Probabilistic Properties of Delay Differential Equations

by

S. Richard Taylor

A thesis
presented to the University of Waterloo
in fulfilment of the
thesis requirement for the degree of

Doctor of Philosophy
in
Applied Mathematics

Waterloo, Ontario, Canada, 2004

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made elecronically available to the public.

Abstract

Systems whose time evolutions are entirely deterministic can nevertheless be studied probabilistically, *i.e.* in terms of the evolution of probability distributions rather than individual trajectories. This approach is central to the dynamics of ensembles (statistical mechanics) and systems with uncertainty in the initial conditions. It is also the basis of ergodic theory—the study of probabilistic invariants of dynamical systems—which provides one framework for understanding chaotic systems whose time evolutions are erratic and for practical purposes unpredictable.

Delay differential equations (DDEs) are a particular class of deterministic systems, distinguished by an explicit dependence of the dynamics on past states. DDEs arise in diverse applications including mathematics, biology and economics. A probabilistic approach to DDEs is lacking. The main problems we consider in developing such an approach are (1) to characterize the evolution of probability distributions for DDEs, *i.e.* develop an analog of the Perron-Frobenius operator; (2) to characterize invariant probability distributions for DDEs; and (3) to develop a framework for the application of ergodic theory to delay equations, with a view to a probabilistic understanding of DDEs whose time evolutions are chaotic. We develop a variety of approaches to each of these problems, employing both analytical and numerical methods.

In transient chaos, a system evolves erratically during a transient period that is followed by asymptotically regular behavior. Transient chaos in delay equations has not been reported or investigated before. We find numerical evidence of transient chaos (fractal basins of attraction and long chaotic transients) in some DDEs, including the Mackey-Glass equation. Transient chaos in DDEs can be analyzed numerically using a modification of the "stagger-and-step" algorithm applied to a discretized version of the DDE.

Acknowledgements

There are many individuals and organizations who have contributed to the success of this project:

I have long discovered that geologists never read each other's works, and that the only object in writing a book is a proof of earnestness, and that you do not form your opinions without undergoing labor of some kind.

—CHARLES DARWIN [28]

Preface

As an undergraduate physics student I discovered Ilya Prigogine's book "The End of Certainty", quite serendipitously, on the new arrivals shelf at the UBC library. I signed the book out and read it through over a very short period, inspired by Prigogine's new (to me) idea that deterministic systems—predictable things in the world of Newtonian mechanics—could and should be discussed in probabilistic terms. This idea, which pointed to a way out of the clockwork universe that classical mechanics usually portrays, resonated with what I had been learning about classical physics, quantum mechanics, and "chaos theory". According to Prigogine the mathematical foundation of his ideas was called ergodic theory. Eager to learn more, I made my first ever visit to the mathematics library and signed out Lasota and Mackey's "Probabilistic Properties of Deterministic Systems". Unfortunately I found that I lacked the mathematical maturity to read the book on my own, and I soon gave up.

A few years later I met Michael Mackey at a summer school in Montréal, where he gave a presentation in which, as an aside, he mentioned that a probabilistic/ergodic approach to delay differential equations was lacking, and that such a theory might have interesting applications. At the time I didn't appreciate the ambitiousness of such a project, but it was the excuse I needed to tackle these ideas again. The result, after some earnest labor, is the present document—a representative subset of my present ideas on a probabilistic approach to delay differential equations.

Contents

List of Figures

Chapter 1

Introduction

Today, some of the most profound unanswered scientific questions are related to the interplay between order and disorder. The physical basis of consciousness and intelligence (the mind-body problem), the origins of life, the nature of turbulence, and the paradox between between thermodynamics and deterministic microscopic dynamics, are but a few examples. These are fundamental problems that have plagued scientists and philosophers for centuries, and each remains largely unresolved. Each of them involves the spontaneous creation of order out of disorder, or disorder out of order, and they share among them the difficulty of explaining how such processes can result from the operation of deterministic physical laws.

Recent decades have seen a renewal of interest and progress in the understanding of the nature of order and disorder, beginning with the discovery in 1963 by Lorenz of the existence of systems that, despite being deterministic, exhibit disordered and essentially unpredictable behavior. Since then, the study of "deterministic chaos" has brought forth a rich corpus of experimental and theoretical results, incorporating the insights of researchers in diverse fields spanning mathematics, physics and engineering. This corpus

comprises what has come to be called "chaos theory"[1].

In particular, the mathematical field of *dynamical systems* (*i.e.*, systems that evolve in time according to deterministic laws) has contributed much to a unified understanding of chaotic systems. This general framework reveals that, despite the differences in origins and physical nature of different deterministic systems, the same underlying mechanisms operate to generate disorder. One of the central results of this theory is the identification of a precise notion of what constitutes a "chaotic system", and the discovery of sufficient conditions that imply the existence of chaos as such [76].

Given the random character of so-called chaotic evolutions, it is not surprising that statistical and probabilistic ideas should be useful tools for their analysis. It turns out that far from being merely descriptive, probabilistic ideas provide fundamental insights into the behavior of dynamical systems [74, 96]. This observation is the basis of the ergodic theory of dynamical systems, a theory that had its origins in foundational issues in statistical mechanics in the late 1800's. Despite the success of ergodic theory as a mathematical endeavor, the physical problems that motivated its development surprisingly remain unresolved [84, 95, 96].

It may seem, at first, that the application of probabilistic ideas to deterministic systems is inherently contradictory. Indeed, probability theory is concerned with the study of inherently random phenomena, which are antithetical to a strictly deterministic conception of physical law. However, experience has shown quite the opposite: probabilistic ideas provide a new and, in the end, quite natural way to view deterministic phenomena. The following discussion is intended to give the flavor of this viewpoint. In Chapter 2 we give a more detailed theoretical presentation.

1.1   Probabilistic Approach to Deterministic Systems

Consider the following, much-studied example of a simple deterministic system whose evolution exhibits a species of random behavior. For a given a real number x_0 (the "initial state" of the system) between 0 and 1, let x_1, x_2, x_3, *etc* be defined by repeated application of the formula

$$x_{n+1} = 4x_n(1 - x_n), \quad n = 0, 1, 2, \ldots \tag{1.1}$$

---

[1]Abuse and misuse of the term "chaos theory" in the popular literature has led some serious people to avoid the term. One comprehensive reference [71] manages not to use the term in its entire 800+ pages.

Once can view this formula as prescribing the evolution of the state of the system, x_n, at discrete times $n = 0, 1, 2, \ldots$. The evolution of this system is deterministic, in that once the initial state is specified equation (1.1) uniquely determines the sequence of values $\{x_0, x_1, x_2, \ldots\}$ (*i.e.*, the *trajectory* of the system) for all time. Thus for example if $x_0 = 0.51$ we obtain

$$x_1 = .9996, x_2 \approx .0026, x_3 \approx .0064, x_4 \approx .025, x_5 \approx .099, \ etc.$$

The qualitative behavior of this system is most easily appreciated graphically, as in Figure 1.1 which plots x_n vs. n for typical trajectories obtained for different choices of x_0. Each of these trajectories is erratic, and random in the sense that no regularity is apparent. Furthermore, it can be seen by comparing the graphs in Figure 1.1 that two near-identical initial states eventually yield radically different time evolutions. This phenomenon, termed "sensitive dependence on initial conditions" [78], imposes strong limits on the predictability of this system over long periods of time: a small error in the determination of the initial condition rapidly becomes amplified to the extent that reliable prediction of the future states of the system eventually becomes impossible. Thus, despite being entirely deterministic, trajectories of this simple system have some hallmarks of essentially *random* phenomena: their behavior is irregular and unpredictable.

The mechanisms underlying the random character of this system are reasonably well understood (see *e.g.* [24]), the key notion being sensitivity to initial conditions and its consequences. However, within this framework it is difficult to approach questions of the type "what is the asymptotic behavior of a *typical* trajectory of this system?" Indeed, the very nature of sensitivity to initial conditions would seem to preclude any notion of "typical" behavior, since even very similar initial conditions eventually lead to their own very particular, uncorrelated evolutions.

However, different conclusions are reached if one takes a probabilistic point of view. Suppose that instead of being precisely determined, the initial state x_0 has associated with it some uncertainty. In particular, suppose we know the initial probability density, ρ, giving the probabilities of all possible values that x_0 can take. Then it makes sense to ask, "what will be the probability density of x_1, the new state after one iteration of the map (1.1)?". A precise answer to this question can be found using analytical methods described in Chapter 2. For an approximate answer, it suffices to simulate a large ensemble of different initial states x_0 distributed according to ρ, evolve these states
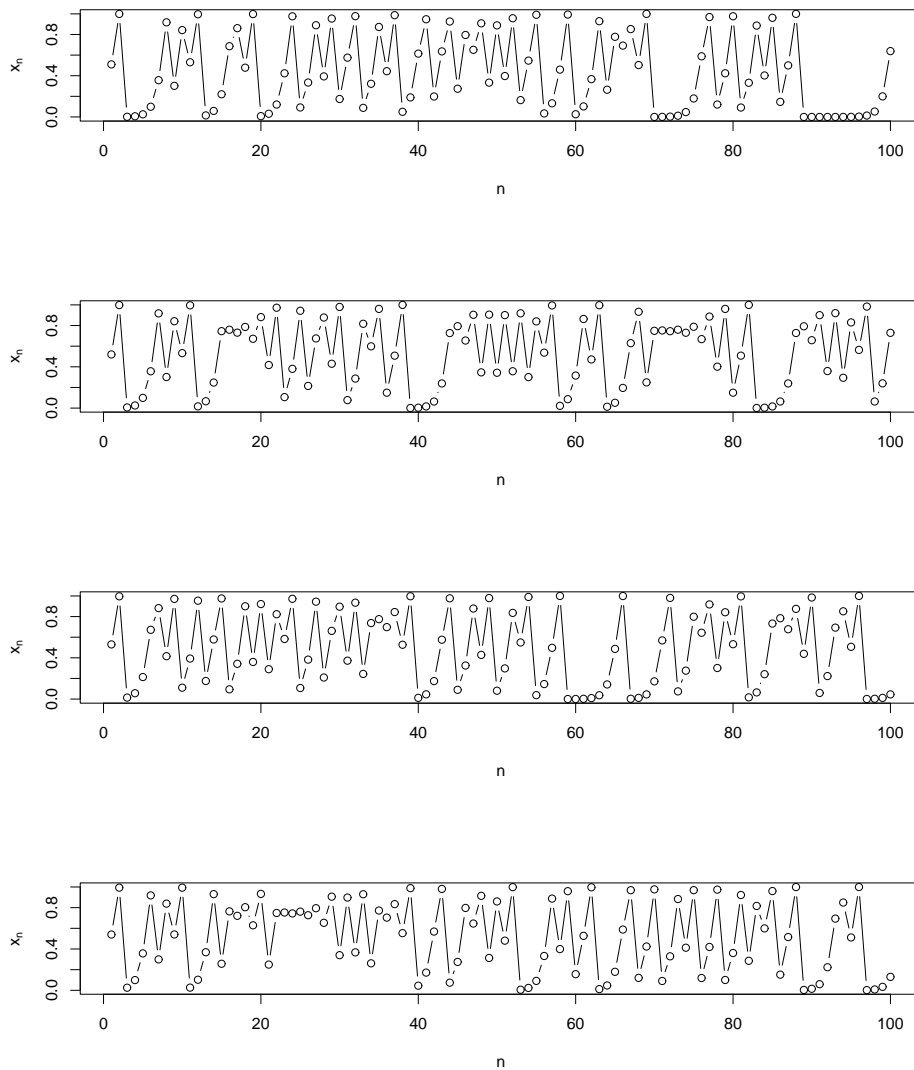
Figure 1.1: Numerical trajectories for the map $x \mapsto 4x(1-x)$. The initial conditions different only slightly for each trajectory.

forward under the map (1.1), and approximate the transformed density of the ensemble by constructing a histogram of the ensemble of values x_1. One can then proceed, in the same fashion, to determine the probability densities of subsequent states x_2, x_3, *etc.* Thus, even if the initial state x_0 is not known precisely, it is at least possible to give a probabilistic description of the system's evolution in terms of the evolution of a probability density.

The graphs in Figure 1.2 show a particular choice for the probability density ρ of the initial state x_0, together with the subsequent densities of states x_1, x_2, x_3, x_4, obtained by numerical simulation of an ensemble of 10^6 initial values distributed according to ρ, iterated forward under the map (1.1). The striking feature of this figure is that the sequence of densities rapidly approaches an equilibrium or invariant density that does not change under further iteration. Moreover, the invariant density appears to be unique. This is supported by Figure 1.3, which shows how a different choice of initial density evolves toward the same equilibrium density as before.

A different but related statistical approach to this system is to focus on the statistics of a single trajectory. For a given initial state x_0, by iterating $x_{n+1} = 4x_n(1 - x_n)$ we obtain an arbitrarily long sequence $\{x_n\}$ like the one illustrated in Figure 1.1. A histogram of this sequence reveals the long-term frequency with which the trajectory visits various parts of the interval $[0, 1]$. Figure 1.4 shows such a histogram, for a trajectory of length 10^6. Remarkably, this histogram reproduces the invariant density shown in Figures 1.2 and 1.3, which arises in a different context. Moreover, the same histogram is obtained for almost *any* choice of initial state.[2] Thus the invariant density describes the behavior of "typical" trajectories, *i.e.* those whose statistics are described by this particular probability density.

A probabilistic or ensemble treatment of dynamical systems provides a point of view complementary to one given in terms of the evolution of individual trajectories. The iterated map (1.1) is just one example of a system that behaves erratically on the level of individual trajectories, but has very regular asymptotic properties when considered at the level of probability densities. This observation appears to hold for many other systems. Moreover, it turns out that the converse holds as well: various regularity properties at the level of probability densities imply various degrees of disorder in the evolution of individual trajectories. Chapter 2 explores these connections further.

---

[2]There are exceptions, such as $x_0 = 0$, that yield trajectories with different (periodic) asymptotic behavior. These exceptions are very rare: in fact they constitute a set of Lebesgue measure 0 (*cf.* Chapter 2).
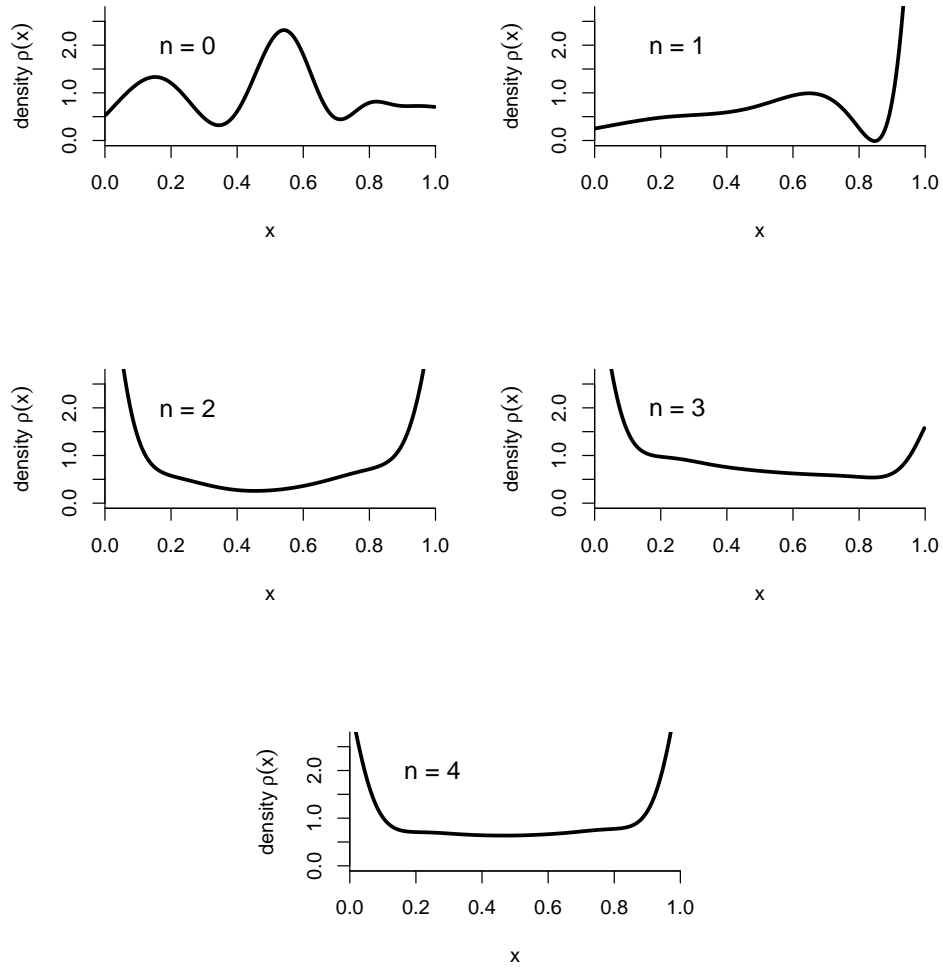
Figure 1.2: Simulated evolution of an ensemble density ρ under iterations of the map $x \mapsto 4x(1-x)$.
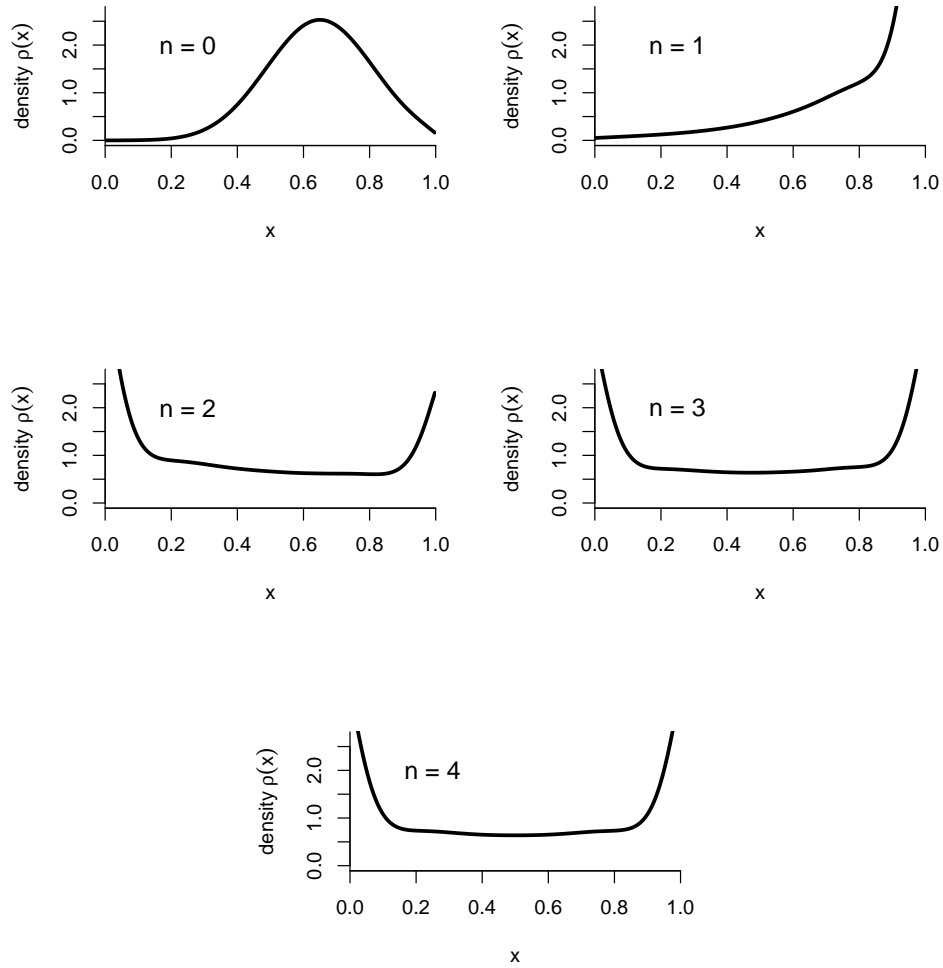
Figure 1.3: As Figure 1.2, with a different initial density.
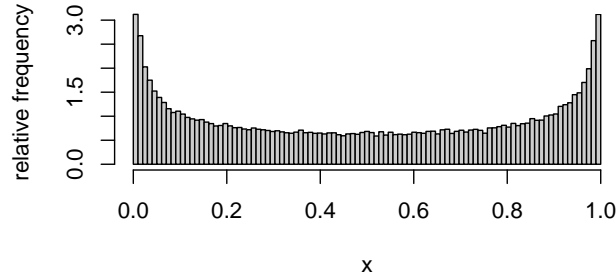
Figure 1.4: Histogram of a trajectory $\{x_n\}$ generated by 10^6 iterates of the map $x_{n+1} = 4x_n(1 - x_n)$.

1.2   Delayed Dynamics

The aim of the present work is to develop a probabilistic (*i.e.* ergodic) approach to systems with delayed dynamics, particularly those systems whose evolution can be described by a *delay differential equation*. Delay differential equations (DDEs) arise in the mathematical description of systems whose time evolution depends explicitly on a past state of the system, as for example in the case of delayed feedback. Neural systems [3], respiration regulation [47], agricultural commodity markets [83], nonlinear optics [46], and neutrophil populations in the blood [47] are but a few systems in which delayed feedback leads naturally to a description in terms of a delay differential equation.

We will restrict our attention to systems modeled by evolutionary delay equations that can be expressed in the form

$$x'(t) = f\big(x(t), x(t - \tau)\big), \quad x(t) \in \mathbb{R}^n, t \geq 0. \tag{1.2}$$

Here the "state" of the system at time t is $x(t)$, whose rate of change depends explicitly, via the function f, on the past state $x(t - \tau)$ where τ is a fixed time delay. More general delay equations might be considered: multiple time delays, variable time delays, continuously distributed delays, and higher derivatives all arise in applications and lead to more complicated evolution equations. Nevertheless, equations of the form (1.2) constitute a sufficiently broad class of systems to be of practical importance, and they will provide

adequate fodder for the types of problems we wish to consider.

Some delay equations give rise to erratic time evolutions; there are numerous examples in the literature [4, 5, 7, 42, 54, 61, 118, 121]. In some cases it is possible to describe precisely the sense in which these systems are chaotic, and to identify generic mechanisms responsible [4, 5, 7, 121]. These works focus on the application of topological notions of chaos, but little has been done on interpreting delay equations in an ergodic or measure theoretic context. We are aware of a handful of published works [4, 19, 41, 79, 80], among which there is no consistent framework for discussing ergodic notions in the context of DDEs. One aim of the present work is to develop such a framework.

A probabilistic approach to delay equations in particular is desirable for a variety of reasons. It has been suggested [30, 88] that brains encode information at the level of large neuron ensembles, rather than at the level of individual neurons. Statistical periodicity [74], in which the ensemble density cycles periodically rather than settling down to an invariant density, provides one possible mechanism for representing information at this level. Statistical periodicity does not (and cannot) arise in *ordinary* differential equations, but it does occur in some delay equations [80, 84]. Moreover, since synaptic and conduction delays introduce explicit delays into the dynamics of neurons, delay equations arise naturally in models of neuron dynamics [6, 3, 23]. Since the dynamics of large ensembles can be treated with probabilistic methods, a probabilistic approach to delay equations (*i.e.* a statistical mechanics of systems with delayed dynamics) would provide the mathematical tools for further developing this theory of brain functioning.

Delay equations also serve as relatively simple models for the study of infinite dimensional systems. Much of the complex dynamics that we experience directly is extended in space as well as in time, for example fluid turbulence and other systems modeled by partial differential equations (PDEs). These systems are infinite dimensional, in that the state at any given time cannot be specified precisely by a finite set of data. There is a rich literature on chaotic dynamics in PDEs, including some rigorous results on the existence of compact attractors and Li and Yorke type chaos, as well as numerous numerical observations, including methods for estimating Lyapunov exponents and dimensions for chaotic attractors (see *e.g.* [22, 45, 50, 57, 82, 115, 123, 126] and references therein). However, to our knowledge, with the exception of [99, 100, 101] and [22] little has been said of infinite-dimensional systems within the context of ergodic theory. Delay equations are also infinite dimensional, but in some respects they are much simpler than PDEs.[3]

---

[3]For example, a delay equation is much easier to solve, since it can often be represented as a sequence

Thus delay equations are a natural place to begin formulating ergodic concepts applicable to a broader class of spatially extended systems.

1.3   Probabilistic Questions for Delay Equations

The following chapters explore various problems related to a probabilistic or ergodic-theoretic treatment of delay differential equations. Questions that arise naturally in this context, and which we propose to consider, are the following:

- How can the language and concepts of ergodic theory be adapted and applied in the context of delay equations?

- How can the evolution of a probability distribution or ensemble density for a given DDE be determined; *i.e.*, can one find an evolution equation for the probability density?

- How can one determine invariant probability distributions for DDEs?

To some extent these problems can be considered independently, and each is the subject of one of the chapters that follow.

Chapter 2 presents the elements of ergodic theory needed in our subsequent discussion. It also serves to illuminate the scope of ergodic concepts that might be considered in the context of DDEs, and to further motivate the considerations of subsequent chapters.

Chapter 3 examines the fundamental theoretical questions posed by an ergodic approach to delay equations, and infinite dimensional dynamical systems in general. Alternative representations of a delay equation as a dynamical system are considered, and the measure-theoretic framework needed for a probabilistic treatment is developed. As it turns out, even these fundamental questions lead to unresolved ambiguities that are endemic to infinite dimensional systems.

Chapter 4 explores the practical issue of how the evolution of probability distributions for delay equations might be determined. In particular it would be nice to have an evolution equation governing the evolution of a probability density under the action of a delay equation. Such an equation would be a useful tool, for example, in making probabilistic predictions for systems governed by DDEs, and for modeling ensembles (such as neuron populations, *cf.* page 9) whose microscopic dynamics are governed by DDEs. A

---

of ordinary differential equations (method of steps), *cf.* Section 3.2.3.

number of different approaches to this problem are explored, including analytical methods and computer algorithms.

Invariant measures, which determine the long-term statistical properties of deterministic systems, are important quantities in statistical mechanics and also play a fundamental role in ergodic theory. Chapter 5 considers the problem of determining invariant measures for delay equations. Here the focus is on finding computer algorithms for computing invariant measures and their densities.

Transient chaos is a phenomenon in which a deterministic system has a chaotic time evolution for a transient period of time after which it asymptotically becomes regular (*e.g.* periodic). Such systems can be described in ergodic theoretic terms, but due to the instability of the "chaotic set" specialized techniques are required for their numerical analysis. Chapter 6 illustrates the existence of transient chaos in some delay equations, and shows how numerical methods can be adapted to their analysis. Until recently, limits on computational resources made it impractical to numerically investigate transient chaos in infinite dimensional systems such as DDEs.

Chapter 2

Ergodic Theory

This chapter introduces the basic concepts of the ergodic theory of dynamical systems. Following a brief review of measure theoretic probability (those familiar with measure theory might like to skip directly to Section 2.3), we consider the problem of determining the evolution of a probability measure or density under the evolution of a dynamical system. This serves to formalize some of the intuitive ideas presented in Chapter 1. In section 2.4 we then turn to the characterization, using ergodic concepts, of dynamical systems that exhibit irregular behavior. The general discussion given here sets the context in which we will consider the problem of an ergodic/probabilistic treatment of delay differential equations.

2.1 Dynamical Systems Formalism

The essential properties of a deterministic evolutionary system are encapsulated in the mathematical notion of a *semigroup*. At any particular time t the state of such a system is identified with an element x_t (the current *phase point*) of a *phase space* X. In practice X is often simply $\mathbb{R}^n$ or $\mathbb{R}$, and the phase point or state x represents the numerical value of some physical quantity, *e.g.* a voltage, displacement, population, *etc.* In some of the following we will require only that X have a topology, but occasionally we will want X to be equipped with some further structure.

The term "deterministic" implies that the phase point x_t is, for all times $t > 0$, uniquely determined by the initial phase point x_0. That is, for each $t > 0$ there is a rule that determines x_t from x_0—in other words, a transformation $S_t : X \to X$ that takes the initial state x_0 to the future state x_t:

$$x_0 \mapsto S_t(x_0) = x_t. \tag{2.1}$$

In many systems, the dynamical law that determines S_t does not change with time (*e.g.* autonomous systems evolving under time-invariant physical laws such as Newton's laws of motion). Then the result of the time evolution depends only on the initial phase point and the amount of time elapsed, not on the particular moment identified as the initial time. The identification of $t = 0$ as the initial time in equation (2.1) is then somewhat arbitrary. Consequently, the evolution of an initial phase point x_0 to time $s + t$ can be accomplished by a sequence of two evolutions, one from time 0 to time s carried out by the transformation S_s, followed by the evolution from time s to time t carried out

by the transformation S_t. This can be expressed by the relation

$$S_{s+t} = S_t \circ S_s. \tag{2.2}$$

That is, the family of transformations $\{S_t : t \geq 0\}$ form a semigroup:

Definition 2.1 (semigroup of transformations). Let either $G = \mathbb{Z}_+ = \{0, 1, 2, \dots\}$ or $G = \mathbb{R}_+ = \{t \in \mathbb{R} : t \geq 0\}$. A one-parameter *semigroup of transformations* $\{S_t : t \in G\}$ on X is a family of transformations $S_t : X \to X$ satisfying

1. $S_0(x) = x \quad \forall x \in X$,

2. $S_t \circ S_{t'} = S_{t+t'} \quad \forall t, t' \in G$.

Thus the evolution of a deterministic system under a time-invariant dynamical law can be described by an evolution equation $x_t = S_t x_0$ where the family of transformations $\{S_t : t \in G\}$ forms a semigroup. For discrete-time systems $G = \mathbb{Z}_+$; for continuous-time systems $G = \mathbb{R}_+$.

If the mapping $x_0 \mapsto x_t$ is also continuous with respect to both x_0 and t, the semigroup is called a semidynamical system:

Definition 2.2 (semidynamical system). Let X be a Banach space with $\{S_t : t \in G\}$ a semigroup of transformations on X. Then $\{S_t\}$ is a *semidynamical system* if the mapping

$$(x, t) \mapsto S_t(x) \tag{2.3}$$

from $X \times G$ into X is continuous.

Definition 2.3 (orbit; trajectory). Let $\{S_t : t \in G\}$ ($G = \mathbb{R}_+$ or $\mathbb{Z}_+$) be a semidynamical system on X, and let $x_0 \in X$. The set

$$\{S_t(x_0) : t \geq 0\} \subset X \tag{2.4}$$

is called the *orbit* or *trajectory* originating at x_0.

In general, the evolution of a semidynamical system cannot be extended uniquely to all times in the past. However, if each of the S_t is one-to-one (hence invertible), the identification $S_{-t} = S_t^{-1}$ extends the family of transformations to all $t < 0$, and the initial phase point does indeed uniquely determine x_t for all times in the past. In this case the

resulting *group* (rather than semigroup) of transformations $\{S_t\}$ is called a *dynamical system*. In the present work we will be concerned with systems that in general are not invertible. Nevertheless, for simplicity's sake we will frequently abuse notation and refer to these systems simply as dynamical systems, as the technical distinction between dynamical and semi-dynamical systems will not play an important role.

