SHIFT EQUIVALENCE AND A COMBINATORIAL-TOPOLOGICAL APPROACH TO DISCRETE-TIME DYNAMICAL SYSTEMS

BY JUSTIN BUSH

A dissertation submitted to the

Graduate School—New Brunswick

Rutgers, The State University of New Jersey

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

Graduate Program in Mathematics

Written under the direction of

Konstantin Mischaikow

and approved by

_____

_____

_____

_____

New Brunswick, New Jersey

January, 2015

ABSTRACT OF THE DISSERTATION

Shift equivalence and a combinatorial-topological approach to discrete-time dynamical systems

by Justin Bush

Dissertation Director: Konstantin Mischaikow

Given a parameterized family of discrete-time dynamical systems, we aim to investigate how the global dynamics depends on the parameters in a way that is meaningful for applications. The discrete Conley index is an algebraic topological invariant of recurrent dynamics that is robust to small changes in parameters. Its definition, however, is given in terms of shift equivalence, which is not straightforward to compute in the category of abelian groups. We discuss the challenge of interpreting shift equivalnce, and give a construction that for every square integer matrix produces an interval map that giving rise to dynamics represented by that matrix. We conclude with applications of this approach to dynamical systems to the logistic map, Newton's method in the plane, and to population models in biology.

Acknowledgements

First and foremost I would like to thank my advisor, Konstantin Mischaikow, for all of his support and guidance. The work in this dissertation is greatly indebted to his influence and counsel.

Thank you to Chuck Weibel, Steve Ferry, and Sarah Day for serving on my thesis committee and for many helpful comments that have improved this dissertation. I am also grateful to Chuck for his insight into the algebra of shift equivalence, which influenced my presentation in Chapter 3.

Selections from Chapter 5 have been published previously in the journal *Chaos*, and a modified version of Chapter 7 has been published in *Entropy*. Other parts of this dissertation are currently in preparation to be submitted, as indicated in the citations throughout. I want to acknowledge all of my coauthors for their contributions. In particular, the figures in Chapter 6 are based on computations performed by Wes Cowan using software written by Shaun Harker, both of whom (along with Konstantin Mischaikow) are coauthors of the preprint on which the chapter is based.

I am an incredibly lucky person to have the family I do, and it is impossible for me to imagine accomplishing anything that I have without them. Thank you all for making my life so wonderful.

Finally, thank you to the many friends that were there while I was working on this dissertation. I appreciate everything you have added to my life.

Dedication

I dedicate this work to Charles Bush, Jr. and Attilio Necciai—the two people who would have been most proud had they been able to see this day—and also to Dave Williams, who would have made it clear that I'm still just a nephew he could tell embarrassing stories about.

Papa, Till, and Uncle Dave, I miss you.

Table of Contents

Chapter 1

Introduction

The impossibility of a complete understanding of most dynamical systems was recognized by Henri Poincaré well over a century ago [39]. One of the foremost achievements to that point in what we would today call dynamical systems was the prediction of the planetary orbits around the sun using Newtonian gravitation, carried out by Newton himself in the *Principia* [37]. As an expedient for solving for the orbits of each planet explicitly, however, he considered only the effect of gravity between the sun and each planet taken separately. This is, of course, a simplification of the full solar system, which would require modeling the effect of gravity of each planet on all the others.

Poincaré was interested the three-body problem: the problem of determining the future trajectories of three objects under Newtonian gravitation. Mathematically this amounts to a solution of a set of nonlinear ordinary differential equations. Poincaré's work showed, among other things, that in general there will be no analytic solution to such a system.

The three body problem is an example of what we would now call a continuous-time dynamical system. We will give a formal definition shortly, but informally a dynamical system is a mathematical model where the state of the world is represented by a point in space, and how the state of the world changes over time is given deterministically by a function. In the case of the three body problem, the state of the system is given by the positions and the velocities of each of the three bodies. This is considered a continuous-time dynamical system because the resulting differential equations give the future state of the system for any positive real-valued time.

With the advent of computational methods, mathematicians and scientists by the

1960s were able to see even more clearly with computer simulations the complicated behavior of dynamical systems that Poincaré was able to demonstrate by hand. Perhaps the most well known example is Lorenz and the discovery of the famous butterfly attractor while simulating a simple climate model [25]. Subsequent decades saw intensive investigation into the ubiquitous "chaos" present in dynamical systems [43].

To be more formal, we start with the following definition:

Definition 1.1. Let X be a metric spaces, and let $\mathbb{T}^+$ denote either the nonnegative reals $\mathbb{R}^+$ or the nonnegative integers $\mathbb{Z}^+$. A *dynamical system* on X is a continuous map

$$\varphi\colon \mathbb{T}^+ \times X \to X$$

subject the following semigroup condition for all $s, t \in \mathbb{T}^+$ and $x \in X$:

$$\varphi(t, \varphi(s, x)) = \varphi(s + t, x)$$

Any map $g\colon \mathbb{T}^+ \to X$ such that $\varphi(t, x_0) = g(t)$ is called a *solution* (or *trajectory*) of the system with initial condition x_0.

We call X the *phase space*, which we think of as representing the possible states of the system, and we almost always think of $\mathbb{T}^+$ as representing time.

Often instead of a single dynamical system in isolation we want to consider a collection of systems which depend continuously on some parameters. In the three body problem, for example, the masses of each of the three bodies affect the dynamics, but do not themselves change over time. Although these can be modeled as part of the state of the system, it is simpler in many cases to model them as parameters in the manner of the forthcoming definition. In many applications, the parameters of a system represent a quantity that is either unknown (in which case we may want to infer the parameters) or "tuneable" (in which case we may want to control the parameters) and so the how the dynamics change as parameters change will be of great interest.

Definition 1.2. Let Z be a metric space, and X and $\mathbb{T}^+$ as in Definition 1.1. Then a *dynamical system* on the parameter space Z is a continuous map

$$\varphi\colon \mathbb{T}^+ \times X \times Z \to X$$

subject to the condition

$$\varphi(t, \varphi(s, x, z), z) = \varphi(s + t, x, z)$$

for all $s, t \in \mathbb{T}^+$, $x \in X$, and $z \in Z$.

Some authors call the functions defined in Definitions 1.1 and 1.2 *semi-dynamical systems* because time is restricted to be positive. If time is instead taken to be all of $\mathbb{R}$ or $\mathbb{Z}$, the semigroup condition implies that the map $\varphi_t := \varphi(t, \cdot)$ is invertible for all t with inverse φ_{-t}. This is not the case with the definiton we use here, so what we call a dynamical system is more general than some other uses of the term.

The above definition could also be further generalized in several ways, for example taking X and Z to be general topological spaces. Because we need metric spaces for some of the theory we develop later (e.g. in order to define the diameter of a grid in Chapter 2), we choose to use this more restrictive definition. Indeed, for the calculations we perform in Chapters 5 and 7 we are computing on subsets of Euclidean space (and in Chapter 6 on Euclidean space with points identified) so that special case is often enough to bear in mind. But this means that a lot of the definitions and theorems we use along the way also hold in a more general setting than we use here.

Finally, a dynamical system where $\mathbb{T}^+ = \mathbb{Z}^+$ is called a *discrete-time* dynamical system, while one where $\mathbb{T}^+ = \mathbb{R}^+$ is called a *continuous-time* dynamical system. These, of course, are both subsumed under Definition 1.2, but in practice are distinct enough to often require separate mathematical methods. In this dissertation we will focus almost exclusively on discrete-time dynamical systems. Note that if we define a function $f_z(x) := \varphi(1, x, z)$, that we can express φ by the t-fold composition

$$\varphi(t, x, z) = f_z \circ f_z \circ \cdots \circ f_z(x) = f_z^t(x)$$

.

For this reason we will typically choose to represent a discrete-time dynamical system φ by the collection of maps $f\colon X \times Z \to X$ where $f(x,z) = f_z(x) = \varphi(1,x,z)$, and use the word *map* in certain contexts as a synonym for discrete-time dynamical system.

In some ways discrete-time systems are simpler and in some ways they are more complicated than continuous-time systems. One way discrete-time systems are more complicated is that chaotic behavior can arise even with a one-dimensional phase space, whereas it requires at least a three-dimensional phase space in continuous time. Moreover, the Conley index, which features prominently throughout this dissertation, is a more complicated invariant in discrete time. On the other hand when studying dynamics on a computer, a map is typically much easier to handle numerically when compared to integrating a flow. So for concrete problems that we wish to implement on a computer there are computational advantages to focusing on discrete time that motivate work on the more difficult theoretical aspects.

Despite the aforementioned hopelessness of completely understanding most dynamical systems, there are many impressive analytic results that go a long way towards giving both a qualitative and quantitative understanding of their complex behavior. As an example, Feigenbaum discovered that the ratios of the interval lengths in the period doubling cascade in the logistic map (about which we will say more in Chapter 2 and Chapter 5) is in fact a universal and regular phenomena [16]. This is an archetypal example of a remarkably general statement true of otherwise extremely complicated dynamics. Although from the point of view of someone using the logistic map or similar unimodal map to model population growth, we can reasonably wonder just how much period-doubling behavior we could expect to observe in practice, and whether the limit in which the Feigenbaum constant holds would even be observable. This relationship between the continuous model and what is mathematically robust and practically observable is a recurring theme.

In this thesis we follow a combinatorial-topological approach to dynamical systems outlined in [2] and [6]. Chapter 2 gives an overview of both the philosophy and the implementation of this approach to dynamics. Specifically, we discuss how we discretize phase and parameter spaces to achieve a combinatorial representation of the dynamics

of $f\colon X \times Z \to X$, and some motivations for approaching the problem in this way. We then discuss how the Conley index, an algebraic-topological invariant, allows us to draw conclusions about the dynamics.

The Conley index has an extensive literature, starting with [9]. Early work dealt mostly with the case of continuous-time flows, where the Conley index can be represented by the homotopy type of a topological space (in a way that we will define in Chapter 2) or, more commonly, by the homology groups of such a space. In discrete time, however, this does not suffice—the same dynamics can give rise to two spaces that are not homotopy equivalent when naïvely extending the continuous-time definition.

The index has been extended to the case of discrete-time dynamical systems in several ways [42, 32, 44, 17]. A general feature of these extensions is that, beyond just a space, one must consider endomorphisms along with an equivalence relation. In this dissertaion we follow the work of Franks and Richeson [17], who show that any two endomorphisms arising from the same dynamics (in a way to be made precise) are related by an equivalence relation called shift equivalence.

The definition of shift equivalence was first given by Williams [48], and has been used extensively in the study of symbolic dynamics. In that context the focus is on shift equivalence of positive integer matrices and much of the literature focuses on this case. But when working with the homological Conley index the entries of our matrices can be any integer (or potentially entries from a different ring entirely). Thus we are interested in establishing results about shift equivalence in the integral case. In Chapter 3 we discuss the problem of computing shift equivalence of matrices. As we will see, shift equivalence is readily computed when the matrices have field entries, but becomes much more delicate when the entries are integers.

Going from a dynamical system to the Conley index is deep theoretically, but thanks to extensive work can be carried out in an automated, algorithmic way [2, 6, 19]. Using the Conley indices to "go backwards" and make a statement about dynamics is far less straightforward, however. One way to try to get a handle on the dynamics associated to a Conley index is through examples. For some Conley indices there is a sensible choice for representative dynamics having that particular Conley index, but it is not

always easy to make explicit and to do so in a way that is conducive to computation. In Chapter 4 we demonstrate that for any Conley index represented by an integer matrix it is possible to construct a dynamical system $f\colon [0,1] \to [0,1]$ realizing that Conley index on the level of first homology. This construction gives an explicit example in the simplest possible setting. Furthermore, it provides a connection between the discrete Conley index as defined by Robbin and Salamon [42] which is extremely general and powerful but difficult to get a handle on, with the discrete Conley index defined by Franks and Richeson [17] which is given in terms of shift equivalence.

Ultimately we are interested in using the information we compute to draw conlcusions about specific dynamical systems. In the final three chapters we look at three systems in particular: the logistic map, Newton's method in the plane, and a family of age-structured population models.

The logistic map is a classical example in discrete dynamics that can be thought of as one of the simplest nonlinear discrete-time population models. The period doubling cascade observed for the logistic map has been extensively studied in the literature. Here we investigate the implications our approach of discretizing the dynamics has for understanding and representing what is happening. This serves as a nice test case for investigating the relationship between the scale of discretization of phase space and parameter space and the tradeoffs involved therein. This is the subject of Chapter 5, and comes from work in [6, 7].

The behavior of Newton's method for real-valued functions is well-understood. We know that the rate of convergence is quadratic, but moreover that the direction in which it will converge is determined by the sign of $f''(x)/f'(x)$ where x is a root of f. For Newton's method in $\mathbb{R}^2$, however, the direction in which the approximations to the root converge is more complicated. In Chapter 6 we model the problem as a dynamical system $S^1 \times Z \to S^1$ and apply our methods with the goal of better understanding the dynamics. The hope is that this could lead to improved numerical performance of algorithms for Newton's method. This is joint work with Wes Cowan, Shaun Harker, and Konstantin Mischaikow [5].

The logistic map is, as we mentioned, one of the simplest nonlinear discrete-time

population models. More realistic population models take into account things like the age-structure of the population, and can use different nonlinearities. Even superficially similar choices of nonlinearity can have implications for the dynamics on parameter space, however. We investigate this phenomena in Chapter 7, which comes from work published with Konstantin Mischaikow [8].

Chapter 2

A coarse approach to dynamics

To motivate the approach we take towards dynamical systems consider the logistic map, a discrete-time dynamical system defined by the function

$$f(x, r) = rx(1 - x) \tag{2.1}$$

Here $x \in [0, 1]$ is the phase variable, and r is a parameter which we take to be in $[0, 4]$ in order to make f is a well defined map on the unit interval.

The logistic map is typically introduced as a simple nonlinear population model and a discrete time analogue to the logistic differential equation $\frac{dx}{dt} = rx(1 - x)$. As a population model, the logistic map determines the population at the next time period by applying f to an initial population x_0. The parameter r controls the growth rate of the population for small values of x. The $(1 - x)$ term in the map produces an overcrowding effect, where for initial populations $0.5 < x_1 < x_2$ the immediately subsequent populations satisfy $f(x_2) < f(x_1)$.

A classical and important approach to studying dynamical systems is to be concerned with the invariant sets of a particular system. The simplest example of an invariant set is an equilibrium. Formally:

Definition 2.1. Let $\varphi\colon \mathbb{T}^+ \times X \to X$ be a dynamical system. Then a subset $S \subseteq X$ is an *invariant set* if for all $t \in \mathbb{T}^+$, the set $\varphi(t, S) := \{\varphi(t, s) : s \in S\}$ satisfies $\varphi(t, S) = S$.

If φ is a discrete-time dynamical system represented by the map $f : X \to X$, this definition is equivalent to $f(S) = S$.

Definition 2.2. A point $x \in X$ is called an *equilibrium* (or *fixed point*) if $\{x\}$ is an invariant set. Furthermore we say x is

a) a *stable equilibrium* if x is an equilibrium and furthermore for every $\epsilon > 0$ there exists $\delta > 0$ such that if $|y - x| < \delta$ then $|\varphi(t, y) - x| < \epsilon$ for all $t \in \mathbb{T}^+$.

b) an *unstable equilibrium* if x is an equilibrium but it is not stable.

c) an *asymptotically stable equilibrium* if x is both a stable equilibrium and there exists $\delta' > 0$ such that if $|y - x| < \delta'$ then $\lim_{t\to\infty} \varphi(t, y) = x$.

As a simple example, the point $\bar{x} = \frac{r-1}{r}$ is a fixed point of the logistic map, but one whose stability depends on r. It is asymptotically stable for $1 < r < 3$, stable but not asymptotically stable for $r = 3$, and unstable for $r > 3$.

For the logistic differential equation, every solution either tends to an equilibrium or diverges to infinity. And for positive initial conditions and values of r, every solution tends to the asymptotically stable equilibrium $x = 1$. For the logistic map, however, there are parameter values in $[0, 4]$, namely $r > 3$, with solutions that do not approach an equilibrium—this, despite the superficial similarity to the logistic differential equation. As an example of different behavior, for some parameters solutions to the logistic map will, instead of approaching an equilibrium, approach a periodic orbit, defined as follows:

Definition 2.3. A solution $g(t)$ to a dynamical system $\varphi : \mathbb{T}^+ \times X \to X$ is a periodic orbit if there exists $T > 0$ such that $g(t + T) = g(t)$. The minimal T for which this holds is called the *period* of the periodic orbit.

For yet other parameters, however, the behavior of the logistic map is extremely complicated and will not approach anything as simple as a periodic orbit. Moreover, the dependence of the dynamics on parameters is extremely sensitive, in a way that also has no analogue in the logistic differential equation. Thus the logistic map serves as important example of how discrete time dynamics can exhibit much more complicated behavior than continuous time dynamics in low dimensions. We discuss implications of these aspects of the logistic map in greater detail in Chapter 5.

From a philosophical perspective, we can think more concretely about what it means to say that the logistic map is a population model. There are lots of simplifications

and idealizations that take place when taking the formal mathematical setup above and making claims about what it represents in the real world. One obvious idealization, for example, is that we are taking our population to be a real number. In some cases this is perhaps easily justified—say in the case of bacteria where mass provides a useful measure of population. In applications where the population is more discrete, however, this assumption requires further examination. Outlining some of the problematic aspects of modeling a real-world system mathematically here is meant to motivate the mathematical approach to dynamics we describe in the remainder of the chapter. The following list of sources of potential error is by no means exhaustive.

- *Continuous approximation* - In some cases, like the population example discussed above, it may not be clear whether the phase space is actually well approximated by a continuous space. Furthermore, the assumption that future states depend continuously on past states may not be reasonable. Even small, imperceptible discontinuities could undermine certain analytic approaches to investigating such systems.

- *Measurement uncertainty* - Suppose a scientist takes the logistic map seriously as a model for the population dynamics of some species. In practice, both the parameter r and the state x of the real-world system are things that have to be either measured or inferred in some way. And measurement is in most cases approximate, if only because there are only so many marks on a metaphorical ruler. The question then becomes to what extent does this imprecision in measurement diminish the usefulness of the mathematical model, especially given well known phenomena like sensitive dependence on initial conditions and sensitive dependence on parameters.

- *Measurement interference* - Suppose the scientist is less certain that the logistic map is the right population model, and wants to consistently measure the population to verify that the dynamics match what the logistic map predicts. To the extent that measuring the system interferes with it, even imperceptibly, it will be impossible to verify that the logistic map is indeed the correct model. Again,

this is a consequence of sensitive dependence on initial conditions, and the fact that any interference will have discontinuous effects. It might be that the best that can be hoped after a measurement is that the system be the same as it was pre-measurement to the extent that our metaphorical ruler allows us to determine.

- *Model uncertainty* - Suppose a scientist isn't sure whether the logistic map is the right model for the population dynamics of some species. We would be interested in a test that could be performed that could determine whether the model is a good one. Or, conversely, we might want a test that could definitively rule out all the possible parameter values for the logistic map, "falsifying" it in some sense [40].

- *Representation on a computer* - When implementing a dynamical system on a computer for simulation, it will in most cases necessarily be an approximation. The same features that make measurement uncertainty so problematic—sensitive dependence on initial conditions and sensitive dependence on parameters—put practical limits on the accuracy and usefulness of such simulations.

None of these considerations by any means invalidate the approach of using a dynamical system like the logistic map to model real-world phenomena. It is an advantage of a mathematical approach, however, insofar as it addresses and assuages these concerns. In what follows for the remainder of this chapter, we describe the background of a computational approach to dynamical systems using the Conley index. This approach we will refer to as the *database of dynamics*, or *database* for short, and has been described in the literature [2, 6] which we follow here. The database approach to dynamics has two principal aspects:

1. Representing continuous dynamics in a discrete, combinatorial way.

2. Computing the Conley index, which rigorously relates the combinatorial dynamics back to the underlying continuous dynamics.

Each of these will be elaborated on further in the remainder of this chapter. It is

worth mentioning, however, in light of the sources of error listed above, the conceptual distinction between treating the discretizations performed as an approximation to an underlying idealized model, and between treating them as the model itself. The relation between discrete and continuous, of course, is the source of much of the mathematical substance of the theory. But from an application perspective, especially in Chapters 5 and 7, we take seriously the idea of the combinatorial model perhaps being more meaningful, because of all the uncertainties and imprecisions required for certain applications.

2.1 Combinatorialization

Consider a discrete-time dynamical system defined by the map $f : X \times Z \to X$, and recall that we are taking X and Z to be metric spaces. In order to study dynamical systems using computational tools, we need a means of representing the dynamics in a finite way. The first step in doing this is to discretize the spaces X and Z into a finite number of pieces. For this purpose, we make the following definition, which is due to Mrozek [33]. In what follows, we use the notation int to denote interior, cl to denote closure, and bd to denote boundary.

Definition 2.4. A *grid* on a metric space X is a finite collection $\mathcal{X}$ of nonempty, compact subsets of X with the following properties:

i) $X = \bigcup_{\xi \in \mathcal{X}} \xi$

ii) $\xi = \mathrm{cl}(\mathrm{int}(\xi))$ for all $\xi \in \mathcal{X}$

iii) $\xi \cap \mathrm{int}(\xi') = \emptyset$ for all $\xi \neq \xi'$

For a subset $\mathcal{A} \subseteq \mathcal{X}$ we denote $\bigcup_{\xi \in \mathcal{A}} \xi$ by $|\mathcal{A}|$. We further define the *diameter* of a grid $\mathcal{X}$ by

$$\mathrm{diam}(\mathcal{X}) = \sup_{\xi \in \mathcal{X}} \mathrm{diam}(\xi).$$

An example of a grid on the n-cube $[0,1]^n \subseteq \mathbb{R}^n$ is the set

$$\left\{ \prod_{i=1}^{n} [c_i 2^{-k}, (c_i + 1)2^{-k}] \mid c_i \in \{0, \ldots, 2^k - 1\} \right\}.$$

This is an example of a cubical grid, as each element of the grid is a product of intervals. For the computations performed in this thesis, all of the grid elements will be represented as cubes in this way. As far as intuition, very little is lost by thinking of phase space X as a subset of $\mathbb{R}^n$ and parameter space Z as a subset of $\mathbb{R}^m$ that can be represented by cubical grids $\mathcal{X}$ and $\mathcal{Z}$, respectively. Note that in this setting we can, by choosing k large enough create a grid of arbitrarily small diameter. Even in settings where there may be no cubical grid, a grid of arbitrarily small diameters can always be found for any compact metric space [21].

Given a parameterized dynamical system, for each $z \in Z$ we can define the map $f_z \colon X \to X$ by $f_z(x) = f(x, z)$. To translate this map into the setting of grids, we use the notion of a *multivalued map* $\mathcal{F}_z : \mathcal{X} \rightrightarrows \mathcal{X}$, where here the double arrow notation denotes that this is a set-valued function. Equivalently, we can think of $\mathcal{F}_z$ as a directed graph on the vertex set $\mathcal{X}$, where there is an edge $\xi_1 \to \xi_2$ if and only if $\xi_2 \in \mathcal{F}_z(\xi_1)$. We will use whichever conceptual framework is more convenient depending on the context.

The idea is to define $\mathcal{F}_z$ in such a way that it captures important information about the underlying dynamics of f_z. In order to apply important theorems pertaining to the Conley index, it is important that our multivalued map satisy certain conditions, namely that it be an outer approximation, defined as follows:

Definition 2.5. A combinatorial multivalued map $\mathcal{F}_z \colon \mathcal{X} \rightrightarrows \mathcal{X}$ is an *outer approximation* of f_z, if

$$f_z(\xi) := \{f_z(x) \mid x \in \xi\} \subseteq \operatorname{int}(|\mathcal{F}_z(\xi)|) \quad \text{for all } \xi \in \mathcal{X}.$$

For a map f_z there will in general be many valid outer approximations. Some of these are trivial—the graph with a directed edge between every pair of grid elements in $\mathcal{X}$ is an outer approximation of every map, for instance. Such an outer approximation cannot possibly convey any interesting information about the underlying dynamics, however, simply because whatever we might prove using such an outer approximation will necessarily be true for *every* dynamical system. This illustrates in an extreme way that the choice of outer approximation has implications for the quality of the information about the underlying dynamics that we can extract.

There are advantages worth noting about the fact that a single outer approximation can be used to represent more than one map. Any outer approximation, for example, will be somewhat robust to perturbations in the map. In particular for any outer approximation $\mathcal{F}_z$ of f_z there exists a $\delta > 0$ such that if $|f_z(x) - g(x)| < \delta$ for all $x \in X$, this implies $\mathcal{F}_z$ is also an outer approximation for g.

The best possible outer approximation for f_z on a grid $\mathcal{X}$ is the minimal outer approximation defined by

$$\widetilde{\mathcal{F}_z}(\xi) := \{\xi' \mid \xi' \cap f_z(\xi) \neq \emptyset\}$$

This is minimal, because any other outer approximation $\mathcal{F}_z$ of f_z satisfies $\widetilde{\mathcal{F}_z}(\xi) \subseteq \mathcal{F}_z(\xi)$ for all $\xi \in \mathcal{X}$ [21].

