} \sum_{j=0}^{i} \frac{1}{C\alpha_{x_0}^2\lambda_1^i} \\ &\le \left(\frac{M_p}{K'_E C\alpha_{x_0}^2}\right)\left(\frac{1}{1-\lambda_1^{-1}}\right). \end{aligned}$$

Thus setting $C_3 = \frac{M_p}{K_E' C(1-\lambda_1^{-1})}$, we have that

$$|h'_{(y_{-i},p_0)}(p)| \leq C_3 \frac{1}{\alpha_{x_0}^2} \tag{D.54}$$

for $0 < i \leq n$ if $p \in V(x_0, p_0)$. Of course, since $x_0 = y_{-n}$, this also implies that

$$|h'_{(x_0,p_0)}(p)| \leq C_3 \frac{1}{\alpha_{x_0}^2}$$

if $p \in V(x_0, p_0)$.

This places the proper bound on the derivative $h'_{(x_0,p_0)}(p)$. Now we need to find a proper bound on the size of $V(x_0, p_0)$. Set

$$\delta p = \min\{\frac{C_2}{2C_3}\alpha_{x_0}^3, C_2\alpha_{x_0}, \sup(I_p) - p_0\}. \tag{D.55}$$

We claim that if $[p_0, p_0 + \delta p] \subset V(y_{-(i-1)}, p_0)$, then $[p_0, p_0 + \delta p] \subset V(y_{-i}, p_0)$. Also, it is clear that $[p_0, p_0 + \delta p] \subset V(c, p_0) = V(y_0, p_0)$. So, by induction on i, this claim implies that $[p_0, p_0 + \delta p] \subset V(y_{-n}, p_0) = V(x_0, p_0)$. Thus if the claim is true, then from (D.55), and since α_{x_0} is bounded above, we know there exists $C_4 > 0$ such that $[p_0, p_0 + \delta p_1] \subset V(x_0, p_0)$ where $\delta p_1 = C_4 \alpha_{x_0}^3$. This proves the lemma. Thus, all that is left to do is to prove the claim.

Suppose that the claim were not true. This means there exists $p_1 \in [p_0, p_0 + \delta p]$ such that $p_1 \notin V(y_{-i}, p_0)$. From our specification of $V(x_0, p_0)$ and the intermediate value theorem, it is apparent that the only way this can happen is if there exists some $p_2 \in [p_0, p_1]$ such that

$$|h_{(y_{-i},p_0)}(p_2) - y_{-i}| = C_2 \alpha_{x_0} \tag{D.56}$$

and $[p_0, p_2] \subset V(y_{-i}, p_0)$.

However, by the mean value theorem, we know that

$$\begin{aligned} |h_{(y_{-i},p_0)}(p_2) - y_{-i}| &= |h_{(y_{-i},p_0)}(p_2) - h_{(y_{-i},p_0)}(p_0)| \\ &= |h'_{(y_{-i},p_0)}(p_3)||p_2 - p_0| \end{aligned} \tag{D.57}$$

for some $p_3 \in [p_0, p_2] \subset V(y_{-(i-1)}, p_0)$. But from (D.54):

$$|h'_{(y_{-i},p_0)}(p_3)| \leq C_3 \frac{1}{\alpha_{x_0}^2} \tag{D.58}$$

Combining (D.57), (D.58), and our choice of δp we find that

$$\begin{aligned} |h_{(y_{-i},p_0)}(p_2) - y_{-i}| &\leq C_3 \frac{1}{\alpha_{x_0}^2} |p_2 - p_0| \\ &\leq C_3 \frac{1}{\alpha_{x_0}^2} \delta p \\ &\leq \frac{1}{2} C_2 \alpha_{x_0} \end{aligned}$$

which contradicts (D.56) and proves the claim.

Lemma D.3.7 *Let $\{f_p : I_x \to I_x | p \in I_p\}$ be a one-parameter family of mappings satisfying (B0) and (B1). Given any $p_0 \in int(I_p)$, $x_0 \in I_x$, $p_1 \in int(I_p)$,, and $x_1 \in I_x$, suppose that $W(x_0) \subset I_x \times I_p$, is a connected set that can be represented in the following way:*

$$W(x_0) = \{(\alpha_{x_0}(t), \beta_{x_0}(t)) | t \in [0,1]\}$$

where $\alpha_{x_0} : [0,1] \to I_x$ and $\beta_{x_0} : [0,1] \to I_p$ satisfy the following properties:

1. *$\alpha_{x_0}(t)$ and $\beta_{x_0}(t)$ are continuous.*
2. *$\beta_{x_0}(t)$ is monotonically increasing with respect to t.*
3. *$\alpha_{x_0}(0) = x_0$, $\alpha_{x_0}(1) = x_1$.*
4. *$\beta_{x_0}(0) = p_0$, $\beta_{x_0}(1) = p_1$.*

Then there exists constants $\delta p > 0$ and $C > 0$ (independent of x_0) such that if $|x_1 - x_0| \geq C|p_1 - p_0|^{\frac{1}{3}}$ and $|p_1 - p_0| < \delta p$, then

$$W(x_0) \cap R(n, y, p_0) \cap (I_x \times [p_0, p_0 + \delta p]) \neq \emptyset$$

for some $n \geq 0$ and $y \in I_x$ such that $y \neq x_0$.