2.2   Measure Theoretic Probability

The evolution of a probability distribution under the action of a dynamical system requires the structures of measure theoretic probability. To this end, the elements of this theory are reviewed in the following. For a more complete review see *e.g.* the standard reference [55].

2.2.1   Measure theory

Basically, a measure is a function that assigns a "size" to a subset $A \subset X$. It turns out that it is not possible, in general, to do this consistently for just any subset of X (see *e.g.* [97] for a construction of an "unmeasurable" set). Rather, such an assignment can only be made consistently for a collection of subsets called a σ-algebra.

Definition 2.4 (σ-algebra, measurable space). Let X be a set. A non-empty collection $\mathcal{A}$ of subsets of X is called a σ-*algebra*, and the pair $(X, \mathcal{A})$ is called a *measurable space*, if each of the following holds.

1. If $A \in \mathcal{A}$ then $X \setminus A \in \mathcal{A}$.

2. If $\{A_i\} \subset \mathcal{A}$ is a finite or countable collection then $\bigcup_i A_i \in \mathcal{A}$.

3. If $\{A_i\} \subset \mathcal{A}$ is a finite or countable collection then $\bigcap_i A_i \in \mathcal{A}$.

4. If $A, B \in \mathcal{A}$ then $A \setminus B \in \mathcal{A}$.

5. $X \in \mathcal{A}$ and $\emptyset \in \mathcal{A}$.

The elements of a σ-algebra are sometimes called *measurable sets*. In the context of probability theory they are sometimes called *events*. Note that the properties above are not independent, *e.g.* properties 3–5 follow from others. We include all five here to give a more thorough and intuitive characterization of σ-algebras.

Finite σ-algebras are easy to construct. For example $\{X, \emptyset\}$ is a trivial σ-algebra. Given $A \subset X$, the collection $\{X, \emptyset, A, X \setminus A\}$ is a σ-algebra. Although less trivial σ-algebras are sometimes extremely difficult to characterize explicitly, the existence of a σ-algebra containing certain sets can be asserted using the following theorem.

Theorem 2.1. *Let X be a set and let $\mathcal{B}$ be a class of subsets of X. Then there is a unique smallest σ-algebra, denoted $\sigma(\mathcal{B})$ (the σ-algebra generated by $\mathcal{B}$), that contains every set in $\mathcal{B}$. That is,*

- *$\mathcal{B} \subset \sigma(\mathcal{B})$, and*

- *if $\mathcal{A}$ is any σ-algebra such that $\mathcal{B} \subset \mathcal{A}$, then $\sigma(\mathcal{B}) \subset \mathcal{A}$.*

Proof. See *e.g.* [55]. $\qquad\square$

We frequently work with a phase space X that comes already equipped with a topology (*e.g.*, a metric space). An important application of the preceding theorem is the identification of a σ-algebra that is compatible with a given topology:

Definition 2.5 (Borel σ-algebra). Let X be a topological space. The σ-algebra $\mathcal{A}$ generated by the collection of all open subsets of X is called the *Borel σ-algebra* on X, and the elements of $\mathcal{A}$ are called *Borel sets*.

The Borel σ-algebra is a natural choice since the sets we typically want to work with— open and closed sets, as well as their finite and countable unions and intersections—are all measurable. In the following, whenever a topology is given or implied the corresponding Borel σ-algebra will be implied also.

Definition 2.6 (measure, measure space). Let $(X, \mathcal{A})$ be a measurable space. A function $\mu : \mathcal{A} \to \mathbb{R}_+$ is called a *measure*, and the triple $(X, \mathcal{A}, \mu)$ is called a *measure space*, if each of the following holds.

1. $\mu(\emptyset) = 0$.

2. $\mu(A) \geq 0 \quad \forall A \in \mathcal{A}$.

3. If $\{A_i\} \subset \mathcal{A}$ is a finite or countable collection of pairwise disjoint sets then
$\mu\left(\bigcup_i A_i\right) = \sum_i \mu(A_i)$.

4. If $A, B \in \mathcal{A}$ and $B \subset A$ then $\mu(A \setminus B) = \mu(A) - \mu(B)$.

Note that 4 is not an independent property, but is a useful consequence of the first three.

In the following, "a measure space X" will be taken to mean "a measure space $(X, \mathcal{A}, \mu)$" wherever this is unlikely to cause confusion.

Definition 2.7 (Borel measure). Given a topological space X, a *Borel measure* on X is a measure defined on the Borel σ-algebra generated by the topology on X.

On the real line, the notion of length provides a natural Borel measure μ, defined on intervals according to

$$\mu\big([a,b]\big) = b - a. \tag{2.5}$$

Similarly, in $\mathbb{R}^n$ the notion of volume also provides a natural Borel measure, defined on n-dimensional "rectangles" according to

$$\mu\big([a_1, b_1] \times \cdots \times [a_n, b_n]\big) = (b_1 - a_1) \cdots (b_n - a_n). \tag{2.6}$$

It can be proved [55] that this formula can be uniquely extended to a measure on the respective Borel σ-algebra of $\mathbb{R}^n$. *Lebesgue measure*, λ, is the completion of this measure.[1]

Definition 2.8 (null set). Let $(X, \mathcal{A}, \mu)$ be a measure space. A set $A \subset X$ is called a *null set* if it is contained in a set $B \in \mathcal{A}$ with $\mu(B) = 0$.

Definition 2.9 (almost everywhere). Let $(X, \mathcal{A}, \mu)$ be a measure space. A property is said to hold μ-almost everywhere if it holds on the complement of a null set.

2.2.2 Probability measures

With the basic concepts of measure theory outlined above, the elements of probability theory that we will need can be expressed in measure theoretic terms.

Definition 2.10 (finite and probabilistic measures). Let X be a measure space. If $\mu(X) < \infty$ then μ is said to be *finite*. If $\mu(X) = 1$ then the finite measure μ is said to be *probabilistic* and $(X, \mathcal{A}, \mu)$ is called a *probability space*.

---

[1]A measure λ is the completion of a Borel measure μ if
- λ is defined on the smallest σ-algebra containing both the Borel sets and all subsets of Borel sets having Borel measure 0.
- λ agrees with μ wherever μ is defined; in particular, $\lambda(A) = 0$ if $A \subseteq B$ and $\mu(B) = 0$.

Note that any finite measure μ can be normalized to yield a corresponding probability measure ν according to $\nu(A) = \mu(A)/\mu(X)$.

Depending on context a probability measure has at least two different interpretations, owing to the similarity in structure between probability and statistics and the fact that probability measures embody this structure:

Probabilistic interpretation A probability measure μ can represent the "probability distribution" of a random variable $x \in X$. For a given measurable set $A \subset X$, the quantity $\mu(A)$ represents $\mathrm{Prob}(x \in A)$. The measurable sets are called *events*, since the statement "$x \in A$" asserts the occurrence of a particular event or outcome. The defining properties of a probability measure can be interpreted in this context as the axioms of probability theory.

Statistical ensemble interpretation A probability measure μ can represent the distribution of an idealized, infinite ensemble of points $\{x_i\} \subset X$. The quantity $\mu(A)$ describes the fraction of the x_i for which $x_i \in A$.

This dual interpretation of probability measures is helpful in solving problems, since we are free to use whichever interpretation is most convenient in a given context.

In the following we will require the notion of measurable transformation:

Definition 2.11 (measurable transformation). Let $(X, \mathcal{A})$ and $(Y, \mathcal{B})$ be measurable spaces. A transformation $S : X \to Y$ is *measurable* if

$$S^{-1}(B) \in \mathcal{A} \quad \forall B \in \mathcal{B}, \tag{2.7}$$

where the pre-image $S^{-1}(B)$ is defined as the set

$$S^{-1}(B) = \{x \in X : S(x) \in B\}. \tag{2.8}$$

In particular, a transformation $S : X \to X$ is measurable if $S^{-1}(A) \in \mathcal{A} \ \forall A \in \mathcal{A}$. A functional $f : X \to \mathbb{R}$ is measurable if $f^{-1}(B) \in \mathcal{A}$ for every Borel set $B \subset \mathbb{R}$. Note that the set $S^{-1}(B)$ is well defined even if S does not have an inverse.

Continuity is a sufficient condition for measurability:

Theorem 2.2. *Let X, Y be topological spaces and $\mathcal{A}$, $\mathcal{B}$ their respective Borel σ-algebras. If $S : X \to Y$ is continuous then S is measurable.*

Proof. See *e.g.* [55]. □

2.2.3   Lebesgue integral

We will require a notion of integration on measure spaces. One such notion that can be defined on an arbitrary measure space (and requires no other structure) is the *Lebesgue integral*. Various developments of the theory of this integral have been given, all somewhat lengthy. The following, based on the presentation given in [74], gives the basic idea; for a more complete treatment see *e.g.* [55].

An intuitive way to approach the Lebesgue integral is to define it first for "simple functions". Given a subset $A \subset X$, denote by 1_A the "indicator function"

$$1_A(x) = \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{if } x \notin A. \end{cases} \tag{2.9}$$

Definition 2.12 (simple function). Let $(X, \mathcal{A}, \mu)$ be a measure space, and suppose the sets $A_i \in \mathcal{A}$, $i = 1, \ldots, n$, are pairwise disjoint. Then a functional $g : X \to \mathbb{R}$ of the form

$$g(x) = \sum_{i=1}^{n} \lambda_i 1_{A_i}(x), \quad \lambda_i \in \mathbb{R} \tag{2.10}$$

is called a *simple function*.

Definition 2.13. The Lebesgue integral of the simple function (2.10) is defined by

$$\int_X g(x) \, d\mu(x) \equiv \sum_{i=1}^{n} \lambda_i \mu(A_i). \tag{2.11}$$

The definition of Lebesgue integral proceeds by approximating such more general functions with simple functions. If $f : X \to \mathbb{R}$ is a non-negative bounded measurable function, then f can be approximated by simple functions f_n such that $\{f_n(x)\}$ converges to $f(x)$ uniformly in x [74, p. 20].

Definition 2.14. If $f : X \to \mathbb{R}$ is a non-negative bounded measurable function, and $\{f_n\}$ is a sequence of simple functions converging uniformly to f then

$$\int_X f(x) \, d\mu(x) \equiv \lim_{n \to \infty} \int_X f_n(x) \, d\mu(x). \tag{2.12}$$

(This limit exists and is independent of the particular sequence $\{f_n\}$.)

The preceding definition can be extended unambiguously to define the integral over X of an arbitrary measurable functional $f : X \to \mathbb{R}$. If the integral is finite then f is said to be *integrable.* For arbitrary $A \in \mathcal{A}$ we define

$$\int_A f(x) \, d\mu(x) \equiv \int_X 1_A(x) f(x) \, d\mu(x). \tag{2.13}$$

Where confusion is unlikely we will sometimes write simply $\int f \, d\mu$ to mean $\int_X f(x) \, d\mu(x)$.

While the Lebesgue integral is more general than the usual Riemann integral, the two notions of integral agree for any Riemann-integrable function if the integral is taken with respect to Lebesgue measure [98, p. 323].

Some important properties of the Lebesgue integral, which we give without proof, are the following.

- $\int |f| \, d\mu = 0$ if and only if $f = 0$ almost everywhere.

- If $\int_A f \, d\mu = \int_A g \, d\mu \; \forall A \in \mathcal{A}$ then $f = g$ almost everywhere.

- If $f : X \to \mathbb{R}$ is integrable and the finite or countable collection of sets $\{A_i\} \subset \mathcal{A}$ is disjoint with $\bigcup_i A_i = A$, then

$$\sum_i \int_{A_i} f \, d\mu = \int_A f \, d\mu. \tag{2.14}$$

Theorem 2.3 (change of variables). *Let (X, μ) be a measure space, $S : X \to Y$ a measurable transformation. If $f : Y \to \mathbb{R}$ is measurable then $\mu \circ S^{-1}$ is a measure on Y, and for any measurable $A \subset Y$,*

$$\int_A f \, d(\mu \circ S^{-1}) = \int_{S^{-1}(A)} (f \circ S) \, d\mu. \tag{2.15}$$

Proof. See *e.g.* [55]. □

The Lebesgue integral leads to an important class of normed vector spaces:

Definition 2.15 (L^p space). Let $(X, \mathcal{A}, \mu)$ be a measure space and p a real number, $1 \leq p < \infty$. The set of functionals $f : X \to \mathbb{R}$ such that $|f|^p$ is integrable, *i.e.,*

$$\int_X |f(x)|^p \, d\mu(x) < \infty, \tag{2.16}$$

is denoted by $L^p(X, \mathcal{A}, \mu)$, or simply $L^p(X)$ if $\mathcal{A}$ and μ are understood. The norm $\| \cdot \|$ on L^p is defined by

$$\|f\| = \left[\int |f|^p \, d\mu \right]^{1/p}. \tag{2.17}$$

2.2.4 Densities

Densities provide a convenient and intuitive way of prescribing probability measures on $\mathbb{R}^n$. Indeed, applied probability and statistics is concerned largely with certain specific types of densities (*e.g.*, the uniform, Gaussian, and Poisson densities, among others). Densities have the additional advantage that they allow many problems in measure theory and probability to be solved using calculus.

Suppose $(X, \mathcal{A}, \mu)$ is a measure space. If $f : X \rightarrow \mathbb{R}$ is a non-negative integrable function then

$$\mu_f(A) = \int_A f \, d\mu \tag{2.18}$$

defines a finite measure on A. For a certain class of measures, this observation can be reversed:

Definition 2.16 (absolutely continuous measure). Let $(X, \mathcal{A}, \mu)$ be a measure space. A measure ν on $\mathcal{A}$ is *absolutely continuous* with respect to μ (written $\nu \ll \mu$) if for every set $A \in \mathcal{A}$ for which $\mu(A) = 0$ we have that $\nu(A) = 0$.

Theorem 2.4 (Radon-Nikodym). *Let X be a σ-finite measure space*[2] *and let μ_f be a finite measure on $\mathcal{A}$ with $\mu_f \ll \mu$. Then there exists a unique, non-negative integrable functional $f \in L^1(X)$ such that*

$$\mu_f(A) = \int_A f \, d\mu \quad \forall A \in \mathcal{A}. \tag{2.19}$$

Proof. See *e.g.* [44, p. 69].   □

Absolute continuity with respect to a given measure μ defines an important class of measures μ_f that can be expressed in the form (2.19). If μ_f is a probability measure, the functional f in equation (2.19) is called the *probability density* or simply the *density* of μ_f.

---

[2]A measure space X is σ-finite if $X = \bigcup_i A_i$ with $\mu(A_i) < \infty$ [55]. Any measure we might wish to consider for practical purposes is σ-finite.

Definition 2.17 (Set of densities). Let $(X, \mathcal{A}, \mu)$ be a measure space. The space

$$D(X) = \{f \in L^1(X) : f \geq 0 \text{ and } \|f\| = 1\} \tag{2.20}$$

is the set of *densities on* X, and an element $f \in D(X)$ is called a *density*.

On $\mathbb{R}^n$ densities are usually given with respect to Lebesgue measure.

2.3 Evolution of Probabilities for Dynamical Systems

2.3.1 Evolution of probability measures

Suppose we have a dynamical system described by a semigroup $\{S_t : t \in G\}$ ($G = \mathbb{R}_+$ or $\mathbb{Z}_+$) of transformations $S_t : X \to X$, such that $x_0 \mapsto S_t(x_0) = x_t$. Suppose further that X is equipped with a σ-algebra $\mathcal{A}$, and that a particular measure ν_0 on $\mathcal{A}$ has the interpretation that $\nu_0(A) = \mathrm{Prob}(x_0 \in A)$. One can then ask, what is the probability measure ν_t that describes x_t, *i.e.*, such that $\nu_t(A) = \mathrm{Prob}(x_t \in A)$?

Consider that for arbitrary $A \in \mathcal{A}$ we have

$$\begin{aligned}
\nu_t(A) &= \mathrm{Prob}\big(S_t(x_0) \in A\big) \\
&= \mathrm{Prob}\big(x_0 \in S_t^{-1}(A)\big) \\
&= \nu_0\big(S_t^{-1}(A)\big).
\end{aligned} \tag{2.21}$$

Thus the probability measure ν_t describing x_t is given by the formula

$$\nu_t = \nu_0 \circ S_t^{-1}. \tag{2.22}$$

It is readily verified that this expression does indeed define a probability measure ν_t on $\mathcal{A}$. This measure is said to be the *image* of ν_0 under the transformation S_t and (abusing notation somewhat) we write $S_t \nu_0 \equiv \nu_0 \circ S_t^{-1}$. Note that each of the S_t must be a measurable transformation. Since $\{S_t\}$ is assumed to be a semidynamical system (*cf.* Definition 2.2), the transformation $x \mapsto S_t(x)$ is indeed measurable since it is continuous.

2.3.2 Evolution of densities: Perron-Frobenius operator

If the initial phase point x_0 of a dynamical system is described by a probability density rather than a probability measure, we can ask "what is the probability density that de-

scribes $x_t = S_t(x_0)$? For this question to be well-posed, we require that S_t be nonsingular:

Definition 2.18 (nonsingular transformation). Let $(X, \mathcal{A}, \mu)$ be a measure space. A measurable transformation $S : X \to X$ is *nonsingular* (with respect to μ) if for every set $A \in \mathcal{A}$ for which $\mu(A) = 0$ we have that $\mu(S^{-1}(A)) = 0$.

Suppose the density of x_0 (with respect to some measure μ) is given by f_0, with corresponding probability measure ν_0 (*cf.* equation (2.18)). If S_t is nonsingular then from equation (2.22) we have that

$$\nu_t \ll \nu_0 \ll \mu. \tag{2.23}$$

Therefore, by Theorem 2.4 (assuming the measure space is σ-finite) there is a unique element $f_t \in L^1(X)$ such that

$$\nu_t(A) = \int_A f_t \, d\mu \quad \forall A \in \mathcal{A}. \tag{2.24}$$

Using equation (2.22) we obtain

$$\int_A f_t \, d\mu = \int_{S_t^{-1}(A)} f_0 \, d\mu \quad \forall A \in \mathcal{A}. \tag{2.25}$$

In fact, for any given $f_0 \in L^1(X)$, this relationship uniquely determines an element $f_t \in L^1(X)$ [74, p. 42] and thus defines a mapping $P_t : f_0 \mapsto f_t$.

Definition 2.19 (Perron-Frobenius operator). Let $(X, \mathcal{A}, \mu)$ be a measure space, $S : X \to X$ a nonsingular transformation. The operator $P : L^1(X) \to L^1(X)$ defined by

$$\int_A (Pf) \, d\mu = \int_{S^{-1}(A)} f \, d\mu \quad \forall A \in \mathcal{A} \tag{2.26}$$

is called the *Perron-Frobenius* operator corresponding to S.

Note that if P_t is the Perron-Frobenius operator corresponding to S_t then the action of P_t on a density f_0 carries out the evolution of this density under the action of S_t.

Some important properties of the Perron-Frobenius operator, which can be verified directly from its definition, are the following [74].

- $P : L^1(X) \to L^1(X)$ is linear.

- $Pf \geq 0$ if $f \geq 0$.

- $\int (Pf)\,d\mu = \int f\,d\mu.$

- If P is the Perron-Frobenius operator with respect to S, then P^n is the Perron-Frobenius operator with respect to S^n.

- If $\{S_t : t \in G\}$ is a semigroup then the corresponding family of Perron-Frobenius operators $\{P_t : t \in G\}$ is also a semigroup.

From the first three of these properties it follows that P is a *Markov operator*, *i.e.* a linear transformation of $L^1(X)$ that maps densities to densities.

Explicit Perron-Frobenius operators

In some cases equation (2.26) can be used to find an explicit representation of the Perron-Frobenius operator. If $S : \mathbb{R} \to \mathbb{R}$ and $A = [a, x]$ then (2.26) becomes

$$\int_a^x (Pf)(s)\,ds = \int_{S^{-1}([a,x])} f(s)\,ds. \tag{2.27}$$

Differentiating then yields

$$Pf(x) = \frac{d}{dx} \int_{S^{-1}([a,x])} f(s)\,ds. \tag{2.28}$$

Consider for example the dynamical system discussed in Chapter 1, defined by iterates of

$$S : x \mapsto 4x(1 - x). \tag{2.29}$$

Since

$$S^{-1}([0, x]) = [0, \tfrac{1}{2} - \tfrac{1}{2}\sqrt{1 - x}] \cup [\tfrac{1}{2} + \tfrac{1}{2}\sqrt{1 - x}, 1], \tag{2.30}$$

equation (2.28) becomes

$$\begin{aligned} Pf(x) &= \frac{d}{dx} \int_0^{1/2 - 1/2\sqrt{1-x}} f(s)\,ds + \frac{d}{dx} \int_{1/2 + 1/2\sqrt{1-x}}^1 f(s)\,ds \\ &= \frac{1}{4\sqrt{1-x}} \left[f\left(\tfrac{1}{2} - \tfrac{1}{2}\sqrt{1-x}\right) + f\left(\tfrac{1}{2} + \tfrac{1}{2}\sqrt{1-x}\right) \right]. \end{aligned} \tag{2.31}$$

Figure 2.1 illustrates the sequence of densities $P^n(f)$, $n = 0, \ldots, 3$, obtained by equation (2.28) for a uniform initial density $f = 1$ on $[0, 1]$; *cf.* Figures 1.2–1.3.
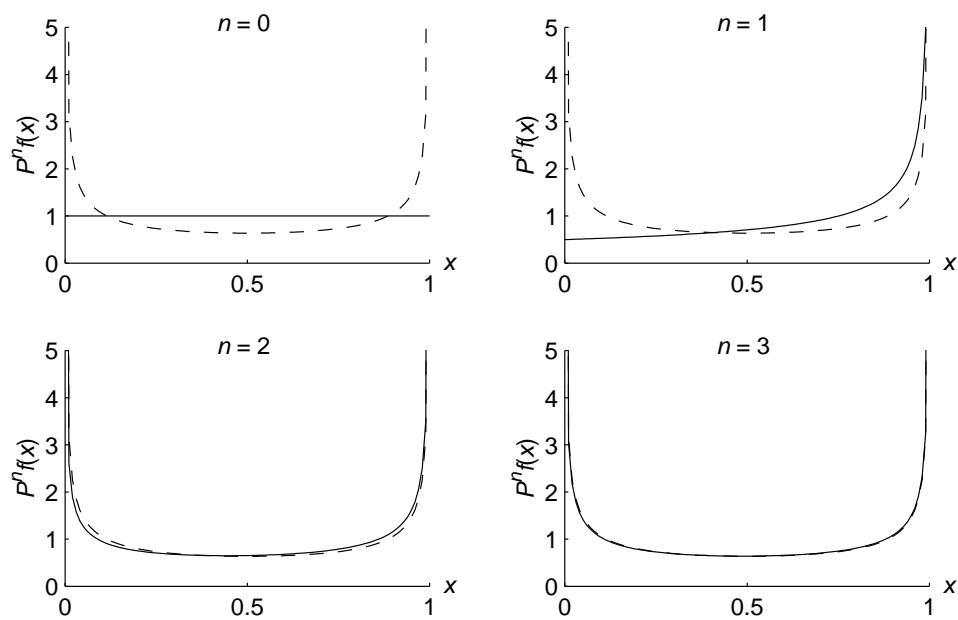
Figure 2.1: Iterates $P^n(f)$ of the Perron-Frobenius operator corresponding to the quadratic map $x \mapsto 4x(1-x)$, for the initial density $f = 1$. Also shown, for comparison, is the invariant density $f_*(x)$ (dashed curve $---$) given by equation (2.38).

Perron-Frobenius operator for flows

Autonomous ordinary differential equations consitute an important class of continuous-time dynamical systems. If the initial value problem

$$\frac{dx}{dt} = F(x), \quad t \geq 0, \quad F : \mathbb{R}^n \rightarrow \mathbb{R}^n$$
$$x(0) = x_0$$

(2.32)

has a unique solution for all $t > 0$ [3] for each initial value x_0, then the family of solution maps $S_t : x_0 \mapsto x(t)$ constitute a differentiable semigroup or *flow* $\{S_t : t \in \mathbb{R}_+\}$ on $\mathbb{R}^n$.

Suppose $u(t) \in L^1(\mathbb{R}^n)$ is the probability density of the solution variable x at time t. If the vector field F is smooth and we interpret $u(x,t)$ as a scalar quantity transported by the flow, then by analogy with the continuity equation for scalar transport in fluid mechanics, we have the following evolution equation for u,

$$\frac{\partial u}{\partial t} = -\nabla \cdot (uF).$$

(2.33)

(See *e.g.* [74, p. 210] for a measure-theoretic justification of this evolution equation.) The semigroup $\{P_t\}$ of Perron-Frobenius operators corresponding to $\{S_t\}$ is then defined by

$$(P_t f)(x) = u(x,t)$$

(2.34)

where $u(x,t)$ is the solution of (2.33) with initial data $u(x,0) = f(x)$.

2.4 Ergodic Theory

Ergodic theory can be described as the study of measure theoretic invariants of dynamical systems. Invariant measures play a fundamental role.

2.4.1 Invariant measures

Chapter 1 gave an example of a discrete-time dynamical system, defined by iterates of the map

$$S : x \mapsto 4x(1-x).$$

(2.35)

---

[3] *e.g.* it suffices to have F continuous with $|F(x)| \leq \alpha + \beta|x|$ for constants α, β [39].

Numerical evidence suggests the existence of a unique density that is unchanged by evolution under the action the map (2.35) (*cf.* Figures 1.2 and 1.3). The concepts introduced in the previous section provide the following characterization of invariant densities and the more general notion of invariant measure.

Definition 2.20 (invariant measure). Let $(X, \mathcal{A})$ be a measurable space, $\{S_t : t \in G\}$ ($G = \mathbb{R}_+$ or $\mathbb{Z}_+$) a dynamical system on X. A measure μ on $\mathcal{A}$ is *invariant* under $\{S_t\}$ if $S_t \mu = \mu$. That is,

$$\mu\big(S_t^{-1}(A)\big) = \mu(A) \quad \forall A \in \mathcal{A}, \ \forall t \in G. \tag{2.36}$$

We say that S_t preserves μ, or that S_t is μ-preserving.

Absolutely continuous invariant measures can be characterized as fixed points of the Perron-Frobenius operator:

Definition 2.21 (invariant density). Let $(X, \mathcal{A}, \mu)$ be a measure space, $\{S_t : t \in G\}$ a dynamical system such that each S_t nonsingular. A probability density f is *invariant* under $\{S_t\}$ if $P_t f = f$, where P_t is the Perron-Frobenius operator with respect to S_t. That is,

$$\int_A f \, d\mu = \int_{S_t^{-1}(A)} f \, d\mu \quad \forall A \in \mathcal{A}, \ \forall t \in G. \tag{2.37}$$

For example, with the Perron-Frobenius operator (2.28) corresponding to the map (2.35), it is straightforward to check that the density

$$f_*(x) = \frac{1}{\pi \sqrt{x(1-x)}} \tag{2.38}$$

is invariant under iterates of S. The graph of this density is shown in Figure 2.1; it agrees with the numerically obtained invariant density shown in Figures 1.2 and 1.3.

One immediate consequence of the existence of an invariant measure is the following.

Theorem 2.5 (Poincaré Recurrence). *Let $\{S_t : t \in G\}$ ($G = \mathbb{Z}_+$ or $\mathbb{R}_+$) be a μ-preserving dynamical system on a measure space X. Then $\forall A \in \mathcal{A}$ and $\forall T > 0$, the set*

$$\{x \in A : S_t x \notin A \ \forall t > T\} \tag{2.39}$$

has measure 0. In other words, if $\mu(A) > 0$ then under the action of $\{S_t\}$, μ-almost every initial phase point in A returns to A infinitely often.

Proof. See *e.g.* [71], or [84] in the simpler case $\mu(X) < \infty$. $\square$

This theorem already has interesting consequences for the orbit structure of the dynamical system $x \mapsto 4x(1-x)$ (typical orbits are shown in Figure 1.1): since the measure with density (2.38) is invariant, under iteration by S, Lebesgue-almost every $x \in [0,1]$ returns arbitrarily close to x infinitely often. This imples a certain kind of irregularity, in fact aperiodicity, thus confirming the intuitive impression of Figure 1.1. However, for some systems the consequences of the Poincaré Recurrence Theorem are quite vacuous. For example the identity map preserves Lebesgue measure on $[0,1]$, but the consequences of the Poincaré Recurrence Theorem in this case are trivial.

2.4.2 Statistical regularity

From the examples above it seems that, for some systems, existence of an invariant measure implies strong statistical properties, while for other systems is does not. It would be nice to have tools for discerning between dynamical systems that exhibit varying degrees of irregularity and have different statistical properties.

As observed in Chapter 1, an orbit $\{x_n\}$ of the map $x \mapsto 4x(1-x)$ appears to have a well defined asymptotic distribution, as seen from a histogram of the trajectory (*cf.* Figure 1.4). The frequency with which an orbit visits a given histogram bin is asymptotically regular, despite the irregularity of the orbit itself. A somewhat stronger notion of statistical regularity requires that an arbitrary continuous function, evaluated along an orbit of $\{S_t\}$, has a well-defined time average [104]:

Definition 2.22 (statistical regularity). Let $\{S_t : t \in G\}$ ($G = \mathbb{R}_+$ or $\mathbb{Z}_+$) be a dynamical system on X. An orbit $\{S_t x : t > 0\}$ is *statistically regular* if the time average

$$\bar{\phi}(x) = \lim_{n \to \infty} \frac{1}{n} \sum_{k=1}^{n} \phi(S_k x) \qquad (\text{if } G = \mathbb{Z}_+) \tag{2.40}$$

or

$$\bar{\phi}(x) = \lim_{T \to \infty} \frac{1}{T} \int_0^T \phi(S_t x) \, dt \qquad (\text{if } G = \mathbb{R}_+) \tag{2.41}$$

exists for every bounded continuous function $\phi : X \to \mathbb{R}$.