The minimal outer approximation is in some sense the best discrete representation of the dynamics for a given grid. In many cases, however, the computational cost of finding the minimal outer approximation is substantial, and it is more efficient to use another outer approximation that is easier to compute. This potentially comes at the price of having a coarser representation of the dynamics in the account we give in the sections that follow.

All of the definitions above are given in terms of a fixed map f_z for a particular parameter value $z \in Z$. In order to stay true to our goal of representing the dynamics in a finite way, we need to perform a similar discretization in parameter space as we did in phase space. Let $\mathcal{Z}$ be a grid on Z as in Definition 2.4. For every grid element $\zeta \in \mathcal{Z}$, an outer approximation $\mathcal{F}_\zeta \colon \mathcal{X} \rightrightarrows \mathcal{X}$ is any multivalued map satisfying

$$f_z(\xi) \subseteq \operatorname{int}(|\mathcal{F}_\zeta(\xi)|) \quad \text{for all } \xi \in \mathcal{X} \text{ and } z \in \zeta$$

In other words an outer approximation $\mathcal{F}_\zeta$ is any multivalued map that is an outer approximation for every $z \in \zeta$.

2.2 Recurrent and nonrecurrent dynamics

Given a choice of grids $\mathcal{X}$ and $\mathcal{Z}$ and an outer approximation $\mathcal{F}_\zeta \colon \mathcal{X} \rightrightarrows \mathcal{X}$ for some $\zeta \in \mathcal{Z}$, we want to extract from the multivalued map information about the recurrent

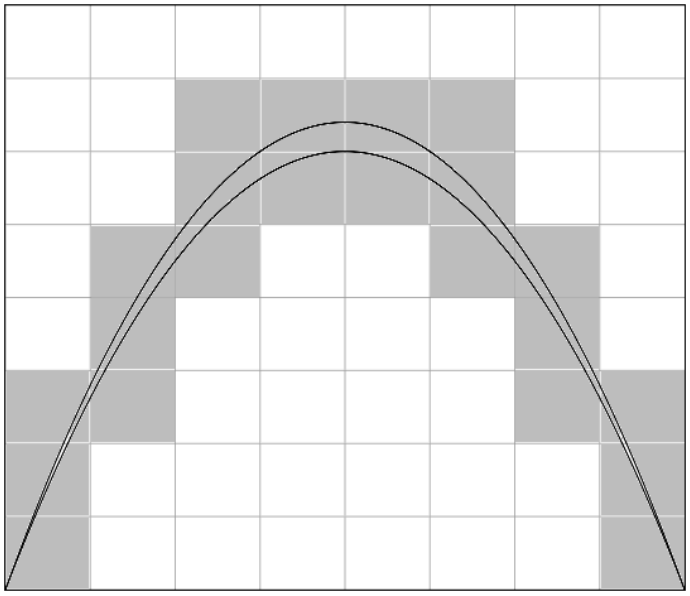


Figure 2.1: An illustration of a minimal outer approximation for the logistic map, Equation 2.1. The lower parabola is the graph of $r = 3$, while the upper is for $r = 3.2$. For each grid element contained in the domain (i.e. one of the eight intervals along the x-axis), the outer approximation of the grid element under f for $r \in [3, 3.2]$ is indicated by the shaded boxes above the domain grid element. A shaded box corresponds to an edge in the multivalued map from the domain grid element to the range grid element. This outer approximation is minimal because removing any shaded box fails to yield an outer approximation.

and nonrecurrent dynamics. The goal is for the theorems we prove about $\mathcal{F}_\zeta$ to be true statements about the dynamics for every $z \in \zeta$. We start with the following definition.

Definition 2.6. A *strongly connected path component* of a directed graph $\mathcal{F}_\zeta$ on the vertex set $\mathcal{X}$ is a subset $\mathcal{Y} \subseteq \mathcal{X}$ such that for $v_1, v_2 \in \mathcal{Y}$ (not necessarily distinct) there exists a directed path from v_1 to v_2.

A strongly connected path component represents recurrence on the graph level. From this definition, there is nothing preventing a strongly connected component from being contained in a larger strongly connected component, but conceptually we want to partition the dynamics into disjoint components reflecting recurrent dynamics, along with the remainder which contains all the nonrecurrent dynamics. To that end, we make the following definitions:

Definition 2.7. A *Morse set* $\mathcal{M}_\zeta \subseteq \mathcal{X}$ is a maximal strongly connected path component of $\mathcal{F}_\zeta$. The *Morse decomposition* of $\mathcal{F}_\zeta$ is the set of Morse sets.

Remark 2.8. In some contexts, a Morse set is defined to be $\operatorname{Inv}|\mathcal{M}_\zeta|$, making it an object of the underlying continuous dynamics. Here we choose instead to define it in terms of the discretization. These are not precisely equivalent: although $\mathcal{M}_\zeta$ clearly determines $\operatorname{Inv}|\mathcal{M}_\zeta|$, the converse is not true, so different outer approximations can yield different Morse sets under our definition while giving the same Morse sets under the invariant set definition.

The choice to define Morse sets at sets of grid elements has a few advantages. First, the Morse sets that result with either definition are necessarily dependent on the choice of discretization and outer approximation. This is a consequence of the fact that the Morse decomposition is not in general unique, so allowing for more possible decompositions is at least in accordance with the dependence on Morse decomposition on various choices. Second, this allows us to avoid some awkwardness when there are two distinct Morse sets $\mathcal{M}_\zeta^1$ and $\mathcal{M}_\zeta^2$ such that $\operatorname{Inv}\mathcal{M}_\zeta^1 = \operatorname{Inv}\mathcal{M}_\zeta^2 = \emptyset$, because we would prefer not to identify the two in this situation. The workaround for the invariant set definition usually involves a separate index set for the partial order, which we are able to omit.

Let $\mathsf{M}_\zeta = \{\mathcal{M}_\zeta^i\}$ be the Morse decomposition for $\mathcal{F}_\zeta$. Because Morse sets are maximal subsets of $\mathcal{X}$ defined by an equivalence relation, we have $\mathcal{M}_\zeta^i \cap \mathcal{M}_\zeta^j = \emptyset$ whenever $i \neq j$. Moreover, $\mathcal{F}_\zeta$ induces a partial order on M_ζ by stipulating $\mathcal{M}_\zeta^i \leq \mathcal{M}_\zeta^j$ if and only if there is a directed path in the graph $\mathcal{F}_\zeta$ from a grid element in $\mathcal{M}_\zeta^j$ to a grid element in $\mathcal{M}_\zeta^i$. Note that because Morse sets are maximal, no distinct elements are equivalent in this partial order.

This partial order encodes information about the underlying non-recurrent dynamics as follows. If $\mathcal{M}_\zeta^i < \mathcal{M}_\zeta^j$ then for any $x_i \in \xi_i \in \mathcal{M}_\zeta^i$ and $x_j \in \xi_j \in \mathcal{M}_\zeta^j$ there is no trajectory from x_i to x_j. (Otherwise such a trajectory would define a path from $\mathcal{M}_\zeta^i$ to $\mathcal{M}_\zeta^j$ in $\mathcal{F}_\zeta$, by the assumption that $\mathcal{F}_\zeta$ is an outer approximation.)

To more concisely represent the partial order on the Morse sets, we often choose to represent it as a directed acyclic graph, using the Hasse diagram [11]. To recover the partial order information from the Hasse diagram simply take the transitive closure of the relation induced by the edges of the directed acyclic graph.

Definition 2.9. The *Morse graph* MG_ζ of $\mathcal{F}_\zeta$ is the Hasse diagram for the Morse decomposition M_ζ of $\mathcal{F}_\zeta$ along with its associated partial order defined by $\mathcal{F}_\zeta$

2.3 Conley index

What is missing thus far is an account of the structure of the recurrent dynamics within each Morse set. This is accomplished with the Conley index. The classical approach to studying dynamical systems focuses on the study of invariant sets (see Defintion 2.1). We denote by $\mathrm{Inv}(N, f)$ the maximal invariant set contained in a neighborhood $N \subseteq X$ under f.

An invariant set S is said to be *isolated* if there exists a compact neighborhood N such that $\mathrm{Inv}(N, f) = S \subseteq \operatorname{int} N$. Such a neighborhood N is called an *isolating neighborhood* for S. Note, for example, that a single point of the identity map on a space is an invariant set (it is a fixed point) but will not be an isolated invariant set.

The Conley index is a topological invariant of isolated invariant sets that is robust to small changes in parameters. The original reference for the Conley index is [9],

which discusses the index (under the name Morse index) in the case of continuous-time dynamical systems. The spirit of Conley's work—hence the name Morse index—is to generalize the Morse index for critical points of vector fields to more general isolated invariant sets. Informally, the Conley index associates to an invariant set the homotopy type of a topological space. As an example, in the case of critical points, an index n critical point is associated to the homotopy type of the sphere S^n.

In the case of maps, some care needs to be taken to generalize the definition of the continuous-time Conley index. There have been several different formulations of a suitable alternative invariant [42, 32, 44]. The Conley index, whether continuous or discrete, is typically defined for a pair of topological spaces (P_1, P_0), $P_0 \subseteq P_1 \subseteq X$, typically called an *index pair*. In the continuous time, the homotopy type of P_1/P_0 suffices to define an invariant, but in discrete-time extra information is needed. The different approached to the discrete Conley index reflect different ways of encoding this additional information, and different levels of generality. Here we follow the definition of Franks and Richeson [17] and give a definition in terms of *filtration pairs*. This allows us to use the characterization of the Conley index as a shift equivalence class, about which we say more in Chapter 3.

Definition 2.10. Let $f\colon X \to X$ be a discrete time dynamical system, and $P_0 \subseteq P_1$ compact subsets of X that are closures of their interiors. Then the pair $P = (P_1, P_0)$ is a *filtration pair* for f if

1. $\mathrm{cl}(P_1 \setminus P_0)$ is an isolating neighborhood of $\mathrm{Inv}(P_1 \setminus P_0)$
2. P_0 is a neighborhood of the exit set $E = \{x \in P_1 \mid f(x) \notin \operatorname{int} X\}$ in P_1.
3. $f(P_0) \cap \mathrm{cl}(P_1 \setminus P_0) = \emptyset$

Furthermore, we say that P is a filtration pair for the isolated invariant set S, if $S = \mathrm{Inv}(\mathrm{cl}(P_1 \setminus P_0))$.

We will use filtration pairs in a topological construction that we use to define the Conley index. Recall that a *pointed topological space* is a pair (V, v_0) where V is a topological space and basepoint $v_0 \in V$ is a distinguished point. A continuous map

between pointed topological spaces $g\colon (V, v_0) \to (W, w_0)$ is a continuous map from V to W with the property that $g(v_0) = w_0$.

Consider a map $g\colon X \to X$ representing a discrete-time dynamical system, and $P = (P_1, P_0)$ a filtration pair. Let $(P_1/P_0, [P_0])$ denote the pointed topological space where $[P_0]$ represents the equivalence class of P_0 under the quotient [34]. Define $g_P\colon (P_1/P_0, [P_0]) \to (P_1/P_0, [P_0])$ by

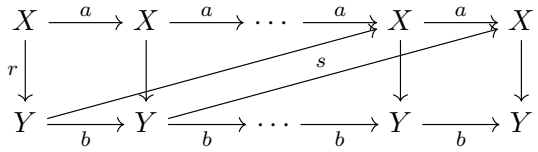
$$g_P([x]) = \begin{cases} g(x) & \text{if } x, g(x) \in P_1 \setminus P_0 \\ [P_0] & \text{otherwise} \end{cases}$$

Ultimately, it is this map g_P that we will use to carry information about the dynamics of g on $P_1 \setminus P_0$, but in order to do this we need to first define shift equivalence

Definition 2.11. Let $\mathcal{C}$ be any category. For endomorphisms a and b on the objects X and Y respectively, we say that a is *shift equivalent* to b, denoted $a \sim b$, if there exist morphisms $r : X \to Y$, $s : Y \to X$ and a positive integer l called the *lag* such that the following four conditions hold:

(i) $r \circ a = b \circ r$ (ii) $a \circ s = s \circ b$ (iii) $s \circ r = a^l$ (iv) $r \circ s = b^l$

These equations represent the commutivity of the following diagram:



The important result for the definition of the Conley index is as follows.

Theorem 2.12 ([17], Theorem 4.3)**.** *Let $S \subseteq X$ be an isolated invariant set for the map $g : X \to X$, and (P_1, P_0), (Q_1, Q_0) filtration pairs for S. Then the induced maps g_P and g_Q are shift equivalent.*

An important question for our approach to dynamics is how to go about finding filtration pairs from our combinatorial representation. The following Proposition establishes the connection.

Proposition 2.13. *Let* $\mathsf{M}_\zeta = \{\mathcal{M}_\zeta^i\}$ *be the Morse decomposition for an outer approximation* $\mathcal{F}_\zeta \colon \mathcal{X} \rightrightarrows \mathcal{X}$*. Then for each* $\mathcal{M}_\zeta^i$ *define*

$$P_1 := \left|\mathcal{F}_\zeta(\mathcal{M}_\zeta^i)\right|, P_0 := \left|\mathcal{F}_\zeta(\mathcal{M}_\zeta^i) \setminus \mathcal{M}_\zeta^i\right|$$

Then (P_1, P_0) *is a filtration pair for every* $z \in \zeta$*.*

Proof. First, we check that both P_1 and P_0 are are closures of their interiors. Let P be any finite union of grid elements $\bigcup_{i=1}^n \xi_i$, where $\xi_i = \mathrm{cl}(\mathrm{int}\, \xi_i)$ by the definition of grid. Because P is closed (as a finite union of compact sets), we trivially have that $\mathrm{cl}(\mathrm{int}\, P) \subseteq P$. To show $P \subseteq \mathrm{cl}(\mathrm{int}\, P)$, note that $\mathrm{int}\, xi_i \subseteq \mathrm{int}\, P$ for every i, hence $\mathrm{cl}(\mathrm{int}\, \xi_i) \subseteq \mathrm{cl}(\mathrm{int}\, P)$ for every i, hence $P = \bigcup_{i=1}^n \xi_i \subseteq \mathrm{cl}(\mathrm{int}\, P)$. Since P_1 and P_0 are unions of grid elements, each is the closure of its interior.

That $\mathrm{cl}(P_1 \setminus P_0)$ is an isolating neighborhood follows from Proposition 2.8 in [6]. To see that P_0 is a neighborhood of the exit set E relative to P_1, suppose for the sake of contradiction that $x \in E$ but $x \notin \mathrm{int}\, P_0$ (relative to P_1. Because $x \in E$, this means that there exists ξ_1 such that $f(x) \in \xi_1$, but $\xi_1 \notin P_1$. Because $\mathcal{F}_\zeta$ is an outer approximation, then for every grid element ξ such that $x \in \xi$, $\xi_1 \in \mathcal{F}_\zeta(\xi)$. At the same time, because $x \notin \mathrm{int}\, P_0$ then $x \in \mathrm{cl}(P_1 \setminus P_0)$, which implies $x \in \xi_2$, for some $\xi_2 \in P_1 \setminus P_0$. By the definitions of P_1 and P_0, this means that $\xi_2 \in \mathcal{M}_\zeta^i$. But this implies $\xi_1 \in \mathcal{F}_\zeta(\xi_2)$ hence $\xi_1 \in P_1$, contradicting the fact that ξ_1 was chosen to not be in P_1.

Finally to verify condition 3, suppose for the sake of contradiction that there exists $x \in P_0$ such that $f_z(x) \in \mathrm{cl}(P_1 \setminus P_0)$ for some $z \in \zeta$. Then in particular $f_z(x) \in P_1$ since P_1 is closed. Because $x \in P_0$ it must be that there exists ξ such that $x \in \xi \in \mathcal{F}_\zeta(\mathcal{M}_\zeta^i \setminus \mathcal{M}_\zeta^i)$. Since $\xi \in \mathcal{F}_\zeta(\mathcal{M}_\zeta^i)$ then there must exist $\xi_1 \in \mathcal{M}_\zeta^i$ such that $\xi \in \mathcal{F}_\zeta(\xi_1)$. But then by our hypothesis that $f_z(x) \in \mathrm{cl}(P_1 \setminus P_0)$, there must exist $\xi_2 \in \mathcal{M}_\zeta^i$ such that $\xi_2 \in \mathcal{F}_\zeta(\xi)$. Since for any grid element η there is a directed path from η to ξ_1, there must therefore be a directed path from η to ξ. Similarly, there is a directed path from ξ_2 to η, hence a directed path from ξ to η. Therefore, by the maximality of Morse sets, it must be the case that $\xi \in \mathcal{M}_\zeta^i$ which contradicts $\xi \notin \mathcal{M}_\zeta^i$ as required by the definition of P_0.

□

To define the discrete Conley index, we need to be explicit about what object, precisely we are associating an index to. Traditionally, the Conley index has been defined for isolated invariant sets. Following Franks and Richeson we can make the following definition:

Definition 2.14. Let S be an isolated invariant set for $f\colon : X \to X$ and $P = (P_1, P_0)$ a filtration pair for S. Let $[f_P]$ denote the homotopy class of the map $f_P\colon (P_1/P_0, [P_0]) \to (P_1/P_0, [P_0])$. Then the *homotopy Conley index* of S is the shift equivalence class of $[f_P]$.

In general, however, homotopy type is too difficult to get a handle on directly. Instead, we typically choose to pass via a functor to something more readily computable. In this dissertation, we typically choose to work with homology.

Definition 2.15. Let S be an isolated invariant set for $f\colon : X \to X$ and $P = (P_1, P_0)$ a filtration pair for S. Then the *homology Conley index* of S is the shift equivalence class of $f_{P*}\colon H_*(P_1/P_0, [P_0]) \to H_*(P_1/P_0, [P_0])$.

It is an important fact about the combinatorial approach we employ that in many cases these homology classes of a map f_z are recoverable just from the discrete representation of the map $\mathcal{F}_\zeta$. This will be true, for instance, under the assumption that $|\mathcal{F}_\zeta(\xi)|$ is acyclic for every $\xi \in \mathcal{X}$. More details are found in [2, 6, 19]. For our purposes here, what is important is that whenever the map on homology can be computed for a Morse set $\mathcal{M}_\zeta$ we are able to assign a Conley index to $\mathcal{M}_\zeta$, namely the Conley index of $S = \operatorname{Inv} |\mathcal{M}_\zeta|$. Attaching this Conley index data to the Morse graph, we get the *Conley-Morse graph.*

In Chapters 5–7 we take a more global view of parameter space, and aim to relate dynamics on different grid elements $\zeta, \zeta' \in \mathcal{Z}$. For the database approach, we compute two different Morse graphs based on the different multivalued maps $\mathcal{F}_\zeta$ and $\mathcal{F}_{\zeta'}$. If $\zeta \bigcap \zeta' \neq \emptyset$ then we say ζ and ζ' are in the same *isolating neighborhood continuation class* if each Morse set of $\mathcal{F}_\zeta$ intersects a unique Morse set of $\mathcal{F}_{\zeta'}$ and each Morse set

of $\mathcal{F}_{\zeta'}$ intersects a unique Morse set of $\mathcal{F}_{\zeta}$. This is an equivalence relation that can be extended to Morse sets in non-adjacent parameter boxes by transitivity. By [29, Theorem 3.10] any two Morse sets in the same isolating neighborhood continuation class have the same Conley index.

The adjacent parameter cubes ζ, ζ' are said to be in the same *Conley-Morse graph continuation class* when their Conley-Morse graphs are isomorphic and the isomorphism gives an isolating neighborhood continuation for each of the nodes. Two classes CMCC_i and CMCC_j are adjacent if there exists $\zeta_i \in \mathsf{CMCC}_i$ and $\zeta_j \in \mathsf{CMCC}_j$ such that $\zeta_i \bigcap \zeta_j \neq \emptyset$. These adjacencies can then be summarized in a graph called the *continuation graph*. In the case where parameter space is high dimensional and therefore difficult to visualize, the continuation graph gives an interpretable representation of parameter space.

Chapter 3

Problem of shift equivalence

Shift equivalence is first discussed extensively by Williams in [48] as an invariant of symbolic dynamics. In that context, edge shift dynamics can be represented by square matrices with nonnegative integer entries, and shift equivalence is an invariant of the dynamics up to conjugacy. For more about shift equivalence and symbolic dynamics see [24].

In this dissertation we are interested in the role of shift equivalence in defining the Conley index for discrete time dynamical systems. In Section 2.3 we defined the homology Conley index for an isolated invariant set S with filtration pair $P = (P_1, P_0)$ by the shift equivalence of $[f_P]$, and the homology Conley index as the shift equivalence class of the map f_{P*}. These various types of shift equivalence are all special cases of the general categorical definition given in Definition 2.11.

A couple general remarks hold for shift equivalence in any category. Let a and b

Remark 3.1. Shift equivalence is, in fact, an equivalence relation as we can check. For reflexivity $a \sim a$, let r be the identity, $s = a$, and $l = 1$. For symmetry just exchange r and s. For transitivity, suppose $a \sim b$ via r, s with lag l and $b \sim c$ via r', s' with lag l'. Then $a \sim c$ via $r'r, ss'$ with lag $l + l'$.

Remark 3.2. If $a \sim b$ with lag l via r, s, then $a \sim b$ with lag $l + 1$ via r, sb.

We can translate our examples above in the language of categories straightforwardly as follows. For symbolic dynamics the relevant category is the set of finitely-generated free abelian groups with morphisms restricted to those that can be represented as positive integer matrices. (One just needs to observe such maps contain the identity and are closed under composition.) For the Conley index as defined by Franks and

Richeson in [17] the category is homotopy classes of pointed topological spaces. The type of shift equivalence we are principally concerned with in computational approaches to the Conley index is the category of abelian groups, since both the homology of a filtration pair and the map induced on homology by the dynamics can be readily computed [19].

Passing to homology, of course, means that the Conley index we choose to work with is coarser than the homotopy Conley index defined by Franks and Richeson. For an example where this means we lose information about the dynamics consider the space $S^1 \vee S^1$. If we take the generators of $\pi_1(S^1 \vee S^1)$ to be a, b, then the map $S^1 \vee S^1 \to S^1 \vee S^1$ defined by $a \mapsto aba^{-1}b^{-1}$ and $b \mapsto 1$ will have the same Conley index as the constant map as we have defined it on the homology level. But because this induces a nontrivial map of fundamental groups these two maps can be distinguished on the homotopy level.

In this chapter we focus on shift equivalence of the restriction of the induced map to the free part of homology, and even more narrowly on maps that can be represented by 2×2 integer matrices. These represent dynamics that commonly arise in applications and that it would be ideal to understand comprehensively. First, note that we are able to recast the problem of determining whether two matrices are shift equivalent into the language of modules. This perspective is due to C. Weibel [47], and related remarks appear in [24]. If we denote the set of $n \times n$ matrices with entries in $\mathbb{Z}$ by $M_n(\mathbb{Z})$, then for any $A \in M_n(\mathbb{Z})$, we can define an $\mathbb{Z}[t]$-module M_A on the underlying set $\mathbb{Z}^n$ by having A act as t. This follows the standard treatment of the classification of linear transformations over a field using the fundamental theorem of finitely generated modules over a PID, but of course $\mathbb{Z}[t]$ is not a PID, making the classification problem more difficult.

The following proposition characterizing shift equivalence requires that we consider the localization $M_A[t^{-1}]$ of the $\mathbb{Z}[t]$-module M_A [15]. Informally, we are inverting the element t in the ring $\mathbb{Z}[t]$ and carrying out all the implications this has for the module M_A. To be concrete about what localization is, we can think about the elements as being "fractions" where the denominator is a power of t. Representing these fractions

as ordered pairs, we can write the elements of $M_A[t^{-1}]$ as (m, t^k) for $m \in M_A$ and $k \in \mathbb{Z}^+$, subject to the relation $(m_1, t^{k_1}) \sim (m_2, t^{k_2})$ iff there exists $j \in \mathbb{Z}^+$ so that $t^j(t^{k_2}m_1 - t^{k_1}m_2) = 0$. This relation is essentially the same one that defines equivalence of different representations of $\mathbb{Q}$. The leading t^j (unnecessary in the case of rational numbers) ensures that any module element that yields 0 when multiplied by a power of t is zero in the localization.

Proposition 3.3. *Let $A \in M_n(\mathbb{Z})$ and $B \in M_m(\mathbb{Z})$ define the $\mathbb{Z}[t]$-modules M_A and M_B. Then $A \sim B$ iff $M_A[t^{-1}] \cong M_B[t^{-1}]$ as $\mathbb{Z}[t, t^{-1}]$-modules.*

Proof. First assume $A \sim B$. Then from the definition of shift equivalence there exist $\mathbb{Z}$-module homomorphisms $r : \mathbb{Z}^n \to \mathbb{Z}^m$ and $s : \mathbb{Z}^m \to \mathbb{Z}^n$. Condition (i) in the definition of shift equivalence gives us that for every $m \in M_A$

$$r(t \cdot m) = r(Am) = A(r(m)) = t \cdot r(m)$$

So r is in fact a $\mathbb{Z}[t]$-module homomorphism. An exactly analogous argument holds for condition (ii) and s. Condition (iii) then tells us $s(r(m)) = t^l \cdot m$. Passing to the localization, $s(r(m)) = t^l \cdot M_A[t^{-1}]$. Therefore the homomorphism $t^{-l} \cdot s : M_B[t^{-1}] \to M_A[t^{-1}]$ is an inverse to r, so $M_A[t^{-1}] \cong M_B[t^{-1}]$.