Proof: We assume that $x_1 > x_0$ and $p_1 > p_0$ (the other cases are similar). From lemma D.2.13, we know that there exist constants $K_0 > 0$ and $\epsilon_0 > 0$ so that for any positive $\epsilon < \epsilon_0$, there is a $y \in (x_0, x_0 + \epsilon)$ such that $f^n(y, p_0) = c$ and $\min_{0 \leq i < n} f^i(y, p_0) > K_0\epsilon$ for some $n \geq 0$. From lemma D.3.6, we know that there exist constants $K_1 > 0$ and $K_2 > 0$ such that if

$$\delta p_\epsilon = K_1(K_0\epsilon)^3 \tag{D.59}$$

then for all $p \in [p_0, p_0 + \delta p_\epsilon]$:

$$|h'_{(y,p_0)}(p)| < K_2(\frac{1}{K_0\epsilon})^2. \tag{D.60}$$

Thus given $x_0 \in I_x$, $x_1 \in I_x$, $p_0 \in int(I_p)$, and $p_1 \in int(I_p)$ choose

$$\epsilon = \frac{1}{K_0}(\frac{p_1 - p_0}{K_1})^{\frac{1}{3}}. \tag{D.61}$$

Also, set $\delta p = K_1(K_0\epsilon_0)^3$. Note that this means $p_1 - p_0 < \delta p$ implies that $\epsilon < \epsilon_0$, so that the results of the previous paragraph hold.

In particular, if we substitute (D.61) into (D.59), we find that $\delta p_\epsilon = K_1(K_0\epsilon)^3 = p_1 - p_0$ so that from (D.60) we have that for all $p \in [p_0, p_1]$:

$$|h'_{(y,p_0)}(p)| < K_2(\frac{1}{K_0\epsilon})^2$$

for some $y \in (x_0, x_0 + \epsilon)$. Consequently:

$$\begin{aligned}
h_{(y,p_0)}(p_1) &< h_{(y,p_0)}(p_0) + (p_1 - p_0) \inf_{p \in [p_0,p_1]} |h'_{(y,p_0)}(p)| \\
&< y + K_2(\frac{1}{K_0\epsilon})^2(p_1 - p_0) \\
&\leq (x_0 + \epsilon) + K_2(\frac{1}{K_0\epsilon})^2(p_1 - p_0) \\
&= x_0 + \frac{1 + K_0K_1K_2}{K_1^{\frac{1}{3}}K_0}(p_1 - p_0)^{\frac{1}{3}} \\
&= x_0 + C(p_1 - p_0)^{\frac{1}{3}} \qquad \text{(D.62)}
\end{aligned}$$

where $C = \frac{1+K_0K_1K_2}{K_1^{\frac{1}{3}}K_0}$.

Now suppose that $(x_1, p_1) \in W(x_0)$ where $x_1 - x_0 \geq C|p_1 - p_0|^{\frac{1}{3}}$. From (D.62) we know that there exists a continuous function, $h_{(y,p_0)}(p)$ such that $(h_{(y,p_0)}(p), p) \in R(n, y, p_0)$ for all $p \in [p_0, p_1]$ where $h_{(y,p_0)}(p_0) = y > x_0$ and $h_{(y,p_0)}(p_1) < x_1$. We are also given that $W(x_0)$ can be represented as $W(x_0) = \{(\alpha_{x_0}(t), \beta_{x_0}(t)) | t \in [0,1]\}$. Using the Intermediate Value Theorem, it can be shown that $h_{(y,p_0)}(\beta(t_1)) = \alpha_{x_0}(t_1)$ for some $t_1 \in [0,1]$. This implies that

$$W(x_0) \cap R(n, y, p_0) \cap (I_x \times [p_0, p_0 + \delta p]) \neq \emptyset \qquad \text{(D.63)}$$

which proves the lemma.

Proof of Theorem 3.3.2: Note that the theorem is trivial if $x_0 = Bd(I_x)$ (where $Bd(I_x)$ denotes the boundary of I_x). Otherwise, fix $p_0 \in int(I_p)$ such that f_{p_0} satisfies (CE1) and suppose there exists $\delta p_1 > 0$ such that $D(f_p, t)$ is monotonically decreasing for $p \in [p_0, p_0 + \delta p_1]$. Given any $x_0 \in int(I_x)$ let:

$$\begin{aligned}
X_n^-(x_0) &= \{x | N(x, f_{p_0}) \leq n \text{ and } x < x_0\} \\
X_n^+(x_0) &= \{x | N(x, f_{p_0}) \leq n \text{ and } x > x_0\}
\end{aligned}$$

Define the following functions $a_{n,x_0}^- : I_p \to I_x$ and $a_{n,x_0}^+ : I_p \to I_x$:

$$\begin{aligned}
a_{n,x_0}^-(p) &= \sup_{x' \in X_n^-(x_0)} \{x | (x,p) \in R(n, x', p_0), p \in I_p\} \qquad \text{(D.64)} \\
a_{n,x_0}^+(p) &= \inf_{x' \in X_n^+(x_0)} \{x | (x,p) \in R(n, x', p_0), p \in I_p\} \qquad \text{(D.65)}
\end{aligned}$$

It is apparent from our specification of $R(n,x,p_0)$ that $a^-_{n,x_0}(p)$ and $a^+_{n,x_0}(p)$ must be continuous with respect to p.