Statistical regularity implies the existence of an invariant probability measure:

Theorem 2.6. *Let $\{S_t : t \in G\}$ be a dynamical system on a metrizable space X. If the time average $\bar{\phi}(x)$ exists for every bounded continuous function $\phi : X \to \mathbb{R}$ then there is a unique invariant Borel probability measure μ_x such that*

$$\bar{\phi}(x) = \int \phi \, d\mu_x. \tag{2.42}$$

Proof. The mapping $T : \phi \mapsto \bar{\phi}(x)$ defined by equations (2.40)–(2.41) is a bounded, linear, positive functional on $C(X)$. Therefore by the Riesz Representation Theorem [55] there is a unique Borel probability measure μ_x such that

$$T(\phi) = \int \phi \, d\mu_x. \tag{2.43}$$

Furthermore (in the discrete-time case),

$$
\begin{aligned}
\int \phi \, d(\mu_x \circ S_t^{-1}) &= \int (\phi \circ S_t) \, d\mu_x \\
&= \overline{\phi \circ S_t}(x) \\
&= \lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} (\phi \circ S_t)(S_k x) \\
&= \lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} \phi(S_{t+k} x) \\
&= \lim_{n\to\infty} \frac{1}{n} \Big[\sum_{k=1}^{n} \phi(S_k x) + \sum_{k=n+1}^{n+t} \phi(S_k x) - \sum_{k=1}^{t} \phi(S_k x) \Big] \\
&= \bar{\phi}(x).
\end{aligned}
\tag{2.44}
$$

It follows that $\mu_x \circ S_t^{-1} = \mu_x$, by uniqueness of μ_x. A similar argument holds in the continuous-time case. $\square$

2.4.3 Ergodicity, mixing, exactness

A century ago, motivated by fundamental problems in statistical mechanics, Boltzmann and Gibbs raised the *ergodic problem*: to determine sufficient conditions under which the time average $\bar{\phi}(x)$ (equations (2.40)–(2.41)) exists and is essentially independent of x. The answer, given by Birkhoff in 1931, is that it is both necessary and sufficient that the dynamical system $\{S_t\}$ be ergodic.

Definition 2.23 (invariant set). Let $S : X \to X$ be a measurable transformation. A set $A \in \mathcal{A}$ is *invariant* under S if[4] $S^{-1}(A) \, \Delta \, A$ has measure 0 (that is, $S^{-1}(A) = A$ mod μ). If $\{S_t\}$ is a semigroup then A is invariant under $\{S_t\}$ if A is invariant under $S_t \, \forall t > 0$.

Definition 2.24 (ergodic dynamical system). Let $\{S_t : t \in G\}$ ($G = \mathbb{R}_+$ or $\mathbb{Z}_+$) be a dynamical system on a probability space X. $\{S_t\}$ is *ergodic* (alternatively μ is ergodic) if for every S_t-invariant set A, either $\mu(A) = 0$ or $\mu(A) = 1$.

Ergodicity is a non-decomposability condition: having $\{S_t\}$ ergodic means there is no non-null set that is invariant under $\{S_t\}$. If it were possible to decompose X into invariant sets $A \cup B = X$ then $\{S_t\}$ could be studied separately on either A or B. An ergodic dynamical system must be studied on essentially the entire space X.

Ergodicity is a sufficient condition for statistical regularity:

Theorem 2.7 (Birkhoff Ergodic Theorem). *Let $\{S_t : t \in G\}$ be an ergodic μ-preserving dynamical system on a measure space X, and $\phi \in L^1(X)$. Then for μ-almost all x,*

$$\int \phi \, d\mu = \begin{cases} \displaystyle \lim_{n \to \infty} \frac{1}{n} \sum_{k=1}^{n} \phi(S_k x) & \text{if } G = \mathbb{Z}_+ \\ \displaystyle \lim_{T \to \infty} \frac{1}{T} \int_0^T \phi(S_t x) \, dt & \text{if } G = \mathbb{R}_+. \end{cases} \tag{2.45}$$

Proof. The proof is somewhat technical, and can be found in any of the standard references on ergodic theory—see *e.g.* [71, 93]. $\qquad \square$

In other words, if $\{S_t\}$ is ergodic then the time average of ϕ along μ-almost every orbit is just a "spatial" average or expectation of ϕ, weighted with respect to μ. Ergodicity also gives an explicit formula for the ergodic measure μ in terms of time averages. Applying the ergodic theorem with $\phi = 1_A$ yields

$$\begin{aligned} \mu(A) &= \int 1_A d\mu \\ &= \lim_{n \to \infty} \frac{1}{n} \sum_{k=1}^{n} 1_A(S_k x) \\ &= \lim_{n \to \infty} \frac{1}{n} \#\{k > 0 : S_k x \in A\}, \end{aligned} \tag{2.46}$$

---

[4] The "symmetric difference" of two sets is defined by $A \, \Delta \, B = (A \setminus B) \cup (B \setminus A)$.

which is just the fraction of time that the orbit $\{S_k x : k > 0\}$ spends in the set A. Thus an ergodic measure describes the asymptotic distribution in phase space of μ-almost every orbit. This helps explain the observation made in Chapter 1 that the long-run distribution of points on any given orbit agrees with the invariant density.

From Theorem 2.7 we have that ergodicity implies the existence of the time average of any L^1 function ϕ evaluated along μ-almost every trajectory, hence ergodicity implies statistical regularity. However, ergodicity remains a difficult property to prove for the dynamical systems that arise in statistical mechanics, and for practical purposes the ergodic problem is still unresolved.

Ergodicity (and hence statistical regularity) does not necessarily imply any kind of "random" behavior of individual orbits. Irrational rotations are a classic example. Let points on the circumference of a circle be parametrized by points $x \in [0, 2\pi)$, and denote rotation through angle ϕ by the map

$$S : x \mapsto x + \phi \pmod{2\pi}. \tag{2.47}$$

This system is ergodic if (and only if) $\phi/2\pi$ is irrational (see *e.g.* [74, p. 75] and [71, Prop 4.2.1]). The ergodic invariant measure is Lebesgue measure on $[0, 2\pi)$, hence Lebesgue almost every orbit of this system is statistically regular and is, asymptotically, distributed uniformly on $[0, 2\pi)$.

To distinguish between degrees of irregularity there are other, stronger statistical properties that we can ask of a given dynamical system. Two such properties are mixing and exactness.

Definition 2.25 (mixing). Let $\{S_t : t \in G\}$ ($G = \mathbb{R}_+$ or $\mathbb{Z}_+$) be a μ-preserving dynamical system on a probability space $(X, \mathcal{A}, \mu)$. $\{S_t\}$ is called *mixing* if $\forall A, B \in \mathcal{A}$

$$\mu\big(A \cap S_t^{-1}(B)\big) \to \mu(A)\mu(B) \quad \text{as } t \to \infty. \tag{2.48}$$

In particular if $\mu(A) \neq 0$ then for a mixing dynamical system,

$$\frac{\mu\big(A \cap S_t^{-1}(B)\big)}{\mu(A)} \to \mu(B) \quad \text{as } t \to \infty. \tag{2.49}$$

In other words, the fraction (with respect to μ) of phase points originating in A that end up in B after time t (for sufficiently large t) is equal to the size of B. This result is

independent of the particular choice of sets A, B.

The mixing property comes closer than ergodicity to characterizing what one would call a "random" process, as the condition (2.49) implies that $S_t x$ eventually becomes uncorrelated with the initial state x. Thus mixing plays much the same role as "sensitivity to initial conditions" in the topological approach to chaotic dynamics. Mixing distinguishes between merely statistically regular systems and systems with stronger statistical properties. Irrational rotations, for example, are ergodic but not mixing. A yet stronger statistical property is exactness:

Definition 2.26 (exactness). Let $\{S_t : t \in G\}$ be a μ-preserving dynamical system on a probability space $(X, \mathcal{A}, \mu)$, such that $S_t(A) \in \mathcal{A} \ \ \forall A \in \mathcal{A}$, $t \geq 0$. $\{S_t\}$ is called *exact* if

$$\lim_{t \to \infty} \mu\big(S_t(A)\big) = 1. \tag{2.50}$$

Non-invertibility is a necessary condition for exactness, since for invertible S_t we have

$$\mu\big(S_t(A)\big) = (\mu \circ S_t^{-1})\big(S_t(A)\big) = \mu(A), \tag{2.51}$$

so the condition (2.50) cannot hold. Thus there are mixing systems (*e.g.* the two-dimensional Baker transformation [74], which is invertible) that are not exact. The transformation $x \mapsto 4x(1 - x)$ is exact [74, p. 167].

The implications between the various ergodic properties of dynamical systems, and their topological counterparts, are summarized in the following diagram (proofs can be found in [71, 74]; definitions of topological notions of complex dynamics are given in [71]).

$$
\begin{array}{ccc}
\text{exactness} & & \\
\downarrow & & \\
\text{mixing} & \longrightarrow & \text{topological mixing} \\
\downarrow & & \downarrow \\
\text{ergodicity} & \longrightarrow & \text{transitivity} \\
\downarrow & & \downarrow \\
\text{statistical regularity} & \longrightarrow & \text{orbit recurrence}
\end{array}
$$
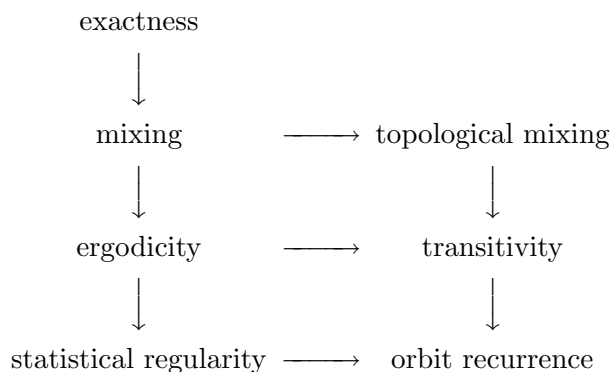
For systems having any of the ergodic properties discussed above, an arbitrary initial density converges—that is, iterates of the Perron-Frobenius operator converge—to a

uniform density with respect to the invariant measure. A classification can be made in terms of the strength of this convergence, and the Perron-Frobenius operator beomes an important classification tool:

Theorem 2.8. *Let $\{S_t : t \in G\}$ ($G = \mathbb{R}_+$ or $\mathbb{Z}_+$) be a dynamical system on a probability space (X, μ), and let $\{P_t : t \in G\}$ be the corresponding semigroup of Perron-Frobenius operators. If $\{S_t\}$ has a unique absolutely continuous measure μ_{f_*} with positive density f_* then*

1. *$\{S_t\}$ is exact iff $\forall f \in L^1(X)$, $\{P_t f\}$ is strongly convergent[5] to f_*;*

2. *$\{S_t\}$ is mixing iff $\forall f \in L^1(X)$, $\{P_t f\}$ is weakly convergent[6] to f_*;*

3. *$\{S_t\}$ is ergodic iff $\forall f \in L^1(X)$, $\{P_t f\}$ is Césaro convergent[7] to f_*.*

Proof. See *e.g.* [74, p. 72], [84, pp. 56, 63, 92]. □

2.4.4 Natural and physical measures

According the Birkhoff Ergodic Theorem 2.7, an ergodic measure μ describes the asymptotic statistics of μ-almost every trajectory of a dynamical system $\{S_t\}$. Unfortunately, this statement need not have much dynamical relevance. For example, the map $x \mapsto 4x(1-x)$ has a fixed point at the origin. The probability measure concentrated at the origin is ergodic, and trivially reflects the statistics of an orbit originating at this point, but it has nothing to say about any other orbit. The same applies to an ergodic measure concentrated on any periodic orbit. The problem here is that the Birkhoff Ergodic Theorem gives a description of orbits only on a set of Lebesgue measure zero.

   In general, a dynamical system may have many ergodic measures—in fact uncountably many [40, 104]—only some of which imply something about the statistical behavior of typical orbits, if by "typical" we mean "Lebesgue almost every". It would be nice to have a way to select, among the ergodic measures present, which measure is natural in the sense of reflecting the dynamics of typical orbits.

   Ergodic measures that are absolutely continuous with respect to Lebesgue measure are a good candidate for a class of "natural" ergodic measures. If μ_f is an ergodic measure

---

[5]$\{f_n\}$ is strongly convergent to f if $\|f_n - f\| \xrightarrow{n \to \infty} 0$ (in the L^1 norm).

[6]$\{f_n\}$ is weakly convergent to f if $\phi(f_n - f) \xrightarrow{n \to \infty} 0$ for any bounded linear $\phi : L^1(X) \to \mathbb{R}$.

[7]$\{f_n\}$ is Césaro convergent to f if $\frac{1}{n} \sum_{k=1}^n \phi(f_k - f) \xrightarrow{n \to \infty} 0$ for every bounded linear $\phi : L^1(X) \to \mathbb{R}$.

with strictly positive density f (with respect to Lebesgue measure) then the statement "μ_f-almost every" implies "Lebesgue almost every".[8] Then the Birkhoff Ergodic Theorem says that μ_f describes the statistics of Lebesgue almost every orbit, hence μ_f is the dynamically relevant ergodic measure. Moreover, if such a natural measure exists then it is unique:

Theorem 2.9. *Let $\{S_t : t \in G\}$ be a nonsingular dynamical system on a measure space X. If $\{S_t\}$ is ergodic then $\{S_t\}$ has at most one invariant density with respect to μ. Furthermore, if $\{S_t\}$ has a unique strictly positive invariant density then $\{S_t\}$ is ergodic.*

Proof. See *e.g.* [74, Thm 4.2.2]; also [71, Prop 5.1.2]. The result follows from the fact that if ν_1, ν_2 are distinct ergodic measures then $\nu_1 \perp \nu_2$[9] [93, p. 94]. Thus ν_1, ν_2 cannot both be absolutely continuous, hence at most one of them can have a density. $\square$

However, absolute continuity is not an adequate criterion for the selection of a natural ergodic measure for a dissipative dynamical system. In a dissipative system, phase space volumes are contracted by the time evolution, typically onto a compact invariant set, or attractor [103, 104]. In this case the relevant ergodic measure is expected to be concentrated on a set of Lebesgue measure zero, and therefore will not be absolutely continuous. Thus, even for the dynamically relevant ergodic measure, the Birkhoff Ergodic Theorem gives a statement only about orbits originating on a set of Lebesgue measure zero.

Nevertheless, in physical experiments and computer simulations there is typically just one invariant measure—the so-called *physical* measure—that describes typical orbits of the system. The existence of such a measure motivates the following definition.

Definition 2.27 (SRB (Sinai-Ruelle-Bowen) measure). Let $\{S_t : t \in G\}$ be a dynamical system on a measure space X. Then μ is an SRB or physical measure for $\{S_t\}$ if for any bounded continuous $\phi : X \to \mathbb{R}$ and for all x in a set of positive Lebesgue measure,

$$\int \phi \, d\mu = \lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} \phi(S_k x) \qquad (\text{if } G = \mathbb{Z}_+) \tag{2.52}$$

---

[8] The relation "$\ll$" is transitive and reflexive, so the condition "$\mu \ll \nu$ and $\nu \ll \mu$" gives an equivalence relation $\mu \sim \nu$. Equivalent measures share the same null sets. All measures with positive density (*e.g.*, with respect to Lebesgue measure) are equivalent. Thus we seek "natural" measures in the equivalence class containing Lebesgue measure.

[9] The relation $\perp$ is the antithesis of absolute continuity. If $\mu \perp \nu$ then essentially the only sets on which μ does not vanish are those on which ν does, and vice versa. See *e.g.* [55].

or

$$\int \phi \, d\mu = \lim_{T \to \infty} \frac{1}{T} \int_0^T \phi(S_t x) \, dt \qquad (\text{if } G = \mathbb{R}_+). \qquad (2.53)$$

Thus an SRB measure is one for which the conclusion of the Birkhoff Ergodic Theorem 2.7 holds, not just on a set of positive μ measure, but on a set of positive *Lebesgue* measure. Thus an SRB measure describes the statistics of Lebesgue almost every orbit originating in some nontrivial set. By definition, existence of an SRB measure requires statistical regularity for all orbits in a set of positive Lebesgue measure. An SRB measure is necessarily invariant under S_t, as can be seen from the proof of Theorem 2.6. An SRB measure need not be ergodic [14].

Other candidates for the notion of "physical" measure have been given. For example, suppose a dynamical system with random perturbations of amplitude ε has a stationary measure μ_ε. The zero-noise limit ($\varepsilon \to 0$) of μ_ε, if it exists, is the *Kolmogorov measure* [104]. For some systems (*e.g.* Axiom A systems[10]) the Kolmogorov measure is known to coincide with the SRB measure. Because of this equivalence, a number of different definitions of SRB measure appear in the literature; see *e.g.* [14, 104]. The definition above is commonly preferred because it is motivated by physical considerations.

Existence of an SRB measure is a strong condition that, although quite natural to define, turns out to be very difficult to prove. Existence is known for Axiom A systems and C^2 flows with hyperbolic attractors [16, 102]. The notion of strange attractor, of which much has been made in chaos studies, has been defined as an attracting invariant set that supports an SRB measure that is mixing [18]. Proving the existence of such a measure is widely recognized as one of the most important outstanding problems in dynamical systems theory [40, 120]. Computer methods are showing promise in this direction, and have recently been used to show the existence of an SRB measure supported on the famous Lorenz attractor [78, 116, 117].

2.5   Dimensions and Lyapunov Exponents

In numerical studies of chaotic systems various numerical parameters—for example Lyapunov exponents, dimensions, and entropy—are frequently used to quantify the degree of "randomness" exhibited by typical orbits. The Birkhoff Ergodic Theorem and its gener-

---

[10]Axiom A systems are a fairly restrictive class of dynamical systems with strong chaotic properties. For details see *e.g.* [104]. Examples are Anosov flows, Smale's horseshoe map, and the solenoid [109].

alizations make it possible to define these quantities rigorously in terms of time averages along trajectories, and to establish that these quantities are identical for almost every trajectory with respect to the ergodic measure:

Lyapunov characteristic exponents are the time averages of the exponential rates of divergence of nearby trajectories, along orthogonal directions in the tangent space of a given orbit [12, 13, 92].

Dimensions help quantify the geometric structure of invariant sets (*e.g.* attractors) of dynamical systems. For sets that support an invariant measure, one can define the information dimension, correlation dimension, and Hausdorff dimension (see *e.g.* [104]) which quantify the average number of "independent directions" on the invariant set.

Entropy is a measure of the average rate of *information creation* along an orbit of a dynamical system (see *e.g.* [108]).

2.6 Reliability of Numerical Simulations

Because of sensitivity to initial conditions, orbits of chaotic dynamical systems cannot be reliably computed using finite precision arithmetic, as in computer simulations. A numerically computed pseudo-orbit approximating a true orbit typically loses any relation to the true orbit after only a few iterations of the system dynamics. The exact map $S : x \mapsto 2x$ mod 1 on $[0, 1]$ makes this point especially clear, since S effects a left-shift on the binary representation of x. That is,

$$S : \ 0.x_1x_2x_3\ldots \ \mapsto \ 0.x_2x_3x_4\ldots \tag{2.54}$$

where x_1, x_2, $\ldots$ are the digits of the binary representation of x. If the computer stores 16 binary digits in its representation of x, then the numerical approximation of any particular orbit becomes meaningless after only 16 iterations of the dynamics.

The situation therefore seems hopeless when we come to using computer simulations to approximate statistical properties of dynamical systems, since very long and reasonably accurate pseudo-orbits are required. Surprisingly, this does not pose a problem for sufficiently well-behaved systems.

Definition 2.28 (pseudo-orbit; shadowing). Let $\{S_n = S^n : n \in \mathbb{Z}_+\}$ be a dynamical system on a metric space X.

- A pseudo-orbit $\{x_n : a \leq n \leq b\}$ is an α-*pseudo-orbit* if

$$d\big(x_{n+1}, S(x_n)\big) < \alpha \quad \forall a \leq n \leq b. \tag{2.55}$$

- A point $y \in X$ β-*shadows* $\{x_n : a \leq n \leq b\}$ if

$$d\big(S^n(y), x_n\big) < \beta \quad \forall n. \tag{2.56}$$

One can think of an α-pseudo-orbit as an orbit of $\{S_n\}$ that is perturbed by an amount smaller than α after each iteration; this models, for example, the round-off error in a numerical simulation. A shadow orbit $\{S^n(y)\}$ is a *true* orbit that is approximated by $\{x_n\}$ within accuracy β at all times.

Definition 2.29 (shadowing property). A dynamical system $\{S_n\}$ has the *shadowing property* if $\forall \beta > 0$, $\exists \alpha > 0$ such that every α-pseudo-orbit is β-shadowed by a point y.

For a system with the shadowing property, a pseudo-orbit (*e.g.*, one found by numerical simulation) is always an accurate representation of *some* nearby true orbit, although perhaps not the particular orbit one was trying to approximate. It is known that Anosov systems[11] have this property [71]. More generally a smooth dynamical system has the shadowing property in a neighborhood of a hyperbolic invariant set; this is the celebrated Shadowing Lemma [51, 71].

It can be shown [11] that for any uniformly continuous functional ϕ and a given $\delta > 0$, one can choose a sufficiently small $\alpha > 0$ so that if $\{x_n\}$ is an α-pseudo-orbit then, $\forall n > 0$,

$$\left| \frac{1}{n} \sum_{k=1}^{n} \phi(x_k) - \frac{1}{n} \sum_{k=1}^{n} \phi(S^k y) \right| < \delta. \tag{2.57}$$

where $\{S^k(y)\}$ β-shadows $\{x_k\}$. Thus the time average of ϕ along the pseudo-orbit differs by less than δ from the time average of ϕ along some true orbit $\{S^k y\}$, hence time averages computed from numerical simulations are in principle reliable.

---

[11] An Anosov system is one for which the entire phase space is a hyperbolic set [71].

The considerations above apply only to systems with the shadowing property. Unfortunately this property is difficult to prove except under fairly restrictive conditions, as in Anosov systems. Nevertheless, numerical studies suggest that numerical simulations of many non-Anosov systems are statistically reliable as well [11]; the shadowing property provides a plausible mechanism that might account for this phenomenon.

Chapter 3

Elements of an Ergodic Theory of Delay Equations

3.1   Introduction

The central aim of this thesis is to apply probabilistic concepts, *e.g.*, from ergodic theory, to the dynamics of delay differential equations (DDEs). Before such a project can proceed, a number of foundational questions must be addressed. For instance,

- In what sense can a DDE be interpreted as a dynamical system, *i.e.*, with a corresponding evolution semigroup?

- What is the phase space for such a system?

- What semigroup of transformations governs the phase space dynamics of a DDE?

The dynamical systems approach to delay equations is well established, and provides standard answers to these questions. This theory is discussed in Sections 3.3–3.4 below. As it happens the phase space for a delay differential equation is infinite dimensional, which complicates matters considerably.

An ergodic approach to DDEs will require a theory of probability in infinite dimensional spaces. The elements of this theory are discussed in Sections 3.5–3.6. Naturally, there are technical and interpretational difficulties with doing probability in infinite dimensions. Indeed, the available mathematical machinery proves to be inadequate to deal with some of the problems that arise. This has important consequences for the remainder of the thesis.

3.2   Delay Differential Equations

Delay differential equations, which are representative of the more general class of *functional* differential equations [53], take a great variety of forms. Delay equations having multiple time delays, time-dependent delays, and even continuous distributions of delays all arise in mathematical models of evolutionary systems [39]. To simplify matters we will restrict our attention to delay equations of the form

$$x'(t) = f\big(x(t), x(t - \tau)\big), \tag{3.1}$$

where $x(t) \in \mathbb{R}^n$, $f : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n$, and $\tau > 0$ is a single fixed time delay. Despite this restriction, the class of delay equations of the form (3.1) provides more than a sufficient arena for the considerations that follow.

3.2.1 Definition of a solution

By a solution of the DDE (3.1) we mean the following: if for some $t_0 \in \mathbb{R}$ and $\beta > t_0$, the function $x : [t_0 - \tau, \beta] \to \mathbb{R}^n$ satisfies (3.1) for $t \in [t_0, \beta]$, then we say x is a solution of (3.1) on $[t_0 - \tau, \beta]$. If $\phi : [t_0 - \tau, t_0] \to \mathbb{R}^n$ and x is a solution that coincides with ϕ on $[t_0 - \tau, t_0]$, we say x is a solution through (t_0, ϕ).[1]

Because equation (3.1) is autonomous (*i.e.*, the right-hand side does not depend explicitly on t), it is invariant under time translation. That is, if $x(\cdot)$ is a solution then, for any $T \in \mathbb{R}$, $x(\cdot + T)$ is also a solution. Consequently the choice of initial time t_0 is arbitrary, and for the sake of convenience we can take $t_0 = 0$. Let $C = C([-\tau, 0])$ be the space of continuous functions from $[-\tau, 0]$ into $\mathbb{R}^n$. Then if $\phi \in C$ and $x : [-\tau, \beta] \to \mathbb{R}^n$, we say x is a solution of (3.1) with *initial function* ϕ, or simply a solution through ϕ, if x is a solution through $(0, \phi)$.

Our intention to consider delay equations as models of deterministic processes imposes some constraints on the equations it makes sense to consider. In order that a given DDE describes an evolutionary process at all, we require the existence of solutions, at least for some subset of initial functions $\phi \in C$. Moreover, since ergodic theory is concerned largely with asymptotic properties, we require *global* existence, *i.e.* existence of solutions on the entire interval $[-\tau, \infty)$. To ensure that the process is deterministic we require that, for given $\phi \in C$, the solution through ϕ should be unique.

These constraints are in fact met under fairly mild restrictions on the right-hand side of (3.1). The following section presents the basic results of this theory that we will require. For further details of the existence and uniqueness theory for delay equations see for example [39, 53].

3.2.2 Existence and uniqueness theory

The existence and uniqueness theory for delay equations can be derived from the more general theory of functional differential equations. Since we intend to consider only equations of the form (3.1) we will not make use of the full generality available. Nevertheless, the more general theory leads to a presentation that is simpler and also benefits from an analogy with similar results in the theory of ordinary differential equations.

In the following, $C([a, b])$ denotes the Banach space of continuous functions from $[a, b]$

---

[1] The difficulty that arises if ϕ does not satisfy (3.1) at t_0 is avoided if x' is interpreted as a right-hand derivative.

into $\mathbb{R}^n$, equipped with the sup norm, and C denotes the space $C([-\tau, 0])$. If $\beta > 0$ and $x \in C([-\tau, \beta])$, let $x_t \in C$ be defined by

$$x_t(s) = x(t+s), \quad s \in [-\tau, 0]. \tag{3.2}$$

Suppose $F : \mathbb{R} \times C \to \mathbb{R}^n$. Then the equation

$$x'(t) = F(t, x_t), \tag{3.3}$$

where x' denotes the right-hand derivative, is called a *retarded functional differential equation* (RFDE) on D.

Equation (3.3) provides for a very general dependence of $x'(t)$ on the retarded values of x on the interval $[t - \tau, t]$. The DDE (3.1) is a special case, with F given by

$$F(t, \phi) = f\big(\phi(0), \phi(-\tau)\big). \tag{3.4}$$

A solution x of (3.3) is defined in the same manner as for the DDE (3.1) (see the preceding section). The basic results on existence and uniqueness of solutions are presented below. We omit the proofs, which are somewhat lengthy and technical, and refer the reader to [39, 53] for details.

Theorem 3.1 (Local existence). *Suppose $\Omega \subset \mathbb{R} \times C$ is open, and $F : \Omega \to \mathbb{R}^n$ is continuous. If $(t_0, \phi) \in \Omega$ then there is a solution of the RFDE (3.3) through (t_0, ϕ).*

Proof. See *e.g.* [53, p. 43]. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ $\square$

Corollary 3.2. *If $f : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n$ is continuous then for any $\phi \in C$ there is a solution of (3.1) through ϕ.*

Proof. Equation (3.1) is equivalent to (3.3) with F defined as in (3.4). Since f is continuous, F is a composition of continuous functions and is therefore continuous on $\mathbb{R} \times C$. The conclusion follows from Theorem 3.1. $\qquad\qquad\qquad\qquad\qquad\qquad$ $\square$

Definition 3.1 (Lipschitzian). Let $\Omega \subset \mathbb{R} \times C$ and $F : \Omega \to \mathbb{R}^n$. F is *Lipschitzian* in ϕ (on Ω) if, for some $K \geq 0$,

$$\big|F(t, \phi_1) - F(t, \phi_2)\big| \leq K|\phi_1 - \phi_2| \quad \forall (t, \phi_1), (t, \phi_2) \in \Omega. \tag{3.5}$$

Theorem 3.3 (Uniqueness). *Suppose $\Omega \subset \mathbb{R} \times C$ is open, $F : \Omega \to \mathbb{R}^n$ is continuous, and $F(t, \phi)$ is Lipschitzian in ϕ on every compact set in Ω. If $(t_0, \phi) \in \Omega$ then there is a unique solution of the RFDE (3.3) through (t_0, ϕ).*

Proof. See *e.g.* [53, p. 44]. □

Corollary 3.4. *If $f : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n$ is Lipschitzian then for any $\phi \in C$ there is a unique solution of (3.1) through ϕ.*

Proof. Equation (3.1) is equivalent to (3.3) with F defined as in (3.4). Since f is Lipschitzian, $F(t, \phi)$ is Lipschitzian in ϕ [39, p. 292]. The conclusion follows from Theorem 3.3. □

Theorem 3.5 (Global existence). *Suppose $F : [t_0, \beta) \times C \to \mathbb{R}^n$ is continuous, and that $F(t, \phi)$ is Lipschitzian in ϕ. If*

$$\big| F(t, \phi) \big| \leq M(t) + N(t) |\phi|, \quad \forall t \in [t_0, \beta), \ \phi \in C, \tag{3.6}$$

for some positive continuous functions M, N on $[t_0, \beta)$, then there is a unique solution of the RFDE (3.3) through ϕ on $[t_0 - \tau, \beta)$.

Proof. See *e.g.* [39, p. 308]. □

Corollary 3.6. *If $f : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n$ is Lipschitzian and satisfies*

$$\big| f(u, v) \big| \leq N(t) \max\{|u|, |v|\} \tag{3.7}$$

for some positive continuous function $N(t)$ on $[0, \beta)$, then $\forall \phi \in C$ there is a unique solution of (3.1) through ϕ on $[-\tau, \beta)$.

Proof. With F defined as in equation (3.4), equation (3.1) is equivalent to the RFDE (3.3). Since f is Lipschitzian it is also continuous, hence F is continuous and $F(t, \phi)$ is Lipschitzian in ϕ [39, p. 292]. Since

$$
\begin{aligned}
\big| F(t, \phi) \big| &= \big| f\big(\phi(0), \phi(-\tau)\big) \big| \\
&\leq N(t) \max\{|\phi(0)|, |\phi(-\tau)|\} \\
&\leq N(t) \sup_{s \in [-\tau, 0]} |\phi(s)| \\
&= N(t) |\phi|,
\end{aligned}
\tag{3.8}
$$

the conclusion follows from Theorem 3.5.   □

The hypotheses of the preceding theorems can be considerably weakened. In particular, continuity of F can be weakened to continuity of $F(t, x_t)$ with respect to t for every continuous x. The Lipschitz condition on F can also be weakened to a local Lipschitz condition, for which it suffices that f in the DDE (3.1) have continuous first partial derivatives [39, p. 261].