Now assume $M_A[t^{-1}] \cong M_B[t^{-1}]$. Let $\phi : M_A[t^{-1}] \to M_B[t^{-1}]$ be an isomorphism, and $m_1, \ldots, m_n$ be generators for M_A as an $\mathbb{Z}$-module. Then $\phi((m_i, 1))$ has a representative (n_i, t^{k_i}) in $M_B[t^{-1}]$. Let $l_1 = \max k_i$, so that $t^{l_1} \cdot (n_i, t^{k_i}) \sim (t^{l_1 - k_i} \cdot n_i, 1)$.

We can then define the $\mathbb{Z}$-module map $r : \mathbb{Z}^n \to \mathbb{Z}^m$ by $r(m_i) = t^{l_1 - k_i} \cdot n_i$ or, briefly, $r = t^{l_1}\phi$. Similarly, define the map $s : \mathbb{Z}^m \to \mathbb{Z}^n$ using ϕ^{-1} and a generating set for M_B. This can be written as $s = t^{l_2}\phi^{-1}$.

The maps r and s commute with multiplication by t (acting as either A or B) because they are defined in terms of the $\mathbb{Z}[t, t^{-1}]$-module isomorphism ϕ. Therefore conditions (i) and (ii) for shift equivalence are met. Furthermore, $s \circ r = t^{l_2}\phi^{-1}(t^{l_1}\phi) = t^{l_1 + l_2}\phi^{-1}\phi = t^l = A^l$, where $l = l_1 + l_2$. Similarly, $r \circ s = t^l = B^l$. So conditions (iii) and (iv) are met, and A and B are shift equivalent.

□

We expect the problem of determining isomorphism classes of $\mathbb{Z}[t, t^{-1}]$ to be too difficult to admit any solution besides in special cases. Naive methods of direct computation will not be efficient, even in the case of 2×2 matrices as the following proposition shows.

Proposition 3.4. *For every positive integer n there exist matrices $A, B \in M_2(\mathbb{Z})$ such that $A \sim B$ with minimal lag n. In other words, there is no shift equivalence between A and B with lag $l < n$.*

Proof. Let λ and p be relatively prime integers not equal to 1 or -1, and consider the matrices

$$A = \begin{bmatrix} \lambda & p \\ 0 & \lambda \end{bmatrix}, B = \begin{bmatrix} \lambda & p\lambda^n \\ 0 & \lambda \end{bmatrix}$$

where n is any positive integer.

We will show that the matrices A and B are shift equivalent with minimum lag n. To show this we directly solve for R and S. Let

$$R = \begin{bmatrix} r_{11} & r_{12} \\ r_{21} & r_{22} \end{bmatrix}, S = \begin{bmatrix} s_{11} & s_{12} \\ s_{21} & s_{22} \end{bmatrix}$$

Then solving $RA = BR$ we get that

$$R = \begin{bmatrix} \lambda^n r_{22} & r_{12} \\ 0 & r_{22} \end{bmatrix}, r_{22} \neq 0$$

and $AS = SB$ yields

$$S = \begin{bmatrix} s_{11} & s_{12} \\ 0 & \lambda^n s_{11} \end{bmatrix}, s_{11} \neq 0$$

Therefore

$$SR = \begin{bmatrix} \lambda^n r_{22} s_{11} & s_{11} r_{12} + r_{22} s_{12} \\ 0 & \lambda^n r_{22} s_{11} \end{bmatrix}$$

In order that $SR = A^l = \begin{bmatrix} \lambda^l & l\lambda^{l-1} p \\ 0 & \lambda^l \end{bmatrix}$, we need $\lambda^l = \lambda^n r_{22} s_{11}$ and therefore $l \geq n$. We can check that equality holds directly by setting $r_{22} = s_{11} = 1$, $r_{12} = p$ and $s_{12} = 0$ so that

$$R = \begin{bmatrix} \lambda^n & n\lambda^{n-1} p \\ 0 & 1 \end{bmatrix}, S = \begin{bmatrix} 1 & 0 \\ 0 & \lambda^n \end{bmatrix}$$

$$SR = \begin{bmatrix} \lambda^n & n\lambda^{n-1} p \\ 0 & \lambda^n \end{bmatrix} = A^n$$

$$RS = \begin{bmatrix} \lambda^n & n\lambda^{2n-1} p \\ 0 & \lambda^n \end{bmatrix} = B^n$$

Thus A and B are shift equivalent with (smallest) lag n.

□

The lesson of Proposition 3.4 is that to determine naïvely whether a pair of 2×2 matrices are shift equivalent there is no *a priori* bound on the number of conditions that potentially need to be checked. So either more sophisticated methods are needed, or we will have to seek out efficiently computable invariants of shift equivalence. In the latter case, it becomes important to understand for each invariant the kinds of dynamics the invariants we use can and cannot distinguish.

One way of thinking about why classification is so difficult is that the ring $\mathbb{Z}[t]$ is not a PID. This suggests a computable invariant for shift equivalence: namely, to view our matrix as having entries in a field $\mathbb{F}$, so that the $\mathbb{F}[t]$-module structure would be characterized by the fundamental theorem of finitely generated modules over a PID. Another way to view this invariant is that we compute homology with coefficients in $\mathbb{F}$ instead of $\mathbb{Z}$. An argument perfectly analogous to Proposition 3.3 shows that shift equivalence in that case is the same as isomorphism of the induced $\mathbb{F}[t, t^{-1}]$-modules.

This will hold for any field $\mathbb{F}$, but starting from an integer matrix a natural choice of field would be either $\mathbb{Q}$, $\mathbb{C}$, or $\mathbb{Z}/p\mathbb{Z}$ for prime p. In Chapters 6 and 7 we opt to compute the Conley index over finite fields, and discuss the considerations that go into making the particular choices we do.

The connection between the $\mathbb{F}[t]$-module viewpoint and more elementary linear algebra is that two matrices will induce isomorphic $\mathbb{F}[t]$-modules iff they are similar. An equivalent way of expressing this is to say that the the matrices have the same invariant factors (i.e. the same rational canonical form) or, in the case of algebraically closed fields, the same elementary divisors (i.e. the same Jordan canonical form). In the general case where our field is not necessarily algebraically closed, we choose to work with invariant factors. To characterize shift equivalence, however, we must localize at t and consider the resulting $\mathbb{F}[t, t^{-1}]$-module. A small modification of the invariant factors suffices to characterize isomorphism (and therefore shift equivalence) in this case, as the following proposition shows.

Recall that the invariant factors of a matrix M are a sequence of monic polynomials $(p_k(x))$ such that $p_i \mid p_{i+1}$ and which characterize M up to similarity. Define the *reduced invariant factors* of M to be $(\tilde{p}_k(x))$ where $\tilde{p}_k(x) = p_k(x)/x^d$ where x^d is the largest power of x dividing p_k.

Proposition 3.5. *Let $\mathbb{F}$ be a field. Then the matrices $A \in M_n(\mathbb{F})$ and $B \in M_m(\mathbb{F})$ are shift equivalent iff they have the same reduced invariant factors (ignoring factors of 1).*

Proof. From [24, Theorem 7.4.10], $A \sim B$ iff they have the same Jordan form away from zero (over the algebraic closure $\bar{\mathbb{F}}$). Therefore the elementary divisors of A and B that are not powers of x are in correspondence. By using the translation between elementary divisors and invariant factors [13] and ignoring these powers of x, we get that the reduced invariant factors must also be identical. □

Since if two endomorphisms are shift equivalent with lag l they are shift equivalent for every lag $l' > l$, the notion of shift equivalence should in some sense "ignore" nilpotent behavior. The reduced invariant factors of a matrix A are a way of characterizing the similarity of A ignoring the eventual kernel.

The problem with choosing to compute Conley indices over fields instead of $\mathcal{Z}$ is that much of the surprising subtlety of shift equivalence is lost. For example a result due to Tollmer discussed in [24, 38] is

$$\begin{bmatrix} 19 & 5 \\ 4 & 1 \end{bmatrix} \not\sim \begin{bmatrix} 19 & 4 \\ 5 & 1 \end{bmatrix}$$

Because a matrix over a field is always similar to its transpose, there is no hope for shift equivalence over fields to capture the same dynamics as shift equivalence over the integers in this case.

As we can see, even in the case of 2×2 full rank integer matrices, the problem is already difficult. We can start by using our obervations about shift equivalence over $\mathbb{Q}$ to note that two matrices shift equivalent over $\mathbb{Z}$ must have the same eigenvalues. This is because the integer matrices R and S giving a shift equivalence over $\mathbb{Z}$ also give a shift equivalence over $\mathbb{Q}$ and consequently the shift equivalent matrices must have the same rational canonical form, hence the same eigenvalues.

Shift equivalence of upper-triangular 2×2 matrices

One special case of particular interest is that of upper-triangular matrices, because of the connection to attractor-repeller pairs in dynamics. For example, the matrix

$$\begin{bmatrix} \lambda & a \\ 0 & \mu \end{bmatrix}$$

arises when there exists an invariant set with Conley index representative $\left[\mu\right]$ and a second invariant set of Conley index representative $\left[\lambda\right]$ along with the (potential) connecting orbits between them, which are represented by a. Conley theory guarantees us that if this matrix is not shift equivalent to the diagonal matrix $\begin{bmatrix} \lambda & 0 \\ 0 & \mu \end{bmatrix}$ that there must be a connecting orbit. Thus, determining the Conley indices in this special case is very useful for extracting dynamical information. In this special setting, we can determine a fair amount of information about the Conley index with a "hands-on"

approach.

Theorem 3.6. *The matrices* $\begin{bmatrix} \lambda & a \\ 0 & \mu \end{bmatrix}$ *and* $\begin{bmatrix} \lambda & b \\ 0 & \mu \end{bmatrix}$ *are shift equivalent if the prime factors of* a *and* b *that do not divide either* λ *or* μ *are identical.*

Proof. If $\lambda = \mu = 0$, then every prime factor of a or b trivially divides λ and μ. In this case the matrices are both nilpotent and therefore shift equivalent to $\begin{bmatrix} 0 \end{bmatrix}$.

If one of λ or μ is zero, again, every prime factor of a or b trivially divides either λ or μ. Suppose $\lambda \neq 0$. Then we have

$$\begin{bmatrix} \lambda & a \\ 0 & 0 \end{bmatrix} \sim \begin{bmatrix} \lambda \end{bmatrix} \text{ via } R = \begin{bmatrix} \lambda & a \end{bmatrix}, S = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

If $\mu \neq 0$ then we have

$$\begin{bmatrix} 0 & a \\ 0 & \mu \end{bmatrix} \sim \begin{bmatrix} \mu \end{bmatrix} \text{ via } R = \begin{bmatrix} 0 & 1 \end{bmatrix}, S = \begin{bmatrix} a \\ \mu \end{bmatrix}$$

Since these shift equivalences do not depend on a we see that the result holds.

If λ, μ are both nonzero, we can assume without loss of generality that $\gcd(a, \lambda) = \gcd(a, \mu) = 1$. Then the result for general a and b will follow by the transitivity of shift equivalence. Furthermore we can write $b = a\gamma$, where every prime factor of γ divides either λ or μ. There must, then, exist an integer n such that $\gamma \mid (\lambda\mu)^n$.

Write $\lambda^n = \gamma_\lambda x_\lambda$, where $\gamma_\lambda = \gcd(\gamma, \lambda^n)$ and $x_\lambda = \lambda^n/\gamma_\lambda$. Next write $\mu^n = \gamma_\mu x_\mu$, where $\gamma_\mu = \gamma/\gamma_\lambda$ and $x_\mu = \mu^n/\gamma_\mu$. (The asymmetry in the definitions of γ_λ and γ_μ is an artifact of an arbitrary choice. Really all we need is a way to express γ as $\gamma = \gamma_\lambda\gamma_\mu$ where $\gamma_\lambda \mid \lambda^n$ and $\gamma_\mu \mid \mu^n$. The terms x_λ and x_μ are then whatever factors of λ^n and μ^n remain.)

Now we construct the following R and S:

$$R = \begin{bmatrix} \lambda^n\gamma_\lambda & a\gamma_\lambda(\mu^{n-1} + \mu^{n-2}\lambda + \cdots + \lambda^{n-1}) \\ 0 & x_\mu \end{bmatrix}$$

$$S = \begin{bmatrix} x_\lambda & a\gamma_\mu(\mu^{n-1} + \mu^{n-2}\lambda + \cdots + \lambda^{n-1}) \\ 0 & \mu^n\gamma_\mu \end{bmatrix}$$

□

In the special case where $\lambda = \mu$ we also have the converse.

Theorem 3.7. *The matrices* $\begin{bmatrix} \lambda & a \\ 0 & \lambda \end{bmatrix}$ *and* $\begin{bmatrix} \lambda & b \\ 0 & \lambda \end{bmatrix}$ *are shift equivalent only if the prime factors of* a *and* b *that do not divide* λ *are identical.*

Proof. Suppose $A = \begin{bmatrix} \lambda & a \\ 0 & \lambda \end{bmatrix}$, $B = \begin{bmatrix} \lambda & b \\ 0 & \lambda \end{bmatrix}$, and $A \sim B$ with lag n. We can assume without loss of generality that $\gcd(a, \lambda) = \gcd(b, \lambda) = 1$. Then we know that R and S can be written as

$$R = \begin{bmatrix} x & r \\ 0 & y \end{bmatrix}, S = \begin{bmatrix} \lambda^n/x & s \\ 0 & \lambda^n/y \end{bmatrix}$$

Looking at the upper-right entry of $SR = A^n$, we have $(\lambda^n/x)r + ys = an\lambda^{n-1}$. Similarly, looking at the upper-right entry of $RS = B^n$ we have $(\lambda^n/y)r + xs = bn\lambda^{n-1}$.

The left hand side of this second equation is x/y times the left hand side of the first, while the right hand side of the second is b/a times the right hand side of the first. These will be equal whenever $ax = by$. We have $\gcd(a, y) = \gcd(b, x) = 1$, since x and y each divide λ^n. Therefore a divides b and b divides a, hence $a = b$.

□

This converse is not true in general for distinct eigenvalues, however. The difference $|\mu - \lambda|$ between them also plays a role as in the following special case:

Proposition 3.8. *The matrices* $\begin{bmatrix} \lambda & a \\ 0 & \lambda + 1 \end{bmatrix}$ *and* $\begin{bmatrix} \lambda & b \\ 0 & \lambda + 1 \end{bmatrix}$ *are shift equivalent for all* a, b.

Proof. Let $R = \begin{bmatrix} \lambda & b(\lambda + 1) - a\lambda \\ 0 & \lambda + 1 \end{bmatrix}$, $S = \begin{bmatrix} 1 & a - b \\ 0 & 1 \end{bmatrix}$.

□

Chapter 4

Constructing Conley indices

Beyond just being an invariant, we are interested in knowing exactly what implications the Conley index has for understanding dynamics. That the Conley index is a robust and coarse invariant makes this a difficult problem. To take a particularly stark example, the Conley index of the empty set is trivial, but so is the Conley index of the Smale horseshoe. Thus the kind of statements we might hope to make about a set with trivial Conley index has to somehow be true for both the absence of recurrent dynamics and the Smale horseshoe.

The easiest kinds of statements to make are negative ones: if the Conley index of a Morse set is nontrivial, then we can conclude that there are, in fact, recurrent dynamics contained within the Morse set. (Furthermore, those recurrent dynamics will not be a Smale horseshoe.)

We can also use the Conley index to make statements about other, perhaps more familiar, invariants. For example, we can use information about the Conley index to put lower bounds on entropy [3].

Not every Conley index can arise on the zero level of homology. Indeed, as [2, Proposition 5.8] shows, the eigenvalues of every dimension zero Conley index must be roots of unity. Since the eigenvalues are an invariant of the Conley index as we discuss in Chapter 3 this shows that any matrix with an eigenvalue that is not a root of unity cannot be a representative of a Conley index in dimension zero.

For the Conley index in level one and higher, however, there is no such restriction. Any map, for example, can be realized as the Conley index of a map on a wedge of Moore spaces. In what follows we go further and show that for Conley indices in dimension 1 that can be represented by an integer matrix there is an interval map

realizing that Conley index. Specifically, given any integer matrix M, we provide an explicit construction of a 1-d map $f : [0,1] \to [0,1]$ and a pair $P = (P_1, P_0)$ such that the map

$$f_P : H_1(P_1/P_0, [P_0]) \to H_1(P_1/P_0, [P_0])$$

is shift equivalent to M.

To simplify this construction, it will be convenient to work with a more general notion than a filtration pair. In particular, we produce an index pair in accordance with the definition due to J. Robbin and D. Salamon [42], and show that the induced map on homology for this index pair is shift equivalent to the induced map on homology for a filtration pair for the same isolated invariant set. Here we state the Robbin-Salamon definition of index pair:

Definition 4.1. Let $f : \mathbb{T}^+ \times X \to X$ be a dynamical system. A compact pair (P_1, P_0) in M is an *index pair* if $P_1 \setminus P_0$ is an isolating neighborhood and $f_P^t : (P_1/P_0, [P_0]) \to (P_1/P_0, [P_0])$ is continuous for all t, where

$$f_P^t(x) = \begin{cases} f(t,x) & \text{if } f(s,x) \in P_1 \setminus P_0, \forall s, \text{ such that } 0 \le s \le t \\ [P_0] & \text{otherwise} \end{cases}$$

Remark 4.2. In the case where $\mathbb{T}^+ = \mathbb{Z}^+$, this is equivalent to

$$f_P := \begin{cases} f(x) & x, f(x) \in P_1 \setminus P_0 \\ [P_0] & \text{otherwise} \end{cases}$$

being continuous.

We start by defining a special class of matrices, meant to capture necessary conditions for an induced map on first homology for disjoint intervals on the real line.

Definition 4.3. Call a matrix B *constructible* if it has the following properties:

1. The entries of B are from $\{0, 1, -1\}$.

2. No column contains both 1 and -1.

3. For each column, the nonzero entries in that column occur in contiguous rows.

Informally, think of the columns of B as indexing disjoint intervals in $[0,1]$ that are ordered by the order of the columns. To incorporate notation we will use shortly, let the i^{th} interval be $[\underline{x_i}, \overline{x_i}]$. If we denote our interval map by f, then if the (i,j)-entry is 1, this means that $f(\underline{x_i}) < \underline{x_j}$ and $\overline{x_j} < f(\overline{x_i})$—in other words, that the image of the i^{th} interval extends over the j^{th} with the order of the endpoints preserved. If the entry is -1, this corresponds to the order of the endpoints of the i^{th} interval being reversed. A zero entry means that the image of the i^{th} interval does not intersect the j^{th}. We make this explicit in part a) of the following theorem with a construction.

Theorem 4.4. *Suppose B is an $n \times n$ constructible matrix. Then*

a) there exists an interval map f and an index pair (P_1, P_0) for an isolated invariant set X of f so that B represents the map $f_* : H_1(P_1/P_0, [P_0]) \to H_1(P_1/P_0, [P_0])$

b) there is a filtration pair $(P_1, \widetilde{P}_0)$ for X, such that

c) the induced map $\tilde{f}_* : H_1(P_1/\widetilde{P}_0, [\widetilde{P}_0]) \to H_1(P_1/\widetilde{P}_0, [\widetilde{P}_0])$ is shift equivalent to $f_* : H_1(P_1/P_0, [P_0]) \to H_1(P_1/P_0, [P_0])$

Proof. a) We want the dynamics we are interested in to be contained in n disjoint intervals. To make things concrete, define $I_k := [\underline{x_k}, \overline{x_k}] := \left[\frac{2k-1}{2n+1}, \frac{2k}{2n+1}\right]$ for $k = 1, \ldots, n$.

In order to define f we will also want to make use of points that lie between each of these intervals. For $k = 0, \ldots, n$ let m_k be the midpoint between $\overline{x_k}$ and $\underline{x_{k+1}}$ (or the midpoint between 0 and $\underline{x_1}$ in the case of m_0, and $\overline{x_n}$ and 1 in the case of m_n). Explicitly $m_k = \frac{4k+1}{4n+2}$. In this way, we ensure that $\overline{x_k} < m_k < \underline{x_{k+1}}$.

The idea is to define f in a piecewise linear way, so that the image of f on the interval I_j extends over the intervals $I_{k_1}, \ldots, I_{k_s}$ where $k_1, \ldots, k_s$ are the (contiguous) nonzero entries of the j^{th} column of B. If these entries are positive we want f to have positive slope, otherwise for f to have negative slope. This choice gives us the appropriate sign for the induced map on homology.

Define

$$\underline{c_k} := \begin{cases} (\min\{i : M_{ik} = 1\}) - 1 & \exists j, M_{jk} = 1 \\ \max\{i : M_{ik} = -1\} & \exists j, M_{jk} = -1 \\ 0 & \forall j, M_{jk} = 0 \end{cases}$$

$$\overline{c_k} := \begin{cases} \max\{i : M_{ik} = 1\} & \exists j, M_{jk} = 1 \\ (\min\{i : M_{ik} = -1\}) - 1 & \exists j, M_{jk} = -1 \\ 0 & \forall j, M_{jk} = 0 \end{cases}$$

We now define f as the linear interpolation between the values $f(\underline{x_k}) = m_{\underline{c_k}}$ and $f(\overline{x_k}) = m_{\overline{c_k}}$.

$$f(x) := \begin{cases} m_{\underline{c_1}} & x < \underline{x_1} \\ (2n+1)(m_{\overline{c_k}} - m_{\underline{c_k}})(x - \underline{x_k}) + m_{\underline{c_k}} & \underline{x_k} \le x < \overline{x_k},\ k = 1, \ldots, n \\ (2n+1)(m_{\underline{c_{k+1}}} - m_{\overline{c_k}})(x - \overline{x_k}) + m_{\overline{c_k}} & \overline{x_k} \le x < \underline{x_{k+1}},\ k = 1, \ldots, n-1 \\ m_{\overline{c_n}} & x \ge \overline{x_n} \end{cases}$$

In order to get a Conley index, we need an index pair (P_1, P_0) in addition to this map. Let $P_1 = \bigcup_k I_k$, $P_0 = \bigcup_k \{\underline{x_k}, \overline{x_k}\}$. We need to check both conditions for an index pair.

First, we verify that $\mathrm{cl}(P_1 \backslash P_0)$ is an isolating neighborhood. Since $\mathrm{bd}\,\mathrm{cl}(P_1 \backslash P_0) = P_0$ and for every $x \in P_0, f(x) = m_i \notin P_1$ for some i, we have that x cannot be in $\mathrm{Inv}(\mathrm{cl}(P_1 \setminus P_0))$. Therefore $\mathrm{Inv}(\mathrm{cl}(P_1 \setminus P_0)) \bigcap \mathrm{bd}(\mathrm{cl}(P_1 \setminus P_0)) = \emptyset$, so $\mathrm{cl}(P_1 \setminus P_0)$ is an isolating neighborhood.

Next, we need to show that $f_\# : (P_1/P_0, [P_0]) \to (P_1/P_0, [P_0])$ defined by

$$f_\# := \begin{cases} f(x) & x, f(x) \in P_1 \setminus P_0 \\ [P_0] & \text{otherwise} \end{cases}$$

is continuous.

This is a consequence of the following theorem:

Theorem 4.5. *([42], Theorem 4.5) Assume $f : M \to M$, $N \subseteq M$ is closed, and*

$$(\operatorname{bd} N) \bigcap f^{-1}(N) \bigcap f(N) = \emptyset$$

Define

$$L_0 = f(N) \bigcap (\operatorname{bd} N)$$

and

$$L_1 = \operatorname{cl}(N \setminus f^{-1}(N))$$

Then $L_0 \subseteq L_1$ and $f_\# : N/L \to N/L$ is continuous for any closed subset L of N satisfying $L_0 \subseteq L \subseteq L_1$.

Theorem 4.5 applies, since $\operatorname{bd} P_1 \bigcap f^{-1}(P_1) = \emptyset$ and because

$$P_0 = L_0 = \operatorname{bd} P_1 \bigcap f(P_1) = \operatorname{bd} P_1.$$

Finally we verify that the induced map on first homology $f_* : H_1(P_1/P_0, [P_0]) \to H_1(P_1/P_0, [P_0])$ is represented by the matrix B. We can give P_1 a CW-structure, by taking the 0-cells to be points of the set P_0 and the 1-cells to be the intervals between the points in P_0. The generators of $H_1(P_1/P_0, [P_0])$ are given by the intervals (1-cells)I_k. By looking at the degree of the map on the quotient, the image of a generator I_k is 1 for every generator I_j such that $f(\underline{x_k}) < \underline{x_j}$ and $\overline{x_j} < f(\overline{x_k})$, and is -1 when $f(\underline{x_k}) > \overline{x_j}$ and $\underline{x_j} > f(\overline{x_k})$. In all other cases the image is 0.