First of all note that $a^-_{m,x_0}(p) \geq a^-_{n,x_0}(p)$ if $m > n$. Furthermore, we claim that for any $n \geq 0$ there exists $m > n$ such that $a^-_{m,x_0}(p) > a^-_{n,x_0}(p)$ for all $p \in [p_0, p_0 + \delta p_1]$. By lemma D.3.2 we know that if $D(t, f_p)$ is monotonically decreasing for $p \in [p_0, p_0 + \delta p_1]$ then $R(n,x,p_0)$ and $R(n,x',p_0)$ do no intersect in the region $I_x \times [p_0, p_0 + \delta p_1]$ provided $x \neq x'$. This is implies that we can rewrite (D.64) as:

$$a^-_{n,x_0}(p) = \sup\{x | (x,p) \in R(n, x^*_n, p_0)\} \tag{D.66}$$

where $x^*_n = \sup\{X^-_n(x_0)\}$. Also we know from lemma D.2.9 that given any $n \geq 0$ there exists some $m > n$ such that $x^*_m > x^*_n$. This proves the claim. Similarly, we also can show that for any $n \geq 0$ there exists $m > n$ such that $a^+_{m,x_0}(p) < a^+_{n,x_0}(p)$ for all $p \in [p_0, p_0 + \delta p_1]$.

Returning to the lemma, we note that since $a^-_{n,x_0}(p)$ is monotonically increasing with respect to n, and bounded above by $\sup I_x = 1$, there exists a function, $a^-_{x_0}(p)$, such that the limit

$$a^-_{x_0}(p) = \lim_{n\to\infty} a^-_{n,x_0}(p) \tag{D.67}$$

converges pointwise. Now set

$$b^-_{x_0}(p) = \limsup_{t\to p} a^-_{x_0}(t) \tag{D.68}$$

and define

$$S^-(x_0) = \{(x,p) | \liminf_{t\to p} b^-_{x_0}(t) \leq x \leq \limsup_{t\to p} b^-_{x_0}(t)\}. \tag{D.69}$$

Similarly we can also define $S^+(x_0)$ as follows:

$$\begin{aligned} a^+_{x_0}(p) &= \lim_{n\to\infty} a^+_{n,x_0}(p) \\ b^+_{x_0}(p) &= \liminf_{t\to p} a^+_{x_0}(t) \\ S^+(x_0) &= \{(x,p) | \liminf_{t\to p} b^+_{x_0}(t) \leq x \leq \limsup_{t\to p} b^+_{x_0}(t)\}. \end{aligned}$$

The next step is to show that

$$S^-(x_0) \cap R(n,x,p) \cap (I_x \times [p_0, p_0 + \delta p_1]) = \emptyset \tag{D.70}$$

for any $x \neq x_0$ and any $n \geq 0$. This will be done in two parts. First we address the case where $x < x_0$. We claim that (D.70) is true if $x < x_0$. Suppose the claim is not true. Then from (D.64) there must exist some $(x',p') \in S^-(x_0)$ and $n \geq 0$ such that $a^-_{n,x_0}(p') \geq x'$

where $p' \in [p_0, p+\delta p_1]$. But we have already seen that for any $n \geq 0$ there exists an $m > n$ such that $a^-_{m,x_0}(p) > a^-_{n,x_0}(p)$ for all $p \in [p_0, p_0+\delta p_1]$. Thus $a^-_{x_0}(p) > a^-_{n,x_0}(p)$ for any $n \geq 0$ if $p \in [p_0, p_0+\delta p_1]$. Consequently since $a^-_{n,x_0}(p)$ is continuous:

$$\begin{aligned} x' &\leq a^-_{n,x_0}(p') = \liminf_{t_1\to p'} \limsup_{t\to t_1} a^-_{n,x_0}(t) \\ &< \liminf_{t_1\to p'} \limsup_{t\to t_1} a^-_{x_0}(t) = \liminf_{t\to p'} b^-_{x_0}(t) \end{aligned}$$

which from (D.69) implies that $(x', p') \notin S^-(x_0)$. This is a contradiction which proves the claim.

We now claim that $S^-(x_0) \cap R(n, x, p) \cap (I_x \times [p_0, p_0+\delta p_1]) = \emptyset$ if $x > x_0$. If this claim is not true, then from (D.65) we can see that there must exist some $(x', p') \in S^-(x_0)$ and $n \geq 0$ such that $a^+_{n,x_0}(p') \leq x'$ where $p' \in [p_0, p_0+\delta p_1]$. Furthermore there exists $m > n$ such that $a^+_{m,x_0}(p) < a^+_{n,x_0}(p)$ for $p \in [p_0, p_0+\delta p_1]$. Thus there exists $\epsilon > 0$ such that $a^+_{m,x_0}(p') \leq x' - 2\epsilon$. Since $a^+_{m,x_0}(p)$ is continuous, this implies that there exists $\delta > 0$ such that

$$a^+_{m,x_0}(p) \leq x' - \epsilon. \tag{D.71}$$

for any p such that $|p - p'| < \delta$. But since $(x', p') \in S^-(x_0)$,

$$\limsup_{t_1\to p'} \limsup_{t\to t_1} \lim_{n\to\infty} a^-_{n,x_0}(t) \geq x'.$$

Since $a^-_{n,x_0}(p)$, is continuous, this implies that for any $\delta > 0$ and $\epsilon > 0$ there is an $n \geq 0$ and p_1 with $|p_1 - p'| < \delta$ such that $a^-_{n,x_0}(p_1) > x' - \epsilon$. Combining this with (D.71) we see that there exists p_2 such that $a^-_{n,x_0}(p_2) = a^+_{n,x_0}(p_2)$. But this is impossible by lemma D.3.2 because it implies that $(x', p') \in R(m, x_1, p_0)$ and $(x', p') \in R(n, x_2, p_0)$ for some $n \geq 0$, $m \geq 0$, $x_1 \neq x_2$, and $p' \in [p_0, p_0+\delta p_1]$. This contradiction proves the claim.