In the following we also require continuous dependence of solutions on initial conditions, for which the following theorem gives a result analogous to that for ordinary differential equations.

Theorem 3.7 (Continuous dependence). *Suppose x is a solution through (t_0, ϕ) of the RFDE (3.3) and that it is unique on $[t_0 - \tau, \beta]$. If $\{(t_n, \phi_n)\} \subset \mathbb{R} \times C$ is a sequence such that $(t_n, \phi_n) \to (t_0, \phi)$ as $n \to \infty$, then for all sufficiently large n every solution x_n through ϕ_n exists on $[t_n - \tau, \beta]$, and $x_n \to x$ uniformly on $[t_0 - \tau, \beta]$.*

Proof. See Theorem 2.2 of [53, p. 43], which proves a stronger result giving continuous dependence on (t_0, ϕ, F). The version given here is a simpler special case.   □

3.2.3   Method of steps

Existence and uniqueness for a given DDE can sometimes be shown indirectly, by representing the DDE as a sequence of ordinary differential equations. This approach, known as the method of steps [39], also furnishes a method of finding explicit solutions.

The DDE problem

$$
\begin{aligned}
x'(t) &= f\big(x(t), x(t - \tau)\big), \quad t \geq 0 \\
x_0 &= \phi,
\end{aligned}
\tag{3.9}
$$

when restricted to the interval $[0, \tau]$, becomes the *ordinary* differential equation

$$
x'(t) = f\big(x(t), x_0(t - \tau)\big) \equiv g_0\big(t, x(t)\big), \quad t \in [0, \tau],
\tag{3.10}
$$

since $x_0 = \phi$ is a known function. Under suitable hypotheses on g, existence and uniqueness of a solution of this equation (hence a solution of (3.9)) on $[0, \tau]$ can be established. Denoting this solution by x_1 and restricting equation (3.9) to the interval $[\tau, 2\tau]$, we

obtain the ordinary differential equation

$$x'(t) = f\big(x(t), x_1(t-\tau)\big) \equiv g_1\big(t, x(t)\big), \quad t \in [\tau, 2\tau], \tag{3.11}$$

for which we can again establish existence and uniqueness of a solution x_2.

Proceeding inductively, considering equation (3.9) as an ordinary differential equation on a sequence of intervals $[n\tau, (n+1)\tau]$, it is sometimes possible to show existence and uniqueness of a solution of the DDE on $[-\tau, \infty)$. This approach is especially simple if $f\big(x(t), x(t-\tau)\big) = f\big(x(t-\tau)\big)$ is independent of $x(t)$, since existence and uniqueness of x_{n+1} then requires only integrability of x_n, hence almost-everywhere continuity of ϕ is sufficient to guarantee existence and uniqueness of a solution on $[-\tau, \infty)$.

3.3   Delay Equation as a Dynamical System

As noted above, to make sense of the DDE (3.1) as prescribing the evolution of a deterministic system, we require that for any ϕ in C, a solution x through ϕ exists and is unique on $[-\tau, \infty)$. We will also require that $x(t)$ depend continuously on ϕ. Thus from now on we will simply assume that sufficient conditions are satisfied to guarantee that these constraints are met, for example the hypotheses of Corollary 3.6.

By a simple rescaling of the time variable in (3.1), the delay time τ can be made equal to 1. For the sake of convenience, and wherever it seems natural, we will assume in the following that such a rescaling has been done. Thus the generic DDE "initial data problem" we consider is the following,

$$\begin{aligned} x'(t) &= f\big(x(t), x(t-1)\big), \quad t \geq 0 \\ x(t) &= \phi(t), \quad t \in [-1, 0], \end{aligned} \tag{3.12}$$

where $\phi \in C = C([-1, 0])$.

Since equation (3.12) specifies the evolution of a variable $x(t) \in \mathbb{R}^n$, it might seem that such a DDE could be regarded simply as a dynamical system on $\mathbb{R}^n$. However, $x(t)$ alone is inadequate as a "phase point", since the initial value $x(0)$ does not provide sufficient information to determine a solution. Indeed, in order that the right-hand side $f\big(x(0), x(-1)\big)$ is well defined for all $t \in [0, 1]$, initial data consisting of values of $x(t)$ for $t \in [-1, 0]$ must be supplied, as in (3.12).

In general, to determine a unique solution of (3.12) for all $t \geq T$, it is necessary and

sufficient to know the retarded values of $x(t)$ for all t in the "delay interval" $[T - 1, T]$. Thus equation (3.1) can only be considered as a dynamical system if the phase point at time t contains information about the solution $x(t)$ on the entire interval $[t - 1, t]$. That this is in fact sufficient to define a dynamical system corresponding to the initial value problem (3.12) is shown in the following.

As before, let C be the Banach space of bounded continuous functions from $[-1, 0]$ into $\mathbb{R}^n$, supplied with the sup norm. For each $t \geq 0$ define a transformation $S_t : C \to C$ by

$$(S_t\phi)(s) \equiv x_t(s) = x(t + s), \quad s \in [-1, 0], \tag{3.13}$$

where $x(t)$ is the solution of (3.12). Then we have (*cf.* [53]):

Theorem 3.8. *The family of transformations S_t, $t \geq 0$, defined by equation (3.13), is a semidynamical system on C (cf. Definition 2.2). That is,*

(a) $S_0\phi = \phi \quad \forall \phi \in C$,

(b) $(S_t \circ S_{t'})\phi = S_{t+t'}\phi \quad \forall \phi \in C$, $t_1, t_2 \geq 0$,

(c) $(t, \phi) \mapsto S_t(\phi)$ is continuous $\forall t \geq 0$.

Proof. (a) is obvious from equations (3.12) and (3.13), since by definition

$$(S_0\phi)(s) = x(s) = \phi(s), \quad s \in [-1, 0]. \tag{3.14}$$

To prove (b), let $x(t)$ be the solution of (3.12). Then by definition of S_t,

$$\begin{aligned}
(S_{t+t'}\phi)(s) &= x(t + t' + s), \\
(S_{t'}\phi)(s) &= x_{t'}(s) = x(t' + s).
\end{aligned} \tag{3.15}$$

By translation invariance of the DDE, $x(t + t')$ is also a solution, corresponding to the initial function $x_{t'}$. Thus by definition of S_t,

$$(S_t x_{t'})(s) = x(t + t' + s). \tag{3.16}$$

Combining with (3.15), we have

$$(S_t \circ S_{t'})\phi = S_t x_{t'} = S_{t+t'}. \tag{3.17}$$

(c) follows from Theorem 3.7, which asserts continuity of $(t_0, \phi) \mapsto x \in C([t_0 - 1, \beta])$. Since x_{t+t_0} is just the restriction of x to $[t + t_0 - 1, t + t_0] \subset [t_0 - 1, \beta]$, we also have continuity of $(t_0, \phi) \mapsto x_{t+t_0} = S_{t+t_0}\phi$. $\qquad \Box$

In terms of the evolution semigroup just defined, the initial data problem (3.12) can be written as an abstract initial value problem,

$$\begin{cases} x_t = S_t(x_0) \\ x_0 = \phi. \end{cases} \tag{3.18}$$

In accordance with the terminology of Section 2.1 we call the function x_t the "phase point" at time t of the corresponding DDE (3.12). The trajectory

$$\{x_t = S_t\phi : t \geq 0\} \tag{3.19}$$

is a continuous curve in the function space C. The relationship of the DDE solution $x(t)$ to this trajectory is simple, and is given by

$$x(t) = x_t(0). \tag{3.20}$$

That is, the solution $x(t)$ "reads off" the right endpoint of the phase point x_t. In other words, $x(t)$ can be interpreted as the projection of x_t under the map $\pi : C \to \mathbb{R}^n$ defined by $\pi(x_t) = x_t(0)$.

The action of S_t has a simple geometric interpretation. Since $(S_t\phi)(\cdot) = x(t + \cdot)$, S_t consists of a translation of the solution x followed by a restriction to the interval $[-1, 0]$. Figure 3.1 illustrates this action, together with the relationship of the state x_t to the DDE solution $x(t)$.

The phase space of the dynamical system $\{S_t\}$ (and hence the phase space of the corresponding DDE (3.12)), being the space of continuous functions on the interval $[-1, 0]$, is infinite dimensional. The infinite dimensionality of the phase space for delay equations complicates their analysis dramatically, and as we will see, it proves to be a serious barrier to developing a probabilistic treatment.
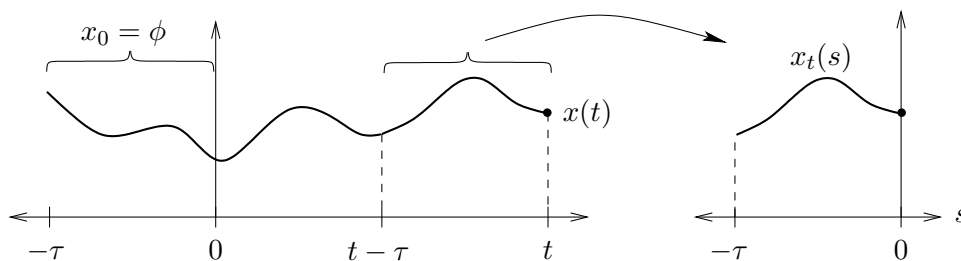
Figure 3.1: Relationship between the solution $x(t)$ of the delay equation (3.12) and the phase point $x_t \in C$ of the corresponding dynamical system.

3.4   Representations of the Semigroup

The previous section illustrates how the delay equation (3.12) can be viewed as a dynamical system in an infinite dimensional phase space. However, the definition of the corresponding semigroup S_t (*cf.* equation (3.13)) is given implicitly in terms of a particular solution of the DDE. That is, in order to evaluate $S_t(\phi)$ we must have the corresponding solution x of (3.12) already in hand. Consequently, the present definition of S_t provides little insight as to how S_t operates as a transformation on C. It is illuminating to consider alternative representations of the semigroup, with a view to making its action on the phase space of continuous functions more transparent. Sections 3.4.1–3.4.3 explore some of the possibilities.

3.4.1   Explicit solution map

For some delay equations it is possible to write the semigroup operator S_t explicitly as an iterated map on C. For example, consider delay equations of the form (3.12) where f is linear in its first argument, *viz.*,

$$x'(t) = -\alpha x(t) + g\big(x(t-1)\big). \tag{3.21}$$

Using the notation introduced in the previous section, let $x_t(\cdot) = x(t + \cdot) \in C$ represent the phase point at time t for the corresponding dynamical system $\{S_t\}$. It is simplest to construct just the time-one map $S = S_1$ for this system, for which the only relevant

phase points are those at discrete times,

$$x_n(\cdot) = x(n + \cdot), \quad n = 0, 1, 2, \ldots. \tag{3.22}$$

In this notation the DDE (3.21) becomes

$$x'_{n+1}(s) = -\alpha x_{n+1}(s) + g\big(x_n(s)\big), \tag{3.23}$$

an ordinary differential equation for x_{n+1} in terms of the (known) previous phase point x_n. Its solution defines the time-one map $S : x_n \mapsto x_{n+1}$. Explicitly (*cf.* [41]),

$$(Su)(s) = u(0)e^{-\alpha(s+1)} + \int_{-1}^{s} e^{\alpha(t-s)} g\big(u(t)\big)\, dt, \quad s \in [-1, 0]. \tag{3.24}$$

This map gives a representation of the DDE (3.21) as a discrete-time dynamical system,

$$x_{n+1} = Sx_n. \tag{3.25}$$

Together with an initial function $x_0 = \phi$, this system defines a trajectory $\{x_n : n = 0, 1, \ldots\} \subset C$. From this trajectory, the solution $x(t)$ of the original delay equation (3.21) can be recovered according to equation (3.22).

It is interesting that, although $\{S_t\}$ is a continuous-time dynamical system, a trajectory of the discrete-time system $\{S^n : n \in \mathbb{Z}_+\}$ is sufficient to construct the solution $x(t)$ of the original DDE for all $t > 0$. The continuous-time family of maps S_t does not provide any additional information about the solution, so it is reasonable to treat the DDE as a truly discrete-time dynamical system in C. This observation does not depend on the special form of the DDE (3.21), as the same conclusion can be drawn for the more general DDE (3.12) where, although we do not have an explicit formula for the time-one map, S can be defined using the method of steps (*cf.* page 46).

3.4.2 Initial boundary value problem

The semigroup of operators S_t on C also has a representation in terms of the solution of an initial boundary value problem. Again, this representation may be more illuminating than an implicit definition of S_t in terms of solutions $x(t)$ of the DDE, and it applies even if an explicit solution map like that in the previous section cannot be obtained.

If the right-hand side f of the DDE (3.12) is continuous, then the solution x is

continuously differentiable on $(0, \infty)$. Therefore, at least for $t > 1$, the phase point $x_t(s)$ is differentiable in both t and s. It follows that x_t, considered as a function

$$u(s, t) = x_t(s) = x(t + s), \tag{3.26}$$

satisfies the partial differential equation

$$\frac{\partial u(s, t)}{\partial t} = \frac{\partial u(s, t)}{\partial s}, \quad s \in [-1, 0], \ t > 1. \tag{3.27}$$

The DDE (3.12) implies a boundary condition on u,

$$\left.\frac{\partial u(s, t)}{\partial s}\right|_{s=0} = f\big(u(0, t), u(-1, t)\big). \tag{3.28}$$

Equations (3.27)–(3.28), together with initial data

$$u(s, 0) = \phi(s), \tag{3.29}$$

constitute an initial boundary value problem describing the evolution of x_t. If the initial function ϕ is differentiable and satisfies the "splicing condition"

$$\phi'(0) = f(\phi(0), \phi(-1)), \tag{3.30}$$

then the domain of (3.27) can be extended to $[-1, 0] \times [0, \infty)$.[2]

This initial boundary value problem evidently has a solution $u(s, t) = x(t + s)$. Moreover, since any solution $u(s, t)$ of (3.27)–(3.29) defines a solution x of the corresponding DDE via (3.26), uniqueness of the solution of the DDE implies uniqueness for the initial boundary value problem.

3.4.3   Abstract differential equation

The connection between the initial boundary value problem (3.27)–(3.29) and the evolution semigroup $\{S_t\}$ can be made more explicit by re-interpreting the initial boundary value problem as an "abstract Cauchy problem", *i.e.*, an initial value problem on the function space C.

---

[2]If the splicing condition does not hold, $u(s, t)$ can be interpreted as a weak solution on $[-1, 0] \times [0, \infty)$ [10].

Recall that the phase point x_t for the DDE (3.12) is given by

$$x_t(s) = x(t+s), \tag{3.31}$$

where x is a solution of the DDE. The phase space trajectory corresponding to this solution is a continuous curve $\{x_t : t \geq 0\} \subset C$. Under suitable hypotheses on the function f in the DDE (3.12) this curve is differentiable. That is, the time derivative

$$\frac{d}{dt}x_t = \lim_{h \to 0} \frac{x_{t+h} - x_t}{h} \tag{3.32}$$

exists, where the limit is taken in the strong sense of convergence in C. In fact, we have:

Theorem 3.9. *Suppose that $f : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n$ is continuous, and that $\phi \in C$ is continuously differentiable and satisfies the splicing condition (3.30). Let x be the corresponding solution of the DDE (3.12), and let $x_t(s) = x(t+s)$, $s \in [-1, 0]$. Then the strong derivative $\frac{d}{dt}x_t$ exists and satisfies $\frac{d}{dt}x_t = \mathcal{A}x_t$ where $\mathcal{A} : C \to C$ is given by $\mathcal{A} : u \mapsto u'$.*

Proof.

$$\lim_{h \to 0} \left\| \frac{u_{t+h} - u_t}{h} - \mathcal{A}u_t \right\|$$
$$= \lim_{h \to 0} \sup_{s \in [-1,0]} \left| \frac{x(t+h+s) - x(t+s)}{h} - x'(t+s) \right|. \tag{3.33}$$

Then by the mean value theorem,

$$\lim_{h \to 0} \left\| \frac{u_{t+h} - u_t}{h} - \mathcal{A}u_t \right\| = \lim_{h \to 0} \sup_{s \in [-1,0]} |x'(t+c(h)+s) - x'(t+s)| \tag{3.34}$$

for some $|c(h)| < |h|$. Under the given hypotheses, x is continuously differentiable on $[-\tau, \infty)$. Thus x' is continuous and hence uniformly continuous on any closed interval containing $[t-1, 1]$, so the limit above is zero. $\qquad\square$

Thus, at least for continuously differentiable initial functions satisfying the splicing condition, the trajectory $\{x_t : t \geq 0\}$ corresponding to $x_0 = \phi$ is differentiable and satisfies

$$\frac{d}{dt}x_t = \mathcal{A}x_t. \tag{3.35}$$

The operator $\mathcal{A}$ is called the *infinitesimal generator* of the semigroup [74, Ch. 7]. Equa-

tion (3.35) can be regarded as an "abstract differential equation", with the mapping $u \mapsto \mathcal{A}u$ acting like a vector field on C. Together with the initial condition $x_0 = \phi$, it constitutes an "abstract Cauchy problem", or initial value problem, on C. The DDE semigroup $\{S_t\}$ furnishes a solution of this initial value problem, $x_t = S_t(\phi)$, $t \geq 0$.[3] Thus the action of the semigroup S_t can be interpreted as carrying the initial function ϕ along a trajectory in C that is an integral curve of the differential equation (3.35).

The theory of abstract differential equations such as (3.35) is most fully developed in the case where the corresponding semigroup turns out to be a family of *linear* operators. This is the case, for instance, when the DDE (3.12) is linear [53, p. 194]. Then there is an existence and uniqueness theory for initial value problems satisfying differential equations like (3.35) (the Hille-Yosida Theorem and its relatives [26, Ch. 2]). For our purposes a detailed discussion of this theory is unwarranted. Instead, in the following we merely sketch its relevance to linear delay equations.

Note that the infinitesimal generator $\mathcal{A}$ is not defined on all of C, so that it is not strictly valid to consider $\mathcal{A}$ as a vector field on C. In fact, it is clear from the proof of Theorem 3.9 that $\mathcal{A}$ is defined only on the domain

$$D(\mathcal{A}) = \{u \in C : u' \in C \text{ and } u'(0) = f\big(u(0), u(-1)\big)\}, \tag{3.36}$$

However, $D(\mathcal{A})$ is dense in C [53, p. 194]. This, together with restrictions on $\mathcal{A}$ that are satisfied if the delay equation (3.1) is linear, implies that the initial value problem

$$\begin{aligned}
\frac{d}{dt}x_t &= \mathcal{A}x_t, \quad x_t \in D(\mathcal{A}), \\
x_0 &= \phi \in D(\mathcal{A}),
\end{aligned} \tag{3.37}$$

has a unique solution $\{x_t = S_t(\phi) : t \geq 0\}$ [36].

3.5   Perron-Frobenius Operator

Having determined how a delay differential equation defines a dynamical system, we are in a position to approach one of the fundamental problems posed in this thesis. That is,

---

[3]If ϕ does not satisfy the splicing condition, $x_t = S_t\phi$ can be interpreted as a mild solution of (3.35), *i.e.*, there is a sequence of functions $\phi_n \in C$, converging to ϕ, that do satisfy the splicing condition, such that $S_t\phi_n$ converges uniformly to $S_t\phi$ [10, 36].

given a system whose evolution is determined by a DDE

$$x'(t) = f\big(x(t), x(t-1)\big), \tag{3.38}$$

and whose initial phase point $\phi \in C$ is not known but is given instead by a probability distribution over all possible initial states, how does the probability distribution for the phase point evolve in time? Alternatively, we could consider the statistical formulation of the problem: given a large ensemble of independent systems, each governed by (3.38), and whose initial functions are distributed according to some density over C, how does this ensemble density evolve in time? It is of particular interest to characterize those probability distributions that are invariant under the action of the DDE.

In a sense, the answer to this problem is simple and is provided by the Perron-Frobenius operator formalism, introduced in Chapter 2. Suppose the initial distribution of phase points is described by a probability measure μ on C. That is, the probability that the initial function ϕ is an element of a given set $A \subset C$ (correspondingly, the fraction of the ensemble whose initial functions are elements of A), is given by $\mu(A)$. Then, after evolution by time t, the new distribution is described by the measure ν given by

$$\nu = \mu \circ S_t^{-1}, \tag{3.39}$$

provided S_t is a measurable transformation on C (*cf.* Section 3.6.1). That is, after time t the probability that the phase point is an element of $A \subset C$ is $\nu(A) = \mu(S_t^{-1}(A))$. If the initial distribution of states u can be described by a density $\rho(u)$ with respect to some measure λ, then after time t the density will have evolved to $P_t\rho$, where the Perron-Frobenius operator P_t corresponding to S_t is defined by

$$\int_A P_t\rho(u)\, d\lambda(u) = \int_{S_t^{-1}(A)} \rho(u)\, d\lambda(u) \quad \forall\ \lambda\text{-measurable } A \subset C. \tag{3.40}$$

Equations (3.39)–(3.40) might appear to answer the problem of the evolution of probability measures for DDEs. However, they amount only to a formal answer—essentially a symbolic restatement of the problem. In fact, everything that is specific to a given DDE is contained in the symbol S_t^{-1}.

Although the DDE can be expressed in terms of an evolution semigroup, in none of its representations (*cf.* Section 3.4) is there an apparent way to invert the resulting transformation S_t. It is almost certain that such an inversion will be non-trivial, since

solutions of delay equations frequently cannot be uniquely extended into the past [39], so that S_t will not have a unique inverse. That is, S_t^{-1} may have numerous branches that need to be accounted for when evaluating $S_t^{-1}(A)$ in the Perron-Frobenius equation (3.40). This is a serious barrier to deriving a closed-form expression for the Perron-Frobenius operator P_t.

There are other subtle issues raised by equations (3.39)–(3.40). The most apparent difficulty is that the integrals in (3.40) are over sets in a function space, and it is not immediately apparent how such integrals can be carried out. More fundamentally, it is not clear what family of measures we are considering, and in particular what subsets $A \subset C$ are measurable (*i.e.*, what is the relevant σ-algebra on C?). Also, in equation (3.40) what should be considered a natural choice for the measure λ with respect to which probability densities are to be defined? For that matter, does it make sense to talk about probability densities over the function space C? These issues are explored in the following section.

3.6   Probability in Infinite Dimensional Spaces

Any discussion of an ergodic theory of delay equations will require a theory of measure and integration on function spaces. In particular we need to discuss probability measures on the space C of continuous functions on the interval $[-1, 0]$, since this is a natural phase space for the DDE (3.12). Colloquially speaking, we need to make precise the somewhat non-intuitive notion of selecting a *random function* from C.

Measure-theoretic probability provides a sufficiently abstract setting to accomplish this. Recall from Chapter 2 that we can represent a random variable $x \in X$ by its associated probability measure μ, with the interpretation that for a given subset $A \subset X$, $\mu(A)$ expresses the probability that $x \in A$. To ensure consistency with the axioms of probability, we cannot assign a probability to just any subset of X. Rather, μ must be defined on an appropriate σ-algebra—a collection of so-called *measurable sets* (*cf.* Section 2.2.1). So choosing an appropriate σ-algebra on C is a necessary starting point.

3.6.1   Appropriate σ-algebra

In real Euclidean spaces, the notion of measure derives from our physical intuition of length, area, volume, and their generalizations to higher dimensions. Thus line segments in one dimension, and rectangles in two dimensions, are natural candidates for inclusion in the σ-algebras of choice for these spaces. The natural choice of σ-algebra would seem

to be the smallest σ-algebra that contains all such sets—that is, the σ-algebra generated by these sets. This is the so-called *Borel σ-algebra*, which happens to coincide with the smallest σ-algebra that contains all open subsets.

A similar approach leads to a natural choice of σ-algebra for infinite dimensional spaces such as C. That is, we take the Borel σ-algebra generated by the metric topology on C. With this choice, many important subsets of C such become measurable, *i.e.* we can assign meaningful probabilities to them:

- any open set in C

- $\{u \in C : u(s) \in (a,b) \ \forall s \in [-1,0]\}$; $a, b \in \mathbb{R}$

- any ϵ-ball $B_\epsilon(v) = \{u \in C : \|u - v\| < \epsilon\}$; $v \in C$, $\epsilon \in \mathbb{R}$

Besides achieving the measurability of important sets for analysis, there is a more fundamental reason for choosing the Borel σ-algebra. Recall that studying the evolution of probability measures under a given transformation makes sense only if the transformation is *measurable*. Therefore, for our study of DDEs it is essential to choose a σ-algebra on which the semigroup S_t defined by equation (3.13) is measurable. The following establishes that the Borel σ-algebra accomplishes this.

Theorem 3.10. *For every $t \geq 0$, $S_t : C \to C$ (cf. equation (3.13)) is a measurable transformation on the Borel σ-algebra on C.*

Proof. S_t is continuous on C, by Theorem 3.8(c), hence measurable, by Theorem 2.2. $\square$

It may be the case that the Borel σ-algebra on C is in fact not the most natural choice in the context of a probabilistic approach to DDEs. Certainly, as demonstrated in the following sections, measures on the Borel sets of infinite dimensional spaces do not behave as we might like. However, in light of the preceding considerations, from now on we will consider only measures defined on the Borel sets of C.

3.6.2 Densities

Recall that if a measure μ is absolutely continuous with respect to a measure λ, then it can be expressed as

$$\mu(A) = \int_A \rho \, d\lambda, \tag{3.41}$$

where the integral is in the sense of Lebesgue, and $\rho \in L^1(X, \lambda)$ is the density of μ with respect to λ. Furthermore, any Lebesgue integrable function $\rho \in L^1(X, \lambda)$ with

$$\int \rho \, d\lambda = 1 \tag{3.42}$$

uniquely determines an absolutely continuous measure μ (*cf.* Section 2.2.4).

Since the relations (3.41)–(3.42) require only a σ-algebra and a measure λ on X, they apply equally well in the more abstract setting of infinite dimensional spaces such as C. That is, if C is equipped with a σ-algebra $\mathcal{A}$ and measure λ on $\mathcal{A}$, then the function space $L^1(C) = L^1(C, \mathcal{A}, \lambda)$ is unambiguously defined (*cf.* Section 2.2.3), and any functional $\rho \in L^1(C)$ determines an absolutely continuous measure on C. However, in this context the intuitive appeal of densities is lacking: it is impossible to draw the graph of such a density functional. Even imagining a density on C seems beyond the power of one's imagination.

The analytical benefits of using densities also appear to be quite limited in infinite dimensional spaces. The connection between measure theory and calculus in finite dimensions owes much to the theory of integration, notably the fundamental theorem of calculus and other theorems that facilitate calculations with integrals. There is no adequate theory of integration on function spaces that makes it possible to evaluate integrals like (3.41) on C (*cf.* comments in [79]). A notable exception to this is Wiener measure, although this does not seem to be adequate for our purposes; see Section 3.6.5, page 62.

Even allowing that a more powerful theory of integration may be available in the future, there remain some inherent difficulties with using densities to specify probability measures on infinite dimensional spaces. Equation (3.40) for the evolution of a probability density ρ under the action of a semigroup S_t is valid only if S_t is non-singular. That is, pre-images under S_t of λ-measure-zero sets must have λ-measure zero. It turns out to be difficult to guarantee this. In fact, on an infinite dimensional space, every absolutely continuous measure fails to remain absolutely continuous under arbitrary translations [122]. That is, for *any* measure λ on C, there is some $v \in C$ for which the translation

$$T : u \mapsto u + v \tag{3.43}$$

is singular (in the measure-theoretic sense), and hence does not map densities to densities. If even translations do not lead to well-defined density evolution, there is little hope of studying delay equations with density functionals.

3.6.3   Lack of a "natural" measure on C

As if the preceding did not complicate matters enough, if we are to work with densities on C there remains the problem of choosing a basic measure λ with respect to which densities are to be defined (*cf.* equation (3.41)). This too turns out to be problematic.

In real Euclidean spaces we are accustomed to taking Lebesgue measure as the "natural" measure with respect to which densities are defined. That is, "a random number distributed uniformly on the interval $[0, 1]$" means "a random variable on $[0, 1]$ distributed according to Lebesgue measure". Why is Lebesgue measure—of all possible measures—the gold standard for representing the concept of "uniformly distributed"?

The property of Lebesgue measure that selects it uniquely as the natural measure on Euclidean spaces is its translation invariance. Given a random variable x uniformly distributed on $[0, 1]$, we expect that adding a constant a to x should result in a new random variable, $x + a$, that is uniformly distributed on $[a, a+1]$, at least according to what seems to be the common intuitive notion of "uniformly distributed". More generally, a random variable uniformly distributed on any set in $\mathrm{I\!R}^n$ should remain uniformly distributed if translated by a constant vector. Formally, the measure λ on $\mathrm{I\!R}^n$ that encapsulates uniform distribution should satisfy

$$\lambda(A) = \lambda(A + a), \quad \forall a \in \mathrm{I\!R}^n, \ \forall \text{ measurable } A. \tag{3.44}$$

Another way to say this is that λ is invariant under the translation group

$$T_a : x \mapsto x - a. \tag{3.45}$$

That is,

$$\lambda = \lambda \circ T_a^{-1} \quad \forall a \in \mathrm{I\!R}^n. \tag{3.46}$$

Equation (3.46) uniquely defines the Borel measure λ on the Borel σ-algebra on $\mathrm{I\!R}^n$(which agrees with Lebesgue measure on the Borel sets). This is a specific instance of Haar measure: every locally compact topological group (*e.g.*, the translation group just considered on $\mathrm{I\!R}^n$) has a unique group-invariant measure on the Borel σ-algebra, called the Haar measure, that is non-zero on any open set [73, p. 313].

In light of these considerations, in choosing a natural measure on C it seems reasonable to seek a translation-invariant measure. After all, we would like that a uniformly distributed ensemble of functions in the unit ball in C should remain uniformly distributed

under translation by any function in C. Unfortunately the existence of a Haar measure on C is not guaranteed, since C is not locally compact.[4] In fact the situation is worse than that, as the following theorem demonstrates.

Theorem 3.11. *Let X be an infinite dimensional separable Banach space. If λ is a non-zero translation-invariant measure on the Borel sets of X, then $\lambda(A) = \infty$ for every open $A \subset X$.*

Proof. (after [58].) Let $B \subset X$ be an open ball of radius $\epsilon > 0$, and suppose $\lambda(B) > 0$ is finite. Because X is infinite dimensional, there is an infinite sequence B_i, $i = 1, 2, \ldots$ of disjoint open balls $B_i \subset B$, each of radius $\epsilon/4$ (*cf.* the proof of Theorem 4.3.3 in [44, p. 134]). Because $\{B_i\}$ is a countable disjoint collection with $\cup_i B_i \subset B$, we have

$$\lambda(B) \geq \lambda(\cup_i B_i) = \sum_{i=1}^{\infty} \lambda(B_i), \tag{3.47}$$

where $\lambda(B_i) = \lambda(B_1)$ by translation invariance. Since $\lambda(B)$ is finite, this implies that $\lambda(B_i) = 0 \,\forall i$. Separability of X implies that X can be covered by a countable collection of $\epsilon/4$-balls, each of which we have just shown must have measure 0. Hence $\lambda(X) = 0$, a contradiction. $\qquad\square$

Since we expect any reasonable measure to be non-zero at least on some open sets, we can conclude that translation-invariance will not suffice to select a natural measure on C.