This is precisely what is achieved by our choice of $\underline{c_k}$ and $\overline{c_k}$, and taking a straight-line interpolation between $f(\underline{x_k}) = m_{\underline{c_k}}$ and $f(\overline{x_k}) = m_{\overline{c_k}}$.

b) Define $\widetilde{P}_0 := \operatorname{cl}(P_1 \setminus f^{-1}(P_1))$. We check that the pair $(P_1, \widetilde{P}_0)$ is a filtration pair. In what follows, let Y^C denote the complement of Y in $[0, 1]$. We need to check

   i) $\operatorname{cl}(P_1 \setminus \widetilde{P}_0)$ *is an isolating neighborhood.*

      Because $\operatorname{cl}(P_1 \setminus \widetilde{P}_0)$ is a finite set of disjoint open intervals, $\operatorname{bd}(\operatorname{cl}(P_1 \setminus \widetilde{P}_0)) \subseteq \widetilde{P}_0$. By our definition, we have that $f(\widetilde{P}_0) \subseteq \operatorname{cl}(P_1^C)$, hence $f(\widetilde{P}_0) \bigcap P_1 \subseteq \operatorname{bd} P_1$. But

$f(\operatorname{bd} P_1) \subseteq \{m_i\}$ by the definiton of f, hence because the m_i are disjoint from P_1, $f(f(\operatorname{bd} \operatorname{cl}(P_1 \setminus \widetilde{P}_0))) \bigcap P_1 = \emptyset$ and $\operatorname{Inv} \operatorname{cl}(P_1 \setminus \widetilde{P}_0) \subseteq \operatorname{int}(\operatorname{cl}(P_1 \setminus \widetilde{P}_0))$, which is the definition of isolating neighborhood.

ii) *$\widetilde{P}_0$ is a neighborhood of the exit set $E = \{x \in P_1 : f(x) \notin P_1\}$ in P_1.*

Because $f^{-1}(P_1)$ is closed, we have that the exit set $E = P_1 \setminus f^{-1}(P_1) = P_1 \bigcap (f^{-1}(P_1))^C$ is open in P_1, hence $\widetilde{P}_0 = \operatorname{cl}(P_1 \setminus f^{-1}(P_1)) = \operatorname{cl} E$ is a neighborhood of E.

iii) $f(\widetilde{P}_0) \bigcap \operatorname{cl}(P_1 \setminus \widetilde{P}_0) = \emptyset$

We have $f(\widetilde{P}_0) = f(\operatorname{cl}(P_1 \setminus f^{-1}(P_1)) \subseteq \operatorname{cl}(P_1^C)$. Therefore $f(\widetilde{P}_0) \bigcap f(P_1) = P_1$. Since $\operatorname{bd} P_1 = (\cup\underline{\{x_k\}}) \cup (\cup\overline{\{x_k\}})$

c) We now need to check that the induced map on homology for $\widetilde{f}_*\colon H_1(P_1/\widetilde{P}_0, [\widetilde{P}_0]) \to H_1(P_1/\widetilde{P}_0, [\widetilde{P}_0])$ is shift equivalent to $f_*\colon H_1(P_1/P_0, [P_0]) \to H_1(P_1/P_0, [P_0])$. We need to check this because shift equivalence is only proven to define the an invariant for filtration pairs and (P_1, P_0) is an index pair but not a filtration pair.

Let B be our constructible matrix representing f_*. We will produce matrices $\widetilde{R}$ and $\widetilde{S}$ such that $\widetilde{S}\widetilde{R} = B$, and such that the product $\widetilde{R}\widetilde{S} = C$ represents $\widetilde{f}_*$.

Let $p = \sum_{i,j} |B_{ij}|$, which is the number of generators of $H_1(P_1/\widetilde{P}_0, [\widetilde{P}_0])$. If $p = 0$, that means both that B is the zero matrix and that $H_1(P_1/\widetilde{P}_0, [\widetilde{P}_0]) = 0$. This is something of a special case, since there is no matrix representing $\widetilde{f}_*$, but B is shift equivalent to the zero map nevertheless. We assume $p > 0$ in what follows.

Let $\xi_1, \ldots, \xi_n$ denote the generators of $H_1(P_1/P_0, [P_0])$, and $\eta_1, \ldots, \eta_p$ the generators of $H_1(P_1/\widetilde{P}_0, [\widetilde{P}_0])$. Let $\widetilde{R}$ be a $p \times n$ matrix defined as follows:

$$\widetilde{R}_{ij} = \begin{cases} 1 & \eta_i \subseteq \xi_j \\ 0 & \text{otherwise} \end{cases}$$

Let $\widetilde{S}$ be a $n \times p$ matrix defined by

$$\widetilde{S}_{kl} = \begin{cases} 1 & \eta_l \text{ maps to } \xi_k \text{ with positive orientation} \\ -1 & \eta_l \text{ maps to } \xi_k \text{ with negative orientation} \\ 0 & \text{otherwise} \end{cases}$$

Then $(\widetilde{S}\widetilde{R})_{ij} = \sum_k \widetilde{S}_{ik}\widetilde{R}_{kj}$. This is nonzero precisely when η_k maps to ξ_i and $\eta_k \subseteq \xi_j$, with sign determined by the orientation of the mapping. Summing over all k, this gives the degree of the map from ξ_j to ξ_i, although by construction there will only be at most one nonzero term in the sum, and it will be 1 if the orientation is positive and -1 if it is negative. Thus $\widetilde{S}\widetilde{R} = B$.

Switching factors around $(\widetilde{R}\widetilde{S})_{ij} = \sum_k \widetilde{R}_{ik}\widetilde{S}_{kj}$. This nonzero precisely when $\eta_i \subseteq \xi_k$ and η_j maps to ξ_k. This happens precisely when η_j maps to η_i, and thus $\widetilde{R}\widetilde{S}$ represents $\widetilde{f}_*$.

□

We have shown that every constructible matrix B can indeed be realized as a representation of the induced map on H_1 for an index pair in $[0, 1]$, and moreover than there exists a filtration pair for the same isolated invariant set with shift equivalent Conley index. This is a step towards relating the extremely general definition of Conley index in terms of index pairs with the definition we give here for filtration pairs.

We can show something much more general still, however, that will allow us to construct examples of nearly any Conley index we want with a map on the unit interval.

Theorem 4.6. *Let M be an $n \times n$ integer matrix. Then M is shift equivalent to a constructible matrix.*

Proof. We will proceed by producing two matrices R and S, such that $SR = M$ and RS is constructible. First, we check that any such maps R and S automatically satisfy conditions (i) and (ii) for shift equivalence. If $SR = M$, then $RM = R(SR) = (RS)R$ and $MS = (SR)S = S(RS)$.

Let $m_{i,j}$ denote the (i, j) entry of M, and let $p = \sum_{i,j} |m_{i,j}|$. If $p = 0$ then M is a zero matrix, which is already constructible, and every matrix is shift equivalent to

itself. Otherwise, take R to be a $p \times n$ matrix and S an $n \times p$ matrix defined as follows.

We will construct the matrices R and S by assembling them blockwise from vectors. For example, suppose v and w are column vectors of length l_v and l_w. Then the vector $\begin{bmatrix} v \\ w \end{bmatrix}$ denotes a column vector of length $l_v + l_w$ with entries from v followed by entries from w. The two-column matrix $\begin{bmatrix} v & w \end{bmatrix}$ is defined so long as $l_v = l_w$. For convenience, we will want to define an *empty vector* ε such that $v = \begin{bmatrix} v \\ \varepsilon \end{bmatrix}$.

We define the vectors we will use as bulding blocks as follows. Whenever $m_{i,j} \neq 0$ let $\mathbf{0}_{i,j}$ denote the column vector of length $|m_{i,j}|$ with all zero entries, $|\mathbf{1}_{i,j}|$ the column vector of length $|m_{i,j}|$ with all entries one, and $\mathbf{1}_{i,j} = \operatorname{sgn}(m_{i,j})\,|\mathbf{1}_{i,j}|$. If $m_{i,j} = 0$ let $\mathbf{0}_{i,j}$, $|\mathbf{1}_{i,j}|$, and $\mathbf{1}_{i,j}$ denote the empty vector ε.

Now let R be a block matrix with n^2 block rows and n block columns. Let $c = ((k-1) \bmod n) + 1$ and $d = \left\lfloor \frac{k-1}{n} \right\rfloor + 1$. Then the k, l block of R is $|\mathbf{1}_{c,d}|$ if $d = l$ and $\mathbf{0}_{c,d}$ otherwise.

$$R = \begin{bmatrix} |\mathbf{1}_{1,1}| & \mathbf{0}_{1,1} & \mathbf{0}_{1,1} & \cdots & \mathbf{0}_{1,1} \\ |\mathbf{1}_{2,1}| & \mathbf{0}_{2,1} & \mathbf{0}_{2,1} & \cdots & \mathbf{0}_{2,1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ |\mathbf{1}_{n,1}| & \mathbf{0}_{n,1} & \mathbf{0}_{n,1} & \cdots & \mathbf{0}_{n,1} \\ \mathbf{0}_{1,2} & |\mathbf{1}_{1,2}| & \mathbf{0}_{1,2} & \cdots & \mathbf{0}_{1,2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{0}_{n,n} & \mathbf{0}_{n,n} & \mathbf{0}_{n,n} & \cdots & |\mathbf{1}_{n,n}| \end{bmatrix}$$

Now we define the block matrix S with n block rows and n^2 block columns. With c and d as before, we have that the l, k block of S is given by $\mathbf{1}_{c,d}^T$ if $c = l$ and $\mathbf{0}_{c,d}^T$ otherwise. It is more notationally convenient to write S^T as follows:

$$S^T = \begin{bmatrix} \mathbf{1}_{1,1} & \mathbf{0}_{1,1} & \mathbf{0}_{1,1} & \cdots & \mathbf{0}_{1,1} \\ \mathbf{0}_{2,1} & \mathbf{1}_{2,1} & \mathbf{0}_{2,1} & \cdots & \mathbf{0}_{2,1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{0}_{n,1} & \mathbf{0}_{n,1} & \mathbf{0}_{n,1} & \cdots & \mathbf{1}_{n,1} \\ \mathbf{1}_{1,2} & \mathbf{0}_{1,2} & \mathbf{0}_{1,2} & \cdots & \mathbf{0}_{1,2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{0}_{n,n} & \mathbf{0}_{n,n} & \mathbf{0}_{n,n} & \cdots & \mathbf{1}_{n,n} \end{bmatrix}$$

We can use block multiplication to compute the product SR (and for convenience allow indices to refer to blocks). In this case the i, j entry in the product is given by

$$\sum_{m=1}^{n} S_{i,m} R_{m,j}$$

Where $S_{i,m}$ is a zero vector unless $(m \bmod n) + 1 = i$ and $R_{m,j}$ is a zero vector unless $\left\lfloor \frac{m}{n} \right\rfloor + 1 = j$. Thus the only nonzero term is when $m = (j-1)n + i - 1$ in which case it is the dot product $\mathbf{1}_{i,j} \cdot |\mathbf{1}_{i,j}| = m_{i,j}$.

It remains only to check that the product RS is constructible by checking that it satisfies the three defining properties of a constructible matrix.

1. *The entries of B are from $\{0, 1, -1\}$.*

   This follows from the fact that there is exactly one nonzero entry in each row of R and all entries of R and S are $0, 1$, or -1.

2. *No column contains both 1 and -1.*

3. *For each column, the nonzero entries in that column occur in contiguous rows.*

   These two properties follow from the fact that the nonzero entries in each column of R are a contiguous block of ones, hence because S has at most one nonzero entry in each column (either ± 1), the columns of the product must contain a single contiguous block of either all ones (in the case the nonzero entry in the column of S is 1), all minus ones (in the case the nonzero entry in the column of S is 1), or all zeroes (in the case that the columns of S is all zeroes).

□

Corollary 4.7. *Let M be a square matrix with integer entries. Then M is a representative for the Conley Index on the first level of homology for an interval map f along with filtration pair $(P_1, \widetilde{P}_0)$.*

Proof. By Theorem 4.6 we can produce a constructible matrix B shift equivalent to M. By Theorem 4.4 there exists a map f and a filtration pair $(P_1, \widetilde{P}_0)$ with Conley index on the first level C shift equivalent to B. By transitivity of shift equivalence, M is shift equivalent to C. □

Chapter 5

Application: the logistic map

In Chapter 2 we suggest ways the combinatorial-topological approach we take has advantages when studying "real-world" dynamics, taking into account factors like imprecision in measurement and computation and uncertainty about parameters. In this chapter and the next two, we look at three specific maps and ways which they illustrate different aspects of interpreting the output of database computations. In a very real sense the computations we perform prove more theorems (of varying levels of interest) than we can reasonably hope to look at. Thus we need heuristics and insight to separate the wheat from the chaff. This particular chapter represents work that appears in [6] and [7].

Recall the definition of the logistic map from Chapter 2:

$$f(x, r) = rx(1 - x)$$

where $x \in [0, 1]$ and $r \in [0, 4]$.

The logistic map is a standard example in the theory of dynamical systems, well known for its period doubling behavior and chaotic dynamics. As the parameter r increases from 3 to approximately 3.57, the attractor changes from a stable fixed point to periodic orbits of period $2, 4, 8, \ldots$ on successively shorter intervals. The ratio of the lengths of consecutive intervals limits to Feigenbaum's constant ($\approx$4.67), a general property of such period doubling bifurcations on unimodal maps with quadratic maximum [16]. Beyond this period doubling cascade, still more complicated behavior is observed up until $r = 4$ beyond which the map is no longer well defined on the unit interval.

Although the study of the logistic map has largely been motivated by these dynamic phenomena, it is typically introduced as an extremely simple nonlinear discrete time

population model. The question naturally arises, then, that if we suppose the logistic map is a good model for the dynamics of some real world phenomena, what exactly could we expect to see in practice bearing in mind the discussion in Chapter 2 of the imperfections of any model?

5.1 Database output and classical theory

The logistic map is a good example of the tenuous connection between the database viewpoint of dynamics, and the more classical, analytic viewpoint that emphasizes things like bifurcation theory and structural stability. Under ideal circumstances—for example, a classical saddle-node bifurcation—changes in the Morse graph for different parameter values relate quite clearly to qualitative changes in the underlying dynamics. In the case of the logistic map, however, the connection is more complicated and equivocal.

However, it is important to note that there is no simple correspondence between edges in the continuation graph and the classical notion of a bifurcation point. For a more complete analysis in the context of the a saddle-node bifurcation see [1].

The transition from region A to region B indicates the first period doubling bifurcation in the cascade. Likewise the transition from region B to region C corresponds to the second period doubling, and the transition from region C to region D the third. But, beyond this, further period doublings are invisible at this level of discretization. In fact, rather than the dynamics becoming finer, Figure 5.1 shows the dynamics becoming increasingly coarse as one proceeds through the period doubling cascade. These transitions do not correspond to bifurcations, but instead to the more global phenomenon of the periodic attractor moving around in a larger area of phase space given the level of discretization. Thus, as the period doublings accelerate we actually see a sort of "mirror image" effect of the first few doublings. The important observation is that changes of continuation class can in some circumstances indicate the presence of classical bifurcations, but they can also be a consequence of dynamics without necessarily corresponding to any change in structural stability, and that the two possibilities can

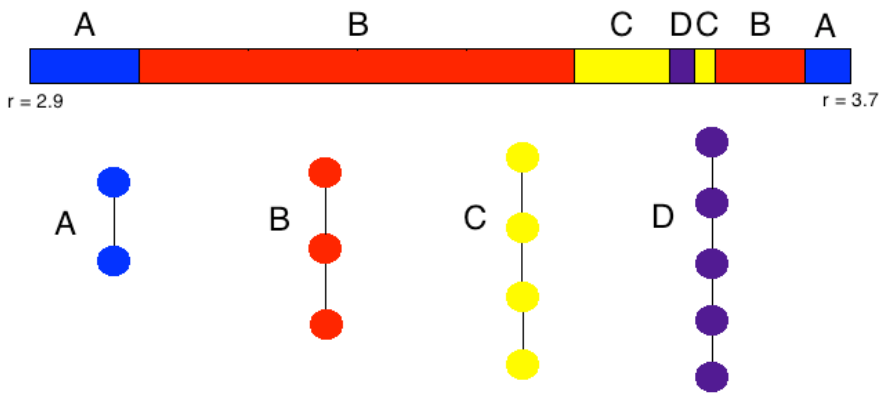


Figure 5.1: Schematic picture of parameter space for the logistic map, $r \in [2.9, 3.7]$. For this computation, phase space was divided into boxes of size 2^{-16} and parameter space was subdivided into a grid of size $\sim .008$. Here the transition from A to B occurs at 3.0078, B to C at 3.4563, and C to D at 3.5500. These transitions correspond to the first three period doublings, and they necessarily occur after the actual period doublings due to discretization in phase and parameter space. The actual bifurcations occur at exactly 3, then approximately 3.4495, 3.5441. The transition from D back to C occurs at 3.5727, C to B at 3.5914, and B to A at 3.6781. These transitions do not correspond to bifurcations, but instead to the way the periodic attractor increasingly spreads out through the discretized phase space, making it impossible to distinguish from the unstable dynamics (except for the origin).

not always immediately be distinguished.

5.2 Phase vs parameter subdivision

Consider still the parameter space $Z = [2.9, 3.7]$ for the logistic map on $X = [0, 1]$, with grids $\mathcal{Z}$ and $\mathcal{X}$ respectively. For each grid element $\zeta \in \mathcal{Z}$ the minimal Morse set $\mathcal{M}^i_\zeta$ in the Morse graph $\mathsf{MG}(\mathcal{F}_\zeta)$ represents the stable dynamics that occur for all $z \in |\zeta|$. (In general a Morse graph can have more than one minimal node in cases where there is more than one attractor. For the logistic map, however, the minimal node will be unique.) Because $\mathcal{F}_\zeta$ is an outer approximation for a continuous map, and because $\mathcal{M}^i_\zeta$ is by definition a strongly connected path component, the map $\mathcal{F}_{\zeta_i}$ must permute the connected components of $|\mathcal{M}^i_\zeta|$. In this sense, the number of connected components of $|\mathcal{M}^i_\zeta|$ gives a measure of the amount of periodicity visible over ζ given the level of discretization in phase space. The number of connected components can be determined from the Conley Index. A consequence of this is that two parameter boxes with isomorphic Conley-Morse graphs have same level of periodicity visible.

By taking the maximum period observed over all $\zeta \in \mathcal{Z}$, we have a measure of the amount of information visible given the level of subdivision in both phase and parameter space. Longer periodic orbits mean that we can see "deeper" into the period-doubling cascade. The choice of $Z = [2.9, 3.7]$ was made above because for any reasonable level of subdivision in parameter space the endpoints ζ_0 and ζ_n will have isomorphic Conley-Morse graphs. This is not to say that the dynamics for arbitrary $r_0 \in \zeta_0$ and $r_n \in \zeta_n$ are at all similar. Indeed, they will be quite different. For $r_0 < 3$ the logistic map has an unstable fixed point at the origin and a stable fixed point isolated away from the origin. The Conley-Morse graph will therefore consist of two nodes: the maximal one representing the unstable origin and the minimal one exhibiting period-1 behavior (i.e. one connected component containing the fixed point).

For $r_n \approx 3.7$, however, the dynamics can depend sensitively on the specific choice of r_n and is quite a bit more complicated. The origin is still unstable and the rest of the dynamics is still isolated away from the origin, but there is no longer an attracting

fixed point, and for many values of r_n there is a chaotic attractor. Because of the way the stable dynamics extends through phase space, however, the Morse graph will again contain just two nodes: the maximal one representing the unstable origin and the minimal one representing the attractor as a single connected component and therefore exhibiting period-1 behavior. These similarities are enough to ensure in this case that the Morse graphs MG_{ζ_0} and MG_{ζ_n} are isomorphic.

Figure 5.2 demonstrates the ability to resolve dynamics as a function of the level of subdivision in phase and parameter space. There is a point of diminishing returns in which increasing the resolution in whichever space is finer does very little to help resolve further dynamics. This point is relevant for the computations in Chapters 6 and 7 where such practical choices need to be made.

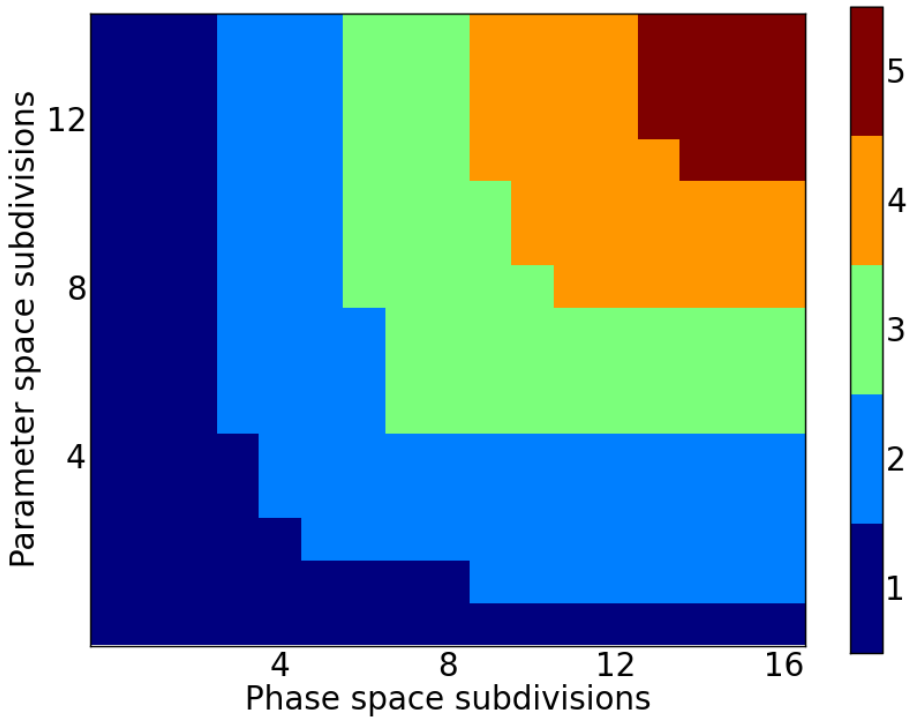


Figure 5.2: Representation of the resolution of dynamics for different levels of subdivision in phase and parameter space. Here the colors represent the number of Morse sets present at the given level of subdivison of phase and parameter space. Finer resolution leads to more finely resolved dynamics, but with diminishing returns for subdividing exclusively in either phase or parameter space.

Chapter 6

Application: Newton's method

This chapter consists of lightly edited selections of [5], work done jointly with Shaun Harker, Wes Cowan, and Konstantin Mischaikow. The contributions of the author of this dissertation are noted.

6.1 Introduction

For a differentiable function $f\colon \mathbb{R}^d \to \mathbb{R}^d$, Newton's method is an iterative procedure for approximating the roots of f. It can be understood by defining a sequence $(x_1, x_2, \ldots)$ by the map

$$x_{n+1} = x_n - Df(x_n)^{-1} f(x_n). \tag{6.1}$$

Under certain hypotheses, this sequence converges to a root of f. There exist instances where globally this algorithm is not well defined (Df must remain invertible throughout the application) or does not converge [4]. In this chapter we are concerned exclusively with local convergence. By translating a root of f to the origin we can assume that $f(0) = 0$, and we further assume that $Df(0)$ is invertible. A theorem of Kantorovich ensures that for an initial condition x_0 sufficiently close to the origin $\lim_{n\to\infty} \|x_n\| = 0$, and moreover that the rate of convergence is quadratic [12, Theorem 2.1]. When $d = 1$ the direction is determined by $f''(0)/f'(0)$. Motivated by the work of Dupont and Scott [14], we consider the question of the direction of convergence when $d = 2$ where things are much more complicated.

The goal is to provide, using the techniques outlined in Chapter 2, a rigorous, finite characterization of the asymptotic dynamics of the direction of convergence for all nondegenerate nonlinear functions. As is demonstrated in Section 6.2 these dynamics

are governed by the following four parameter family of maps

$$\begin{aligned} q\colon S^1 \times Z &\to S^1 \\ (\theta; a, b, c, \phi) &\mapsto \arctan\left(\frac{a\cos^2(\theta+\phi) + b\sin^2(\theta+\phi)}{c\cos^2\theta + \sin^2\theta}\right) \end{aligned} \tag{6.2}$$

where $Z = [-1, 1]^3 \times S^1$. This system undergoes a wide variety of bifurcations across parameter space, making it impractical to give a description of the dynamics in terms of the classical notion of invariant sets in a way that it robust to small changes in parameters. Instead we provide a database for the global dynamics as outlined Chapter 2.