The next step is to show that $S^-(x_0) \cup S^+(x_0)$ is invariant under f. We claim that if $(x, p) \in S^-(x_0)$ then either $(f(x,p), p) \in S^-(f(x_0, p_0))$ or $(f(x,p), p) \in S^+(f(x_0, p_0))$. For any $x_0 \in int(I_x)$, there exists an $\epsilon > 0$ such that $(x_0 - \epsilon, x_0) \subset (I_x \setminus \{c\})$. Let $J = (x_0 - \epsilon, x_0)$. Then, since f_{p_0} is a diffeomorphism on J, for any $y_1 \in f(J, p_0)$ such that $n(y_1) = N(y_1, f_{p_0}) < \infty$, there exists $y_0 \in J$ such that $y_1 = f(y_0, p_0)$ and $N(y_0, f_{p_0}) = n(y_1) + 1$. Consequently, from (D.66) we know that there exists $N > 0$ such that for all $n > N$:

$$f(a^-_{n,x_0}(p), p) = \begin{cases} a^-_{n,f(x_0,p_0)}(p) & \text{if } D_x f(x, p_0) > 0 \text{ on } J \\ a^+_{n,f(x_0,p_0)}(p) & \text{if } D_x f(x, p_0) < 0 \text{ on } J \end{cases}$$

for any $p \in [p_0, p_0+\delta p_1]$ if $x \in int(I_x)$. This result combined with our specification of $S^-(x_0)$ in (D.67), (D.68), and (D.69) proves the claim. Using the analogous result for $S^+(x_0)$ gives us that $S^-(x_0) \cup S^+(x_0)$ is invariant under f.

Finally, from the formulation of $S^{-}(x_0)$ in (D.69), it is apparent that there exists a $W^{-}(x_0) \subset S^{-}(x_0)$ such that $W^{-}(x_0)$ can be represented in the following way:

$$W^{-}(x_0) = \{(\alpha_{x_0}(t), \beta_{x_0}(t)) | t \in [0,1]\}$$

where $\alpha_{x_0} : [0,1] \to I_x$ and $\beta_{x_0} : [0,1] \to I_p$ are continuous functions and $\beta_{x_0}(t)$ is monotonically increasing with respect to t with $\beta_{x_0}(0) = p_0$ and $\beta_{x_0}(1) = p_0 + \delta p_1$. Of course, a similar $W^{+}(x_0) \subset S^{+}(x_0)$ also exists.

Putting it all together, we have now shown that: (1) $S^{-}(x_0) \cup S^{+}(x_0)$ is invariant under f and (2) $(S^{-}(x_0) \cup S^{+}(x_0)) \cap R(n, x, p_0) \cap (I_x \times [p_0, p_0 + \delta p_1]) = \emptyset$ for any $n \geq 0$ and any $x \neq x_0$. From property (2) above, lemma D.3.7, and since $W^{-}(x_0) \subset S^{-}(x_0)$, it is apparent that there exists $\delta p_2 > 0$ and $C > 0$ (independent of x_0) such that if $(x,p) \in W^{-}(x_0)$ then $|x - x_0| \leq C(p - p_0)^{\frac{1}{3}}$. Set $\delta p = \min\{\delta p_1, \delta p_2\}$ and let $W(x_0) = W^{-}(x_0)$ for $p \in [p_0, p_0 + \delta p]$. Then property (1) implies that given any $x_0 \in int(I_x)$, if $(x,p) \in W(x_0)$ and $p \in [p_0, p_0 + \delta p]$, then $|f^n(x,p) - f^n(x_0, p_0)| < C(p - p_0)^{\frac{1}{3}}$ for any $n \geq 0$. This proves the theorem.

Appendix E

Proof of theorem 3.4.2

This appendix contains the proof for theorem 3.4.2. For reference, the conditions, (CE1) and (CE2), can be found in the beginning of appendix D.

Theorem 3.4.2 *Let* $I_p = [0, 4]$, $I_x = [0, 1]$, *and* $f_p : I_x \to I_x$ *be the family of quadratic maps such that* $f_p(x) = px(1-x)$ *for* $p \in I_p$. *Then there exist constants* $\delta > 0$, $C > 0$, $K > 0$, *and set* $E(\gamma) \subset I_p$ *with positive Lebesgue measure for every* $\gamma > 1$ *such that:*

(1) If $\gamma > 1$ *and* $p_0 \in E(\gamma)$, *then* f_{p_0} *satisfies (CE1).*

(2) If f_{p_0} *satisfies (CE1), then for any* $\epsilon > 0$ *sufficiently small, any orbit of* f_{p_0} *can be* $\epsilon-$*shadowed by an orbit of* f_p *for* $p \in [p_0, p_0 + C\epsilon^3]$.

(3) If $\gamma > 1$ *and* $p_0 \in E(\gamma)$, *then for any* $\epsilon > 0$, *almost no orbits of* f_{p_0} *can be* $\epsilon-$*shadowed by any orbit of* f_p *for* $p \in (p_0 - \delta, p_0 - (K\epsilon)^\gamma)$.

That is, the set of possible initial conditions, $x_0 \in I_x$, such that the orbit $\{f_{p_0}^i(x_0)\}_{i=0}^\infty$ can be $\epsilon-$shadowed by some orbit of f_p comprises at most a set of Lebesgue measure zero on I_x if $p \in (p_0 - \delta, p_0 - (K\epsilon)^\gamma)$.

Proof of Theorem 3.4.2: We first address parts (1) and (3) of theorem and come back to part (2) at the end of the proof.

The basic idea behind parts (1) and (3) is to apply theorem 3.3.1 to theorem 3.4.1. There are four major steps. We first set lower bounds on the return time of the orbit of the turning point, $c = \frac{1}{2}$, to neighborhoods of c. Next we show that f_p satisfies (CP1) and favors higher parameters on a positive measure of parameter values. This allows us to apply theorem 3.3.1. Finally we show that almost every orbit of these maps approach arbitrarily close to c so that if the orbit, $\{f_{p_0}^i(c)\}_{i=0}^\infty$, cannot be shadowed then almost all other orbits of f_{p_0} cannot be shadowed either.