Aside from making the definition of densities on C ambiguous, the absence of a natural measure undermines one of the most important concepts in ergodic theory. Recall from Section 2.4.4 that an SRB measure μ for a dynamical system S_t on X is one such that, for any functional $\varphi \in L^1(X)$,

$$\lim_{T \to \infty} \frac{1}{T} \int_0^T \varphi(S_t x)\, dt = \int \varphi\, d\mu \tag{3.48}$$

for Lebesgue almost every x. Thus the time average of φ along almost every trajectory is equal to the spatial average of φ weighted with respect to μ. Because $\varphi(x)$ represents an arbitrary observable of the system, and μ encapsulates the asymptotic statistical behavior

---

[4]A normed vector space is locally compact iff it is finite dimensional [73, p. 39].

of $\varphi(x(t))$ on almost every orbit of the system, is it widely accepted that an SRB measure is *the* relevant physical measure—the one that nature reveals to the experimentalist.

The notion of "almost every" in (3.48) is always unquestioningly taken to mean "Lebesgue almost every". As we have seen, for infinite dimensional systems, and for delay equations in particular, we have no natural analog of "Lebesgue almost every", since there is no translation invariant measure to take the place of Lebesgue measure.

That this ambiguity emerges at all is somewhat amusing, since the notion of SRB measure was introduced on purely *physical* grounds. The very definition of SRB measure requires that we make precise the notion of "physically relevant"—but for DDEs this leads to considerations in the decidedly non-physical setting of infinite dimensional geometry, where it appears to be an inherently ambiguous term.

3.6.4 Genericity and prevalence

Without a natural measure on C to characterize a physically relevant notion of "almost every", the definition of SRB measure for a delay differential equation is problematic. One way out of this dilemma is to introduce a notion of almost every that does not depend on a specific measure, such as the topological concept of *genericity*. A property is said to be generic if it holds on a residual set, that is a countable intersection of open dense sets. The complement of a residual set is a set of "first category", hence first category sets are topological analogs of sets of measure zero. Although genericity provides one way to quantify the notion of almost every in infinite dimensional spaces, it lacks the probabilistic interpretation that we would like to have in the context of ergodic theory. More importantly, even in $\mathbb{R}^n$ residual sets can have measure zero [58], so using genericity in the definition of SRB measure would be inconsistent with the accepted definition for finite dimensional systems.

A more promising alternative is a translation-invariant probabilistic notion of almost every called *prevalence* [58]:

Definition 3.2. Let X be a Banach space equipped with its Borel σ-algebra $\mathcal{A}$. A Borel set $A \in \mathcal{A}$ is called *shy* if there is a measure μ on $\mathcal{A}$ such that

- $0 < \mu(U) < \infty$ for some compact $U \subset X$, and

- $\mu(A + x) = 0 \;\; \forall x \in X$.

A is called *prevalent* if it is the complement of a shy set.

Roughly speaking, a set is shy if for some nontrivial measure on X, every translate of A has measure zero. Two key properties make prevalence an attractive candidate for a notion of "almost every" appropriate to a definition of SRB measure for infinite dimensional systems (for proofs see [58]):

1. If A is prevalent then any translate of A is prevalent; *i.e.* prevalence is a translation-invariant property.

2. $A \subset \mathbb{R}^n$ is shy if and only if A has Lebesgue measure zero.

The first property means prevalence is a natural or physical notion of almost every in the sense discussed in the previous section. The second property guarantees that, in finite dimensions, a property holds on a prevalent set if and only if it holds on a set of positive Lebesgue measure. Thus for finite dimensional systems the definition of SRB measure (*cf.* Definition 2.27, page 35) is unchanged is we substitute "a prevalent set" for "a set of positive Lebesgue measure". The novelty and significance of this alternative definition is that it applies equally well to *infinite* dimensional systems.

Tools for proving shyness and prevalence are developed in [58]. The following interesting results have been proved (here we use "almost every" in the sense of "in a prevalent set"):

- If X is infinite dimensional then every compact subset of X is shy.

- Almost every element of C is nowhere differentiable.

- For $1 \le p \le \infty$ almost every C^p map on $\mathbb{R}^n$ has the property that all of its periodic points are hyperbolic.

- Almost every $f \in C([0,1], \mathbb{R})$ satisfies $\int_0^1 f(x)\,dx \ne 0$.

We are unaware of any applications of prevalence to the concept of SRB measure. This appears to be a promising direction for further investigation.

3.6.5   Wiener measure

As already noted, an adequate theory of integration on infinite dimensional spaces in lacking. Such a theory is needed if we are to further develop the Perron-Frobenius operator formalism to characterize the evolution of densities for DDEs, which requires a theory of

integration of functionals on the space C. This difficulty also arises in [79], in the context of a different approach to the evolution of densities for DDEs.

However, there is a notable exception worth mentioning. There is one probability measure (or family of measures) on a function space, called Wiener measure, for which there is a substantial theory of integration [68]. This measure plays an important role in quantum field theory (see *e.g.* [105]), and is central to the theory of stochastic differential equations [74, Ch. 11].

Let

$$C_0 = \{u \in C([0,1], \mathbb{R}^n) : u(0) = 0\}. \tag{3.49}$$

A Brownian motion[5] is a stochastic process that generates a random path or "random function" $w \in C_0$ such that for a given $t \in [0,1]$, $w(t)$ has Gaussian probability density [79]

$$\rho(x_1, \ldots, x_n) = \frac{1}{(\sqrt{2\pi t})^n} \exp\left[-(x_1^2 + \cdots + x_n^2)/(2t)\right]. \tag{3.50}$$

Then, roughly speaking, Wiener measure μ_w assigns to a given subset $A \subset C_0$ a measure equal to the probability that a Brownian motion generates an element of A.

With Wiener measure it is possible to prove strong ergodic properties (*e.g.* exactness) for a certain class of partial differential equations [17, 99, 100, 101]. The success of these investigations, together with the considerable machinery that has been developed around the Wiener measure, suggests that Wiener measure might be a good choice for the measure of integration in the study of other infinite dimensional systems such as delay equations. However, in contrast with the quantum field equations and the PDEs mentioned above, the dynamical system $\{S_t\}$ corresponding to a delay equation does not leave the space C_0 invariant. That is, we cannot study S_t on C_0 alone. Thus Wiener measure does not seem to be adequate for our purposes. Nevertheless, an approach based on Wiener measure might be still possible, and this suggests a fruitful avenue for further investigation.

3.7 Conclusions

In this chapter we have developed a framework in which an ergodic treatment of delay differential equations might be developed. This provides a setting and terminology that

---

[5]A Brownian motion is a continuous-time analog of a random walk starting at the origin. See *e.g.* [74].

will be needed for our subsequent discussions of the ergodic properties of DDEs.

However, as far as the possibilities for the rigorous development of an ergodic theory of DDEs are concerned, the main results of this chapter are somewhat pessimistic. The picture that emerges is a characterization of DDEs as infinite dimensional dynamical systems on the phase space C of continuous functions on the interval $[-1, 0]$. With this characterization, an ergodic theory of DDEs is possible in principle. In such a theory the mathematical objects of primary interest are probability measures on C. This entails a theory of measure and probability on infinite dimensional spaces. As we have seen, the foundations of this theory run aground on a number of technical and interpretational difficulties including the following.

- Non-invertibility of the evolution semigroup $\{S_t\}$.

- Likely singularity of S_t with respect to most measures on C.

- Lack of an adequate theory of integration on infinite dimensional spaces.

- Non-existence of a natural (*i.e.* translation-invariant) measure on C.

- Ambiguity in the definition of SRB measure for infinite dimensional systems.

Some of these difficulties (*e.g.*, with integration in infinite dimensions) appear to require significant new mathematical tools that are beyond the scope of this thesis. Others (*e.g.*, with the choice of a natural measure on C and the definition of SRB measure) are simply ambiguities that arise when dynamical systems theory developed with only finite-dimensional systems in mind is carried over to an infinite dimensional setting. Nevertheless, in the absence of criteria by which these ambiguities could be resolved, we must content ourselves with having carefully discussed the available alternatives.

In light of the foregoing the following chapters focus less on ergodic formalism, in order to pursue more fruitful lines of inquiry. In the next chapter we turn to the practical problem of computing the evolution of probability densities for the state $x(t) \in \mathbb{R}^n$ rather than an abstract phase point in C.

Chapter 4

Density Evolution for Delay Equations

Figure 4.1: An ensemble of 100 solutions of the Mackey-Glass equation (4.2), correspond-
ing to an ensemble of constant initial functions with values uniformly distributed on the
interval $[0.3, 1.3]$.

In light of the results of the previous chapter, a comprehensive treatment of delay
equations within the ergodic theory of dynamical systems is out of reach. Nevertheless, a
probabilistic treatment is feasible if the dynamical systems formalism is abandoned, and
this is the approach taken in the present chapter.

This chapter again considers systems that can be modeled by a DDE of the form

$$x'(t) = f\big(x(t), x(t-1)\big), \quad t \geq 0, \quad x(t) \in \mathbb{R}^n, \tag{4.1}$$

where without loss of generality the "delay time" has been scaled to one. In contrast with
the previous chapter, we now take the point of view of an experimentalist, interpreting
this equation as prescribing the evolution of an observable quantity $x(t) \in \mathbb{R}^n$, rather
than a phase point in an abstract function space. Thus we imagine an experimental
setting in which an ensemble of independent systems evolves according to (4.1), and
seek a probabilistic description of this ensemble in terms of the evolution of the density
$\rho(x, t)$ of the ensemble of solution values $x(t)$. (Alternatively we can think of $\rho(x, t)$ as
a *probability* distribution that quantifies our uncertain knowledge of the state of a single
system governed by (4.1).)

Figure 4.1 illustrates the problem we wish to consider. The figure depicts an ensemble of 100 solutions[1] of the Mackey-Glass equation [86],

$$x'(t) = -\alpha x(t) + \beta \frac{x(t-1)}{1 + x(t-1)^{10}},$$

$$\alpha = 2, \quad \beta = 4, \quad n = 10,$$

(4.2)

which was originally introduced to model oscillations in neutrophil populations. This equation has been the subject of much study because of the variety of dynamical phenomena it exhibits. The parameter values chosen here correspond to the existence of a chaotic attractor. The solutions shown in Figure 4.1 correspond to an ensemble of 100 constant initial functions, whose values are uniformly distributed on the interval $[0.3, 1.3]$. From the density of solution curves on this graph, one can form an idea of the density $\rho(x,t)$ of solution values $x(t)$ at any given time t. For example, at $t = 3$ solutions are particularly dense near $x = 0.45$, $x = 0.6$, $x = 0.9$ and $x = 1.2$.

The main question this chapter attempts to answer in the context of Figure 4.1 is the following. If the density $\rho_0(x) = \rho(x; 0)$ of constant initial values x at $t = 0$ is known, how can one determine (*i.e.*, predict) the density $\rho(x; t)$ for times $t > 0$? The following section develops an appropriate framework for the analysis of this problem. In Sections 4.2–4.4 this framework is used to develop various approaches to the evolution of densities. Each of these approaches is essentially independent of the others, but they are presented in an order that takes advantage of the interplay between them. Analytical techniques are considered in Sections 4.2 and 4.5; Sections 4.3 and 4.4 focus on computational approaches.

---

[1]Numerical solutions were computed here using the solver DDE23 [107]

4.1   Probabilistic Framework

Although equation (4.1) describes the evolution of a finite-dimensional vector $x(t) \in \mathbb{R}^n$, the space of initial conditions for this equation—the space C of continuous functions from $[-1, 0]$ into $\mathbb{R}^n$—is infinite dimensional. This is the main source of difficulty in our attempts so far to develop a probabilistic approach. To make the problem more intuitive as well as mathematically tractable, it is necessary to somehow restrict the dimension of the set of "allowable" initial conditions.

4.1.1   Restricted initial value problem

The simplest such restriction would be to allow only initial functions from some n-dimensional subspace of C, such as the space of constant initial functions (*cf.* Figure 4.1). Given the plethora of different finite-dimensional subspaces available in C, this restriction might seem excessive. However, there is a physical justification for such a restriction, since in an experimental setting the initial preparation of the ensemble is typically in an equilibrium state. In this case we expect each of the units in the ensemble will have a constant initial history, and thus the subspace of constant initial functions is naturally selected by the experiment.

There are a number of other ways that an experimental setting might naturally select a finite-dimensional set of allowable initial functions for (4.1). Since our hypothetical ensemble has presumably not been in existence for all time, there must be some process by which the individual initial histories are generated. Since this process cannot be described by the governing delay equation, it is reasonable to posit some other process that does govern the initial histories $x(t)$ on the interval $[-1, 0]$, and to describe this process by an *ordinary* differential equation.

For convenience, let the initial time for the DDE (4.1) be $t = 1$ rather than $t = 0$. Thus we consider the DDE

$$x'(t) = f\big(x(t), x(t-1)\big), \quad t \in [1, \infty) \tag{4.3}$$

with initial function specified on the interval $[0, 1]$. Then the corresponding initial value

problem can be written

$$x'(t) = \begin{cases} g\big(x(t)\big) & \text{if } t \in [0,1) \\ f\big(x(t), x(t-1)\big) & \text{if } t \geq 1, \end{cases} \tag{4.4}$$

$$x(0) = x_0,$$

where $g : \mathbb{R}^n \to \mathbb{R}^n$ describes the process by which the initial function is determined by the initial value $x(0)$. Of course we require that $x' = g(x)$ with initial value $x(0) = x_0$ have a unique solution[2] on $[0,1]$, so that (4.4) describes a deterministic process on $[0, \infty)$. In (4.4) the set of allowable initial functions selected by g is just the set of solutions of the ODE $x' = g(x)$ on $[0,1]$. This is a one-dimensional set parametrized by the initial value x_0. For example the space of constant initial functions corresponds to $g = 0$.

Even if the initial function is not determined by an ODE, we still would like the set of allowable initial functions to be parametrized by the initial value x_0, since specifying an ensemble of initial values x_0 then determines an ensemble of initial functions, and hence an ensemble of solutions of the given DDE. Thus in the most general case we wish to consider the DDE (4.3) with an initial function specified by

$$x(t) = \psi(t, x_0), \quad t \in [0,1] \tag{4.5}$$

for some function $\psi : \mathbb{R} \times \mathbb{R}^n \to \mathbb{R}^n$ which should have the following properties:

- The function $t \mapsto \psi(t, x_0)$ (*i.e.* the initial function corresponding to the initial value x_0) is continuous.

- The mapping $x_0 \mapsto \psi(t, x_0)$ is a measurable, non-singular transformation of $\mathbb{R}^n$, so if x_0 is distributed with density ρ_0 then the density of $x(t) = \psi(t, x_0)$ is well defined for each $t \in [0,1]$ (*cf.* Section 2.3.2, page 23).

- $\psi(0, x_0) = x_0$, so the parameter x_0 defines the initial value $x(0)$.

Every such ψ determines a particular one-parameter family of allowable initial functions in C. If $\psi(t, x_0)$ is the solution map for an ordinary differential equation $x' = g(x)$ (*i.e.*, the function $t \mapsto \psi(t, x_0)$ is the solution of the ODE with $x(0) = x_0$), then it satisfies

---

[2]*e.g.*, it suffices to have g bounded and continuously differentiable [39].

the conditions above. For example the family of constant initial functions corresponds to $\psi : (t, x) \mapsto x$, which in turn corresponds to $g = 0$.

Having parametrized the set of allowable initial functions according to a particular function ψ, the initial value problem corresponding to (4.3) becomes

$$
\begin{aligned}
x'(t) &= f\big(x(t), x(t-1)\big), \quad t \geq 1 \\
x(t) &= \psi(t, x_0), \quad t \in [0, 1], \\
x(0) &= x_0.
\end{aligned}
\tag{4.6}
$$

Under suitable mild restrictions on f (*cf.* Chapter 3), for each $x_0 \in \mathbb{R}^n$ this problem uniquely determines the evolution of $x(t)$ for $t \in [0, \infty)$.

In the following we restrict our attention to systems in which the initial function is determined by an ordinary differential equation. Thus the remainder of this chapter is concerned with probabilistic approaches to the initial value problem (4.4) which we will call the "augmented DDE", as distinguished from the corresponding DDE (4.3) with no restriction on the set of allowable initial functions.

4.1.2   Perron-Frobenius operator

Let $S_t : \mathbb{R}^n \mapsto \mathbb{R}^n$ be the solution map for the augmented DDE (4.4). That is,

$$
S_t : x_0 \mapsto x(t), \quad t \in [0, \infty),
\tag{4.7}
$$

where $x(t)$ is the (presumed unique) solution of (4.4). If an ensemble of initial values x_0 is specified with density ρ_0, then the evolution of this density under the action of S_t is given, in principle, by the corresponding Perron-Frobenius operator $P_t : L^1(\mathbb{R}^n) \to L^1(\mathbb{R}^n)$ (*cf.* Section 2.3.2). This operator carries the initial density ρ_0 to the density $P_t \rho_0$ at time t, and is defined by the relation

$$
\int_A P_t \rho_0(x) \, dx = \int_{S_t^{-1}(A)} \rho_0(x) \, dx \quad \forall \text{ Borel } A \subset \mathbb{R}^n.
\tag{4.8}
$$

Recall that P_t is well defined only if S_t is a measurable, nonsingular transformation. In fact measurability is guaranteed because for each t, S_t is a continuous map on $\mathbb{R}^n$ (this follows from continuity with respect to initial conditions for both the ODE and the DDE, *cf.* Theorem 3.10 page 57). However, non-singularity of S_t is not guaranteed for

all t—indeed, Section 4.2 presents a counter-example.

Note that the family of transformations $\{S_t : t \geq 0\}$ is not a semigroup. Consequently, neither is the family of Perron-Frobenius operators $\{P_t\}$. This is not a consequence of restricting the allowable set of initial functions, but rather comes from viewing the DDE as specifying an evolution in $\mathbb{R}^n$ (rather than the function space C). In $\mathbb{R}^n$ the augmented DDE (4.4) is non-autonomous, in that an explicit time dependence appears in the term $x(t-1)$ (which acts as a forcing term). This destroys the time-invariance required by the semigroup property. In short, the value of $x(t)$ at a particular time is not sufficient to uniquely determine its subsequent evolution—an obvious consequence of delayed dynamics.

The absence of the semigroup property for S_t and P_t has important consequences. For instance, it is not possible to express $P_{t+t'}$ as a composition $P_t \circ P_{t'}$. With the semigroup property, to find P_n for any integer n it suffices to find P_1 and then express $P_n = (P_1)^n$. Without the semigroup property this construction fails, and finding P_t for arbitrarily large t becomes far less trivial.

The remainder of this chapter is concerned with the evolution of densities for the augmented DDE (4.4). This amounts to finding the corresponding Perron-Frobenius operator P_t. Sections 4.2 and 4.5 are concerned with finding an analytical formula for P_t. Sections 4.3 and 4.4 present numerical approaches to approximating $P_t \rho_0$ for given initial densities ρ_0.

4.2 Explicit Solution Map

For some delay equations it is possible to find an explicit formula for the Perron-Frobenius operator defined by equations (4.4) and (4.7)–(4.8). This can be accomplished by first finding an explicit formula for the transformation $S_t : x_0 \mapsto x(t)$, which requires that the general solution to the given DDE be found. Equation (4.8) is then used to derive a formula for P_t. The following examples illustrate this procedure.

Example 4.2.1. Consider the linear DDE

$$x'(t) = \alpha x(t-1), \quad t \geq 1, \tag{4.9}$$

with the set of allowable initial functions on $[0, 1]$ restricted to constant functions, *i.e.*,

$$x(t) = x_0, \quad t \in [0, 1], \tag{4.10}$$

with x_0 distributed according to a given initial density ρ_0. Define the family of solution maps $\{S_t : t \geq 0\}$ by

$$S_t : x_0 \mapsto x(t), \tag{4.11}$$

where $x(t)$ is the solution of (4.9)–(4.10). Since the DDE does not depend explicitly on $x(t)$, the method of steps (*cf.* Section 3.2.3 page 46) reduces to iterating the following integral for $n = 1, 2, \ldots$,

$$x(t) = x(n) + \int_n^t \alpha x(s - 1)\, ds, \quad t \in [n, n + 1]. \tag{4.12}$$

Thus we obtain

$$S_t(x_0) = \begin{cases} x_0 & t \in [0, 1] \\ (\alpha t - \alpha + 1)x_0 & t \in [1, 2] \\ (\frac{1}{2}\alpha^2 t^2 - 2\alpha^2 t + \alpha t + 2\alpha^2 - \alpha + 1)x_0 & t \in [2, 3] \\ \vdots & \\ \beta_n(t)x_0 & t \in [n, n + 1] \end{cases} \tag{4.13}$$

$$= \beta(t)x_0,$$

where, from equation (4.12), $\beta_n(t)$ is a polynomial of degree n. Recall that the Perron-Frobenius operator corresponding to S_t is defined by

$$\int_A P_t\rho_0(x)\, dx = \int_{S_t^{-1}(A)} \rho_0(x)\, dx. \tag{4.14}$$

Taking $A = [0, x]$ we have

$$S_t^{-1}(A) = \begin{cases} [0, x/\beta(t)] & \text{if } \beta(t) > 0 \\ [x/\beta(t), 0] & \text{if } \beta(t) < 0, \end{cases} \tag{4.15}$$

and equation (4.14) becomes

$$\int_0^x P_t \rho_0(s)\, ds = \begin{cases} \displaystyle\int_0^{x/\beta(t)} \rho_0(s)\, ds & \text{if } \beta(t) > 0 \\[2ex] \displaystyle\int_{x/\beta(t)}^0 \rho_0(s)\, ds & \text{if } \beta(t) < 0. \end{cases} \tag{4.16}$$

Differentiating on both sides yields the explicit formula

$$(P_t \rho_0)(x) = \rho(x, t) = \frac{1}{|\beta(t)|} \rho_0\left(\frac{x}{\beta(t)}\right). \tag{4.17}$$

Notice that S_t is non-singular (hence P_t is well defined) if and only if $\beta(t) \neq 0$, which does not necessarily hold for all t. In particular, for any $\alpha \leq -1/2$ there is a time $t_* = 1 - 1/\alpha \in [1, 2]$ at which $\beta(t_*) = 0$ and therefore

$$S_{t_*}(x) = 0 \quad \forall x. \tag{4.18}$$

That is, all solutions of (4.9)–(4.10) pass through 0 at $t = t_*$. In general this occurs whenever $\beta(t) = 0$. At these times S_t is singular and P_t is undefined, though it is clear that the ensemble of solutions is described by a point mass concentrated at $x = 0$. In such cases it is possible to give the interpretation $P_t \rho_0 \to \delta$ (the Dirac delta function) as $\beta(t) \to 0$ (*cf.* [89] and [74, p. 398]).

In the previous example the Perron-Frobenius operator was easy to construct because the solution map S_t was one-to-one and easy to invert. The following example shows what happens for even slightly more interesting DDEs, where the solution map is not necessarily one-to-one.

Example 4.2.2. Consider the DDE

$$x'(t) = -x(t-1)^2, \quad t \geq 1, \tag{4.19}$$

where again we allow only constant initial functions on $[0, 1]$, so that

$$x(t) = x_0, \quad t \in [0, 1]. \tag{4.20}$$

With the solution map $S_t : x_0 \mapsto x(t)$ defined as before, the method of steps yields

$$
S_t(x) = \begin{cases}
x & t \in [0,1] \\
x - (t-1)x^2 & t \in [1,2] \\
x - (t-1)x^2 + (t^2 - 4t + 4)x^3 & \\
\quad + (-\frac{1}{3}t^3 + 2t^2 - 4t + \frac{8}{3})x^4 & t \in [2,3] \\
\ \vdots &
\end{cases}
\tag{4.21}
$$

For $t \in [0,1]$, we have simply $P_t \rho_0 = \rho_0$ since S_t is the identity transformation. For $t \in [1,2]$, take $A = (-\infty, x]$. Then if $x \le \frac{1}{4(t-1)}$,

$$
S_t^{-1}(A) = \left(-\infty, \frac{1 - \sqrt{1 - 4(t-1)x}}{2(t-1)} \right] \cup \left[\frac{1 + \sqrt{1 - 4(t-1)x}}{2(t-1)}, \infty \right),
\tag{4.22}
$$

and otherwise $S_t^{-1}(A) = \mathbb{R}$. Differentiating with respect to x on both sides of (4.14) then yields

$$
(P_t \rho_0)(x) = \frac{1}{\sqrt{1 - 4(t-1)x}} \left[\rho_0 \left(\frac{1 - \sqrt{1 - 4(t-1)x}}{2(t-1)} \right) \right.
$$
$$
\left. + \rho_0 \left(\frac{1 + \sqrt{1 - 4(t-1)x}}{2(t-1)} \right) \right]
\tag{4.23}
$$

if $x \le \frac{1}{4(t-1)}$, and $(P_t \rho_0)(x) = 0$ otherwise. Inverting S_t becomes extremely difficult for $t \in [2,3]$, and impossible for $t > 3$ (since it would require explicit roots of a fifth-order polynomial), so that it is not possible to derive an explicit formula for P_t.

In each of the preceding examples, the solution map S_t is a differentiable transformation on $\mathbb{R}$. For such transformations the corresponding Perron-Frobenius operator P_t can be expressed as

$$
(P_t \rho_0)(x) = \sum_{y \in S_t^{-1}\{x\}} \frac{\rho_0(y)}{|S_t'(y)|},
\tag{4.24}
$$

where $S_t'(y)$ is understood to mean $\frac{d}{dy} S_t(y))$. Indeed, this is frequently given as the *definition* of the Perron-Frobenius operator in studies of transformations of the real line (see *e.g.* [71, 93]). The results of examples 4.2.1–4.2.2 are in fact specific cases of this result.

For more complicated delay equations than those considered in the examples above, the difficulties in finding an explicit formula for the Perron-Frobenius operator are twofold:

- It can be difficult to determine the general solution, and hence the solution map S_t. This was made easier in the examples by lack of explicit dependence on $x(t)$, but in general the problem can be difficult.

- Determining pre-images S_t, as in equation (4.14) and (4.24), can be quite difficult.

The second of these difficulties is the more imposing, especially as S_t is generally not a one-to-one transformation. Even the simple example 4.2.2 results in a solution map S_t for which it is impossible to find an expression for the pre-image $S_t^{-1}\{x\}$ that occurs in (4.24). For these reasons it is not practical, in general, to construct Perron-Frobenius operators for DDEs by the direct means of first constructing the solution map.

For many applications, an analytical solution of the problem will not be possible, whereas a numerical approximation of the density $\rho(x,t) \equiv (P_t\rho_0)(x)$ might suffice. The following section presents a simple method of computing such an approximation, by directly simulating an ensemble of solutions.

4.3 Ensemble Simulation

The simplest approach to approximating $P_t\rho_0$ for particular initial densities ρ_0 is the "brute force" method of simulating an actual ensemble of solutions. That is, a large ensemble of initial values $\{x_0^{(1)}, \ldots, x_0^{(N)}\}$ is chosen at random from a distribution with density ρ_0. For each $x_0^{(i)}$ the corresponding solution $x^{(i)}(t) = S_t(x_0^{(i)})$ of (4.4) is constructed (numerically, or by some analytical formula). Then for any given t the density $\rho(x,t) = (P_t\rho_0)(x)$ is approximated by a histogram of the set of values $\{x^{(i)}(t), \ldots, x^{(N)}(t)\}$. With reference to Figure 4.1, this amounts to constructing a histogram of solution values $x(t)$ plotted above a given value of t.

Figure 4.2 shows the results of such a computation applied to the Mackey-Glass equation (4.2). As in Figure 4.1, the initial ensemble consists of constant functions (hence $g = 0$ in (4.4)), with values uniformly distributed on the interval $[0.3, 1.3]$ (all initial values in this interval are eventually attracted to the same chaotic attractor). That is,

$$\rho_0(x) = \begin{cases} 1 & \text{if } x \in [0.3, 1.3] \\ 0 & \text{otherwise.} \end{cases} \tag{4.25}$$
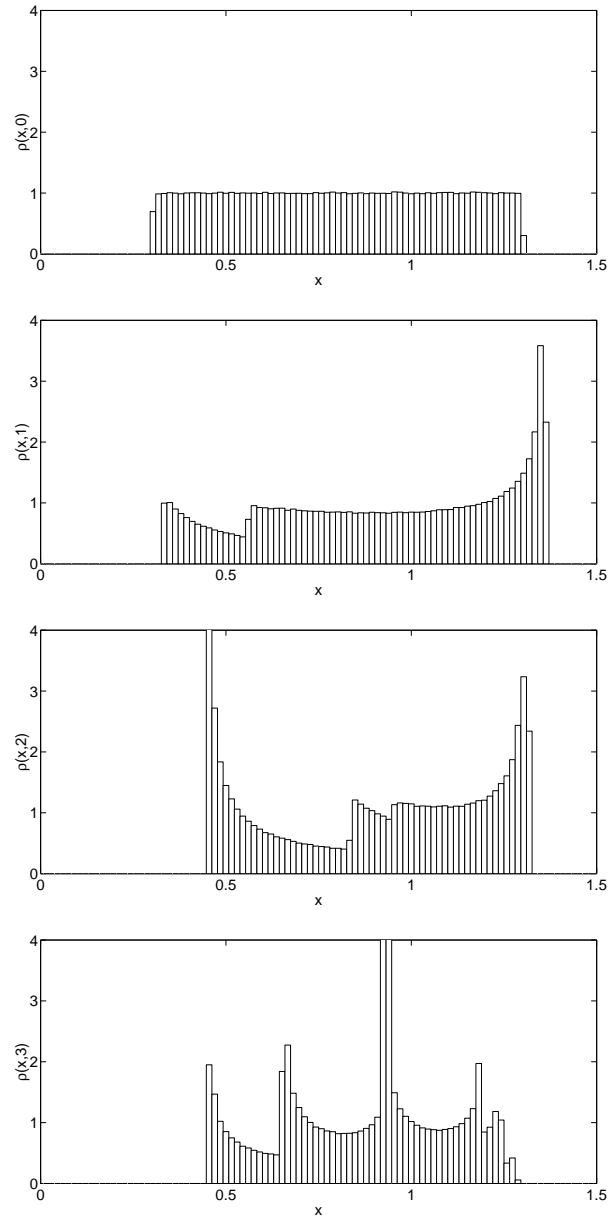
Figure 4.2: Normalized histograms of $x(t)$, at times $t = 0, 1, 2, 3$, for an ensemble of 10^6 solutions of the Mackey-Glass equation (4.2). The solutions correspond to an ensemble of constant initial functions with values uniformly distributed on $[0.3, 1.3]$ (*cf.* Figure 4.1).