Thus we begin in Section 6.2 with the derivation of (6.2). Section 6.3 provides a brief discussion of how the computational methods we discuss in Chapter 2 apply in this particular example. We include a simple table and discussion that is meant to provide intuition as to the proper interpretation of the Conley indices that appear in the Newton map. In Section 6.4 we provide synopses of results about the dynamics of (6.2) as a function of different parameter values. In particular, we exhibit a number of instances of the classically interesting dynamical behavior, discovered via exploration of the computed database.

6.2 The angular dynamics

Definition. The *angular dynamics map* induced by performing Newton's method on a function f is defined (when it exists) by the map $\Theta_f\colon S^1 \to S^1$ with

$$\Theta_f(x, y) := \lim_{r\to 0^+} \arg\left(s_f(rx, ry)\right), \tag{6.3}$$

where $s_f : \mathbb{R}^2 \to \mathbb{R}^2$ describes the Newton iteration (6.1) applied to f, i.e.

$$s_f(x, y) := (x, y) - Df(x, y)^{-1} f(x, y).$$

Here "arg" refers to the angle formed by a vector from the origin to a point on the plane with the positive x-axis.

For some functions f, the map (6.3) is not well-defined—e.g., in cases where f is not differentiable, Df is singular, s_f is not defined, or $s_f(rx, ry) = 0$ infinitely often as $r \to 0^+$ for some (x, y). Furthermore, in cases where it is well-defined it nevertheless

might fail to be continuous. Accordingly, we define the following class of functions:

Nondegeneracy assumption. We say a C^2 function $f : \mathbb{R}^2 \to \mathbb{R}^2$ is *nondegenerate* provided $f(0) = 0$, $Df(0)$ is invertible, and the following condition holds of the Hessians A and B at the origin of f_1 and f_2, respectively:

$$\text{For all } v \in \mathbb{R}^2, \begin{bmatrix} v^T A v \\ v^T B v \end{bmatrix} = 0 \implies v = 0.$$

(Here f_1 and f_2 denote the 1st and 2nd coordinate functions of f, and the Hessian is the 2×2 matrix of second derivatives evaluated at the origin.) We denote the collection of such nondegenerate functions as $\mathfrak{F}$.

The following theorems first characterize the angular dynamics of non-degenerate functions, and then provide a parameterization of all possible angular dynamics from $\mathfrak{F}$. For proofs refer to [5], as these are not principally the work of the author of this dissertation.

Theorem 6.1. *Each $f \in \mathfrak{F}$ induces a continuous angular dynamics map Θ_f. Moreover, letting $v = (x, y)$, $J = Df(0)$, and A, B be the Hessians of f_1, f_2 at the origin, we have*

$$\Theta_f(v) = \arg\left(J^{-1} \begin{bmatrix} v^T A v \\ v^T B v \end{bmatrix} \right). \tag{6.4}$$

Theorem 6.2. *For every $f \in \mathfrak{F}$, there exists $z = (a, b, c, \phi) \in [-1, 1]^3 \times S^1$ such that the angular dynamics map Θ_f is topologically conjugate to the map $\theta \mapsto q_z(\theta)$ defined by:*

$$q_z(\theta) = \arctan\left(\frac{a\cos^2(\theta + \phi) + b\sin^2(\theta + \phi)}{c\cos^2\theta + \sin^2\theta} \right), \tag{6.5}$$

defining the arctan *function over the extended reals. Further, defining $Z_s \subset [-1, 1]^3 \times S^1$ to be the set of $z \in [-1, 1]^3 \times S^1$ for which q_z is discontinuous or ill-defined,*

$$Z_s = \{(a, b, c, \phi) \in [-1, 1]^3 \times S^1 \mid ((a+bc)\cos^2\phi + (b+ac)\sin^2\phi)^2 = 4abc, c \leq 0\}. \tag{6.6}$$

taking $Z = [-1, 1]^3 \times S^1 \setminus Z_s$, we may define a parameterized map $q : S^1 \times Z \to S^1$ by $q(\theta, z) = q_z(\theta)$ such that for any $f \in \mathfrak{F}$, there exists $z \in Z$ such that Θ_f is topologically conjugate to the map $\theta \mapsto q(\theta, z)$.

Theorem 6.3. *Let Z be the space $([-1,1]^3 \times S^1) \setminus Z_s$, where*

$$Z_s := \{(a,b,c,\theta) \in [-1,1]^3 \times S^1 \mid ((a+bc)\cos^2\phi + (b+ac)\sin^2\phi)^2 = 4abc\}, \quad (6.7)$$

and let $q\colon S^1 \times Z \to S^1$ be a parameterized family of maps given by and let $q\colon S^1 \times (Z \setminus Z_s) \to S^1$ be a parameterized family of maps given by

$$\begin{aligned} q\colon S^1 \times Z &\to S^1 \\ (\theta; a,b,c,\phi) &\mapsto \arctan\left(\frac{a\cos^2(\theta+\phi) + b\sin^2(\theta+\phi)}{c\cos^2\theta + \sin^2\theta}\right). \end{aligned} \quad (6.8)$$

Then for each $f \in \mathfrak{F}$ there exists $z \in Z$ such that the angular dynamics map induced by $f \in \mathfrak{F}$ is topologically conjugate to the map $\theta \mapsto q(\theta, z)$.

Considering only functions from the class $\mathfrak{F}$ is sufficient to ensure the angular dynamics map is continuous, but there are functions not in $\mathfrak{F}$ where this is still the case. Furthermore, some of these maps will have dynamics that do not arise when we restrict our attention to $\mathfrak{F}$. For example, as we will see in Proposition 6.10, all of the angular maps $\theta \mapsto q(\theta, z)$ as above will have winding number 0, -2, or 2. Yet, for $N \in \mathbb{N}$, taking

$$f_N(x,y) = \begin{pmatrix} x + \mathrm{Re}((x \pm iy)^N) \\ y + \mathrm{Im}((x \pm iy)^N) \end{pmatrix}, \quad (6.9)$$

it can be shown that the angular dynamics of f_N are given by $\theta \mapsto \pm N\theta$, a map of winding number $\pm N$. However, $f_N \notin \mathfrak{F}$ for $N > 2$ because the Hessians vanish at the origin. In general, the angular dynamics of Θ are governed by the lowest degree nonlinear terms in the expansion of f. As such, our present analysis is when the dynamics are governed by the quadratic terms.

6.3 The dynamics database

This section gives more detail on how we apply the techniques of Chapter 2 to this particular dynamical system.

Computing the database

The purpose of the database, as outlined in Section 6.1, is to compute and store a rigorous account of the global dynamics of the map $q\colon S^1 \times Z \to S^1$ over the entire parameter space. To do this, we discretize parameter space Z it into a cubical grid $\mathcal{Z}$. The appropriate choice of scale for the grid is application dependent. In this Chapter we chose the level of subdivision in terms of the time required for computation as well as the size of the output. The most refined computation of the full parameter space is subdivided uniformly in each of the four dimensions into 2^6 equal size intervals, forming a grid with $(2^6)^4 = 2^{24}$ cubes, see Section 6.4.

To give a sense of the scale of the computation, the full database for the Newton map was computed on a cluster consisting of 896 processor cores. The computation required approximately 4 hours. The resulting output is a roughly 17 GB file. Post-processing this raw output to extract only continuation information results in a 20 MB file.

As we did with $\mathcal{Z}$, we start with a cubical grid $\mathcal{X}$ on the phase space S^1. For the full 4D parameter space computation we subdivide S^1 into 2^{12} intervals of equal length. The question of what level of subdivision in phase space will optimally extract information about the dynamics given the level of subdivision in parameter space is a subtle one, discussed with respect to the logistic map in Section 5.2. If phase space is too coarsely subdivided some important information about the dynamics will be lost. But in practice we observe that beyond a point there are diminishing returns to further subdivision in phase space without accompanying subdivisions in parameter space. With no good theoretical estimates for the appropriate relationship, we choose the level of subdivision in phase space through experimentation. Finally, given the grids $\mathcal{Z}$ and $\mathcal{X}$, we can for each $\zeta \in \mathcal{Z}$ compute a multivalued map $\mathcal{Q}_\zeta\colon \mathcal{X} \rightrightarrows \mathcal{X}$ that is an outer approximation of q_z for all $z \in \zeta$.

For this application, we choose to compute the induced map on homology over the finite field $\mathbb{Z}_5$ and represent the shift equivalence class of each of the two linear transformations by their invariant factors with coefficients taken mod 5. Moreover, we

adopt the convention in the database of reporting all polynomials as monic, with other coefficients chosen to be smallest in absolute value (considered as integers).

This choice is made for computational convenience, but potentially comes at a price. For example, the uniform winding map on the circle of degree n has just one Morse set with induced map on H_1 represented by the polynomial $x - n$. Because $x - 4 \equiv x + 1 \mod 5$, it's impossible to distinguish a degree 4 map from a degree -1 map without further analysis. However, as shown in Proposition 6.10 for the Newton map the only winding numbers that appear are 2, 0, and -2, and thus using $\mathbb{Z}_5$ coefficients yields the "expected" output of $x - 2$ and $x + 2$ in the cases of winding number 2 and -2 respectively.

Interpreting the database

The database provides a wide variety of information about the underlying dynamical system. For example, as discussed in Section 2.2, the Morse graph gives a characterization of the gradient like structure of the dynamics. To see how to make use of this information, assume that $\mathcal{M}$ denotes a minimal node in a Conley-Morse graph $\mathsf{CMG}(\zeta)$. For each $z \in |\zeta|$, $\mathcal{Q}_\zeta$ is an outer approximation of $q_z \colon S^1 \to S^1$. Since $\mathcal{M}$ is a strongly connected path component and a minimal Morse set, for any $\xi \in \mathcal{M}$, we have $\mathcal{Q}_\zeta(\xi) \subseteq \mathcal{M}$. Thus,

$$q_z(|\mathcal{M}|) \subseteq \operatorname{int}(|\mathcal{M}|)$$

and hence for all $z \in \zeta$, $|\mathcal{M}|$ is a trapping neighborhood for q_z. In particular, we obtain the following proposition.

Proposition 6.4. *Let $\zeta \in \mathcal{Z}$ and let $\mathcal{M}$ denote a minimal node for the Conley-Morse graph $\mathsf{CMG}(\zeta)$. For each $z \in |\zeta|$, the basin of attraction of $\operatorname{Inv}(|\mathcal{M}|, q_z)$ is a set of positive measure.*

To obtain a more detailed understanding of the dynamics associated with the Morse sets we turn to the Conley index. As indicated in Section 6.3, the Conley index is an algebraic topological invariant of $\operatorname{Inv}(|\mathcal{M}|, q_z)$, thus the dynamics associated with Morse sets with different Conley indices must be fundamentally different. Table 6.1 provides

Invariant set	Conley index (H_0 , H_1)
$\emptyset$	(trivial , trivial)
stable fixed point	$(x-1$, trivial)
unstable fixed point, orientation preserving	(trivial , $x-1)$
unstable fixed point, orientation reversing	(trivial , $x+1)$
stable period-T orbit	$(x^T-1$, trivial)
unstable period-T orbit, orientation preserving	(trivial , $x^T-1)$
unstable period-T orbit, orientation reversing	(trivial , $x^T+1)$
S^1 under $f(\theta)=2\theta$	$(x-1$, $x-2)$
S^1 under $f(\theta)=-2\theta$	$(x-1$, $x+2)$

Table 6.1: This table shows the Conley indices corresponding to specific invariant sets that can arise from a map $S^1 \to S^1$. For the last two entries the invariant sets are S^1. As discussed in the text, the converse does not hold, i.e. the presence of a Conley index on the right does not imply the existence of the corresponding invariant set. All coefficients are mod 5.

a "one-way" dictionary: the invariant sets on the left are guaranteed to produce the Conley indices on the right, but the Conley indices on the right, while suggestive of behavior on the left, do not provide any such guarantee. Perhaps the most important consequence of this is that any Morse set with nontrivial Conley index necessarily implies that there is a nonempty invariant set in the corresponding region of phase space. The converse, however, is not true: a Morse set with trivial Conley index need not have empty invariant set.

As the previous comments indicate, given the Conley index of a Morse set Table 6.1 is at best a suggestion of what the associated dynamics may be. However, it is possible to use the Conley index to obtain concrete descriptions of the dynamics.

Definition 6.5. *A map $f\colon X \to X$ exhibits an* attracting T-cycle set *in $N \subset X$ if there exist non-empty, disjoint, compact regions $N_i \subset X$, $i = 1, \ldots, T$, such that $N := \cup_{i=1}^{T} N_i$ and $f(N_{i-1}) \subset N_i$ for $i = 1, \ldots, T$, where $N_0 := N_T$.*

The following result is an immediate consequence of [2, Proposition 5.8].

Proposition 6.6. *Let $\mathcal{M}$ be a Morse set for a Conley-Morse graph $\mathsf{CMG}(\zeta)$. If the H_0 invariant factor for the Conley index of $\mathcal{M}$ is $x^T - 1$, then for any $z \in \zeta$, q_z exhibits an attracting T-cycle set in $|\mathcal{M}|$.*

In a convenient abuse of notation we refer to the Morse set $\mathcal{M}$ as an *attracting T-cycle* if it satisfies the hypotheses of Proposition 6.6.

The following proposition is due to the author of this dissertation.

Proposition 6.7. *Assume $\mathcal{M}$ is a Morse set for a Conley-Morse graph $\mathsf{CMG}(\zeta)$ with $\mathbb{Z}_5$ Conley indices of the form $(x-1, x \pm 2)$. If $z \in \zeta$ then $q_z \colon S^1 \to S^1$ has winding number $\mp 2 (mod\ 5)$.*

Proof. $\mathcal{M}$ is a subset of the grid $\mathcal{X}$ representing the circle S^1, hence it must either be a union of intervals or all of $\mathcal{X}$. We will show that $\mathcal{M}$ must, in fact, be all of $\mathcal{X}$.

Let $P = (P_1, P_0) := (|\mathcal{Q}_\zeta(\mathcal{M})|, |\mathcal{Q}_\zeta(\mathcal{M}) \setminus \mathcal{M}|)$ be the index pair we produce for $\mathcal{M}$. Because the Conley index of $\mathcal{M}$ is nontrivial on the H_0 level, there must be some connected component of P_1 that does not intersect P_0, because every connected component of P_1 intersecting P_0 becomes trivial in $H_0(P_1/P_0, [P_0])$.

Now consider the set of all connected components of P_1 that do not intersect P_0. Each of these is a generator of $H_0(P_1/P_0, [P_0])$. Since $P_1 \setminus P_0 = |\mathcal{M}|$ these are also connected components of $|\mathcal{M}|$. Furthermore, since $\mathcal{Q}_\zeta$ is an outer approximation of a continuous map and $|\mathcal{Q}_\zeta(\mathcal{M})| = P_1$, the image of each generator under the map q_z must be contained in the same connected component of P_1 for every $z \in \zeta$.

Choose an arbitrary $z \in \zeta$ and consider the map on homology q_{z*}. Each generator of $H_0(P_1/P_0, [P_0])$ either maps to another generator (occurring when the image lies in a component of P_1 not intersecting P_0) or it maps to zero. If every generator eventually mapped to zero by iterating the map q_{z*} then the shift equivalence class would be trivial, so there must be at least one generator that does not eventually map to zero.

Consider the set of generators that are not eventually zero under q_{z*}. Because $\mathcal{M}$ is by definition a maximal strongly connected path component of the graph $\mathcal{Q}_\zeta$

and each generator is represented by grid elements in $\mathcal{M}$, the map q_{z*} must cyclically permute the generators. If there are T generators, this results in an H_0 Conley index of $x^T - 1$. Thus, if the H_0 Conley index is $x - 1$ this means there is exactly one nontrivial generator that is not eventually zero, and moreover that the generator must map to itself. Furthermore, since $\mathcal{M}$ is a strongly connected path component, this generator must be all of $|\mathcal{M}|$. Indeed, the index pair $(P_1, P_0) = (|\mathcal{M}|, \emptyset)$.

A connected subset of S^1 is either an interval or all of S^1. Because the H_1 Conley index of $\mathcal{M}$ is nontrivial and $\mathcal{M}$ has index pair $(|\mathcal{M}|, \emptyset)$, it follows that $|\mathcal{M}| = S^1$ and $\mathcal{M} = \mathcal{X}$.

Using integral homology, the map $q_{z*} : H_1(S^1) \xrightarrow{\times k} H_1(S^1)$ implies q_z has H_1 Conley index $x - k$ and winding number k. The proposition follows taking coefficients mod 5.

□

The main Corollary of [30] and [26, Theorem 1] immediately imply the following result.

Corollary 6.8. *Assume $\mathcal{M}$ is a Morse set for a Conley-Morse graph $\mathsf{CMG}(\zeta)$ with $\mathbb{Z}_5$ Conley indices of the form $(x - 1, x \pm 2)$. If $z \in \zeta$ then $q_z \colon S^1 \to S^1$ is chaotic and its topological entropy is bounded below by $\ln 2$.*

Conley-Morse graphs of the following form appear frequently in the databases discussed in Section 6.4.

$$[\mathcal{M}_2 : (\text{trivial}, x^{T-1} + x^{T-2} + \cdots + x + 1)] \to [\mathcal{M}_1 : (x^T - 1, \text{trivial})]$$

This diagram should be read as follows: there exist two Morse sets, $\mathcal{M}_2$ and $\mathcal{M}_1$, such that $\mathcal{M}_2 > \mathcal{M}_1$ (with respect to the partial order on the Morse sets) and they have the given Conley indices. To shed some light on the dynamics associated with Morse set $\mathcal{M}_2$ we investigate the special case where $T = 3$.

Using the database computation on the entire parameter space $[-1, 1]^3 \times S^1$ discussed in Section 6.4 we know that the case $T = 3$ arises at the parameter value $z = (a, b, c, \phi) = (-1, 1, -0.104, 1.597)$. In particular, the Conley index of the Morse set $\mathcal{M}_2$ is $x^2 + x + 1$ on H_1 and trivial on H_0. Figure 6.1 shows a graph of q_z, along with an example of

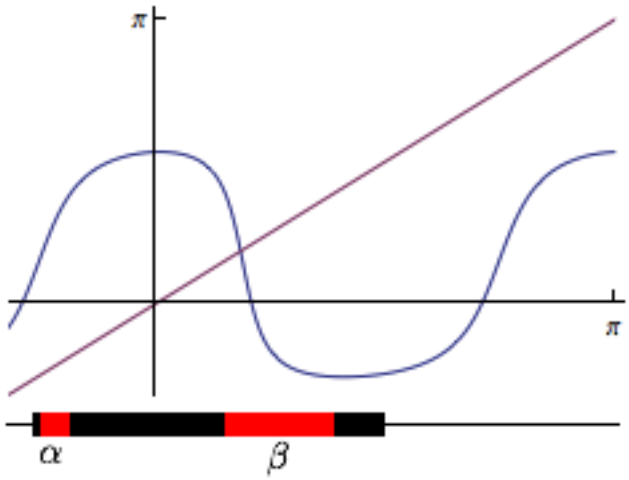


Figure 6.1: The blue line is the graph of the Newton map q_z at parameter $z = (a, b, c, \phi) = (-1, 1, -0.104, 1.597)$. The purple line is the line $y = x$. From the database computation we are assured that at this parameter value there is an index pair (P_1, P_0) with Conley index $x^2 + x + 1$ on H_1. The black regions are the exit set P_0 and the black and red regions together are P_1. The dynamics we are interested in occur in the red set $P_1 \setminus P_0$. Here, after quotienting out the exit set there are two generators of first homology, α and β, each corresponding to an interval in $\mathrm{cl}(P_1 \setminus P_0)$.

an index pair $P = (P_1, P_0)$ for $\mathcal{M}_2$. There are two generators of first homology on $(P_1/P_0, [P_0])$ corresponding to the red intervals in $P_1 \setminus P_0$, labeled here α and β.

Under the map q, the image of the generator α extends across β. In the pointed quotient space $(P_1/P_0, [P_0])$ this corresponds to a wrapping around the loop β, hence $\alpha \mapsto \beta$ on homology. The image of β, on the other hand extends across β and α, but with a change in orientation. This change in orientation is reflected by a minus sign, so that $\beta \mapsto -\alpha - \beta$. Thus, the induced map q_{P*} on first homology is given by the matrix

$$\begin{bmatrix} 0 & -1 \\ 1 & -1 \end{bmatrix}$$

The only invariant factor of this matrix is the characteristic polynomial $x^2 + x + 1$, and since this polynomial is not divisible by x it represents the H_1 Conley index for this index pair.

Up to this point the comments of this section have focussed on interpreting the

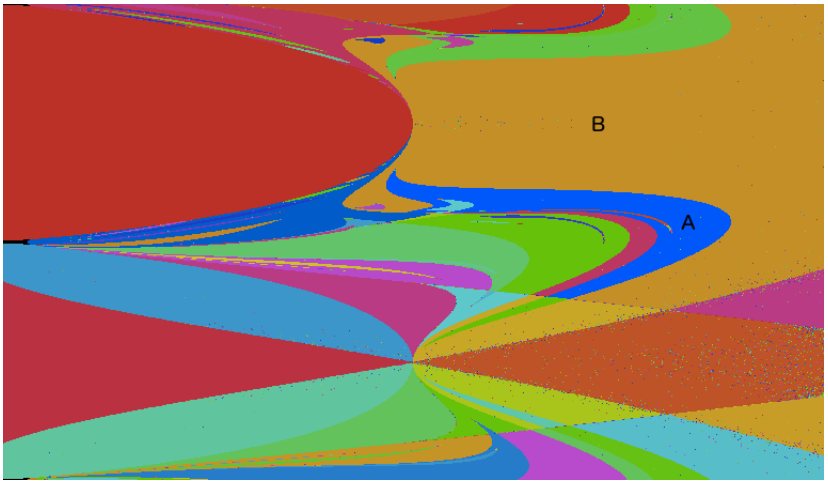
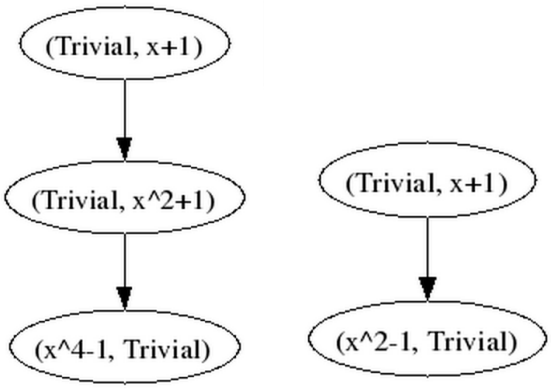


Figure 6.2: A picture of the parameter space for the map q restricted to the 2-dimensional slice $Z_0 = \{-1\} \times \{1\} \times [-1, 1] \times [0, \pi] \subset Z$. Black regions correspond to parameter boxes containing parameters for which the map is singular. Each non-black color represents a distinct Conley-Morse graph continuation class. Conley-Morse graphs for the regions labeled A and B are given in Figure 6.3.

dynamics associated with the Morse sets of the Conley-Morse graphs. The power of the database approach is that it provides insight into the dynamics over the entire parameter space. The "big picture" view of a database computation is given by its continuation graph. Especially for high-dimensional parameter spaces where visualization is difficult or impossible, the continuation graph gives an accessible representation of parameter space, showing which dynamics are adjacent.

In cases where parameter space is two-dimensional or where we are interested in a particular two-dimensional slice of parameter space, however, it is possible to visualize the Conley-Morse graph continuation classes directly. In particular, the Database Explorer is software that provides a picture of parameter space with the Conley-Morse graph continuation classes as color coded regions. Figure 6.2 provides an example of such an image extracted from the Database Explorer for the Newton map where the parameter space is $Z_0 = \{-1\} \times \{1\} \times [-1, 1] \times [0, \pi] \subset Z$.

The two regions labeled A and B in Figure 6.2 have Conley-Morse graphs shown in Figure 6.3. Here Proposition 6.6 allows us to conclude that region A exhibits an



(a) Parameter region A (b) Parameter region B

Figure 6.3: Conley-Morse graphs for the respective parameter space regions in Figure 6.2. Each oval represents a Morse set with ordering given by the directed edges. The Conley index at a given Morse set is represented by the tuple (H_0 Conley index, H_1 Conley index) contained within the oval.

attracting 4-cycle set and region B an attracting 2-cycle set.

What's not indicated in Figure 6.3 is that the Morse sets in regions A and B with Conley index (Trivial, $x+1$) are actually both part of the same isolating neighborhood continuation class. By looking at the dictionary in Table 6.1 we see this continuation class of Morse sets all have the Conley index of an orientation-reversing unstable fixed point. Of course, this is reading the dictionary in the 'wrong way' and does not imply that there is an isolated fixed point, but it is suggestive of the behavior at such Morse sets.

6.4 The computed dynamics database

In this section we present more visual representations of the global angle dynamics over parameter space, with the goal of exhibiting the most interesting dynamic phenomena.