We first show that there is a set of parameters of positive measure such that orbits of the turning point, $\{f_p^i(c)\}_{i=0}^{\infty}$, do not return too quickly to neighborhoods of c. This can be seen from the construction used to prove theorem 3.4.1. In [5] it is shown that for any $\alpha > 0$, if $S(\alpha) \subset I_p$, is the set of parameters such that f_{p0} satisfies both (CE1) and:

$$|f_{p_0}^i(c) - c| > e^{-\alpha i} \tag{E.1}$$

for all $i \in \{0, 1, 2, \dots\}$, then $S(\alpha)$ has a density point at $p = 4$.

We now show that (CP1) is also satisfied on a positive measure of parameter values. First consider what happens if $p = 4$:

$$\begin{aligned} D_p f(c, p = 4) &= \frac{1}{4} && \text{(E.2)} \\ D_p f(f^n(c, p = 4), p = 4) &= 0 \text{ for any } n > 1 && \text{(E.3)} \\ |D_x f(f^n(c, p = 4), p = 4)| &= 4 \text{ for any } n \geq 1 && \text{(E.4)} \\ |D_x f^n(c, p = 4)| &= 4^{n-2} \text{ for any } n \geq 1. && \text{(E.5)} \end{aligned}$$

It also a simple matter to verify that f_p favors higher parameters at $p = 4$. Note that from the chain rule we have that:

$$D_p f^n(c, p) = D_x f(f^{n-1}(c, p), p) D_p f^{n-1}(c, p) + D_p f(f^{n-1}(c, p), p) \tag{E.6}$$

for any $n \geq 1$ and any $p \in I_p$. Consequently, using continuity arguments we can see that for any $N > 0$ and $\delta > 0$ there exists $\epsilon_1 > 0$ such that $p \in [4 - \epsilon_1, 4]$ implies that both of the following hold:

$$\begin{aligned} |D_p(c, p)| &> \frac{1}{4} - \delta && \text{(E.7)} \\ |D_p(f^n(c, p))| &< \delta \text{ for any } n \in \{2, 3, \dots, N\}. && \text{(E.8)} \end{aligned}$$

From (E.6) we can see that:

$$\begin{aligned} D_p f^n(x, p) &= D_p f(f^{n-1}(c, p), p) + \sum_{i=0}^{n-2} [D_p f(f^i(c, p), p) \prod_{j=i+1}^{n-1} D_x f(f^j(c, p), p)] \\ &= \prod_{j=1}^{n-1} D_x f(f^j(c, p), p) [\frac{D_p f(f^{n-1}(c, p), p)}{\prod_{j=1}^{n-1} D_x f(f^j(c, p), p)} + D_p f(c, p) + \\ &\quad \sum_{i=1}^{n-2} \frac{D_p f(f^i(c, p), p)}{\prod_{j=1}^{i} D_x f(f^j(c, p), p)}] \\ &= \prod_{j=1}^{n-1} D_x f(f^j(c, p), p) [D_p f(c, p) + \sum_{i=1}^{n-1} \frac{D_p f(f^i(c, p), p)}{\prod_{j=1}^{i} D_x f(f^j(c, p), p)}] && \text{(E.9)} \end{aligned}$$

for any $n \geq 1$. But from theorem 3.4.1, we also know that there exists $K_E > 0$ and $\lambda_E > 1$ and a set $E \subset I_p$ of positive measure such that if $p \in E$, then (CE1) is satisfied for f_p :

$$|\prod_{j=1}^{n} D_x f(f^j(c,p),p)| = |D_x f^n(f(c,p),p)| > K_E \lambda_E^n.$$

Substituting this into (E.9) we have:

$$|D_p f^n(x,p)| \;>\; K_E \lambda_E^{n-1}[|D_p f(c,p)| - \sum_{i=1}^{n-1} \frac{|D_p f(f^i(c,p),p)|}{K_E \lambda_E^i}]$$

Substituting (E.7) and (E.8):

$$\begin{aligned} |D_p f^n(x,p)| &> K_E \lambda_E^{n-1}[(\frac{1}{4} - \delta) - \sum_{i=1}^{N} \frac{\delta}{K_E \lambda_E^i} - \sum_{i=N+1}^{n-1} \frac{1}{4K_E} \lambda_E^i] \\ &> K_E \lambda_E^{n-1}[\frac{1}{4} - \delta - \frac{\delta}{K_E(1-\lambda_E^{-1})} - \frac{\lambda_E^{-(N+1)}}{4K_E(1-\lambda_E^{-1})}] \end{aligned}$$

for any $n \geq 1$. Now if if we set

$$C_E = [\frac{1}{4} - \delta - \frac{\delta}{K_E(1-\lambda_E^{-1})} - \frac{\lambda_E^{-(N+1)}}{4K_E(1-\lambda_E^{-1})}]$$

we see that $C_E > 0$ if $\delta > 0$ is sufficiently small and $N > 0$ is sufficiently large. From (E.7) and (E.8) we know that we have full control of $\delta > 0$ and $N > 0$ with our choice of ϵ_1. So choose $\epsilon_1 > 0$ small enough so that $C_E > 0$ for any $p \in [4 - \epsilon_1, 4]$. Then we have that:

$$|D_p f^n(x,p)| \;>\; K_E C_E \lambda_E^{n-1} \tag{E.10}$$

for all $n \geq 1$ if $p \in [4-\epsilon_1, 4]$ and f_p satisfies (CE1) (ie, $|D_x f^n(f(c,p),p)| > K_E \lambda_E^n$ for all $n \geq 1$). Looking at (E.6), it is also apparent that if (E.10) is satisfied, then since $|D_p f(f^{n-1}(c,p),p)| < \frac{1}{4}$, the sign of $D_p f^n(x,p)$ is governed by the signs of $D_x f(f^{n-1}(c,p),p)$ and $D_p f^{n-1}(c,p)$ for $n \geq 1$ sufficiently large. Thus, since f_p favors higher parameters at $p = 4$, there exists some $\epsilon > 0$ with $\epsilon < \epsilon_1$ such that f_p favors higher parameters if $p \in [4 - \epsilon, 4]$ and f_p satisfies (CE1).