For each of 10^6 initial values $x_0 = x^{(i)}(0)$ sampled from this distribution, an approximate solution $x^{(i)}(t)$ was computed numerically.[3] The sequence of graphs shown in Figure 4.2 depict the resulting histograms of the solution values $\{x^{(i)}(t) : i = 1, \ldots, 10^6\}$ at times $t = 0, 1, 2, 3$ (here we take the initial time for the DDE to be $t = 0$, so the initial function is specified on the interval $[-1, 0]$). The relationship between these densities and the corresponding ensemble of solutions shown in Figure 4.1 is apparent on brief inspection. For example, jump discontinuities in the densities occur at boundaries where the solutions in Figure 4.1 overlay one another, *e.g.*, near $x = 0.5$ at $t = 1$. High peaks, apparently integrable singularities in the density, occur where the ensemble of solutions in Figure 4.1 "folds over" on itself, *e.g.* near $x = 1.3$ at $t = 1$. Some of these features are artifacts resulting from discontinuities in the initial density, but others are not. See *e.g.* Figure 4.7 which illustrates the evolution of a Gaussian initial density.

An interesting property of the Mackey-Glass equation is revealed when the evolution of densities is carried to large times. Figure 4.3 shows a sequence of histograms constructed at times $t = 0, 20, 50, 100$, for the same ensemble considered in Figure 4.2. It appears from this sequence that the density $\rho(x, t)$ approaches a limiting density $\rho_*(x)$ as $t \to \infty$. That is, there appears to be an asymptotically stable invariant density for this system. The invariant density observed is in fact independent of the initial density. We will return to the problem of characterizing such invariant densities for DDEs in Chapter 5.

Note that convergence to the invariant density is relatively slow, for example compared to maps on the interval where statistical convergence occurs after only a few iterations of the Perron-Frobenius operator (*cf.* Figures 1.2–1.3). The behavior seen here is not typical of dynamical systems considered elsewhere, and may have implications for the statistical mechanics of systems with delayed dynamics, for example the neural ensemble encoding mechanism proposed in [88] where rapid statistical convergence plays an important role.

The brute force approach to densities has the tremendous advantage of being easy to implement—it requires only a method for numerically solving DDEs—and it is the obvious "quick and dirty" solution to the problem. However, it is a naïve approach, in that it provides no insight into the process by which $\rho(x, t)$ evolves. For example, the method offers only a heuristic explanation of the discontinuities and singularities that appear in Figure 4.2. Moreover, as shown below, constructing an accurate histogram can require millions of samples $\{x^{(i)}(t)\}$, hence millions of solutions of the DDE must

---

[3]Numerical solutions were computed using the solver DDE23 [107].
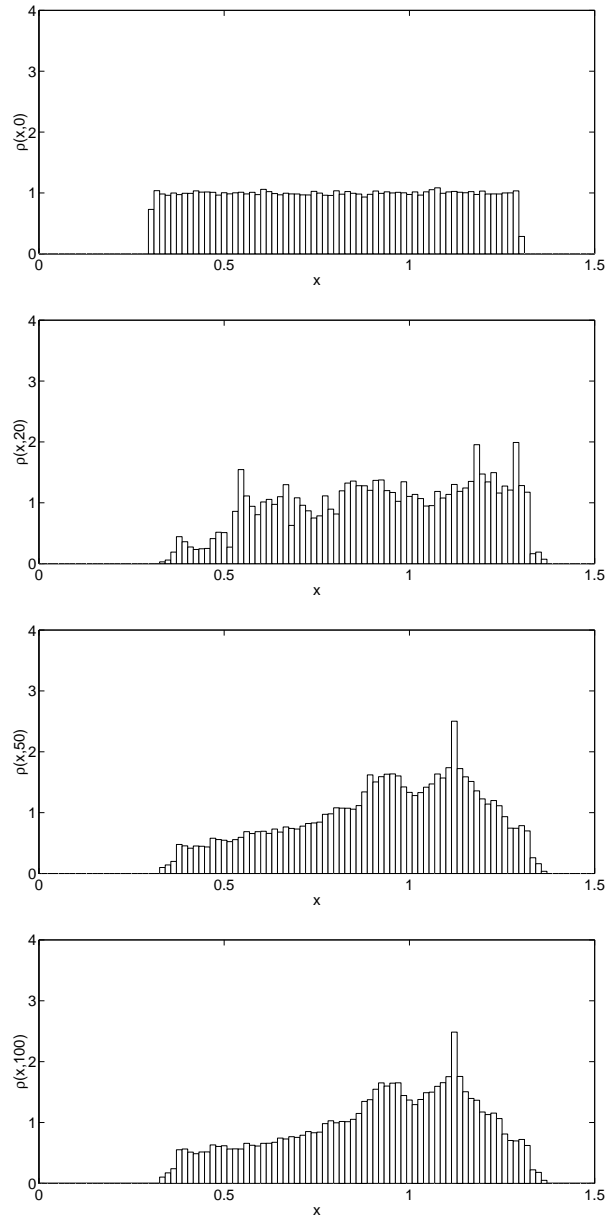
Figure 4.3: Normalized histograms of $x(t)$, at times $t = 0, 20, 50, 100$, for the same ensemble of solutions considered in Figure 4.2.

be computed. Especially when approximating the evolution of densities for large t, the amount of computation necessary can render the method practically useless.

Sampling requirements

In order for the histogram of a solution ensemble $\{x^{(1)}(t), \ldots, x^{(N)}(t)\}$ to provide an accurate approximation of the actual density $\rho(x, t)$, the ensemble must be sufficiently large. Suppose an interval $B \subset \mathbb{R}$ is one of the histogram bins. The height of the histogram on B is given by

$$y = \#\{x^{(i)} \in B\}, \tag{4.26}$$

i.e., the number of the $x^{(i)}$ that lie in B. The $x^{(i)}$ are independent samples from a distribution with density ρ, so for sufficiently large n the fraction

$$\frac{y}{N} = \frac{1}{N} \sum_{i=1}^{N} 1_B(x_i) \tag{4.27}$$

estimates (by the weak law of large numbers) the probability

$$p = \int_B \rho(x, t) \, dx \tag{4.28}$$

that a random number x selected from this distribution will lie in B.

The random variable y takes integer values between 0 and N, with binomial probability distribution

$$P(y) = \frac{N!}{y!(N-y)!} p^y (1-p)^{N-y}. \tag{4.29}$$

For sufficiently large N, $P(y)$ can be approximated by a Gaussian density with mean Np and standard deviation $\sqrt{Np(1-p)}$. Thus the quantity y/N (equation (4.27)) will be distributed with mean p and standard deviation

$$\sigma = \sqrt{\frac{p(1-p)}{N}}. \tag{4.30}$$

Equation (4.30) predicts $O(1/\sqrt{N})$ convergence of the "sample mean" y/N to the "population mean" p—a standard result in sampling and measurement theory [8, p. 36]. Suppose we wish y/N to approximate p within fractional error δ, with 95% confidence.

Then we require that $2\sigma < \delta p$, yielding (via equation (4.30)) the sampling requirement

$$N > \frac{4(1-p)}{\delta^2 p}. \tag{4.31}$$

Thus, for a moderately high-resolution histogram (say, with 100 bins, so p is on the order 10^{-2}), 95% confidence of accuracy within fractional error $\delta = 10^{-2}$ would require a sample of size

$$N \gtrsim \frac{4}{(10^{-2})^2 10^{-2}} = 4 \times 10^6. \tag{4.32}$$

For some delay equations, and particularly when densities are to be obtained for large t, this sample size requirement entails a prohibitive computational cost (for example, constructing Figure 4.3 required about 10 hours of computer time). In such situations, the brute force approach to density evolution becomes impractical. This motivates the following section, which develops a more efficient numerical method for computing the evolution of densities for DDEs.

4.4   Approximate Solution Map

The brute force approach to approximating densities is computationally expensive, so a more efficient numerical method is desirable. Developing such a method is the aim of the present section. To simplify the development, the method is presented only in the context of one-dimensional delay equations (*i.e.*, with solution variable $x(t) \in \mathbb{R}$). The generalization to higher dimensions is straightforward, but requires more elaborate notation.

4.4.1   Approximate Perron-Frobenius operator

Consider the DDE initial value problem (4.4) for $x(t) \in \mathbb{R}$, with solution map $S_t : x_0 \mapsto x(t)$, and suppose an ensemble of initial values x_0 is specified with density ρ_0. Since S_t is a continuous transformation of $\mathbb{R}$, it can be approximated by a piecewise linear function. Thus, suppose $I = (a, b)$ is an interval containing the support of ρ_0, and define a mesh of points $a = x_0 < x_1 < \cdots < x_k = b$ spanning I. Let $\tilde{S}_t : I \to \mathbb{R}$ be the transformation whose graph is a straight line on each interval (x_i, x_{i+1}), and satisfies

$$\tilde{S}_t(x_i) = S_t(x_i) \equiv y_i, \quad i = 1, \ldots, k. \tag{4.33}$$

Then $\tilde{S}_t$ furnishes a piecewise linear approximation of S_t, and agrees with S_t at each of the x_i.

Because $\tilde{S}_t$ is piecewise linear, it is almost-everywhere differentiable. Therefore the corresponding Perron-Frobenius operator $\tilde{P}_t$ can be expressed as [93, Ch. 12]

$$(\tilde{P}_t \rho_0)(x) = \sum_{z \in \tilde{S}_t^{-1}\{x\}} \frac{\rho_0(z)}{|\tilde{S}_t'(z)|}, \tag{4.34}$$

Since only those pre-images of $\{x\}$ that lie in the support of ρ_0 give a non-zero contribution to the sum, we need consider only those $z \in \tilde{S}_t^{-1}\{x\}$ that lie in some interval $[x_i, x_{i+1})$. On each such interval we have simply

$$\tilde{S}_t'(z) = \frac{y_{i+1} - y_i}{x_{i+1} - x_i}. \tag{4.35}$$

Furthermore, since $\tilde{S}_t$ is piecewise linear each $z \in \tilde{S}_t^{-1}\{x\}$ can be found by linear interpolation. Thus, for each interval $[y_i, y_{i+1})$ that contains x, there is exactly one element $z \in \tilde{S}_t^{-1}\{x\}$, given by

$$z = x_i + \frac{x_{i+1} - x_i}{y_{i+1} - y_i}(x - y_i). \tag{4.36}$$

Figure 4.4 illustrates this procedure for determining the set of pre-images of $\{x\}$ under a piecewise linear transformation.

4.4.2 Algorithm

The considerations above suggest the following algorithm for computing an approximation of the transformed density $\rho(x, t) = (P_t \rho_0)(x)$.

1. Specify a grid of closely spaced points $x_1 < x_2 < \cdots < x_k \in \mathbb{R}$, such that the interval (x_1, x_k) contains the support of ρ_0.

2. Compute (*e.g.*, by numerical solution of (4.4)) the sequence of values $\{y_i = S_t(x_i) : i = 1, \ldots, k\}$.

3. Specify a grid of points $\tilde{x}_1 < \tilde{x}_2 < \cdots < \tilde{x}_p \in \mathbb{R}$, at which the density $\rho_i = \rho_0(\tilde{x}_i, t)$ is to be approximated.

4. Initialize $\rho_i = 0$ for $i = 1, \ldots, p$.
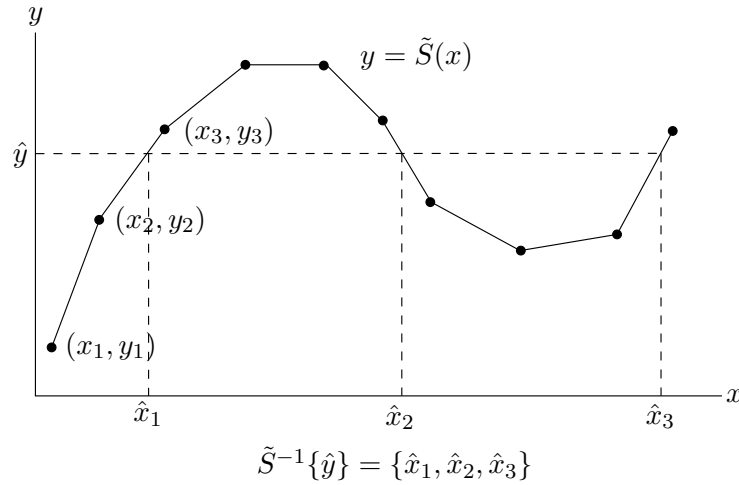
Figure 4.4: Construction of the set of pre-images of a point $\hat{y}$ under a piecewise linear transformation $\tilde{S}$.

5. For each $i \in \{1, 2, \ldots, (k-1)\}$ determine which, if any, of the $\tilde{x}_j$ lie in the interval $[y_i, y_{i+1})$ (or the interval $(y_{i+1}, y_i]$ if $y_{i+1} < y_i$). For each such $\tilde{x}_j$, compute

$$z = x_i + \frac{x_{i+1} - x_i}{y_{i+1} - y_i}(\tilde{x}_j - y_i), \qquad (4.37)$$

and increment ρ_j by the quantity

$$\rho_0(z)\frac{|x_{i+1} - x_i|}{|y_{i+1} - y_i|}. \qquad (4.38)$$

This algorithm constructs a vector $(\rho_i : i = 1, \ldots, p)$ approximating the density $\rho(x, t)$ at points $\{\tilde{x}_i\}$. Steps 1–2 define the piecewise linear approximation $\tilde{S}_t$. Steps 3–5 evaluate (4.34) at each of the points $\tilde{x}_i$. Notice that steps 1–2 are decoupled from 3–5 in that the initial density ρ_0 enters only in steps 3–5, after the approximating transformation $\tilde{S}_t$ has already been determined.

Example

Figures 4.5–4.6 illustrate the results of applying this algorithm to the Mackey-Glass equation (4.2). As before, the equation is restricted to constant initial functions (hence $g = 0$ in (4.4)), and the initial density ρ_0 corresponds to an ensemble of initial values uniformly

distributed on the interval $[0.3, 1.3]$. Figure 4.5 shows graphs of the approximating trans-formation $\tilde{S}_t \approx S_t$, at times $t = 1, 2, 3, 4$. These were obtained by using the numerical solver DDE23 [107] to compute values $y_i = S_t(x_i)$ for a uniform grid of 1000 initial values x_i in the interval $[0, 1.5]$.

4.4.3 Discussion

The algorithm presented here is superior in a number of respects to the "brute force" ensemble simulation approach of Section 4.3. In the brute force approach, a sufficient number of solutions of the DDE must be computed to ensure adequate statistical sam-pling. Here, one solution of the given DDE is computed for each point on the grid $\{x_i\}$ used to define the piecewise linear approximation of S_t. Figure 4.6 was generated using $k = 1000$ such points. In fact even $k = 100$ yields an approximate $\rho(x, t)$ with accuracy on the order of that obtained by the brute force approach with an ensemble of 10^6 solu-tions. This is a dramatic computational saving, and is the primary benefit of the method developed here.

The present method has the further advantage that the computation of $\rho(x, t)$ (page 81, steps 3–5) is decoupled from the solution of the DDE (steps 1–2). Since steps 1–2 are in-dependent of the initial density, a set of solutions of the DDE only needs to be computed once to construct the approximation $\tilde{S}_t$. Subsequently, the evolution of any number of different initial densities can be computed by steps 3–5. By contrast, in the brute force ap-proach, computing the evolution of each different initial density requires the computation of a new ensemble of solutions of the DDE.

Unfortunately it is not possible to evolve densities arbitrarily far forward in time in this way, at least for delay equations with chaotic dynamics. Because of the stretching and folding of phase space typical of chaotic systems, for large t the solution map S_t acquires a very complex structure. This is illustrated in Figure 4.8, which shows the approximate solution map $\tilde{S}_{50}$ for the Mackey-Glass equation (4.2). Owing both to this fine structure and sensitivity to initial conditions, as t increases it eventually becomes impossible to obtain a reasonable approximation of S_t using finite-precision arithmetic. This difficulty is not a function of the accuracy of the numerical method for integrating the DDE, but is rather a consequence of the complex dynamics of the DDE itself.

Surprisingly, even though $\tilde{S}_t$ is a poor approximation of S_t for large t, it nevertheless appears to retain information about the ensemble dynamics. Figure 4.9 shows a density $\rho(x, t)$ evolved forward to time $t = 50$, again for the Mackey-Glass equation (4.2) restricted
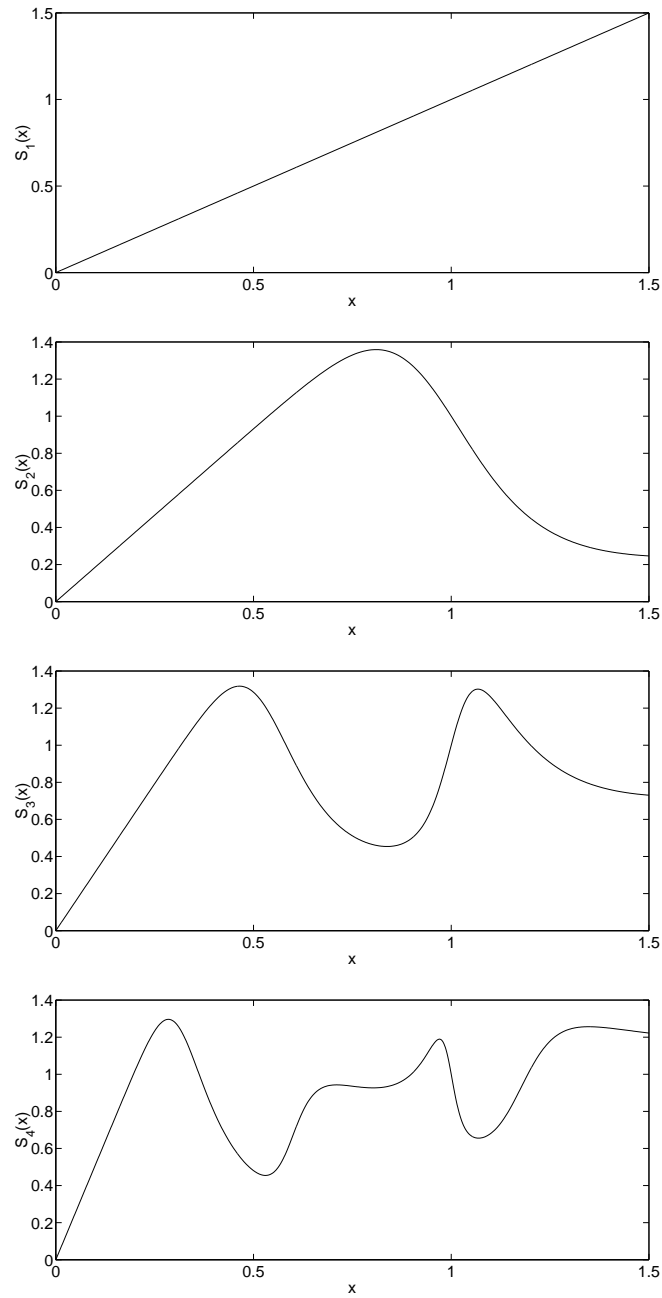
Figure 4.5: Approximate solution maps $\tilde{S}_t \approx S_t$ at times $t = 1, 2, 3, 4$, for the Mackey-Glass equation (4.2) restricted to constant initial functions on $[0, 1]$.
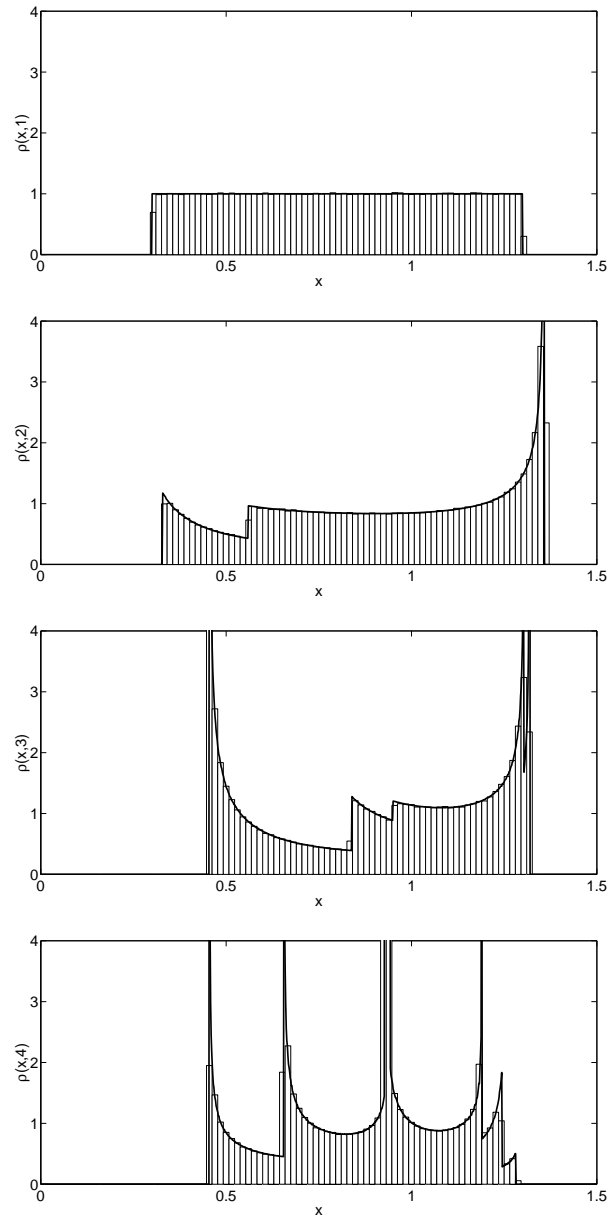
Figure 4.6: Computed densities $\rho(x, t)$ for the Mackey-Glass equation (4.2) restricted to constant initial functions. Densities were computed using the algorithm on page 81, and are shown (heavy curves) together with the corresponding histograms from Figure 4.2, obtained by "brute force" ensemble simulation.
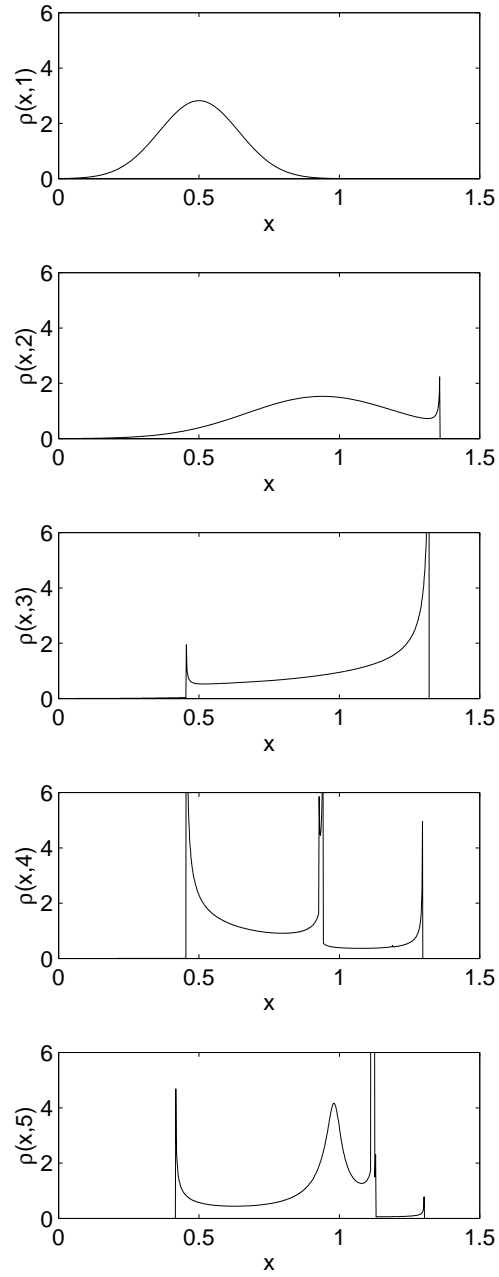
Figure 4.7: As Figure 4.6, with a different initial density.
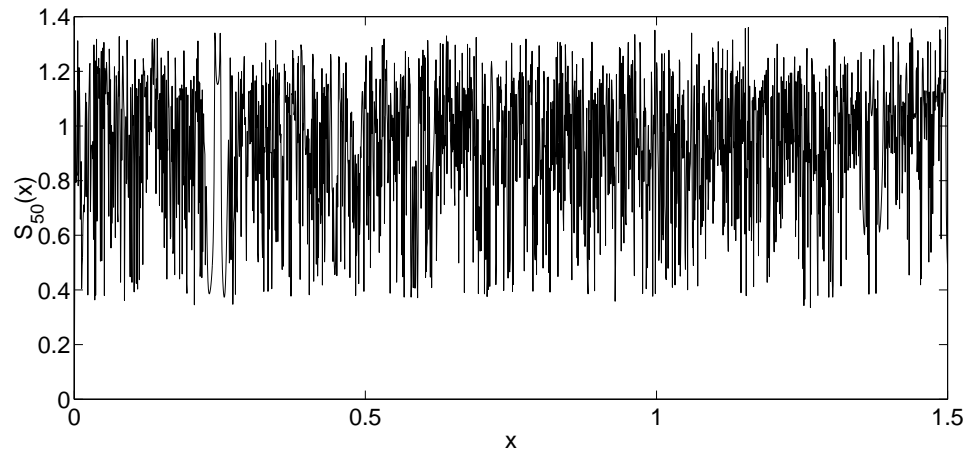
Figure 4.8: Approximate solution map $\tilde{S}_{50}$ for the Mackey-Glass equation (4.2) restricted to constant initial functions (evaluated on a grid of 2000 initial values x).
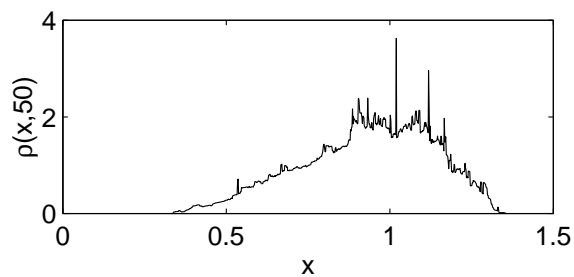
Figure 4.9: Computed density $\rho(x,t)$ at time $t = 50$, for the Mackey-Glass equation (4.2) restricted to constant initial functions.

to constant initial functions. This density was computed using the algorithm above, for the solution map $\tilde{S}_{50}$ shown in Figure 4.8, and the same initial density as in Figure 4.7. The result shows a remarkable agreement with the corresponding density computed by "brute force" ensemble simulation, shown in Figure 4.3, page 78. Thus it appears that the present method can provide an approximation of the same asymptotic density as that found by direct ensemble simulation, while requiring about 2 orders of magnitude less computation time than the ensemble simulation approach.

4.5 Evolution Equation for Densities

In section 4.2 Perron-Frobenius operators were derived by first finding an explicit formula for the solution map S_t. This is an awkward and difficult intermediate step. Rather it would be nice if, in the spirit of equation (2.33) for the evolution of a density under the action of a flow defined by an ordinary differential equation, one could derive an evolution equation for the density itself. This approach to density evolution for DDEs is the subject of the present section.

Consider the augmented DDE initial value problem

$$x'(t) = \begin{cases} g\big(x(t)\big) & t \in [0,1) \\ f\big(x(t), x(t-1)\big) & t \geq 1 \end{cases} \tag{4.39}$$
$$x(0) = x_0,$$

with $x(t) \in \mathbb{R}$, and suppose that an ensemble of initial values x_0 is specified with density ρ_0. We would like to derive an evolution equation for the density $\rho(x,t)$ of the corresponding ensemble of solutions $x(t)$.

There is an important preliminary observation to be made. Ideally, we would like to derive an evolution equation of the form

$$\frac{d\rho}{dt} = \{\text{some operator}\}(\rho). \tag{4.40}$$

However, ρ cannot satisfy such an equation. This is because the family of solution maps $\{S_t\}$ for equation (4.39) does not form a semigroup (*cf.* remarks at the end of Section 4.1.2). That is, the density ρ cannot be sufficient to determine its own evolution, as in (4.40), because the values $x(t)$ in the ensemble it describes are insufficient to determine *their* own evolution. This difficulty arises because ρ does not contain information about the past states of the ensemble, which is necessary to determine the evolution of the ensemble under (4.39). Thus, any solution to the problem must take a form other than (4.40).

4.5.1 ODE system

The method of steps is sometimes used to write a DDE as a system of ordinary differential equations. This is a promising connection, as we already know how densities evolve for
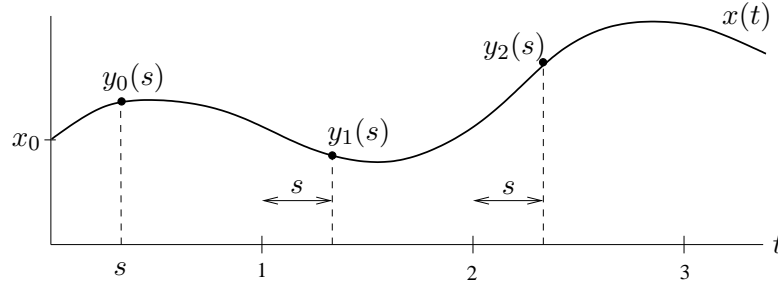
Figure 4.10: Relationship between the DDE solution $x(t)$ and the variables $y_n(s)$ defined in the method of steps, equation (4.41).

ODEs (*cf.* Section 2.3.2, page 27).

Method of steps

Let $x(t)$ be a solution of (4.39), and define for $n = 0, 1, 2, \ldots$ the functions

$$y_n(s) = x(n + s), \quad s \in [0, 1]. \tag{4.41}$$

Figure 4.10 illustrates this relationship between the y_n and x. Since $x(t)$ satisfies (4.39), it follows that for $n \geq 1$,

$$y_n'(s) = f\big(y_n(s), y_{n-1}(s)\big), \quad n = 1, 2, \ldots, \tag{4.42}$$

and y_0 satisfies

$$y_0' = g(y_0). \tag{4.43}$$

Thus the augmented DDE becomes a system of evolution equations for the y_n, together with the set of compatibility or boundary conditions

$$\begin{aligned} y_0(0) &= x_0 \\ y_n(0) &= y_{n-1}(1), \quad n = 1, 2, \ldots \end{aligned} \tag{4.44}$$

The ODE system (4.42)–(4.43), together with these compatibility conditions, can be solved sequentially to yield the solution $x(t)$ of the DDE up to any finite time. This is essentially the method of steps for solving the DDE (*cf.* Section 3.2.3, page 46).