A Conley-Morse Database over $[-1,1]^3 \times S^1$

We apply the Conley-Morse Database methods to the parameterized dynamical system $q : S^1 \times Z \to S^1$ defined by (6.2). In particular, we grid the 4-dimensional parameter

space $Z = [-1, 1]^3 \times [0, \pi)$ via uniform rectangular subdivision 2^6 on a side. This implies that $\mathcal{Z}$, the parameter space grid, consists of $(2^6)^4 = 2^{24} \approx 16$ million rectangular parameter regions. As is indicated at the end of Section 6.2, $\mathcal{Y} \subset \mathcal{Z}$ denotes the set of parameter boxes at which we do not compute Conley-Morse graphs. At this resolution, $\mathcal{Y}$ consists of 3,223,104 boxes. This corresponds to approximately 19.2% of the parameter space.

In this section we will analyze these results of these computations. We will be able to replicate the finding of [14] that the dynamics can be more interesting than a globally attracting fixed point. We also will identify parameter regions where there are multiple basins of attraction, attracting T-cycles for $T > 1$, and chaotic behavior. Finally we consider the structure of components of parameter space corresponding to different winding numbers.

Interesting Dynamics

We begin by identifying the parameter regions for which our computations do not rule out a globally attracting fixed point. By Table 6.1 the Conley-Morse graph of such a dynamical system should have a single nontrivial Morse set with Conley index $(x - 1, \text{trivial})$ (which will be minimal). There may be other Morse sets, but they must all have trivial Conley index. Let $\mathcal{A}_1 \subseteq \mathcal{Z}$ denote the set of grid elements for which this holds. We find $\mathcal{A}_1$ contains approximately 63.2% of the elements of $\mathcal{Z} \setminus \mathcal{Y}$ (equivalently, at most 51.2% of $\mathcal{Z}$). Noting that Conley-Morse graphs do not distinguish attracting 1-cycle sets from stable fixed points, we realize these figures as upper bounds for how often this dynamics can occur.

On the complement (in $\mathcal{Z} \setminus \mathcal{Y}$) of these parameter values we have "interesting" dynamics (i.e. dynamics other than a globally attracting fixed point). Our previous discussion implies we get such behavior on *at least* 48.8% of $\mathcal{Z}$. We consider three ways (not all necessarily mutually exclusive) in which we can obtain interesting dynamics. The first two ways involve attracting T-cycle sets and multiple basins of attraction:

Proposition 6.9. *Define*

$$\mathcal{A}_T := \{\zeta \in \mathcal{Z} \mid \textit{there is an attracting } T\textit{-cycle, for } T > 1\}$$

and

$$\mathcal{A}_{MMN} := \{\zeta \in \mathcal{Z} \mid \mathsf{MG}_\zeta \textit{ has multiple minimal nodes}\}.$$

Let $z \in \zeta$. If $\zeta \in \mathcal{A}_T$, then there exists a set of initial conditions with positive measure such that the asymptotic dynamics is not a fixed point. If $\zeta \in \mathcal{A}_{MMN}$ then there exists two sets of initial conditions with positive measure that converge to along different directions. Furthermore, $\mathcal{A}_T$ and $\mathcal{A}_{MMN}$ make up 18.79% and 1.56% of $\mathcal{Z}$, respectively (approximately, rounding down).

Proof. Consider the case that $\zeta \in \mathcal{A}_{\mathrm{T}}$. Let $\mathcal{M}$ be a Morse note of $\mathsf{CMG}(\zeta)$ which is attracting T-cycle. By Proposition 6.6 for any $z \in \zeta$, $|\mathcal{M}|$ is an attracting T-cycle set. Since $\mathcal{M}$ is made up of elements of $\mathcal{X}$, $\mathcal{M}$ has positive measure. By definition of an attracting T-cycle set if $x \in |\mathcal{M}|$, then $\omega(x, q_z) \subset \mathrm{Inv}(|\mathcal{M}|, q_z)$ and $\mathrm{Inv}(|\mathcal{M}|, q_z)$ does not contain any fixed points. The argument for $\zeta \in \mathcal{A}_{\mathrm{MMN}}$ is similar.

The percentages come from counting the parameter grid elements in both $\mathcal{A}_{\mathrm{T}}$ and $\mathcal{A}_{\mathrm{MMN}}$.

□

We now show the existence of chaotic dynamics indicated by our database computations. Corollary 6.8 establishes chaotic dynamics on the following parameter regions:

$$\mathcal{U}_{\pm 2} := \{\zeta \in \mathcal{Z} \mid \mathsf{CMG}(\zeta) \text{ has the form } [\mathcal{M}_1 : (x-1, x \mp 2)]\}.$$

From the full database computation we conclude that $\mathcal{U}_2$ and $\mathcal{U}_{-2}$ consist of at least 4.62% and 3.88% of $\mathcal{Z}$, respectively.

Winding Numbers

By Proposition 6.7 if z is a parameter value associated with $\mathcal{U}_{\pm 2}$, then the winding number of q_z is ± 2. We remark that if z is associated with $\mathcal{A}_1$, then the winding number of q_z is 0. In fact, using the database computations we have discovered all

possible winding numbers. For the proof see [5]—it is omitted here as it is not the work of the author of this dissertation.

Proposition 6.10. *The only possible winding numbers of q_z are 0 and ± 2.*

Conley-Morse Databases over Two-Dimensional Slices of Parameter Space

To obtain finer information about the structure of the dynamics of Newton's method we restrict ourselves to computations over two-dimensional regions of phase space. In contrast to the full four-dimensional computations these computations can be performed fairly rapidly. We computed a 2D database from which we were able to extract examples demonstrating classically described dynamical phenomena.

A Full 2D Section

We apply the database methods over the cross-section of $[-1,1]^3 \times [0,\pi)$ given by $a = -1, b = 1$. We subdivide this 2D parameter space $[-1,1] \times [0,\pi)$ uniformly with 2^{11} subdivisions on a side, and 2^{16} divisions in phase space. The resulting parameter space grid $\mathcal{Z}$ consists of $(2^{11})^2 \approx 4$ million rectangular parameter regions. Additionally, we compute over subregions of this parameter space, but still dividing into 2^{11} subdivisions per side. The scale this translates to therefore depends on the region we are computing on. The corresponding cross-section in the 4D computation discussed earlier is comprised of only $(2^6)^2 = 4096$ parameter boxes. Hence we are examining the system at a much higher resolution. Figure 6.2 shows this high-resolution cross-section.

The finer resolution results in finer Conley-Morse graphs being computed compared to the coarse 4D computations. Since the results in the cross-section do not necessarily reflect what occurs over the rest of parameter space, we are less interested in reporting observed fractions of parameter space in which we see certain behaviors. Rather, we are interested in what behaviors happen at all; these tell us the richness of what dynamics are possible under (6.2).

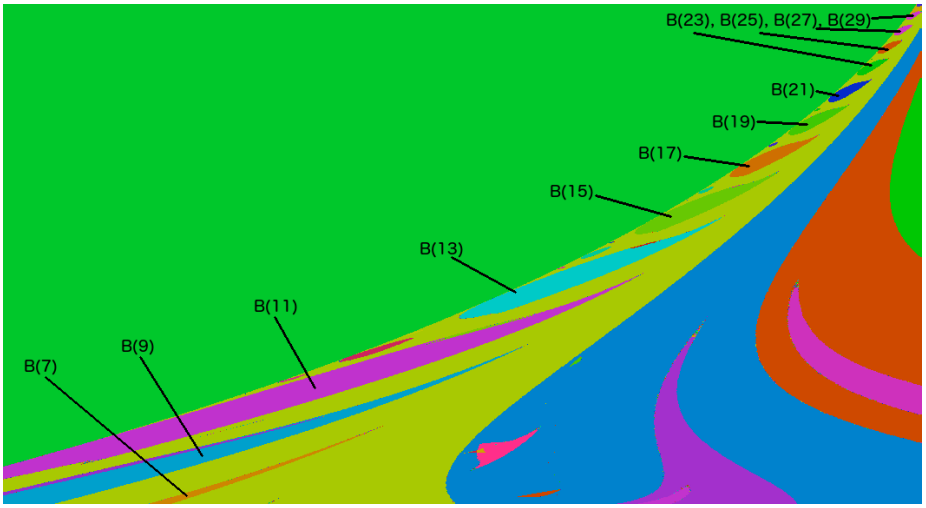


Figure 6.4: A visualization parameter space restricted to $a = -1, b = 1$, with $-0.32 \leq c \leq -0.02$, $1.81 \leq \phi \leq 2.21$. Here c is on the x-axis and ϕ on the y-axis, and each axis is subdivided into 2^{11} boxes. The regions $B(T)$ indicate continuation classes with a stable T-cycle. The surrounding yellow region has a stable 1-cycle. Some unlabeled continuation classes appear, labeled, in Figure 6.5.

Ascending sequence of T-cycles

At a higher resolution, one interesting discovery was a sequence of continuation classes with stable T-cycles of increasing degrees. This behavior is represented in Figure 6.4. Each of these regions is surrounded by a continuation class with dynamics given by an attracting 1-cycle.

For each of the indicated regions, the continuation class has a corresponding Conley-Morse graph of consistent structure. For a specified period T, the Conley-Morse graph of $B(T)$ is given by

$$[\mathcal{M}_1 : (\text{trivial}, x^{T-1} + x^{T-2} + \ldots + x + 1)] \to [\mathcal{M}_2 : (x^T - 1, \text{trivial})]. \tag{6.10}$$

The largest period found was $T = 52$.

To investigate whether we might find more with the same level of subdivision on a smaller region, we computed the cross-section ($a = -1$, $b = 1$, $-0.15 \leq c \leq -0.10$, and $1.98 \leq \phi \leq 2.07$) located roughly between $B(13)$ and $B(15)$. This computation is shown in Figure 6.5.

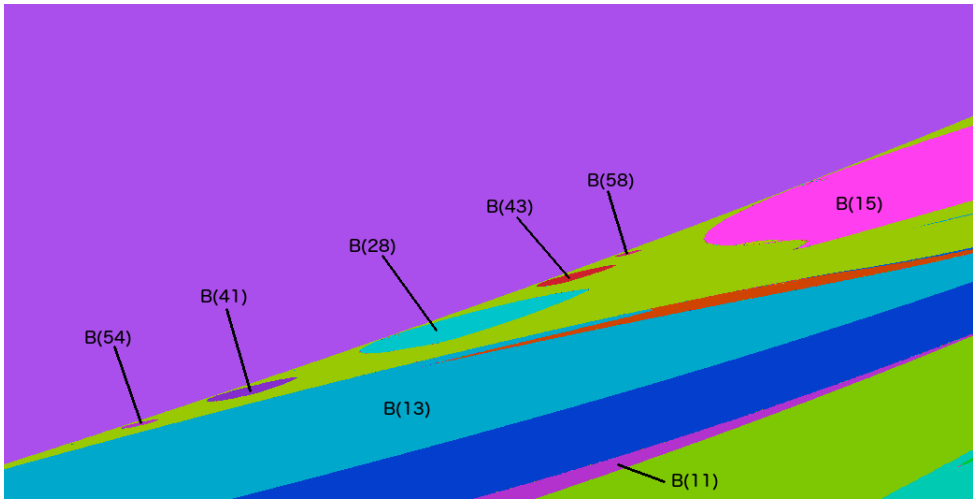


Figure 6.5: A visualization of parameter space restricted to $a = -1$, $b = 1$, with $-0.15 \leq c \leq -0.10$, $1.98 \leq \phi \leq 2.07$. Here c is on the x-axis and ϕ on the y-axis, and each axis is subdivided into 2^{11} boxes. Compare with Figure 6.4.

We suspect further investigation at finer resolution would reveal even more structural detail.

Period Doubling Cascade

In the calculation represented by Figure 6.6 we see a sequence of adjacent continuation classes with attracting cycle sets of periods 2, 4, 8, and 16. This suggests a period-doubling cascade, as discussed in the context of the logistic map in Chapter 5 and [6, Section II.E].

Observing Figure 6.6, the labeled continuation classes have Conley-Morse graphs

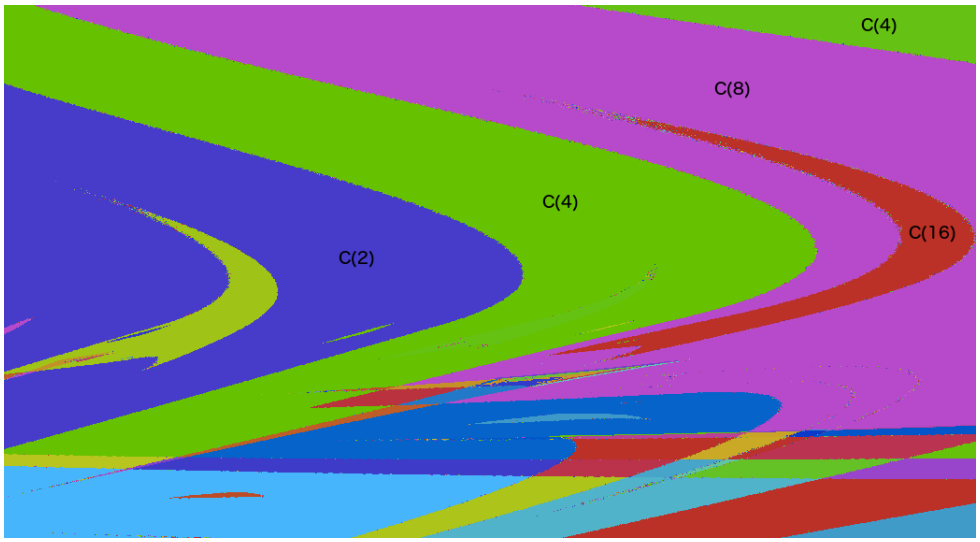


Figure 6.6: A visualization of parameter space restricted to $a = -1$, $b = 1$, with $-0.022 \leq c \leq 0.10$, $1.75 \leq \phi \leq 1.88$. Here c is on the x-axis and ϕ on the y-axis, and each axis is subdivided into 2^{11} boxes. The regions $C(T)$ indicate continuation classes with attracting T-cycles.

given by the following:

$$\begin{aligned}
\text{C}(2)\ :&[\mathcal{M}_1 : (\text{trivial}, x+1)] \to [\mathcal{M}_2 : (x^2-1, \text{trivial})] \\
\text{C}(4)\ :&[\mathcal{M}_1 : (\text{trivial}, x+1)] \to [\mathcal{M}_2 : (\text{trivial}, x^2+1)] \to [\mathcal{M}_3 : (x^4-1, \text{trivial})] \\
\text{C}(8)\ :&[\mathcal{M}_1 : (\text{trivial}, x+1)] \to [\mathcal{M}_2 : (\text{trivial}, x^2+1)] \\
&\to [\mathcal{M}_3 : (\text{trivial}, x^4+1)] \to [\mathcal{M}_4 : (x^8-1, \text{trivial})] \\
\text{C}(16)\ :&[\mathcal{M}_1 : (\text{trivial}, x+1)] \to [\mathcal{M}_2 : (\text{trivial}, x^2+1)] \\
&\to [\mathcal{M}_3 : (\text{trivial}, x^4+1)] \to [\mathcal{M}_4 : (x^8+1, \text{trivial})] \to [\mathcal{M}_5 : (x^{16}-1, \text{trivial})]
\end{aligned} \tag{6.11}$$

As outlined in the indicated discussions, caution in interpretation is warranted, however, as such behavior need not arise as a period doubling bifurcation. In this example, the database is suggestive of where to conduct further confirmatory analysis.

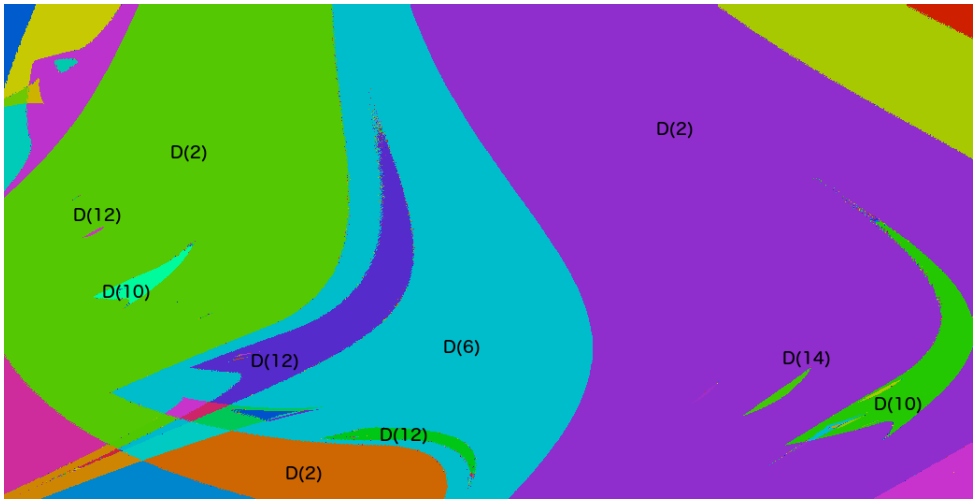


Figure 6.7: A visualization of parameter space restricted to $a = -1, b = 1$ with $-0.17 \leq c \leq 0.012$, $1.78 \leq \phi \leq 1.87$. Here c is on the x-axis and ϕ on the y-axis, and each axis is subdivided into 2^{11} boxes. The regions $D(T)$ indicate continuation classes with attracting T-cycles.

Isolated Period 8 Cycles

Attracting T-cycles of period given by a power of 2 are not by themselves indicative a period-doubling cascade. As an example, in the **IsolatedPeriodEight** database we find the continuation classes indicated in Figure 6.7. The indicated parameter regions have Conley-Morse graphs given by the following:

$$\begin{aligned}
\text{D}(6) :&[\mathcal{M}_1 : (\text{trivial}, x+1)] \to [\mathcal{M}_2 : (\text{trivial}, x^4+x^2+1)] \to [\mathcal{M}_3 : (x^6-1, \text{trivial})] \\
\text{D}(8) :&[\mathcal{M}_1 : (\text{trivial}, x+1)] \to [\mathcal{M}_2 : (\text{trivial}, x^6+x^4+x^2+1)] \\
&\to [\mathcal{M}_3 : (x^8-1, \text{trivial})] \\
\text{D}(10) :&[\mathcal{M}_1 : (\text{trivial}, x+1)] \to [\mathcal{M}_2 : (\text{trivial}, x^8+x^6+x^4+x^2+1)] \\
&\to [\mathcal{M}_3 : (x^{10}+1, \text{trivial})].
\end{aligned} \tag{6.12}$$

Here we see evidence of attracting T-cycles with even T that do not appear to arise as part of a period-doubling cascade. This is very much of the same flavor as the previously described ascending sequence of T-cycles.

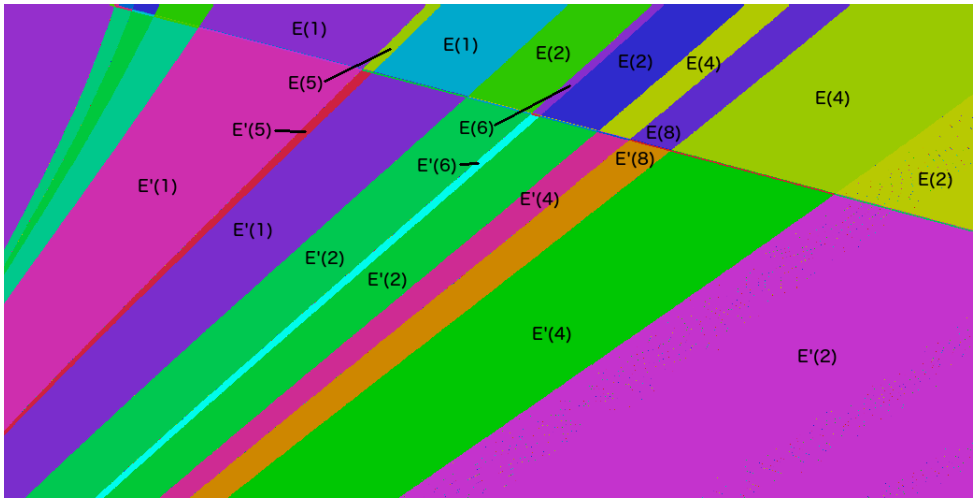


Figure 6.8: A visualization of parameter space restricted to $a = -1$, $b = 1$, and $0.15 \leq c \leq 0.43, 1.10 \leq \phi \leq 1.26$. Here c is on the x-axis and ϕ on the y-axis, and each axis is subdivided into 2^{11} boxes. Regions labeled $E(T)$ have stable T-cycle attractors, while $E'(T)$ regions have stable T-cycle attractors, and an additional attractor with a Conley index of a fixed point.

Multiple Basins of Attraction

Multiple basins of attraction were observed in the full database computation. A more refined computation shown in Figure 6.8 demonstrates a sequence of continuation classes that exhibit an apparent period-doubling cascade, and then a bifurcation across which the corresponding regions exhibit the same cascade, but with a second attractor with the Conley index of a fixed point.

Chapter 7

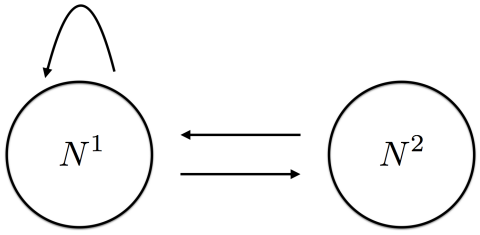
Application: population models

This chapter is a lightly edited version of work originally published in [8], done jointly with Konstantin Mischaikow.

The logistic map discussed at the beginning of Chapter 2 and again in Chapter 5 is one of the simplest nonlinear discrete time models in population biology, but not one of the most realistic. As effective as it is at demonstrating important dynamical phenomena, e.g. a period-doubling cascade, in a very simple setting, we nevertheless are interested in the practical analysis of functions with stronger biological motivation, and many such models have been proposed [27, 41].

As much work of the last century has demonstrated, the invariant structures of nonlinear systems can be extremely sensitive to arbitrarily small perturbations. For example, bifurcations can occur on a Cantor set in parameter space [35, 36], making it hopeless to try to give an explicit account of the dynamics for every parameter value.

The goal of this chapter is to provide a concrete demonstration of the potential importance of this approach. This is done by considering a classical example from population biology: a nonoverlapping overcompensatory two age class model. From the perspective of modeling dynamical systems as networks this is an extremely simple system



where the first age class N^1 produces offspring (the self edge) and becomes the second age class N^2, and the second age class produces offspring that belong to the first age class.

A key decision affecting the dynamics that we investigate here is the functional form of the nonlinear interactions associated with the edges.

In Section 7.1 we provide justification for using the following map as an analytic representation for this network

$$\begin{aligned} N^1_{t+1} &= \left(\sum_{i=1}^{2} \theta_i N^i_t \right) e^{-0.1\left(\sum_{i=1}^{2} (s\theta_i + 1 - s) N^i_t \right)} \\ N^2_{t+1} &= 0.7 N^1_t \end{aligned} \tag{7.1}$$

For the moment we remark that it is obtained by generalizing the models considered in [23, 45] and that that there are three parameters: $0 \leq \theta_i$, $i = 1, 2$ which represent reproduction rates and $0 \leq s \leq 1$ which characterizes competition between age classes.The purpose of exploring the dynamics is to either verify that this is an adequate model or, having accepted the validity of the model, to understand the expected observed behavior as parameters change.

Equation (7.1) makes use of the Ricker nonlinearity [41] which is unimodal and thus it is reasonable to expect a period doubling cascade, e.g., the existence of infinitely many bifurcations on arbitrary small scales with respect to parameters. In the real world, however, except for carefully controlled laboratory settings [10], population measurements are typically inaccurate and are subject to significant random perturbations. Therefore, from an experimental perspective detecting the occurrence of bifurcations is nontrivial under the best conditions, which calls into question whether we should attach great practical significance to analyzing bifurcations.

Finally, the 'verifiability' of the different parameters are not the same. We might suspect, for example, that the reproductive rates represented by θ_i are more easily quantified than the degree or mechanisms of inter- and intra-species competition that we are modeling by s. In Section 7.1 we provide a rationale for the use of s as a parameter to model competition, but to a large extent this is speculative.

With these observations in mind the goals of this chapter are twofold: explain how

the database framework outlined in Chapter 2 can provide robust, rigorous information about important dynamical structures, and at the same time demonstrate that model assumptions (in particular regarding what we are understanding as intra-species competition) play a crucial role in determining the expected observable dynamics and, perhaps more significantly, have an important impact on the relative roles of the reproduction rates in achieving certain dynamics. Note that the complete results of performing the database computations on Equation (7.1) can be accessed at [28].

In Section 7.2 we discuss three biologically relevant issues:

(1) biennial population dynamics,

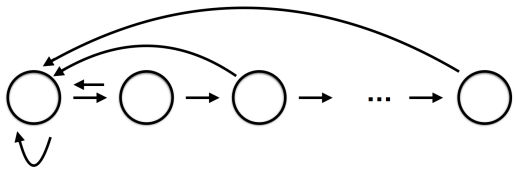
(2) bistability, and

(3) permanence or persistence,

where we focus on how the choice of model, i.e. the value of s, affects the relative significance of the reproduction rates θ_i.