Consequently, (CP1) must be satisfied and f_{p_0} favors higher parameters for any $p_0 \in [4 - \epsilon, 4]$ such that f_{p_0} satisfies (CE1). But recall that for any $\alpha > 0$, $S(\alpha)$ has a density point at $p = 4$ and $p_0 \in S(\alpha)$ implies that f_{p_0} satisfies (CE1). So let $S_*(\alpha) = S(\alpha) \cap [4 - \epsilon, 4]$. Then for any $\alpha > 0$ we can see that if $p_0 \in S(\alpha)$, then condition (E.1) is satisfied, f_{p_0} satisfies (CE1), and f_p satisfies (CP1) and favors higher parameters at $p = p_0$. Furthermore, $S_*(\alpha)$ has a density point at $p = 4$.

Now recall from section 3.3.1 that $n_e(c, \epsilon, p_0)$ is defined to be the smallest integer $n \geq 1$ such that $|f^n(c, p_0) - c| \leq \epsilon$. Thus, if (E.1) is satisfied, then

$$n_e(c, \epsilon, p_0) > -\frac{1}{\alpha} \log \epsilon. \tag{E.11}$$

But from theorem 3.3.1, we know that if f_{p_0} satisfies (CE1) and f_p satisfies (CP1) and favors higher parameters at $p = p_0 \in I_p$, then there exist constants $\delta > 0$, $K_0 > 0$, $K_1 > 0$ and $\lambda > 1$ such that there are no orbits of f_p which $\epsilon-$shadow the orbit, $\{f_{p_0}^i(c)\}_{i=0}^{\infty}$, if $p \in (p_0 - \delta, p_0 - K_0 \epsilon \lambda^{-n_e(c, K_1\epsilon, p_0)})$. Substituting in the condition (E.11) we find that:

$$K_0 \epsilon \lambda^{-n_e(c, K_1 \epsilon, p_0)} = K_0 (K_1 \epsilon)^{1 + \frac{1}{\alpha}\log \lambda}. \tag{E.12}$$

Now suppose we are given any $\gamma > 1$. We can see that if $\alpha < \frac{1}{\gamma - 1} \log \lambda$ then

$$1 + \frac{1}{\alpha} \log \lambda > \gamma. \tag{E.13}$$

So let $E(\gamma) = S(\frac{1}{2(\gamma-1)} \log \lambda)$. Note that E_γ has positive Lebesgue measure and a density point at $p = 4$. For any $\gamma > 1$, we also see that if $p_0 \in E(\gamma)$ then f_p satisfies (CP1) and (CE1) at $p = p_0$. Thus by theorem 3.3.1 and from (E.12) and (E.13) we have that if $p_0 \in E(\gamma)$ then no orbits of f_p $\epsilon-$shadow the orbit, $\{f_{p_0}^i(c)\}_{i=0}^{\infty}$, for any $p \in (p_0 - \delta, p_0 - K_0(K_1\epsilon)^\gamma)$. But since $\gamma > 1$, if we set constant $K = \max\{K_0 K_1, K_1\} > 0$ we see that $p_0 - K_0(K_1 \epsilon)^\gamma > p_0 - (K\epsilon)^\gamma$ for any $\epsilon > 0$. Thus, no orbits of f_p may $\epsilon-$shadow $\{f_{p_0}^i(c)\}_{i=0}^{\infty}$, if $p \in (p_0 - \delta, p_0 - (K\epsilon)^\gamma)$.

The final step is to show that almost any orbit of f_p comes arbitrarily close to c. This can be seen from the following two lemmas:

Lemma E.0.8 *Let U be a neighborhood of c. For any $p \in I_p$, if $E_U = \{x \mid f_p^n(x) \in I \setminus U$ for all $n \geq 0\}$ contains no non-trivial intervals, then the Lebesgue measure of E_U is zero.*

Proof of lemma E.0.8: See Theorem 3.1 in Gukkenheimer [26].

Lemma E.0.9 *If $p_0 \in I_p$ and f_{p_0} satisfies (CE1), then the set of preimages of c, $C_p = \cup_{i \geq 0} f_{p_0}^{-i}(c)$, is dense on I_x.*

Proof of lemma E.0.9: See corollary II.5.5 in Collet and Eckmann [14].

From these two lemmas we can see that for almost all $x_0 \in I_p$, the orbit, $\{f_{p_0}^i(x_0)\}_{i=0}^{\infty}$, approaches arbitrarily close to c if $p \in E(\gamma)$, for any $\gamma > 1$. Thus for almost all $x_0 \in I_p$, there are arbitrarily long stretches of iterates where the orbit, $\{f_{p_0}^i(x_0)\}_{i=0}^{\infty}$, looks arbitrarily close to the orbit, $\{f_{p_0}^i(c)\}_{i=0}^{\infty}$. This means that if there are no orbits of f_p

that can shadow $\{f^i_{p_0}(c)\}_{i=0}^{\infty}$, there can be no orbits of f_p that can shadow $\{f^i_{p_0}(x_0)\}_{i=0}^{\infty}$. Consequently for any $\gamma > 1$ if $p_0 \in E(\gamma)$ then f_{p_0} satisfies (CE1) and almost no orbits of f_{p_0} can be shadowed by any orbit of f_p if $p \in (p_0 - \delta, p_0 - (K\epsilon)^{\gamma})$. This proves parts (1) and (3) of theorem 3.4.2.