If the system of ODEs (4.42)–(4.43) could be taken together as a vector field F in

$\mathbb{R}^{N+1}$, then an ensemble of solutions of the DDE could be represented via (4.41) as an ensemble of vectors $y = (y_0, \dots, y_N)$, each carried along the flow induced by F. The density $\eta(y, t)$ of such an ensemble would evolve according to a continuity equation (*cf.* Section 2.3.2)

$$\frac{\partial \eta}{\partial t} = -\nabla \cdot (\eta F). \tag{4.45}$$

[85] suggests this as an avenue to a probabilistic treatment of DDEs. However, it is unclear how to ensure the compatibility conditions (4.44) are satisfied by every vector in the ensemble, or how to determine the initial $(N + 1)$-dimensional density $\eta(y, 0)$ of this ensemble in terms of a given density of initial values x_0 in (4.39). In short there is no obvious way to treat equations (4.42)–(4.43) simultaneously rather than sequentially. The following modified setup is one way to avoid these difficulties.

Modified method of steps

Any solution of the DDE problem (4.39) can be extended unambiguously to all $t < 0$ by setting

$$x(t) = x_0, \quad t < 0, \tag{4.46}$$

so that $x'(t) = 0$ for all $t < 0$. For $n = 0, 1, 2, \dots$ let functions y_n be defined by

$$y_n(t) = x(t - n), \quad t \geq 0. \tag{4.47}$$

Figure 4.11 illustrates the relationship between the y_n and x. On substitution into equation (4.39) we find that for $t \in [m, m+1]$, $m = 0, 1, 2, \dots$, the $y_n(t)$ satisfy

$$y_n' = \begin{cases} f(y_n, y_{n+1}) & \text{if } n < m \\ g(y_m) & \text{if } n = m \\ 0 & \text{if } n > m. \end{cases} \tag{4.48}$$

Thus, for fixed N the vector $y(t) = \big(y_0(t), \dots, y_N(t)\big)$ satisfies an ordinary differential equation

$$y' = F(t, y), \tag{4.49}$$

PSfrag replacements
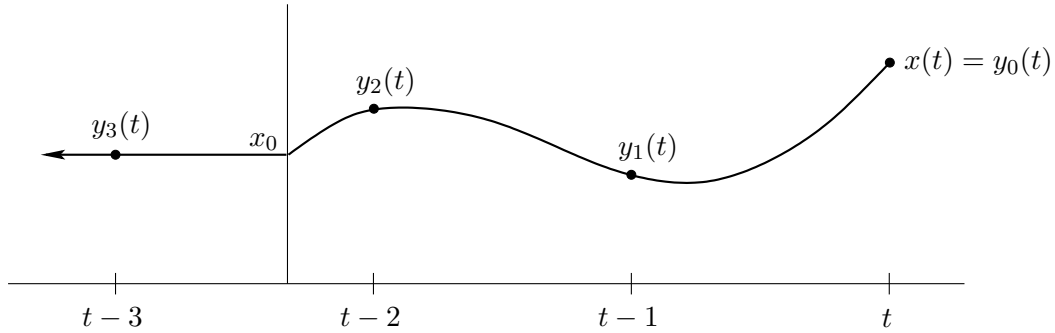
Figure 4.11: Relationship between the DDE solution $x(t)$ and the variables $y_n(t)$ defined in the modified method of steps, equation (4.47).

where the vector field $F : \mathbb{R} \times \mathbb{R}^{N+1} \to \mathbb{R}^{N+1}$ is given by

$$F(t, y_0, \dots, y_N) = (y_0', \dots, y_N'),$$

$$\text{with} \quad y_n' = \begin{cases} 0 & \text{if } t < n \\ g(y_n) & \text{if } t \in [n, n+1) \\ f(y_n, y_{n+1}) & \text{if } t \geq n+1. \end{cases} \tag{4.50}$$

Some remarks about the ODE system (4.49)–(4.50) are in order:

- For fixed $N > 0$, the right-hand side $F(t, y)$, and hence the solution $y(t)$, is defined only for $0 \leq t \leq N + 1$.

- For any given initial vector $y(0) = (y_0(0), \dots, y_N(0))$, equation (4.49) has a unique solution $y(t)$ defined for $0 \leq t \leq N + 1$, provided we have existence and uniqueness for the original DDE problem (4.39).

- $F(t, y)$ is piecewise constant in time. That is, for each $m = 0, 1, 2, \dots$, the vector field $F(t, y) = F(y)$ is independent of $t \in [m, m + 1)$, and induces a flow in $\mathbb{R}^{N+1}$ that carries the solution $y(t)$ forward from $t = m$ to $t = m + 1$. Thus we can speak of (4.49) as defining a sequence of flows in $\mathbb{R}^{N+1}$.

The ODE system (4.49) gives a representation of the method of steps as an evolution equation in $\mathbb{R}^{N+1}$. Indeed, the solution $x(t)$ of the DDE is given, up to time $t = N + 1$, by $x(t) = y_0(t)$ where $y(t)$ is the solution of (4.49) corresponding to the initial condition

$y(0) = (x_0, \ldots, x_0)$. Thus the DDE problem (4.39) is equivalent to the initial value problem

$$
\begin{aligned}
y' &= F(y, t), \quad y(t) \in \mathbb{R}^{N+1}, \ 0 \le t \le N + 1, \\
y(0) &= (x_0, \ldots, x_0),
\end{aligned}
\tag{4.51}
$$

with the identification $x(t) = y_0(t)$.

4.5.2 Continuity equation

Having established the equivalence of the DDE system (4.39) with the ODE system (4.51), we can proceed to the probabilistic treatment of DDEs using techniques developed for ODEs.

Suppose an ensemble of initial vectors $y \in \mathbb{R}^{N+1}$ is given, with $(N + 1)$-dimensional density $\eta(y, t)$. Then under the sequence of flows induced by the vector field $F(y, t)$, this density evolves according to a continuity equation (*cf.* Section 2.3.2),

$$
\frac{\partial \eta(y, t)}{\partial t} = -\nabla \cdot \big(\eta(y, t) F(y, t) \big),
\tag{4.52}
$$

where $\nabla = (\partial/\partial y_0, \ldots, \partial/\partial y_N)$. The initial density $\eta(y, 0)$ derives from the density ρ_0 of initial values x_0 for the DDE (4.39). That is, an ensemble of initial values x_0 with density ρ_0 corresponds to an ensemble of initial vectors $y = (x_0, \ldots, x_0)$ in $\mathbb{R}^{N+1}$, with "density"

$$
\eta(y_0, \ldots, y_N; 0) = \rho_0(y_0)\delta(y_0 - y_1)\delta(y_0 - y_2) \cdots \delta(y_0 - y_N),
\tag{4.53}
$$

where δ is the Dirac delta function. This corresponds to a line mass concentrated on the line $y_0 = y_1 = \cdots = y_N$ with linear density $\rho_0(y_0)$. Thus the problem of density evolution for DDEs becomes a problem of determining how this line mass is redistributed by the flow induced by (4.49).

With singular initial data such as (4.53), strong solutions of the continuity equation (4.52) do not exist. However, (4.52) can be interpreted in a weak sense that makes it possible to define "solutions" that satisfy initial conditions like (4.53) (*cf.* Section 2.3.2). Such a weak solution can be obtained using the method of characteristics.

4.5.3   Method of characteristics

Consider the initial value problem

$$\frac{\partial \eta(y,t)}{\partial t} + \nabla \cdot \big(F(y,t)\eta(y,t)\big) = 0, \quad t \geq 0,$$
$$\eta(y,0) = g(y) \tag{4.54}$$

for the unknown function $\eta(y,t)$, with $y = (y_0, \ldots, y_N)$. Assume provisionally that F and η are differentiable in y, so the divergence operator can be expanded (by the product rule) to yield

$$\frac{\partial \eta}{\partial t} + \eta \nabla \cdot F + F \cdot \nabla \eta = 0. \tag{4.55}$$

Let a curve $\Gamma \subset \mathbb{R}^{N+2}$ (that is, in $(y_0, \ldots, y_N, t)$-space) be parametrized by smooth functions

$$y_0 = y_0(t), \; \ldots \; , y_N = y_N(t) \tag{4.56}$$

defined for all $t \geq 0$, and parametrize the value of η on Γ by

$$\eta(t) = \eta\big(y_0(t), \ldots, y_N(t), t\big). \tag{4.57}$$

(This slight abuse of notation helps clarify the following development.) Differentiating (4.57) yields

$$\frac{d\eta}{dt} = \nabla \eta \cdot \frac{dy}{dt} + \frac{\partial \eta}{\partial t}. \tag{4.58}$$

Thus, if the functions $y(t)$, $\eta(t)$ satisfy

$$\frac{dy}{dt} = F\big(y(t), t\big) \tag{4.59}$$

$$\frac{d\eta}{dt} = -\eta(t) \nabla \cdot F\big(y(t), t\big) \tag{4.60}$$

for all $t \geq 0$, then η as given by equation (4.57) satisfies the PDE (4.55) at every point on Γ. In fact any solution of the ODE system (4.59)–(4.60) furnishes a solution of the PDE (4.55) on a particular curve Γ. In particular, if y, η are solutions of this system corresponding to initial values

$$y(0) = (y_0, \ldots, y_N)$$
$$\eta(0) = g(y_0, \ldots, y_N), \tag{4.61}$$

PSfrag replacements
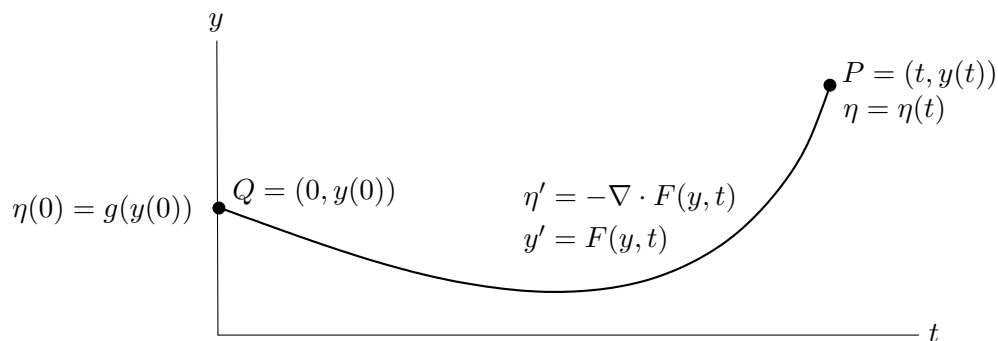
Figure 4.12: Method of characteristics for the continuity equation (4.54): propagation of initial data along an integral curve of $y' = F(y, t)$.

then Γ intersects the hyperplane $t = 0$ at the point $y = (y_0, \ldots, y_N)$, where $\eta(0)$ agrees with the initial data given by g, so that $\eta(t)$ gives the solution of (4.54) at every point of Γ.

The family of curves Γ that satisfy (4.59) are called the *characteristics* of the PDE (4.55). As we have seen, the characteristics have the geometric interpretation that an initial datum specified at $(y_0, \ldots, y_N, 0)$ is propagated along the characteristic that passes through this point. Not surprisingly, the characteristic curves of (4.54) coincide with the integral curves of the vector field F (*cf.* equation (4.59)). That is, initial data are propagated along streamlines of the induced flow.

If the characteristics foliate $\mathbb{R}^{N+2}$, every point $(y_0, \ldots, y_N, t) \in \mathbb{R}^{N+2}$ has a characteristic curve passing through it. Then the solution of (4.54) can found at any point, by using (4.59)–(4.60) to obtain the solution on the characteristic curve through that point. The usual procedure for obtaining the solution $\eta(y, t)$ at $P = (y_0, \ldots, y_N, t)$ is as follows.

1. Determine the characteristic curve Γ through P and follow it "backward" in time to find the point Q on Γ at $t = 0$.

2. Evaluate g at Q to determine the initial value $\eta(0)$ on Γ.

3. With this value of $\eta(0)$, integrate equation (4.60) forward along Γ to P, at which point the value of $\eta(t)$ is the solution of (4.54).

Figure 4.12 gives a schematic illustration of the method.

Although the derivation assumes differentiability of η (hence also of g), the method itself does not rely on any special properties of these functions—it requires only integration of the vector field F and evaluation of g. Hence the method can be applied even if g is discontinuous, or singular as in (4.53). However, in such cases the resulting function η must be interpreted as a weak solution [20, 125].

Supposing the solution $\eta(y, t)$ of the initial value problem (4.52)–(4.53) to have been found—*e.g.*, by the method of characteristics—the corresponding density $\rho(x, t)$ of DDE solutions $x(t)$ can be determined as follows. Since $x(t) = y_0(t)$ (*cf.* equation (4.47)), the density of x is identified with the density of y_0. This density is determined by integrating η over all components except y_0, *i.e.*,

$$\rho(y_0, t) = \int \cdots \int \eta(y_0, y_1, \ldots, y_N; t) \, dy_1 \cdots dy_N. \qquad (4.62)$$

Alternative formulation

There is another way to formulate the evolution of the density $\eta(y, t)$, that turns out to be equivalent to the method of characteristics and serves to illuminate the method above. It also provides an explicit formula (actually a codification of the algorithm on page 95) for the solution $\eta(y, t)$.

Recall that the vector $y(t)$ evolves according to a system of ODEs (4.49)–(4.50). Let $y(t)$ be the solution of this system with initial value $y(0)$, and define the corresponding family of solution maps $\hat{S}_t : \mathbb{R}^{N+1} \to \mathbb{R}^{N+1}$ by

$$\hat{S}_t : y(0) \mapsto y(t) \qquad (4.63)$$

(to be distinguished from the solution map S_t for the DDE, defined by (4.7)). As $y(t)$ evolves under the action of $\hat{S}_t$, the density $\eta(y, t)$ evolves according to the corresponding Perron-Frobenius operator $\hat{P}_t : \eta(y, 0) \mapsto \eta(y, t)$, defined by

$$\int_A \hat{P}_t \eta(y, 0) \, d^{N+1} y = \int_{\hat{S}_t^{-1}(A)} \eta(y, t) \, d^{N+1} y \quad \forall \text{ Borel } A \subset \mathbb{R}^{N+1}. \qquad (4.64)$$

Recall that $\hat{S}_t$ can be represented as a composition of flows on the intervals $[m, m+1]$, $m = 0, 1, 2, \ldots$, so it is one-to-one on $\mathbb{R}^{N+1}$ and has an inverse (which can be found by reversing the sequence of flows). This makes possible the change of variables $z = S_t(y)$

in (4.64), which by Theorem 3.2.1 of [74] becomes

$$\int_A \hat{P}_t \eta(y,0)\, d^{N+1}y = \int_A \eta\big(\hat{S}_t^{-1}(z),t\big) J_t^{-1}(z)\, d^{N+1}z. \tag{4.65}$$

Since A is arbitrary, this implies the following explicit formula for P_t,

$$P_t \eta(y,0) = \eta(y,t) = \eta\big(\hat{S}_t^{-1}(y),0\big) J_t^{-1}(y). \tag{4.66}$$

Here J_t^{-1} is the density of the measure $\lambda \circ S_t^{-1}$ with respect to Lebesgue measure λ [74, p. 46] (also [71, Section 5.1]). If $\hat{S}_t$ and $\hat{S}_t^{-1}$ are differentiable transformations[4] of $\mathbb{R}^{N+1}$ then J_t^{-1} is just the determinant of the Jacobian matrix $D\hat{S}_t^{-1}$,

$$\begin{aligned} J_t^{-1}(y) &= \det\big(D\hat{S}_t^{-1}(y)\big), \\ J_t(y) &= \det\big(D\hat{S}_t(y)\big). \end{aligned} \tag{4.67}$$

In this case the formula (4.66) can be seen as a multi-dimensional analog of (4.24), in the case of invertible S_t.

Notice that $\hat{S}_t$ effects the translation of a point $(y(0),0)$ along a characteristic curve Γ to $(y(t),t)$. Similarly $\hat{S}_t^{-1}$ effects a translation backward along Γ to $t=0$. This draws the connection between (4.66) and the method of characteristics (page 95): the point Q (where the initial density is evaluated) is identified with the point $(\hat{S}_t^{-1}(y),0)$ in (4.66).

The factor $J_t^{-1}(y)$ also has a geometric interpretation: it is the factor by which the volume of an infinitesimal volume element at $y(t)$ increases under transportation by $\hat{S}_t^{-1}$. This factor can equivalently be understood as resulting from step 3 of the method of characteristics algorithm, since $\nabla \cdot F(y,t)$ is the instantaneous growth rate of an infinitesimal volume at y as it is transported by the flow S_t induced by F. Conservation of mass requires that the density supported on an infinitesimal volume element decrease in proportion to the volume growth, *i.e.*, by the factor $J_t^{-1}(y)$. This provides a geometrical explanation of the term $J_t^{-1}(y)$ in (4.66).

Examples

To illustrate this approach to the evolution of densities for DDEs, we revisit examples 4.2.1 and 4.2.2, for which analytical solutions are obtainable for the densities $\eta(y,t)$ and $\rho(x,t)$.

---

[4]It suffices that the vector field $F(t,y)$ be smooth in y [72, p. 19].

Example 4.5.1. Consider again the linear DDE of example 4.2.1 (page 71),

$$x'(t) = \alpha x(t-1), \quad t \geq 1, \tag{4.68}$$

with initial data restricted to constant initial functions on $[0,1]$, and the initial value $x(0)$ distributed with density $\rho_0(x)$. The vector $y = (y_0, y_1, y_2)$ defined by (4.47) satisfies a differential equation $y' = F(y,t)$ for $t \in [0,3]$, with

$$y'_n = \begin{cases} 0 & \text{if } t < n+1 \\ \alpha y_{n+1} & \text{if } n+1 \leq t \leq 3. \end{cases} \tag{4.69}$$

Under the action of this system, the density $\eta(y,t)$ of an ensemble of vectors $y \in \mathbb{R}^3$ evolves according to the continuity equation (4.52). whose characteristic curves are streamlines of the flow induced by (4.69). The solution of this system is readily obtained (*e.g.*, using Maple), and the solution map $\hat{S}_t : y(0) \mapsto y(t)$ found to be

$$\hat{S}_t(y) = \begin{cases} (y_0, y_1, y_2) & \text{if } t \in [0,1) \\ (y_0 + \alpha(t-1)y_1, \; y_1, \; y_2) & \text{if } t \in [1,2) \\ (y_0 + \alpha(t-1)y_1 + (\frac{1}{2}\alpha^2 t^2 - 2\alpha^2(t-1))y_2, \\ \quad y_1 + \alpha(t-2)y_2, \; y_2) & \text{if } t \in [2,3]. \end{cases} \tag{4.70}$$

This linear transformation is easily inverted to yield

$$\hat{S}_t^{-1}(y) = \begin{cases} (y_0, y_1, y_2) & \text{if } t \in [0,1) \\ (y_0 - \alpha(t-1)y_1, \; y_1, \; y_2) & \text{if } t \in [1,2) \\ (y_0 - \alpha(t-1)y_1 + \alpha^2(\frac{1}{2}t^2 - t)y_2, \\ \quad y_1 - \alpha(t-2)y_2, y_2) & \text{if } t \in [2,3]. \end{cases} \tag{4.71}$$

The Jacobian of this transformation is

$$J_t^{-1}(y) = \det\left(D\hat{S}_t^{-1}(y)\right) = 1 \quad \forall t \in [0,3] \tag{4.72}$$

(*i.e.*, $\hat{S}_t^{-1}$ is volume-preserving). The initial ensemble of vectors y has "density" $\eta(y,0)$ is given by

$$\eta(y_0, y_1, y_2, 0) = \rho_0(y_0)\delta(y_0 - y_1)\delta(y_0 - y_2), \tag{4.73}$$

hence equation (4.66) gives

$$
\eta(y,t) = \begin{cases}
\rho_0(y_0)\delta(y_0 - y_1)\delta(y_0 - y_2) & \text{if } t \in [0,1) \\[4pt]
\rho_0\big(y_0 - \alpha(t-1)y_1\big)\delta\big(y_0 - \alpha(t-1)y_1 - y_1\big)\cdot & \\
\quad \delta\big(y_0 - \alpha(t-1)y_1 - y_2\big) & \text{if } t \in [1,2) \\[4pt]
\rho_0\big(y_0 - \alpha(t-1)y_1 + \alpha^2(\tfrac{1}{2}t^2 - t)y_2\big)\cdot & \\
\quad \delta\big([y_0 - \alpha(t-1)y_1 + \alpha^2(\tfrac{1}{2}t^2 - t)y_2] - & \\
\quad\quad [y_1 - \alpha(t-2)y_2]\big)\cdot & \\
\quad \delta\big([y_0 - \alpha(t-1)y_1 + \alpha^2(\tfrac{1}{2}t^2 - t)y_2] - y_2\big) & \text{if } t \in [2,3].
\end{cases}
\tag{4.74}
$$

Integrating over y_1, y_2 with $x = y_0$ we obtain for $t \in [0,1)$,

$$
\begin{aligned}
\rho(x,t) &= \iint \rho_0(x)\delta(x - y_1)\delta(x - y_2)\, dy_1\, dy_2 \\
&= \rho_0(x),
\end{aligned}
\tag{4.75}
$$

(hence $P_t\rho = \rho$ as expected, since S_t is just the identity transformation). For $t \in [1,2)$,

$$
\begin{aligned}
\rho(x,t) &= \iint \rho_0\big(x - \alpha(t-1)y_1\big)\delta\big(x - \alpha(t-1)y_1 - y_1\big)\cdot \\
&\qquad\qquad \delta\big(x - \alpha(t-1)y_1 - y_2\big)\, dy_1\, dy_2 \\
&= \int \rho_0\big(x - \alpha(t-1)y_1\big)\delta\big(x - \alpha(t-1)y_1 - y_1\big)\, dy_1 \\
&= \frac{1}{|1 + \alpha(t-1)|} \int \rho_0\Big(x - \alpha(t-1)\frac{x - z}{1 + \alpha(t-1)}\Big)\delta(z)\, dz \\
&= \frac{1}{|1 + \alpha(t-1)|}\rho_0\Big(x - \alpha(t-1)\frac{x}{1 + \alpha(t-1)}\Big) \\
&= \frac{1}{|1 + \alpha(t-1)|}\rho_0\Big(\frac{x}{1 + \alpha(t-1)}\Big).
\end{aligned}
\tag{4.76}
$$

This agrees with the result (4.17) of example 4.2.1, which was obtained by a different method. For $t \in [2,3]$ the integral for $\rho(x,t)$ becomes too complicated to be worth writing out fully here, but its result also agrees with (4.17).

Example 4.5.2. Consider again the DDE of example 4.2.2 (page 73),

$$
x'(t) = -x(t-1)^2, \quad t \geq 1,
\tag{4.77}
$$

with initial data restricted to constant initial functions on $[0, 1]$, and the initial value $x(0)$ distributed with density $\rho_0(x)$. The vector $y = (y_0, y_1, y_2)$ defined as in (4.47) satisfies a differential equation $y' = F(y, t)$ for $t \in [0, 3]$, with

$$
y_n' = \begin{cases} 0 & \text{if } t < n + 1 \\ -y_{n+1}^2 & \text{if } n + 1 \le t \le 3. \end{cases}
\tag{4.78}
$$

Solving this system (*e.g.*, using Maple), and defining the solution map $\hat{S}_t : y(0) \mapsto y(t)$ yields

$$
\hat{S}_t(y) = \begin{cases} (y_0, \ y_1, \ y_2) & \text{if } t \in [0, 1) \\ \left(y_0 - (t-1)y_1^2, \ y_1, \ y_2 \right) & \text{if } t \in [1, 2) \\ \left(y_0 - (t-1)y_1^2 + (t^2 - 4t + 4)y_1 y_2^2 - \right. \\ \quad \left. (\frac{1}{3}t^3 - 2t^2 + 4t - \frac{8}{3})y_2^4, \ y_1 - (t-2)y_2^2, \ y_2 \right) & \text{if } t \in [2, 3]. \end{cases}
\tag{4.79}
$$

Inverting this transformation yields

$$
\hat{S}_t^{-1}(y) = \begin{cases} (y_0, \ y_1, \ y_2) & \text{if } t \in [0, 1) \\ \left(y_0 + (t-1)y_1^2, \ y_1, \ y_2 \right) & \text{if } t \in [1, 2) \\ \left(y_0 + (t-1)y_1^2 + (t^2 - 2t)y_1 y_2^2 + (\frac{1}{3}t^3 - t^2 + \frac{4}{3})y_2^4, \right. \\ \quad \left. y_1 + (t-2)y_2^2, \ y_2 \right) & \text{if } t \in [2, 3]. \end{cases}
\tag{4.80}
$$

The Jacobian of this transformation is again

$$
J_t^{-1}(y) = \det \left(D\hat{S}_t^{-1}(y) \right) = 1 \quad \forall t \in [0, 3].
\tag{4.81}
$$

The initial density $\eta(y, 0)$ is given by

$$
\eta(y, 0) = \rho_0(y_0)\delta(y_0 - y_1)\delta(y_1 - y_2),
\tag{4.82}
$$

so equation (4.66) gives

$$
\eta(y,t) = \begin{cases}
\rho_0(y_0)\delta(y_0 - y_1)\delta(y_1 - y_2) & \text{if } t \in [0,1) \\[2mm]
\rho_0\big(y_0 + (t-1)y_1^2\big)\delta\big(y_0 + (t-1)y_1^2 - y_1\big)\delta(y_1 - y_2) & \text{if } t \in [1,2) \\[2mm]
\rho_0\big(y_0 + (t-1)y_1^2 + (t^2 - 2t)y_1 y_2^2 + (\tfrac{1}{3}t^3 - t^2 + \tfrac{4}{3})y_2^4\big)\cdot \\
\quad \delta\big([y_0 + (t-1)y_1^2 + (t^2 - 2t)y_1 y_2^2 + (\tfrac{1}{3}t^3 - t^2 + \tfrac{4}{3})y_2^4] \\
\quad\quad -[y_1 + (t-2)y_2^2]\big)\cdot \\
\quad \delta\big([y_1 + (t-2)y_2^2] - y_2\big) & \text{if } t \in [2,3].
\end{cases}
$$

(4.83)

Finally, $\rho(x,t)$ is obtained by integrating $\eta(y,t)$ over y_1, y_2. For $t \in [0,1)$,

$$
\begin{aligned}
\rho(x,t) &= \iint \rho_0(x)\delta(x - y_1)\delta(y_1 - y_2)\, dy_1\, dy_2 \\
&= \rho_0(x),
\end{aligned}
$$

(4.84)

as expected. For $t \in [1,2)$,

$$
\begin{aligned}
\rho(x,t) &= \iint \rho_0\big(x + (t-1)y_1^2\big)\delta\big(x + (t-1)y_1^2 - y_1\big)\delta(y_1 - y_2)\, dy_1\, dy_2 \\
&= \int \rho_0\big(x + (t-1)y_1^2\big)\delta\big(x + (t-1)y_1^2 - y_1\big)\, dy_1 \\
&= \sum_{\{y_1 : x + (t-1)y_1^2 - y_1 = 0\}} \frac{\rho_0\big(x + (t-1)y_1^2\big)}{|2(t-1)y_1 - 1|} \\
&= \frac{1}{\sqrt{1 - 4(t-1)x}}\Big[\rho_0\Big(\frac{1 - \sqrt{1 - 4(t-1)x}}{2(t-1)}\Big) \\
&\qquad\qquad\qquad + \rho_0\Big(\frac{1 + \sqrt{1 - 4(t-1)x}}{2(t-1)}\Big)\Big].
\end{aligned}
$$

(4.85)

This is identical to the result of example 4.2.2. For $t \in [2,3)$ it would be extremely difficult to find an explicit formula for $\rho(x,t)$, since the final integration requires solving a quartic equation. For $t > 3$, a quintic equation must be solved, so finding an explicit formula for $\rho(x,t)$ appears to be impossible.

Remark 4.5.1. In each of the examples above, the transformation $\hat{S}_t$ was found to be volume-preserving, so that the Jacobian $J_t^{-1}(y) = 1$ in equation (4.66). This could have been anticipated from equation (4.60) for the evolution of the density along a character-

istic curve of the continuity equation, since by equation (4.48),

$$\nabla \cdot F(y, t) = D_1 f(y_0, y_1) + \cdots + D_1 f(y_{m-1}, y_m) + g'(y_m) \qquad (4.86)$$

for $t \in [m, m+1)$. Thus, if f is independent of its first argument and $g' = 0$ (as in the examples above), then $\nabla \cdot F(y, t) = 0$, so that the sequence of flows constituting S_t are all volume-preserving.

4.5.4   Geometric interpretation

As the examples above suggest, for all but the simplest delay equations an analytical treatment of the density evolution problem is difficult, and perhaps impossible. Nevertheless, the approach developed above does provide some geometrical insight even when an analytical approach fails.

Recall that up to any finite time $t \leq N+1$ the DDE problem (4.4) can be represented by an ordinary differential equation

$$y' = F(y, t), \quad y(t) \in \mathrm{I\!R}^{N+1}, \quad t \geq 0, \qquad (4.87)$$

with F defined by (4.50). An ensemble of initial values with density ρ_0 corresponds to an ensemble of initial vectors y with $(N+1)$-dimensional "density"

$$\eta(y_0, \ldots, y_N; 0) = \rho_0(y_0)\delta(y_0 - y_1)\delta(y_1 - y_2) \cdots \delta(y_{N-1} - y_N), \qquad (4.88)$$

representing a line mass concentrated on the line $y_0 = y_1 = \cdots = y_N$ in $\mathrm{I\!R}^{N+1}$. Under evolution by (4.87), *i.e.*, under transformation by the solution map $\hat{S}_t$, this line mass is redistributed. This transportation of a line mass under $\hat{S}_t$ is illustrated in Figure 4.13. After evolution by time t, $\eta(y, t)$ is supported on a one-dimensional curve that is the image of this line under $\hat{S}_t$. We will call this curve the "density support curve". It is a continuous, non-self-intersecting curve in $\mathrm{I\!R}^{N+1}$, owing to continuity and invertibility of $\hat{S}_t$.