7.1 Plant and Fish models

The database for dynamics is an especially useful tool in the modeling of complex systems where models cannot be completely determined from first principles. Population biology, in this case, is an archetypical example of a discipline where modeling must necessarily be tolerant of imprecision. To keep the discussion simple we make use of discrete time models and consider the dynamics of the age structured population of a single species. The network diagram for this type of system takes the form



which indicates that with time each age class ages into the successively older cohort, and that the new youngest age class is formed by the offspring of all age classes taken

Invariant set	Conley index (H_0 , H_1)
$\emptyset$	(trivial , trivial, trivial)
stable fixed point	($x-1$, trivial, trivial)
fixed point, 1-d unstable, orientation preserving	(trivial , $x-1$, trivial)
fixed point, 1-d unstable, orientation reversing	(trivial , $x+1$, trivial)
fixed point, 2-d unstable, orientation preserving	(trivial, trivial , $x-1$)
fixed point, 2-d unstable, orientation reversing	(trivial, trivial , $x+1$)
stable period-T orbit	(x^T-1 , trivial, trivial)
period-T orbit, 1-d unstable, orientation preserving	(trivial , x^T-1, trivial)
period-T orbit, 1-d unstable, orientation reversing	(trivial , x^T+1, trivial)
period-T orbit, 2-d unstable, orientation preserving	(trivial, trivial , x^T-1)
period-T orbit, 2-d unstable, orientation reversing	(trivial, trivial , x^T+1)
stable invariant circle	($x-1$, $x-1$, trivial)
invariant circle, 1-d unstable manifold	(trivial, $x-1$, $x-1$)

Table 7.1: This table describes the Conley indices of certain elementary invariant sets. All coefficients are mod 3.

together. While this network provides a framework in which to consider the problem, any actual dynamics requires modeling assumptions about the function the edges in the graph are meant to represent. Most of what follows in this section is classical. However, we include it as it emphasizes the philosophy of how the database approach can be employed to identify sensitive aspects of the modeling process.

Let $N_{t+1} \in \mathbb{R}^n$ be the population vector representing the age structure of the population at time $t+1$ where N_t^i denote the population of age class i. We assume that the future population is determined by a continuous function $N_{t+1} = f(N_t)$. The classical Leslie matrix provides a linear model [22], but fails to explain observed complicated population dynamics and populations levels typically become extinct or unbounded. This suggests that a nonlinear function f may be a more appropriate choice. Certain biological assumptions will constrain the choice of f—for example in the absence of migration and abiogenesis, we want $f(0) = 0$. Furthermore, even in very simple models it is desirable to reflect the reality that a population cannot grow without bound. This can be accomplished by insisting for a sufficiently large population $|N|$ that $|f(N)| \leq |N|$.

There are many biologically plausible functions that meet these criteria [27]. A common way to ensure that the population remains bounded is to introduce an overcrowding effect that make the reproduction rates density-dependent in such a way that the reproduction rate grows more slowly for larger populations, perhaps even decreasing beyond a certain point.

One particular choice of unimodal nonlinearity for discrete time models is the Ricker nonlinearity [41], which in the simple case of only one age class N is given by

$$f(N) = Ne^{r(1-N/K)} \tag{7.2}$$

where the parameters r and K represents the growth rate for small populations and carrying capacity of the environment, respectively.

Depending on the kinds of data available to the ecologist, however, it may be more convenient to express this same nonlinearity using different parameters. Two standard reparameterizations are

$$f(N) = \theta Ne^{-bN} \tag{7.3}$$

(see [45]) and

$$f(N) = \theta N e^{-b\theta N} \tag{7.4}$$

(see [23]).

Although these are just reparameterizations, they offer some advantages for interpretation. The parameter θ in each of these formulations, for example, represents the reproduction rate when the population is small and overcrowding effects are negligible. Perhaps more importantly, the exponent makes explicit what quantity is responsible for the overcrowding effect, whereby at sufficiently high populations the number of offspring actually decrease. Taking b to be a constant, in Equation (7.3) the exponent is proportional to the number of adults in the population, while in Equation (7.4) the exponent is proportional to the reproduction rate (in the absence of overcrowding) times the population.

Both Equations (7.3) and (7.4) can be extended to the case of multiple age classes (see [23, 45]) following the approach of the Leslie matrix. This introduces parameters θ_i for the reproduction rate of age class i as well as ρ_i, the survival rate of the proportion of age class i to age class $i+1$. In particular, Equation (7.3) generalizes to

$$\begin{aligned} N_{t+1}^1 &= \left(\sum_{i=1}^{k} \theta_i N_t^i \right) e^{-b \sum_{i=1}^{k} N_t^i} \\ N_{t+1}^i &= \rho_i N_t^{i-1} \qquad\qquad i = 2, \ldots, k \end{aligned} \tag{7.5}$$

and Equation (7.4) generalizes to

$$\begin{aligned} N_{t+1}^1 &= \left(\sum_{i=1}^{k} \theta_i N_t^i \right) e^{-b \sum_{i=1}^{k} \theta_i N_t^i} \\ N_{t+1}^i &= \rho_i N_t^{i-1} \qquad\qquad i = 2, \ldots, k \end{aligned} \tag{7.6}$$

As discussed above, each of these models can be given a different biological interpretation based on the exponent in the equation determining N_{t+1}^1. Because the exponent in Equation (7.5) is proportional to the total population, this can be understood as representing the youngest members of the species being overcrowded by—or competing with—the existing adult population for resources. This might model, for example, the way that saplings compete with taller, more mature trees for sunlight. For this reason we will refer to Equation (7.5) as a "plant model".

The plant model is discussed from a more classical point of view in [45] and from our database point of view in [2]. In particular, ([45, Figure 1]) gives a picture of the bifurcation diagram along the line indicated in Figure 7.1. (Note that to make this comparison requires the choice of $\rho_2 = 0.7$). This bifurcation diagram shows the presence of tremendously complicated changes in dynamics on fine scales that the database approach summarizes in a much coarser way. In particular, the database reduces all of this complexity to a relatively small number of Conley-Morse graphs.

In Equation (7.6), overcrowding depends on the potential number of recruits in the absence of any density-dependent effects. Holding the adult population constant and increasing the reproduction rate will increase the overcrowding in this model but not in the plant model. Biologically, this can be taken to represent the youngest members of the species being in competition with themselves. In [23] this is taken to be a model of a striped bass population, so for that reason we will refer to Equation (7.6) as a "fish model".

We want to emphasize that both the fish and plant models have been extensively studied [18, 23, 27, 41, 45] and both models exhibit similar and similarly wide range of dynamics: global stable equilibria, stable periodic dynamics, bistability, chaotic dynamics, etc. However, these models are based on different assumptions concerning overcrowding. Furthermore, it is reasonable to suspect that in reality the youngest age group experiences competition *both* from themselves and from the older age groups. Thus a natural series of questions is the following:

- Do the similarities observed in the dynamics of Equations (7.5) and (7.6) depend on the reproduction parameters in the same way?

- If so, can we understand this similarity in terms of the similar network structure of the two models?

- If not, can we understand the differences by putting these models in a wider context of models with the same network structure?

With this in mind it is natural to consider a possible continuum of intermediate cases where the young members of the species compete with both the adults and the offspring

to different degrees. The simplest way to realize this continuum, in the absence of any specific biological model, is to take a linear interpolation between the two exponents using a new parameter $s \in [0,1]$, *i.e.*,

$$\begin{aligned} N_{t+1}^1 &= \left(\sum_{i=1}^{k} \theta_i N_t^i\right) e^{-b\left(s\sum_{i=1}^{k}\theta_i N_t^i + (1-s)\sum_{i=1}^{k} N_t^i\right)} \\ N_{t+1}^i &= \rho_i N_t^{i-1} \qquad\qquad i = 2,\ldots,k \end{aligned} \tag{7.7}$$

This reduces to the plant model when $s = 0$ and the fish model with $s = 1$. In this larger framework we can observe how the dynamics change as the nature of competition changes by varying s.

Given this crude level of modeling, a detailed description of the dynamics on the level of invariant sets is somewhat ambitious and, given the difficulty of obtaining such results, perhaps even counterproductive.

Note that these results are based on the dynamics that can be extracted from the computed outer approximations $\mathcal{F}$. Assume that g is an alternative model to Equation (7.7) that is for example based on more information about the form of competition. Furthermore, assume that $\mathcal{F}$ is an outer approximation of g, i.e.

$$g(\xi,\zeta) \subseteq \mathrm{int}(|\mathcal{F}_\zeta(\xi)|), \tag{7.8}$$

then the database information provided by the computations based on f are valid for g.

The obvious question is, how plausible is the assumption of Equation (7.8)? Since the computations we perform are based on rather crude grids in both phase space $\mathcal{X}$ and parameter space $\mathcal{Z}$, and since we are using interval arithmetic to approximate the dynamics, we claim that this is not an unrealistic assumption. In particular, the computations could be performed at a lower resolution to gain further confidence in the assumption.

7.2 Computational Results

In this section, we provide the results of the database of dynamics computations as outlined in Chapter 2. Details about the algorithms we use are given in [6]. For the

full database computation on the three parameters θ_i, $i = 1, 2$ and s see [28].

We begin by discussing our choice of phase space X and parameter space Z. We then describe the database results for $s = 0$ (plant model) and $s = 1$ (fish model). This information is used to formally state the three questions of interest. We then discuss these questions based on the above mentioned computations of intermediate values of s.

Following [2, 45] we restrict our computations to Equation (7.1) that is obtained from Equation (7.7) by restricting to two age classes and setting $\rho_1 = 0.7$. The parameter b is a scale parameter that has no effect on the dynamics, so we follow [2] and set $b = 0.1$. Even in this simplified setting the difference between the dynamics of the plant and fish models is apparent.

To apply the database computations requires us to restrict to a compact set of parameter values. Conceptually this is not a problem since there always exists an upper bound on the rate of reproduction, however appropriate values for this upper bound is a modeling question that is problem specific. In this case we have chosen

$$Z = \{(\theta_1, \theta_2) \in [0, 50] \times [0, 50]\}$$

for various values of $s \in [0, 1]$, since this range of reproduction rates provides a diverse set of dynamics.

As in the case of parameter space, to apply the database computations we need to choose a compact phase space X. We note that Equation (7.1) has a global compact attractor and hence there exists a compact forward invariant region in $[0, \infty)^2$ that contains all the relevant dynamics. Explicit bounds on regions of this form depend on s and can be determined by bounding the number of new recruits:

$$N_{t+1}^1 = \left(\sum_{i=1}^{k} \theta_i N_t^i \right) e^{-b\left(s \sum_{i=1}^{k} \theta_i N_t^i + (1-s) \sum_{i=1}^{k} N_t^i\right)}.$$

The maximum of this function over the first quadrant is

$$\frac{10\overline{\theta}}{e(1 + (\overline{\theta} - 1)s)},$$

where $\overline{\theta} = \max\{\theta_1, \theta_2\}$. This is a nondecreasing function of $\overline{\theta}$ for all $s \in [0, 1]$, so to determine a the maximum value for all choices of θ_1, θ_2 we set $\overline{\theta} = 50$. This gives an

upper bound on the value of N^1_{t+1} that roughly corresponds to the scale of the dynamics. From this we can bound $N^2_{t+2} = 0.7N^1_{t+1}$.

To allow for numerical error when computing the outer approximation, we round these values up when determining the size of phase space. For $s = 0$, for example, we take $X = [0, 200] \times [0, 140]$, while for $s = 1$ we take $X = [0, 5.0] \times [0, 3.5]$.

As we have emphasized, the scale at which X and Z are discretized into grids is an important consideration that will very much depend on the problem at hand. The precision with which the parameters can be measured, for example, gives one natural and meaningful choice of grid size. Because we are working with an abstract mathematical model in this paper, we have chosen the scale of our grids in phase and parameter space to best illustrate, given our computational resources, the change in dynamics between the plant and fish models. To this end, for each value of s we subdivide Z into a 128×128 parameter grid $\mathcal{Z}$. Similarly in phase space: for each value of s we use an adaptive grid described in [2, 6] that is equivalent to a 4096×4096 grid $\mathcal{X}$ on X. We further allow for additional subdivisions of Morse sets with trivial Conley index into smaller grid elements, to attempt to rule out numerical artifacts.

For each parameter grid element $\zeta \in \mathcal{Z}$, we must ensure the graph $\mathcal{F}_\zeta$ on $\mathcal{X}$ is an outer approximation for the dynamics as described in Chapter 2. This can be accomplished with a variety of numerical techniques. We make use of the most straightforward approach, interval arithmetic. Details concerning the theory and implementation of interval arithmetic can be found in [31]. For our purposes it is sufficient to note that given a real valued function $r(a_1, \ldots, a_k) \colon \mathbb{R}^k \to \mathbb{R}$ and any assignment of intervals $[\underline{a}_i, \overline{a}_i] \subset \mathbb{R}$ to the variable a_i, $i = 1, \ldots, k$, interval arithmetic returns an interval $[\underline{B}, \overline{B}]$ such that $r(\alpha_1, \ldots, \alpha_k) \in [\underline{B}, \overline{B}]$ as long as $\alpha_i \in [\underline{a}_i, \overline{a}_i]$ for all i.

In order to compute the image $\mathcal{F}_\zeta(\xi)$ of a grid element $\xi \in \mathcal{X}$, we represent each parameter and phase variable as an interval using the intervals defining ζ and ξ, and then use interval arithmetic [31] to compute the image of the map. Observe that the image is a rectangle in X. In general this rectangle will not correspond precisely to a union of grid elements. Therefore, to ensure $\mathcal{F}_\zeta$ is an outer approximation we cover this rectangle by grid elements. For any grid element $\xi' \in \mathcal{X}$ that intersects this rectangle

we add the edge $\xi \to \xi'$ to $\mathcal{F}_\zeta$.

A description of eleven important Conley-Morse graphs, that arise in all of the above mentioned computations, is presented in Table 7.2. Each circle corresponds to a Morse set, and the directed edges between Morse sets indicate the partial order on the Morse sets as described in Chapter 2. Inside each circle are the polynomials constituting the Conley index for each Morse set computed with $\mathbb{Z}_3$ coefficients.

To obtain intuition about the dynamics associated with each Conley index see Table 7.1. Observe that there is a Morse set in the Conley-Morse graph titled pink with Conley index $x^2 + x + 1$ in dimension 1. This index does not appear in Table 7.1 so is worth an additional comment. The database allows us to recognize this Morse set arises as the result of a Morse set with Conley index $x - 1$ in dimension 2 merging with a Morse set with Conley index $x^3 - 1$ in dimension 1. Thus we take as our representative dynamics an unstable fixed point with two unstable directions connecting to an unstable period 3 orbit with one unstable direction.

As is indicated in Table 7.1 the Conley index of the empty set is trivial. However, as mentioned in Chapter 3 with regard to the Smale horseshoe, nontrivial invariant sets can also have trivial Conley index. In particular, for the parameter values being considered in this paper the origin is always an unstable fixed point for Equation (7.1). However, because we are restricting the phase space to a compact subset of $[0, \infty)^2$, the Conley index of the origin is always trivial. This is explicitly seen in all the Conley-Morse graphs of Table 7.2 except for the Conley-Morse graph labeled gray and dark gray.

This has interesting implications, however, for the gray and dark gray Conley-Morse graphs. Namely, in these graphs our computations based on the outer approximation must not isolate the origin. In the case of the dark gray Conley-Morse graph, the fact that there is only one node implies that the computed isolating neighborhood for the globally stable dynamics in the interior of $[0, \infty)$ also contains the origin. Reviewing the explicit results for the gray Conley-Morse graph leads to the same conclusion. Since in this biological system the origin plays the special role of extinction, this strongly suggests that for these parameter values stochastic perturbations can easily lead to

extinction. More on this below.

Having established interpretations for the individual Conley-Morse graphs, we now turn to how they are related. Figure 7.1 displays a picture of parameter space output by the database for the plant model $s = 0$. Each color region indicates a subset of phase space with the same Conley-Morse graph. We hasten to add that this image represents a simplification of the continuation classes defined in Section 2.3. In particular, in Figure 7.1 the classification is based on just the Conley-Morse graph, a simplification done for visual clarity.

The major color regions in Figure 7.1 correspond to the labeling of Conley-Morse graphs in Table 7.2. Small changes in shading represent the presence of additional Morse sets with trivial Conley index in the Conley-Morse graph. As discussed previously, a trivial Conley index does not necessarily mean the invariant set of the underlying continuous dynamics is empty. Moreover, given our philosophy of working with the graph $\mathcal{F}_\zeta$ as a discrete representation of the dynamics on the parameter grid element ζ, we do not want to completely disregard the presence of Morse sets with trivial Conley index, since they indicate recurrent behavior that is observed at the level of discretization chosen. There are at least two mechanisms that can lead to trivial Conley indices. The first is the presence of slow dynamics that numerically manifests itself as a kind of recurrent behavior. It should be noted that what is meant by "slow" will depend on the level of discretization. The second is the presence of invariant dynamics, the structure of which cannot be identified at the given level of resolution. In either case the existence of a recurrent set with trivial Conley index is an indicator of a region in phase space of possible interest, depending on the application.

Figure 7.2 shows the same picture as Figure 7.1 except for the fish model $s = 1$. From the image alone we can see the answer to our question from Section 7.1: that despite sharing many of the same dynamics, the way that those dynamics are situated in parameter space is quite different between the plant and fish models. Thus, our objective now is to understand what connection, if any, there is between the similar dynamics as we vary the competition parameter s that interpolates between the two models.

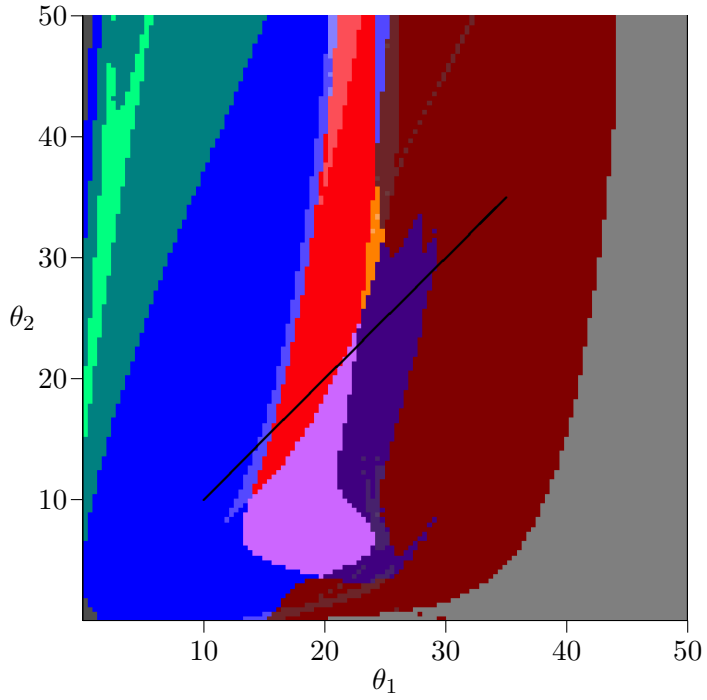


Figure 7.1: Picture of parameter space for database computation with $s = 0$, i.e., the plant model. As in all subsequent figures, the x-axis corresponds to θ_1 (young age class reproduction rate) and the y-axis θ_2 (old age class reproduction rate). The bold diagonal line corresponds to the bifurcation diagram in [45, Figure 1]. See Table 7.2 for the Conley-Morse graphs corresponding to each color.

To do the comparison we have chosen to focus on three distinct dynamical phenomena that have biological interest.

(1) Biennial Population Dynamics. The teal and green Conley-Morse graphs exhibited in Table 7.2 indicate the presence of stable 2-cycle sets and stable 4-cycle sets, respectively. Referring to Figure 7.1 for the plant model, the teal region is in the upper-left corner corresponding to small θ_1 and large θ_2. Thus we can observe robust biennial behavior in the presence of a small amount of reproductive capacity by the younger age cohort.

In the corresponding figure for the fish model, Figure 7.2, the teal stable 2-cycle region is larger, but it is oriented along the θ_1 axis. Hence the dependence on stable biennial dynamics on the rates of reproduction appear to be opposite. An interesting question is if there is any connection between these two regions that can be observed in the intermediate models. In other words, as s is varied is there a choice of reproductive rates that preserves stable biennial population dynamics?

Choosing a particular parameter value and examining the associated stable 2-cycle Morse sets in phase space shows that this biennial behavior corresponds to alternating high-low populations between the two age cohorts. Of course, the actual values, i.e. location in phase space, is parameter dependent and thus not directly accessible from the database.

(2) Bistability. Bistability is easily detected in the Conley-Morse graphs; it occurs if there is more than one minimal node in the Morse graph. Thus, the red, orange, and yellow Conley-Morse graphs imply the existence of bistable dynamics. The dynamics within attractor can be identified via the Conley index. Thus, the red Conley-Morse graph indicates that one attractor is a stable 3-cycle set while the other attractor is a 1-cycle set. Referring to Figure 7.1 for the plant model, the red region is vertically oriented, so that it is sensitive to small changes in θ_1 but robust to changes in θ_2. In the fish model, Figure 7.2, this is reversed, and the bistable region is much more sensitive to changes in θ_2 compared to θ_1. As in the case of the biennial population dynamics, there is the question of whether these dynamics are related by the intermediate models.

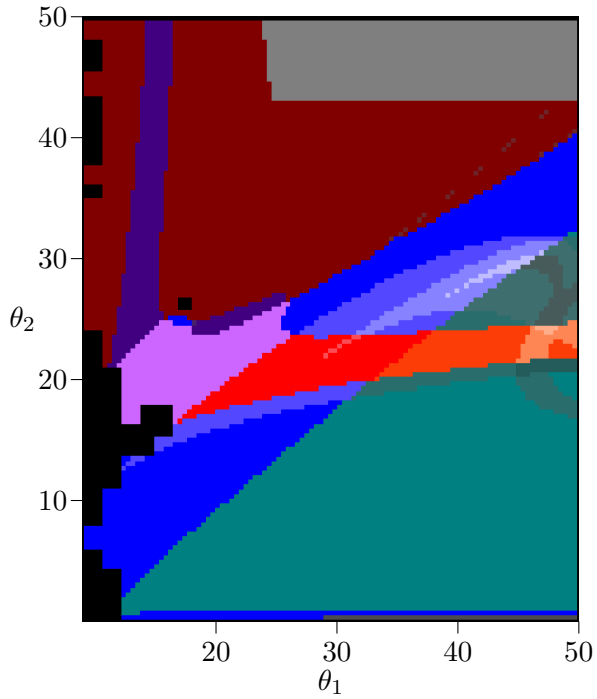


Figure 7.2: Picture of parameter space for database computation with $s = 1$, *i.e.*, the fish model. This is an image of the subset $[9, 50] \times [0, 50]$ of the full parameter space. The black regions along with the region $\theta_1 \leq 9$ were not computed due to memory constraints.

Observe that in the fish model the teal region and the red region intersect in the region that is colored orange. As the orange Conley-Morse graph indicates this region exhibits characteristics of both the teal and the red regions—bistability where one attractor is a stable 3-cycle set and the a stable 2-cycle set. This dynamics is not detected by the database computations in the plant model.

(3) Lack of Permanence. We are using a continuous deterministic model of discrete populations which for at least two reasons is suspect for small population levels: (i) the model can predict population levels below a single unit; and (ii) when the origin is unstable the model does not allow a positive population to become extinct, a phenomenon that one would expect to occur due to stochastic perturbations. The concepts of permanence [20] and persistence [46], which in our setting is equivalent to the attractor of the positive orthant begin bounded away from the origin, were introduced to address this concern. An advantage of the database approach is that it allows one to quantify these ideas.

Suppose that the phase space grid is chosen at a scale to address concerns (i) and (ii). For example, suppose that the grid element containing the origin is large enough to both contain all points in phase space smaller than a single unit of population, and to contain all points that the assumed random perturbations might take to extinction within one time step. Then the fact that the database computations use an outer approximation guarantees that every Morse set not containing the origin is bounded away from the origin, and that any such Morse sets which are stable can be said to exhibit permanence. In this direction the database provides a proof of this fact.

On the other hand, if the origin is not isolated by its own Morse set—for example, if the only attractor contains the origin—then it suggests that at these scales permanence is not achieved. This could be an artifact due to the choice of a poor outer approximation that prevents the database from separating stable dynamics away from the origin. However, if this is an important modeling issue, then it can be resolved, at a computational cost, by using a better outer approximation to determine the multivalued map. And in every case the database indicates all regions of parameter space with possible

lack of permanence.

As is discussed above, the origin is not isolated in the gray and dark grey Conley-Morse graphs. We focus on the gray Conley-Morse graph where the minimal Morse set has the Conley index associated with a stable invariant circle (see Table 7.1). In fact, the dynamics on the associated Morse set includes both the origin and large oscillations in the population. It is reasonable to expect that due to stochastic fluctuations or less than single unit population levels extinction will occur.

As indicated in Figure 7.1 for the plant model the gray Conley-Morse graph occurs for large θ_1, *i.e.*, if the first age class produces large numbers of seeds then extinction is expected (It is interesting to note that biennials typically produce no seeds the first year). Furthermore, for θ_1 sufficiently large, this phenomenon is essentially independent of θ_2. Figure 7.2 shows that for the fish model again this dependency is reversed. Sufficiently high θ_2 given a moderate θ_1 means extinction, and these dynamics are relatively insensitive to θ_1.