Part (2) of theorem 3.4.2 is a direct result of Corollary 3.3.1, Theorem 3.4.1, and the following result, due to Milnor and Thurston:

Lemma E.0.10 *The kneading invariant, $D(f_p, t)$, is monotonically decreasing with respect to p for all $p \in I_p$.*

Proof of lemma E.0.10: See theorem 13.1 in [34].

Thus if $p_0 \in E(\gamma)$ satisfies (CE1), there exists constant $C > 0$ such that if $p_0 \in E(\gamma)$ then any orbit of f_{p_0} can be $\epsilon-$shadowed by an orbit of f_p if $p \in [p_0, p_0 + C\epsilon^3]$. This is exactly part (2) of the theorem.

This concludes the proof of theorem 3.4.2.

Bibliography

[1] M. Adams and V. Guillemin, *Measure Theory and Probability*, Wadsworth, Belmont, CA (1986).

[2] D. Alspach and H. Sorenson, Nonlinear Bayesian estimation using Gaussian sum approximation, *IEEE Trans. Automatic Control*, Vol. AC-17, No. 4, (1972), 439-448.

[3] B. D. Anderson and J. Moore, *Optimal Filtering*, Prentice Hall, Englewood Cliffs, NJ (1979).

[4] D. V. Anosov, Geodesic flow on a riemannian manifold with negative curvature, *Trudy Math Just.*, Stekhlov, Moscow (1967).

[5] M. Benedicks and L. Carleson, The dynamics of the Henon map, *Ann. Math.*, **122** (133), (1991) 73-169.

[6] R. Bowen, ω-limit sets for Axiom A diffeomorphisms, *Journal of Differential Equations*, **18** (1975), 333-339.

[7] R. Bowen, *On Axiom A Diffeomorphisms*, CMBS Regional Conference Series in Mathematics, Vol. 35, AMS Publications, Providence (1978).

[8] E. Bradley, Taming chaotic circuits, Technical Report AI-TR 1388, MIT Artificial Intelligence Lab, September 1992.

[9] E. Bradley, Using chaos to broaden the capture range of a phase-locked-loop, *IEEE Transactions on Circuits and Systems*, November 1993.

[10] M. Casdagli, S. Eubank, J. D. Farmer, and J. Gibson, State space reconstruction in the prescence of noise, *Physica D*, **51** (1991), 52-98.

[11] K. Cuomo, A. Oppenheim, Circuit implementation of synchronized chaos with applications to communications, *Phys. Rev. Letters*, **71**, (1993), 65.

[12] L. Chen, Linking and the shadowing property for piecewise monotone maps, *Proc. Amer. Math. Soc.*, **113** (1991), 251-263.

[13] B. Chirikov, A universal instability of many dimensional ocillator systems, *Phys. Rep.*, **52**, (1979) 263-379.

[14] P. Collet and C. Eckmann, *Iterated maps on the interval as dynamical systems*, Birkhauser, Boston (1980).

[15] P. Collet and C. Eckmann, Positive Liapunov exponents and absolute continuity for maps of the interval, *Ergod. Th. and Dynam. Sys.* **3** (1983), 13-46.

[16] C. Conley, Hyperbolic invariant sets and shift automorphisms, *Dynamical Systems: Theory and Applications*, Lecture notes in physics, Vol. 38, Springer-Verlag, NY (1975), 539-549.

[17] E. Coven, I. Kan, and J. Yorke, Pseudo-orbit Shadowing in the family of tent maps, *Trans. Amer. Math. Soc.*, **308** (1988), 227-241.

[18] E. Coven and I. Mulvey, Transitivity and the center for maps of the circle, *Ergod. Th. and Dynam. Sys.*, **6** (1986), 1-8.

[19] R. Devaney, *An Introduction to Chaotic Dynamical Systems*, Menlo Park, CA, Addison-Wesley (1987).

[20] J.-P. Eckmann and D. Ruelle, Ergodic theory of chaos, *Reviews of Moden Physics*, **57**, (1985) 373-412.

[21] J. Franks, Absolutely structurally stable diffeomorphisms, *Proc. Amer. Math. Soc*, **37** (1973), 293-296.

[22] K. Gauss, *Theory of the motion of heavenly bodies.* Dover, NY, (1963).

[23] A. Gelb, *Applied Optimal Estimation*, Analytic Sciences Corp, Reading, MA (1974).

[24] C. Gregobi, S. Hammel, J. Yorke, and T. Sauer, Shadowing of physical trajectories in chaotic dynamics: containment and refinement, *Phys. Rev. Letters*, **65**, (1990) 1527-1530.

[25] J. Gukkenheimer and P. Holmes, *Nonlinear Oscillations, Dynamical Systems, and Bifurcation of Vector Fields*, Springer-Verlag, NY (1983).

[26] J. Gukkenheimer, Sensitive dependence to initial conditions for one dimensional maps, *Commun. Math. Phys.*, **70** (1979), 133-160.

[27] M. Henon, A two-dimensional mapping with a strage attractor, *Commun. Math. Phys.*, **50** (1976), 69-77.

[28] M. W. Hirsch and C. Pugh, Stable manifolds and hyperbolic sets, in *Global Analysis*, Vol. XIV (Proceedings of Symposia in Pure Mathematics), American Mathematical Society, Providence (1970), 133-163.

[29] A. Jazwinski, *Stochastic processes and filtering theory*, Academic Press Inc, NY, (1970).

[30] R. Kalman, New methods in wiener filtering, in *Proc. of the first symposium on engineering applications of random function theory and probability,*, John Wiley and Sons, NY, (1963).

[31] P. Kaminski, A. Bryson, and J. Schmidt, Discrete square root filtering: a survey of current techniques, *IEEE Trans. Automatic Control*, Vol AC-16, No. 6, (1971) 727-736.

[32] R. Lozi, Un attracteur etrange? du type attracteur de Henon, *J. Phys*, bf 39 (C5), 9-10.

[33] W. de Melo and S. van Strien, *One dimensional dynamics*, Springer-Verlag, NY (1993).

[34] J. Milnor and W. Thurston, On iterated maps of the interval, *Lecture notes in mathematics Vol. 1342*, Springer-Verlag, NY (1988), 465-563.

[35] R. Mane, *Ergodic Theory and Differentiable Dynamics*, Springer-Verlag, NY (1987).

[36] M. Misiurewicz, The Lozi mapping has a strange attractor, in *Nonlinear Dynamics,* R. H. G. Helleman (ed), New York Academy of Sciences: NY, (1980), 348-358.

[37] J. Moser, On a Theorem of Anosov, *Journal of Differential Equations*, **5**, (1969), 411-440.

[38] J. R. Munkres, *Elementary Differential Topology*, Princeton UP (1966).

[39] H. Nusse, J. Yorke, Is every approximate trajectory of some process near an exact trajectory of a nearby process, *Commun. Math. Phys*, **114**, (1988), 363-379.

[40] L. Ljung, *System Identification: Theory for the User*, Prentice Hall, Englewood Cliffs, NJ (1987).

[41] L. Ljung, Asymptotic behavior of the extended Kalman filter as a parameter estimator for linear systems, *IEEE Trans. Automatic Control*, Vol AC-24, No. 1, (1979), 36-50.

[42] S. Newhouse, Lectures in dynamical systrems, In *Dynamical Systems*, C.I.M.E Lectures Bressanone, Italy, June 1978, pp 1-114. Progress in Mathematics, No. 8, Birkhauser, Boston (1980).

[43] Z. Nitecki, *Differentiable Dynamics*, Cambridge, MIT Press (1971).

[44] T. Nowicki, A positive Liapunov exponent for the critical value of an S-unimodal mapping implies uniform hyperbolicity, *Ergod. Th. and Dynam. Sys.*, **5** (1988), 611-616.

[45] T. Nowicki and S. van Strien, Hyperbolicity properties of C^2 multimodal Collet-Eckmann maps without Schwarzian derivative assumptions, *Trans. Amer. Math. Soc.*, **321** (1990), 793-810.

[46] T. Nowicki and S. van Strien, Invariant measures exist under a summability condition for unimodal maps, *Invent. Math.*, **105** (1991), 123-136.

[47] D. S. Ornstein and B. Weiss, Statistical properties of chaotic systems, *Bull. Amer. Math. Soc.* (new series) **24** (1991), 11-116.

[48] V. Oseledec, A multiplicative ergodic theorem. Lyapunov characteristic numbers for dynamical systems, *Trans. Moscow Math. Soc*, *19*, (1968) 197-231.

[49] E. Ott, C. Grebogi, and J. Yorke, Controlling Chaos, *Phys. Rev. Letters* **64**, (1990) 1196-1199.

[50] J. Palis and W. de Melo, *Geometric Theory of Dynamical Systems: An Introduction*, Springer-Verlag, NY (1982).

[51] W. Parry, Symbolic dynamics and transformation of the unit interval, *Trans. Amer. Math. Soc.*, **122** (1964), 368-378.

[52] Ya. Pesin, Characteristic Lyapunov exponents and smooth ergodic theory, *Russ. Math. Surv*, **32**, 55-114.

[53] C. Pugh and M. Shub, Differentiability and continuity of invariant manifolds, in *Nonlinear Dynamics,* R. H. G. Helleman (ed), New York Academy of Sciences: NY, (1980), 322-329.

[54] D. Ruelle, Ergodic theory of differentiable dynamical systems, *Publ. Math. IHES*, **50**, (1979) 275-306.

[55] M. Shub, *Global Stability of Dynamical Systems*, Springer-Verlag, NY (1987).

[56] A. Siapas, *A global approach to parameter estimation of chaotic dynamical systems*, Master's thesis, Dept. of Electrical Engineering and Computer Science, MIT, 1992.

[57] J. Shapiro, A. Willsky, *6.432: Stochastic processes, detection, and estimation: Supplementary notes*, 1991.

[58] D. Singer, Stable orbits and bifurcations of maps of the interval, *SIAM J. Appl. Math*, **35** (1978), 260-267.

[59] S. Smale, Differentiable dynamical systems, *Bull. Amer. Math. Soc.*, **73** (1967), 747-817.

[60] S. van Strien, Hyperbolicity and invariant measures for general C^2 interval maps satisfying the Misiurewicz condition, *Commun. Math. Phys.*, **128** (1990), 437-496.

[61] G. Swiatek, Hyperbolicity is dense in the real quadratic family, Prepint, Stony Brook.

[62] P. Walters, On the pseudo orbit tracing property and its relationship to stability, *Lecture notes in mathematics Vol. 668*, Springer-Verlag, NY (1978), 191-210.