We are interested in understanding how these phenomena (biennial population dynamics, bistability and lack of permanence) depend upon the type of intra-species competition. Thus we perform the database computations using Equation (7.1) at the parameter values $s = 0.01$, 0.02, 0.05, and 0.10. To explain the nonuniform choice of s it should be noted that while s is taken to be a linear interpolation parameter between the exponents of the plant and fish models, it does not follow that the dynamics changes in a uniform manner. In fact, because for $\theta_i > 1$ the exponent in the fish model is larger than that of the plant model, the dynamics at $s = 0.1$ already begin to strongly resemble the dynamics of the fish model at $s = 1$.

Figures 7.3–7.6 identify regions of parameter space with the Conley-Morse graphs of Table 7.2 for the values $s = 0.01$, 0.02, 0.05, and 0.10, respectively. Looked at in sequence this gives a picture of how the dynamics over parameter space can be expected to change in response to a change in the nature of competition. In particular, we revisit the contrasts observed between the dynamics of the plant and fish models to see how these phenomena behave over the transition.

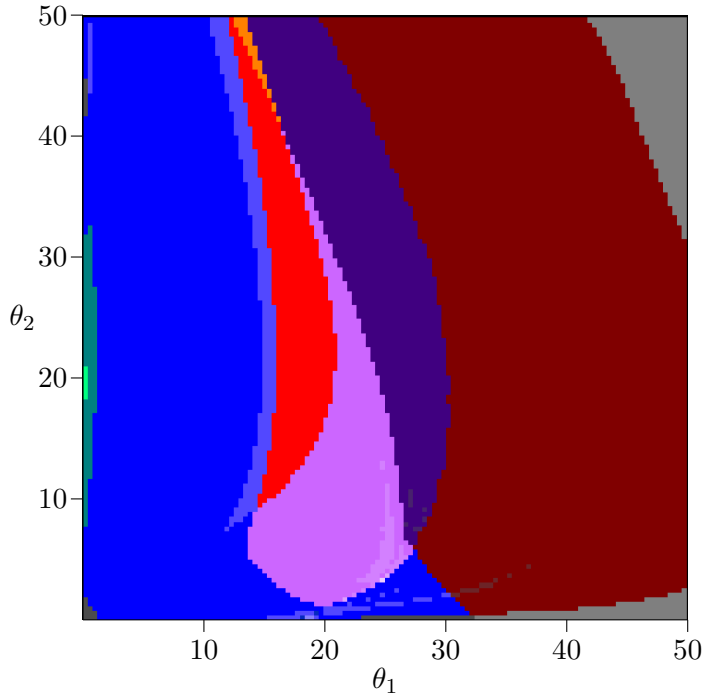


Figure 7.3: Picture of parameter space for database computation with $s = 0.01$.

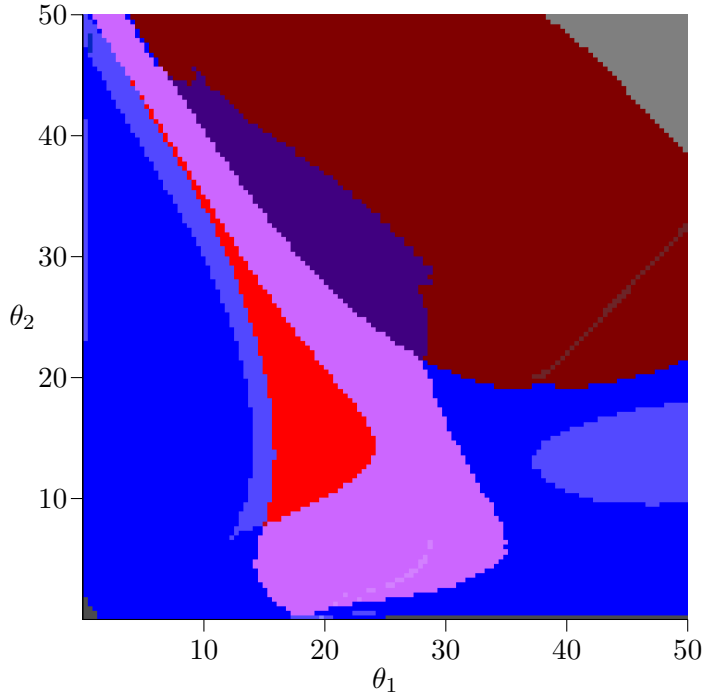


Figure 7.4: Picture of parameter space for database computation with $s = 0.02$.

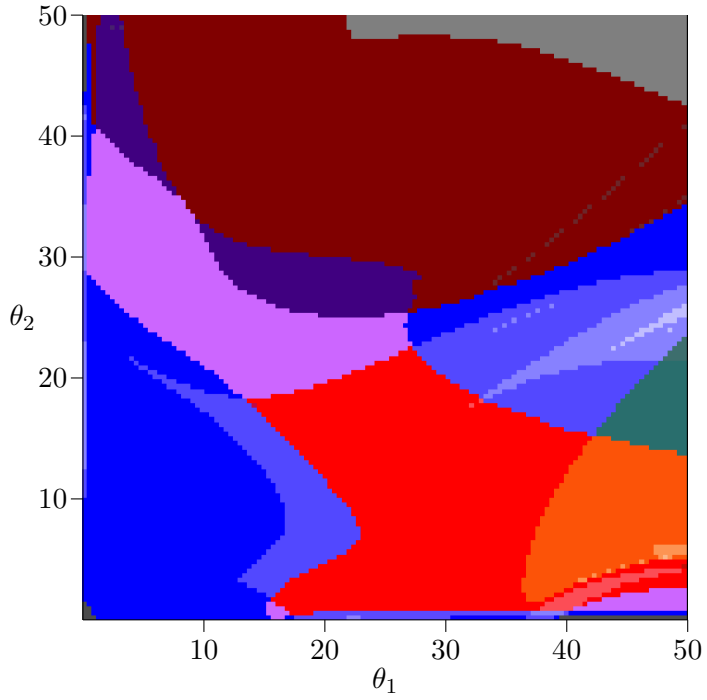


Figure 7.5: Picture of parameter space for database computation with $s = 0.05$.

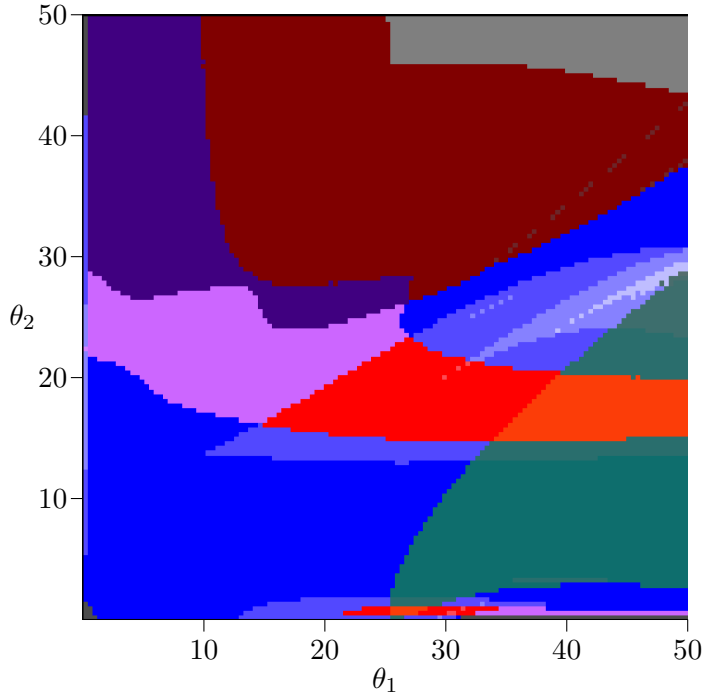


Figure 7.6: Picture of parameter space for database computation with $s = 0.1$.

(1) Mixed Competition and Biennial Population Dynamics. The teal Conley-Morse graph (Table 7.2) which indicates the existence of a stable 2-cycle and occupies most of the upper left corner of parameter space for the plant model (Figure 7.1) is already much less robust by $s = 0.01$ (Figure 7.3) and invisible by $s = 0.02$ (Figure 7.4). It reappears at $s = 0.05$ (Figure 7.5) for large values of θ_1 and moderate values of θ_2. We can see that this 2-cycle region moves closer to the origin as s increases from 0.05 to 1 (Figures 7.5 and 7.6).

From this, we observe that the stable 2-cycle dynamics of the plant model do not appear to continue to the stable 2-cycle dynamics of the fish model. In other words, if the nature of competition changes within a biennial population from plant-like to fish-like, then there is a point at which the biennial dynamics will be lost, regardless of what happens to reproduction rates.

(2) Mixed Competition and Bistability. The red, orange, and yellow Conley-Morse graphs indicate the existence of bistable dynamics. These colors are present in Figures 7.3–7.6, which suggests that bistability persists under continuous changes in s. The ecological interpretation is that bistable dynamics is robust to changes in the type of competition experienced by the young members of the species.

Although bistability is present in each slice of parameter space, the shape of the bistable region changes with s. For $s < 0.05$ the region is more extended in the θ_2 direction, although by $s = 0.02$ the tail that extends primarily in the θ_2 direction is very narrow. By $s = 0.05$ things look quite different—the bistable region is much larger and extends much further along the θ_1 direction than the θ_2 direction. For $0.5 < s \leq 1$ this orientation is preserved and the bistable region narrows further in the θ_2 direction.

(3) Mixed Competition and Lack of Permanence. While in the plant and fish models a lack of persistence (represented by the gray regions) could be reasonably described in terms of thresholds—large θ_1 in the plant model and a combination of large θ_2 and moderate θ_1 in the fish model—in the intermediate models this no longer holds. In the case of $s = 0.01$, for example, large values of θ_1 do not exhibit persistence if θ_2 is sufficiently large or small, but do exhibit persistence for moderate values of θ_2. And

even this non-montone behavior is not described by a threshold: looking at the upper right corner when $s = 0.01$ or 0.02, there is a tradeoff between θ_1 and θ_2 that determines whether the population will persist, given by the slope of the boundary.

7.3 Conclusions

Typically, network diagrams are used to model complex systems in settings in which mechanisms associated with the interactions are poorly understood. It is reasonable to expect that different modeling assumptions concerning these mechanisms will lead to different dynamics, since the structure of invariant sets can be extremely sensitive to small perturbations. However, in the context of limited information one might prefer to understand those dynamic structures that are robust with respect to parameters. In addition, for many applications, the precision with which measurements can be taken will limit the observable dynamics. The database approach to dynamical systems allows us to rigorously investigate qualitative dynamics in a way that is robust to these limitations of precision in modeling.

There are other approaches to modeling dynamical systems that aim to reflect this inherent uncertainty. As alluded to in our discussion on permanence, the inclusion of stochastics into the model is a fairly common technique. How our approach compares with a stochastic dynamical system depends on the kind of noise that is being modeled.

In the case in which it is important to include unbounded noise, any rigorous statement hoping to summarize the dynamics needs to allow for low probability events that, for example, can move systems between different basins of attractions. Invariant measures provide a means by which one can hope to describe such systems. At the moment we know of no method for relating the structure of the Conley-Morse graphs to invariant measure. Nevertheless, if we believe that the system that we are exploring has an underlying deterministic model, the coarse features of which we have captured analytically, then database computations that indicate multiple basins of attraction over a set of parameter values might suggest that the associate invariant measure is not unimodal.

If it is sufficient to assume bounded noise (in either phase or parameter space), then

the database approach can be used to make rigorous statements. Observe that if the noise in the dynamical system is bounded, then this can be incorporated into the outer approximation when the dynamics is combinatorialized. Thus, if we take our model to be a deterministic map along with bounded noise, the database can be used to prove that there can be no trajectory between two basins of attraction of the deterministic system even in the presence of noise.

Finally, there is another interpretation of the database that is broader than merely investigating the underlying deterministic dynamics, but that also does not fit neatly into any stochastic model. We might call this the "ruler" interpretation. Suppose we have instruments capable of measuring either the state or parameters of a system with a certain precision, e.g., with a ruler we might be confident we can measure to the nearest millimeter. The database computation we perform is still valid even if the act of measurement disrupts the system, as long as the disruption does not alter the measured value. In other words, the computation we perform is valid even if the ruler changes the length of what we measure, just so long as it changes in a way that rounds to the same value. This is distinct from typical stochastic models of noise for a couple reasons. First and most importantly, there need be no assumption about the distribution of the perturbations introduced. Second, the types of noise permitted depend on the grid chosen—intuitively, points near the center of a grid element can be moved in any direction, whereas points near the boundary can only be moved very small distances in the direction of that boundary, but much larger distances away from the boundary.

This kind of investigation can have potentially important implications for the practice of modeling. To the extent that the qualitative dynamics is preserved over parameter space, we can conclude that the particular choice of model is less important than the general structure given by the network diagram. On the other hand, to the extent that we see important changes in the structure of dynamics over parameter space we can say that it is important to supplement the network diagram with further modeling assumptions, even though the dynamics of interest may be very coarse.

In the example of overcompensatory age-structured population models, we see that

two classical approaches using the same form of nonlinearity can exhibit similar dynamics while having those dynamics situated quite differently in parameter space. The two different models already have different *a priori* justifications in terms of the biological understanding of competition. The database computations confirm that even looking at the dynamics on a coarse level we can distinguish between the models, despite the fact they arise from the same network diagram, use the same nonlinearity and locally exhibit the same dynamic structures.

By embedding these models into a larger model that incorporates an additional parameter s that represents intraspecies competition, we indicate how the database can be used to investigate empirically the relative importance and impact of the structure of competition. It should be emphasized that the parameter s is really an ordinal parameter—we have no basis to assign any significance to particular values beyond their ordering and furthermore it is not clear how it could be directly measured. However, the database can be used to organize and characterize the changes in the qualitative dynamics corresponding to different levels of competition. In principle, the associated database information could suggest experimental tests to determine positions along this scale.

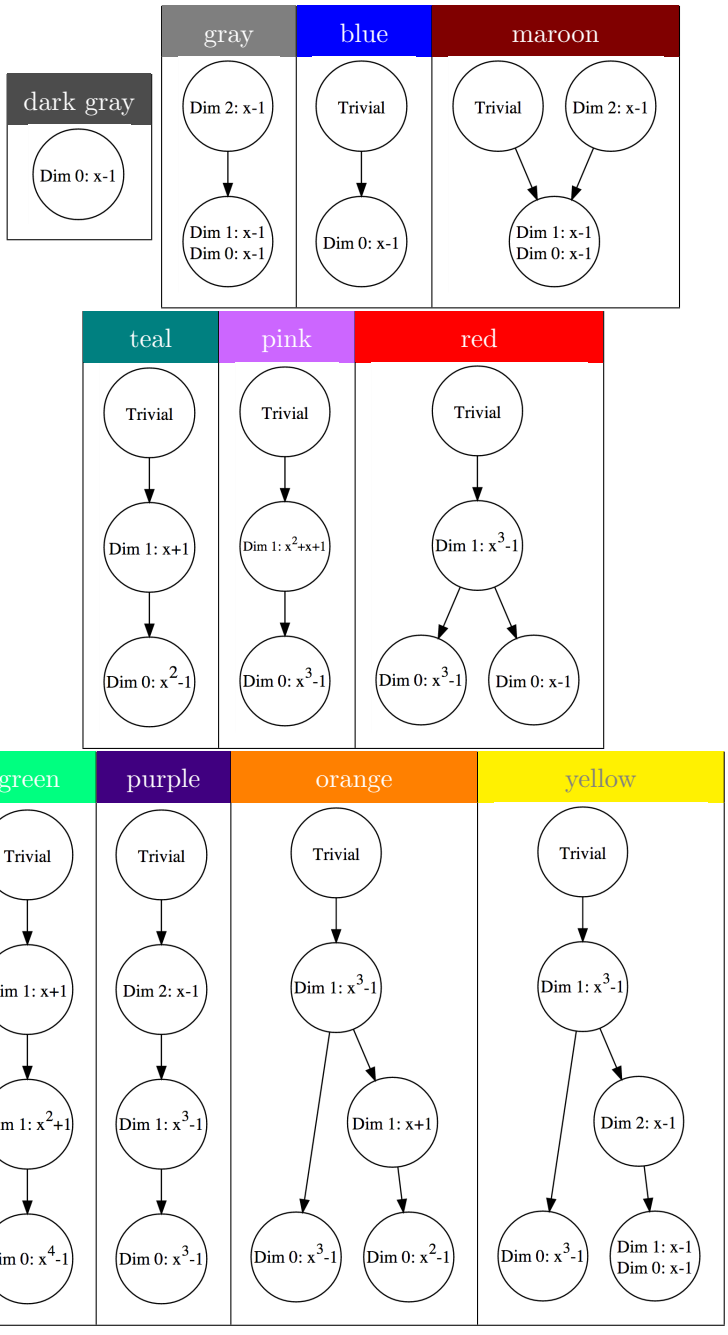


Table 7.2: Table of Conley-Morse graphs corresponding to each color

References

[1] Zin Arai, Marcio Gameiro, Tomas Gedeon, Hiroshi Kokubu, Konstantin Mischaikow, and Hiroe Oka. Graph-based topological approximation of saddle-node bifurcation in maps. In *RIMS Kokyuroku Bessatsu B31: Far-From-Equilibrium Dynamics*, pages 225–241. 2012.

[2] Zin Arai, William D. Kalies, Hiroshi Kokubu, Konstantin Mischaikow, Hiroe Oka, and Paweł Pilarczyk. A database schema for the analysis of global dynamics of multiparameter systems. *SIAM J. Appl. Dyn. Syst.*, 8(3):757–789, 2009.

[3] Anthony W. Baker. Lower bounds on entropy via the conley index with application to time series. *Topol. Appl.*, 120:333–354, 2002.

[4] Paul Blanchard. The dynamics of Newton's method. In *Complex dynamical systems (Cincinnati, OH, 1994)*, volume 49 of *Proc. Sympos. Appl. Math.*, pages 139–154. Amer. Math. Soc., Providence, RI, 1994.

[5] Justin Bush, Wes Cowan, Shaun Harker, and Konstantin Mischaikow. Limiting orientations of newton's method in $\mathbb{R}^2$. *preprint*, 2014.

[6] Justin Bush, Marcio Gameiro, Shaun Harker, Hiroshi Kokubu, Konstantin Mischaikow, Ippei Obayashi, and Paweł Pilarczyk. Combinatorial-topological framework for the analysis of global dynamics. *Chaos*, 22, 2012.

[7] Justin Bush, Tomáš Gedeon, Shaun Harker, Konstantin Mischaikow, and Kelly Spendlove. Coarsening theorems for constructing the conley-morse database. *preprint*, 2014.

[8] Justin Bush and Konstantin Mischaikow. Coarse dynamics for coarse modeling: An example from population biology. *Entropy*, 16:3379–3400, 2014.

[9] Charles Conley. *Isolated invariant sets and the Morse index*, volume 38 of *CBMS Regional Conference Series in Mathematics*. American Mathematical Society, Providence, R.I., 1978.

[10] R.A. Costantino, R.F.Desharnais. Chaotic dynamics in an insect population. *Science*, 275(5298):389 – 391, 1997.

[11] Brian A. Davey and Hillary A. Priestley. *Introduction to Lattices and Order*. Cambridge University Press, 2002.

[12] Peter Deuflhard. *Newton methods for nonlinear problems*, volume 35 of *Springer Series in Computational Mathematics*. Springer-Verlag, Berlin, 2004. Affine invariance and adaptive algorithms.

[13] David S. Dummit and Richard M. Foote. *Abstract Algebra*. Wiley, 3rd edition, 2004.

[14] Todd Dupont and L. Ridgeway Scott. The end-game for Newton iteration. *preprint*, TR-2010-10, 2010.

[15] David Eisenbud. *Commutative Algebra: with a View Toward Algebraic Geometry*. Springer-Verlag, 1995.

[16] M.J Feigenbaum. Quantitative universality for a class of non-linear transformations. *J. Stat. Phys.*, 19:25–52, 1978.

[17] John Franks and David Richeson. Shift equivalence and the conley index. *Trans. Am. Math. Soc.*, 352:3305–3322, 2000.

[18] J. Guckenheimer, G. Oster, and A. Ipaktchi. The dynamics of density dependent population models. *J. Math. Biology*, 4(2):101–147, 1977.

[19] Shaun Harker, Konstantin Mischaikow, Marian Mrozek, and Vidit Nanda. Discrete morse theoretic algorithms for computing homology of complexes and maps. *Foundations of Computational Mathematics*, pages 1–34, 2013.

[20] Vivian Hutson and Klaus Schmitt. Permanence and the dynamics of biological systems. *Math. Biosci.*, 111(1):1–71, 1992.

[21] William Kalies, Konstantin Mischaikow, and Robert C.A.M. Vandervorst. An algorithmic approach to chain recurrence. *Found. Comput. Math.*, 5(4):409–449, 2005.

[22] P.H Leslie. The use of matrices in certain population mathematics. *Biometrika*, 33(3):183–212, 1945.

[23] Simon A. Levin and C. Phillip Goodyear. Analysis of an age-structured fishery model. *J. Math. Biology*, 9:245–274, 1980.

[24] Douglas Lind and Brian Marcus. *An Introduction to Symbolic Dynamics and Coding*. Cambridge University Press, 1995.

[25] Edward N. Lorenz. Deterministic nonperiodic flow. *J. Atmos. Sci.*, 20:130–141, 1963.

[26] Anthony Manning. Topological entropy and the first homology group. In *Dynamical systems—Warwick 1974 (Proc. Sympos. Appl. Topology and Dynamical Systems, Univ. Warwick, Coventry, 1973/1974; presented to E. C. Zeeman on his fiftieth birthday)*, pages 185–190. Lecture Notes in Math., Vol. 468. Springer, Berlin, 1975.

[27] Robert M. May and George F. Oster. Bifurcations and dynamic complexity in simple ecological models. *Amer. Natur.*, 110(974):573–599, 1976.

[28] Konstantin Mischaikow. CHomP. `http://chomp.rutgers.edu/Archives/Databases_for_the_Global_Dynamics/Networks_to_Dynamics.html`, 2014.

[29] Konstantin Mischaikow and Marian Mrozek. Conley index. In *Handbook of dynamical systems, Vol. 2*, pages 393–460. North-Holland, Amsterdam, 2002.

[30] Megumi Miyazawa. Chaos and entropy for circle maps. *Tokyo Journal of Mathematics*, 25(2):453–458, Dec 2002.

[31] R.E. Moore. *Interval Analysis*. Prentice Hall, Inc., 1966.

[32] Marian Mrozek. Leray functor and cohomological conley index for discrete dynamical systems. *Trans. Amer. Math. Soc.*, 318:149–178, 1990.

[33] Marian Mrozek. An algorithmic approach to the conley index theory. *J. Dynam. Differential Equations*, 11(4):711–734, 1999.

[34] James Munkres. *Topology*. Prentice Hall, 2nd edition, 2000.

[35] Sheldon E. Newhouse. The abundance of wild hyperbolic sets and nonsmooth stable sets for diffeomorphisms. *Inst. Hautes Études Sci. Publ. Math.*, 50(1):101–151, 1979.

[36] Sheldon E. Newhouse. New phenomena associated with homoclinic tangencies. *Ergodic Theory Dynam. Systems*, 24(5):1725–1738, 2004.

[37] Issac Newton. *The Mathematical Principles of Natural Philosophy*. Benjamin Motte, 1729.

[38] W. Parry and S. Tuncel. *Classification Problems in Ergodic Theory*, volume 67 of *LMS Lecture Note Series*. Cambridge University Press, 1982.

[39] Henri Poincaré. Sur le problème des trois corps et les équations de la dynamique. *Acta Mathematica*, 13:1–270, 1890.

[40] Karl Popper. *The Logic of Scientific Discovery*. Hutchinson & Co, 1959.

[41] W.E. Ricker. Stock and recruitment. *J. Fish. Res. Board. Can.*, 11:559–623, 1954.

[42] Joel Robbin and Dietmar Salamon. Dynamical systems, shape theory, and the conley index. *Ergod. Th. & Dynam. Sys.*, 8:375–393, 1988.

[43] Steven Strogatz. *Nonlinear Dynamics and Chaos*. Perseus Books Publishing, 1994.

[44] Andrej Szymczak. The conley index for discrete semidynamical systems. *Top. App.*, 66(3):215–240, 1995.

[45] I. Ugarcovici and H. Weiss. Chaotic dynamics of a nonlinear density dependent population model. *Nonlinearity*, 17:1689–1711, 2004.

[46] Paul Waltman. A brief survey of persistence in dynamical systems. In *Delay differential equations and dynamical systems (Claremont, CA, 1990)*, volume 1475 of *Lecture Notes in Math.*, pages 31–40. Springer, Berlin, 1991.

[47] Charles Weibel. Shift equivalence and $\mathbb{Z}[t]$-modules, March 2012. Presentation.

[48] R.F. Williams. Classification of subshifts of finite type. *Ann. Math.*, 98:120–153, 1973.