Example 4.5.3. Explicit representations can be found for the density support curves in the previous examples. Consider the linear DDE of example 4.5.1 (page 98), for which the transformation $\tilde{S}_t$ is given explicitly by equation (4.70). The initial density support
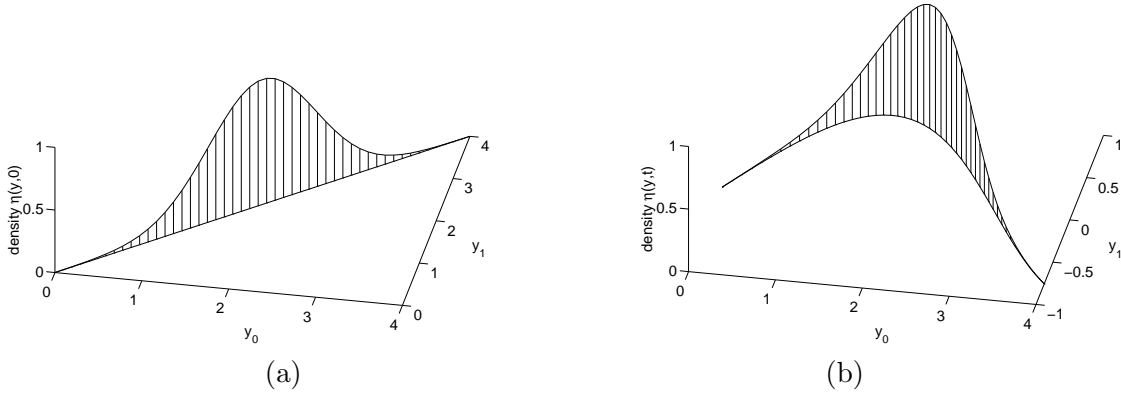
Figure 4.13: Transportation of a line mass under a transformation of the (y_0, y_1)-plane. (a) Initial mass distributed on the line $y_0 = y_1$. (b) Mass distribution after transformation.

curve $y_0 = y_1 = y_2$ can be represented parametrically as

$$\Gamma = \{(s, s, s) : s \in \mathbb{R}\}. \tag{4.89}$$

Then

$$\begin{aligned} \tilde{S}_t(\Gamma) &= \left\{ \tilde{S}_t(y) : y \in \Gamma \right\} \\ &= \left\{ y(s) = (y_1, y_2, y_3)(s) : s \in \mathbb{R} \right\} \end{aligned} \tag{4.90}$$

where

$$(y_1, y_2, y_3)(s) = \begin{cases} (s, s, s) & t \in [0, 1) \\ \left(s + \alpha(t-1)s, \ s, \ s\right) & t \in [1, 2) \\ \left(s + \alpha(t-1)s + \left(\frac{1}{2}\alpha^2 t^2 - 2\alpha^2(t-1)\right)s, \\ \quad s + \alpha(t-2)s, \ s\right) & t \in [2, 3). \end{cases} \tag{4.91}$$

For this DDE, at any given time the density support curve is a straight line, a consequence of linearity of $\tilde{S}_t$.

Example 4.5.4. Consider the quadratic DDE of example 4.5.2 (page 99), for which the transformation $\tilde{S}_t$ is given explicitly by equation (4.79). Representing the initial density

support curve Γ parametrically as in the previous example, we have

$$\tilde{S}_t(\Gamma) = \big\{y(s) = (y_1, y_2, y_3)(s) : s \in \mathbb{R}\big\} \tag{4.92}$$

where

$$(y_1, y_2, y_3)(s) = \begin{cases} (s, \ s, \ s) & t \in [0, 1) \\ \big(s - (t-1)s^2, \ s, \ s\big) & t \in [1, 2) \\ \big(s - (t-1)s^2 + (t^2 - 4t + 4)s^3 - \\ \quad (\frac{1}{3}t^3 - 2t^2 + 4t - \frac{8}{3})s^4, \ s - (t-2)s^2, \ s\big) & t \in [2, 3]. \end{cases} \tag{4.93}$$

Thus, the density evolution method developed in section 4.5.3 amounts to keeping track of the evolution of the density support curve under the action of (4.87). For the purposes of a numerical implementation, this curve can be represented by a set of points $\{y^{(1)}(t), y^{(2)}(t), \ldots, y^{(k)}(t)\} \subset \mathbb{R}^{N+1}$ (*e.g.*, representing a piecewise-linear approximation of the curve). The $y^{(i)}(t)$ are images under $\hat{S}_t$ of points $y^{(i)}(0)$ of the form

$$y^{(i)}(0) = \big(x_0^{(i)}, \ldots, x_0^{(i)}\big) \in \mathbb{R}^{N+1}, \tag{4.94}$$

which lie on the initial density support curve $y_0 = y_1 = \cdots = y_N$. Thus the points $y^{(i)}(t)$ can be determined by integrating (numerically) each of these initial points forward under (4.87). With sufficiently closely spaced points $x_0^{(i)}$ in the support of the initial density $\rho_0(x)$, the resulting set of $y^{(i)}(t)$ should provide a good approximation of the density support curve, and the mass distribution on it (see Figure 4.14).

Figure 4.15 illustrates the results of applying this idea to the Mackey-Glass equation (4.2), for an ensemble of constant initial functions with values distributed on the interval $[0.3, 1.3]$, as in Figures 4.1–4.2. Thus the initial density support curve is the part of the line $y_0 = y_1 = \cdots$ with $0.3 \leq y_0 \leq 1.3$.

The first row of figure 4.15 shows the sequence of density support curves obtained at times $t = 1, 2, 3, 4$, projected onto the (y_0, y_1)-plane (as a result of this projection, some of the curves intersect themselves). The second row shows the corresponding densities $\rho(x, t)$ from Figure 4.2. These densities can be interpreted as resulting from projecting the mass supported on the corresponding density support curve onto the y_0-axis. With this interpretation, the density support curves provide an obvious geometrical interpretation

PSfrag replacements
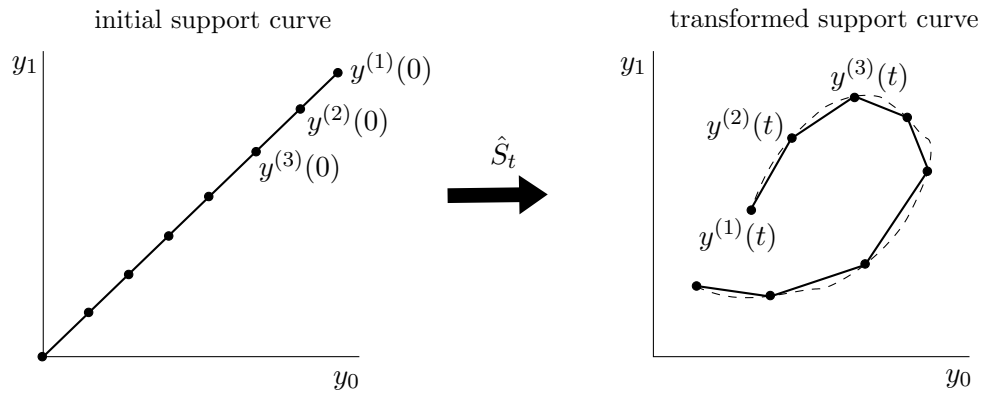
Figure 4.14: Approximating the image under $\hat{S}_t$ of the density support curve, by following the evolution under (4.87) of a set of points $\{y^{(i)}(t)\}$ that represent the nodes of a piecewise linear curve. Here the curve is shown projected onto the (y_0, y_1)-plane.

of the structures observed in the corresponding densities $\rho(x,t)$. Discrete jumps in the density occur at the endpoints of the transformed density support curve, and the maxima (singularities) correspond to turning points of the transformed density support curve.
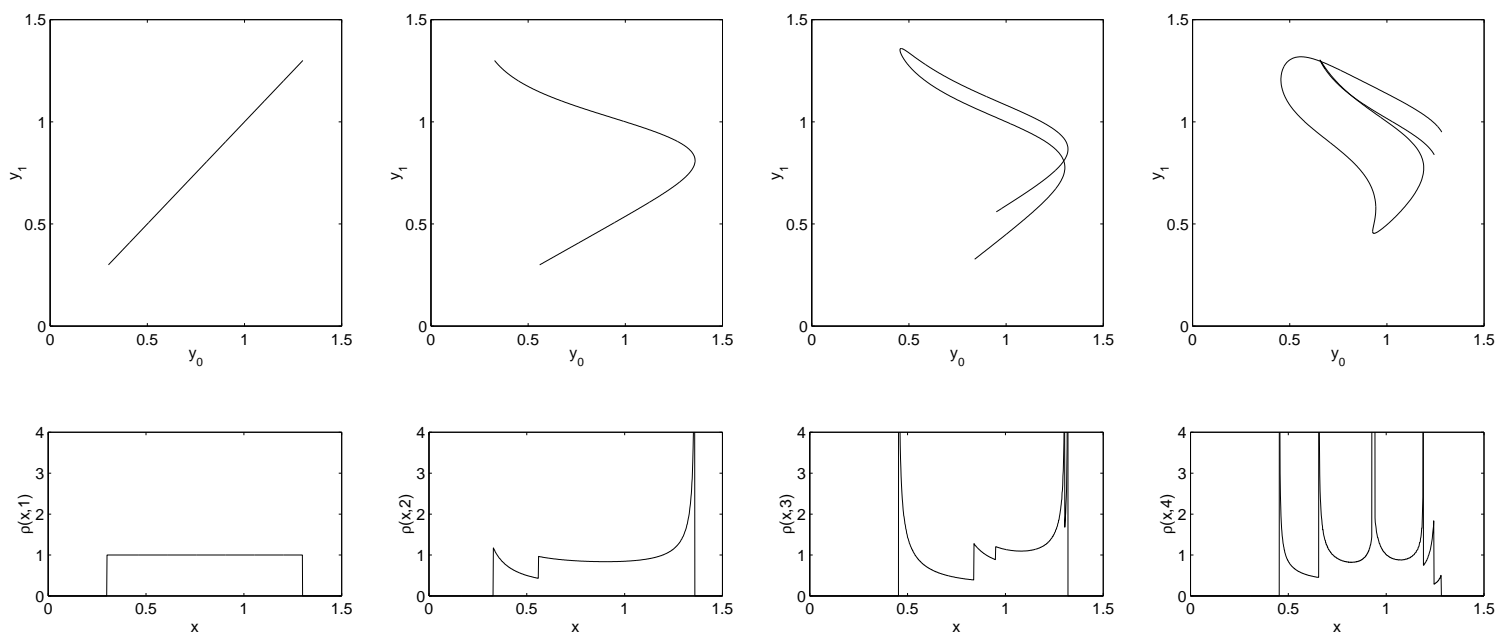
Figure 4.15: Density support curves (projected onto the (y_0, y_1)-plane) for the Mackey-Glass equation (4.2) restricted to constant initial functions. The second row of graphs shows the corresponding computed densities from Figure 4.6.

For large t the transformed density support curve becomes very complicated. Figure 4.16 shows the density support curve at time $t = 20$, which follows in the sequence of Figure 4.15. As with the solution map shown of Figure 4.8, the complexity of this curve results from the repeated stretching and folding that occurs under the dynamics of the DDE. Because of this complexity it is difficult to provide a clear geometric interpretation of the corresponding density, as was possible for small times as in Figure 4.15. Also, just as with the solution map, determining the density support curve numerically becomes problematic (in fact impossible using finite precision) for large times.
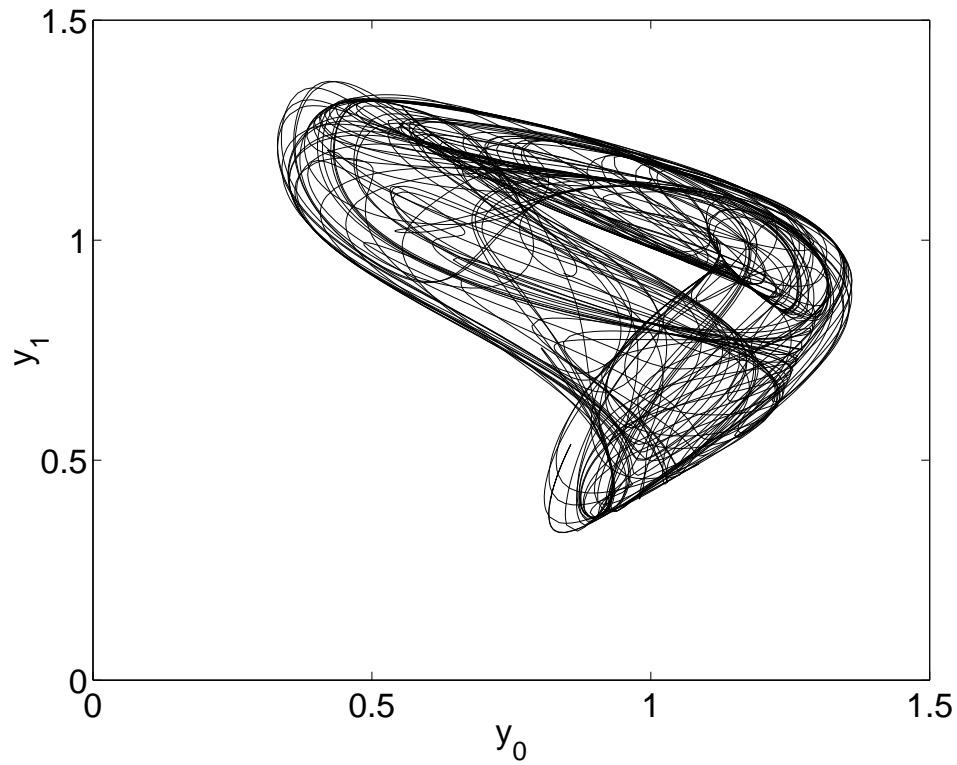
Figure 4.16: Density support curve (projected onto the (y_0, y_1)-plane) for the Mackey-Glass equation (4.2) restricted to constant initial functions, evolved forward in time to $t = 20$.

4.6   Conclusions

This chapter has developed a number of approaches to the evolution of densities for delay equations. To place the problem in a more intuitive and mathematically tractable setting, we have considered delay equations for which ensembles of initial functions are restricted to some finite-dimensional set. This results in a family of measurable solution maps $S_t : \mathbb{R}^n \to \mathbb{R}^n$ for which a Perron-Frobenius operator can be defined. However, this restriction destroys the semigroup structure of the family of solution maps, so any results that are derived for such systems will have limited application to an ergodic theory of delay equations.

At least for simple DDEs, the analytical techniques considered in sections 4.2 and 4.5 can be used to derive explicit formulae for the Perron-Frobenius operator corresponding to S_t (examples 4.2.1–4.2.2). For more complicated equations, non-invertibility of S_t makes both methods difficult to apply. As such, there appears to be little hope, using these methods, of an analytical approach to the evolution of densities for delay equations with interesting statistical properties, such as the Mackey-Glass equation and others that exhibit chaotic solutions. Nevertheless, Section 4.5 provides an intuitive model for the evolution of densities for DDEs, in terms of the transportation of a line mass by a sequence of flows in $\mathbb{R}^N$. This model gives some geometrical insight into results of numerical approaches applied to more complicated DDEs (*e.g.*, Figure 4.15).

In the absence of a generally applicable analytical method, it is desirable to have an effective computational (numerical) approach to the evolution of densities for DDEs. Of the numerical methods considered (Sections 4.3 and 4.4), the simplest is the "brute force" method of simulating a large ensemble of solutions, for an ensemble of initial values chosen at random in accordance with the initial density. Because this method relies on adequate statistical sampling to obtain accurate results, it is computationally intensive to the point of being impractical for many applications. This limitation motivates Section 4.4, which develops a method based on a piecewise linear approximation of the solution map S_t, and provides a much more efficient approach to computing the evolution of densities.

Much of the interest in a probabilistic approach to delay equations is with regard to their asymptotic statistical properties. Asymptotic densities, such as those observed for the Mackey-Glass equation in Figure 4.3, quantify states of statistical equilibrium. That is, they describe the long-term equilibrium distribution of ensembles of systems governed by DDEs. The same densities also characterize the long-time statistics of individual so-

lutions. Thus invariant densities are important from the points of view of both statistical mechanics and ergodic theory.

Asymptotic densities can be found by evolving an initial density forward to large time until the asymptotic statistics become apparent. Unfortunately none of the methods considered in this chapter, other than the "brute force" ensemble simulation method, is well suited to evolving densities to large times. As the time increases, so does the complexity of the solution map for the DDE (*cf.* Figures 4.8 and 4.16, and the examples of Section 4.2). The dimension of the system increases with time as well (*cf.* Section 4.5). These seem to be fundamental obstacles to developing effective (*i.e.* fast) numerical techniques for evolving densities to large times and thereby obtaining asymptotic densities. None of the methods developed so far provides a viable alternative to the computationally intensive ensemble simulation approach. Other approaches to obtaining asymptotic densities, which do not rely on evolving densities to large times, are investigated in next chapter.

Chapter 5

Asymptotic Densities

Chapter 4 considered density evolution for the augmented DDE problem

$$
x'(t) = \begin{cases} g\big(x(t)\big) & t \in [0,1] \\ f\big(x(t), x(t-1)\big) & t \geq 1, \end{cases} \tag{5.1}
$$
$$
x(0) = x_0,
$$

which determines the evolution of a quantity $x(t) \in \mathrm{I\!R}^n$.[1] The role of the function g here is to restrict the DDE to a particular family of allowable initial functions parametrized by x_0, each a solution of $x' = g(x)$ on $[0,1]$. An ensemble of initial values x_0 with density ρ_0 generates a corresponding ensemble of solutions x of (5.1), with density $\rho(x,t)$ at time t. Chapter 4 considered the problem of determining the evolution of this density.

Numerical simulation of solution ensembles for the Mackey-Glass equation (*cf.* Figure 4.3, page 78) suggests that $\rho(x,t)$ approaches a limiting density $\rho_*(x)$ as $t \to \infty$. In fact, in multiple simulations with different initial densities and different families of allowable initial functions (corresponding to different choices of the function g), this same limiting density is observed. It appears, then, that the asymptotic density ρ_* is an intrinsic property of the DDE. This same phenomenon can be observed in other delay equations, of which section 5.1 presents some examples. These observations motivate the present chapter, the purpose of which is to formulate an interpretation of asymptotic densities for DDEs and to investigate methods for computing such densities.

Section 5.2 suggests a theoretical framework to account for the existence of asymptotic densities for delay equations. We find that the existence of ρ_* is consistent with the existence of an SRB measure (Definition 2.27), μ, for the corresponding infinite dimensional dynamical system. In this interpretation ρ_* can be seen as the projection of μ onto the finite-dimensional space $\mathrm{I\!R}^n$ in which the solution variable (*i.e.*, the physical state) $x(t)$ is observed.

The existence of an asymptotic density has practical significance in that it characterizes the asymptotic behavior of any ensemble governed by a given DDE. Moreover, ρ_* also appears to characterize the asymptotic statistics of every "typical" solution of the DDE. In light of the important role played by asymptotic densities, it is desirable to characterize and if possible compute them. Sections 5.3–5.4 explore the problem of computing asymptotic densities for DDEs. Two methods are considered, both based

---

[1]As before, we assume that sufficient conditions are satisfied to guarantee existence and uniqueness of solutions, as well as continuity of solutions with respect to the initial value.

on previously published techniques that have proved successful in the context of some finite-dimensional dynamical systems. Section 5.3 presents an adaptation of the well-studied "Ulam's method", while section 5.4 develops a "self-consistent Perron-Frobenius operator" method.

5.1   Existence of Asymptotic Densities: Ensemble Simulation

The simplest and most direct approach to estimating asymptotic densities is to actually simulate large ensembles of solutions and investigate their asymptotic statistics, as we have done already in Figure 4.3 for the Mackey-Glass equation. Even if finite-precision numerical simulations do not give meaningful predictions of the fate of individual solutions, there is reason to be optimistic about the accuracy of statistics collected on large ensembles of solutions (see section 2.6).

In each of the following examples an ensemble of 10^6 solutions has been simulated for a given delay equation[2], and histograms constructed from the ensemble of solution values $x(t)$ for some large t. In each case this histogram, which approximates the ensemble density $\rho(x,t)$, appears to approach a limiting density $\rho_*(x)$ as $t \to \infty$. This same limiting density is obtained independent of the initial density and the particular family of allowable initial functions (determined by g in (5.1)).

It must be emphasized that asymptotic regularity of the ensemble dynamics does not imply regularity of individual solutions. On the contrary, statistical regularity is closely tied to disordered behavior of individual trajectories (*cf.* section 2.4). The asymptotic density characterizes a statistical rather than a dynamical equilibrium. To emphasize this point, in each example below we illustrate a single long-time solution typical of other solutions represented in the ensemble, confirming that the solutions themselves appear to exhibit a random character, despite the eventual invariance the ensemble density.

Example 5.1.1 (Mackey-Glass equation). Densities for the Mackey-Glass equation [86]

$$x'(t) = -\alpha x(t) + \beta \frac{x(t-1)}{1 + x(t-1)^n},$$
$$\alpha = 2, \quad \beta = 4, \quad n = 10,$$

(5.2)

---

[2]Numerical solutions have been performed using the numerical DDE solver DDE23 [107]

were estimated by ensemble simulation in Chapter 4 (*cf.* Figures 4.2–4.3). The histogram approximating the limiting density ρ_*, together with part of a typical asymptotic solution, is shown in Figure 5.1

Example 5.1.2 (Piecewise-constant nonlinearity). The delay equation

$$x'(t) = -\alpha x(t) + F\big(x(t-1)\big),$$
$$F(x) = \begin{cases} c & \text{if } x \in [x_1, x_2] \\ 0 & \text{otherwise.} \end{cases} \tag{5.3}$$

has been studied previously in [4, 5, 7, 81]. Despite the simplicity of the piecewise-constant feedback term, solutions of this equation are known to exhibit a wide variety of behaviors as the parameters are varied.

Equation (5.3) can be reduced to an ordinary differential on a sequence of intervals, on each of which it easy to construct an analytical solution, *viz.*,

$$x(t) = \begin{cases} \frac{c}{\alpha} + \big(x(t_0) - \frac{c}{\alpha}\big)e^{-\alpha(t-t_0)} & \text{if } x(t-1) \in [x_1, x_2] \\ x(t_0)e^{-\alpha(t-t_0)} & \text{otherwise.} \end{cases} \tag{5.4}$$

The dependence on $x(t-1)$ occurs only through the "crossing times" at which $x(t-1) = x_1$ or x_2 (where the forcing term switches on or off). In fact, the solution of (5.3) for $t \geq 1$ is uniquely determined by the values $x(0)$, $x(1)$, and the set of crossing times in the interval $[0, 1]$. This simplification facilitates an analytical treatment, to the extent that the existence of limit cycles, a homoclinic orbit, and chaos (in the sense of Li and Yorke [76]) have been proved for certain parameters [5]. It has also been proved [4] that, for certain parameters, the map governing the evolution of the crossing times is exact (*cf.* definition 2.26, page 33). To date this is the only rigorous result on strong ergodic properties for a delay differential equation.

Figure 5.2 shows the asymptotic density obtained by a histogram of 10^6 long-time solutions of (5.3), with parameter values $x_1 = 1$, $x_2 = 2$, $\alpha = 6$, $c = 24$ (which were also considered in [5]). The typical form of an asymptotic solution is also shown.

Example 5.1.3 ("Tent map" nonlinearity). The delay equation

$$\varepsilon x'(t) = -x(t) + 1 - 1.9|x(t-1)|. \tag{5.5}$$
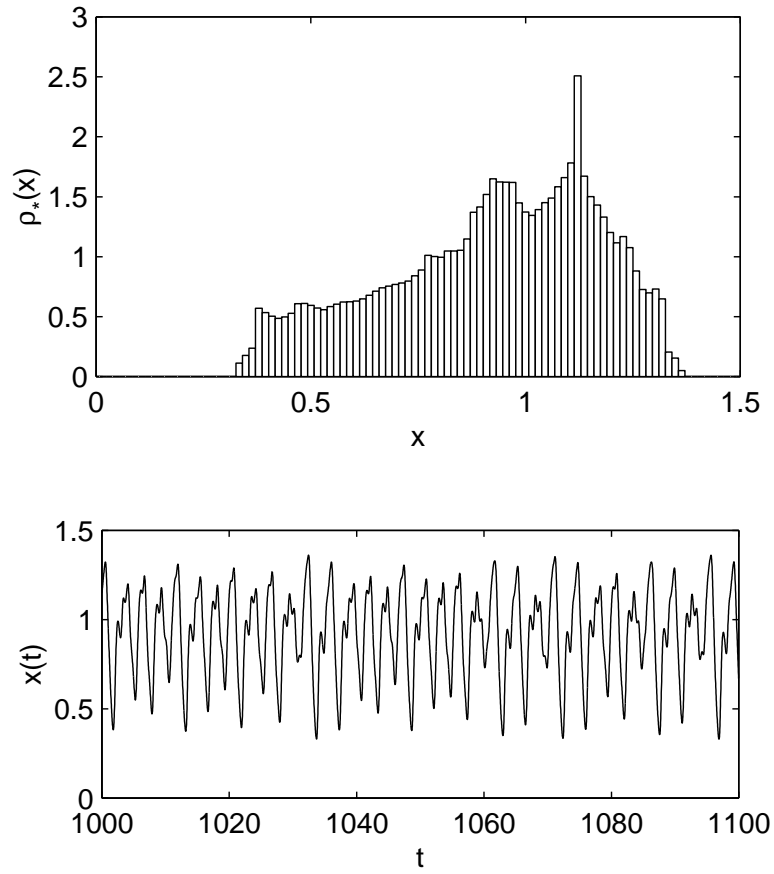
Figure 5.1: Histogram approximating the asymptotic density for the Mackey-Glass equation (5.2), obtained from a simulated ensemble of 10^6 large-time numerical solutions. Also shown is a single solution typical of those represented in the ensemble at large times.
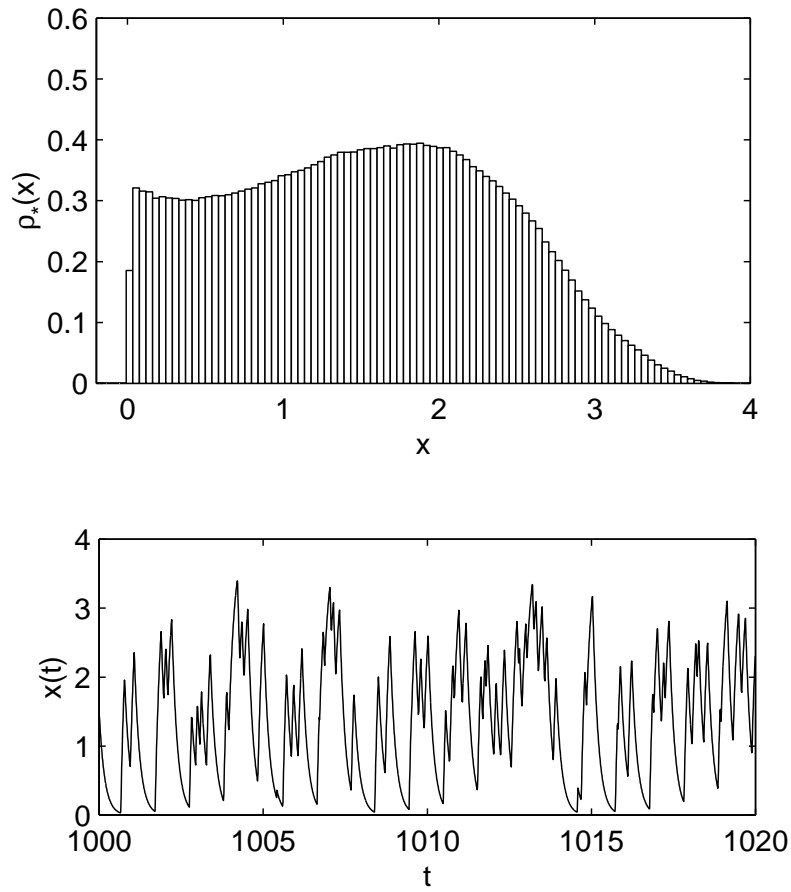
Figure 5.2: Histogram approximating the asymptotic density for the delay equation (5.3) with piecewise-constant feedback, obtained from a simulated ensemble of 10^6 large-time numerical solutions. Also shown is a typical solution represented in the ensemble.
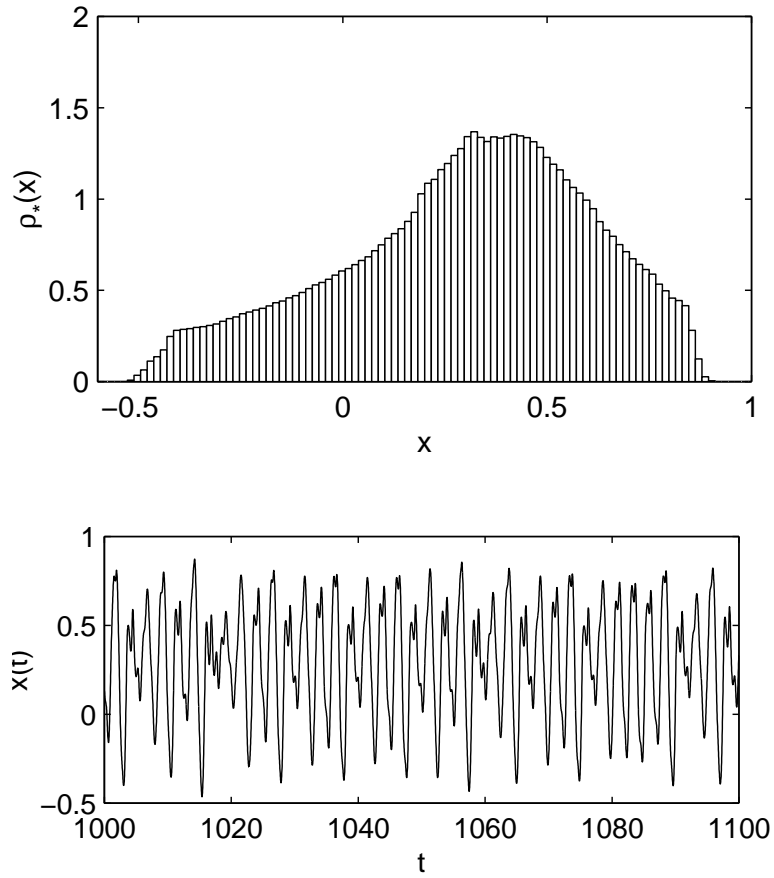
Figure 5.3: Histogram approximating the asymptotic density for the delay equation (5.5) with piecewise-linear feedback ($\varepsilon = 0.3$), obtained from a simulated ensemble of 10^6 large-time numerical solutions. Also shown is a typical solution represented in the ensemble.

has been previously been studied from a probabilistic point of view in [41]. For sufficiently small ε the solutions of DDE (5.5) appear to have chaotic solutions. Figure 5.3 shows the asymptotic density obtained by a histogram over 10^6 large time solutions, together with a segment of a typical long-time solution when $\varepsilon = 0.3$.

5.2 Ergodic Theoretic Interpretation

The existence of asymptotic densities in the examples above can be explained in terms of ergodic properties of the corresponding dynamical system on $C([-1, 0])$. While we are not in a position to prove that a given DDE actually possesses strong ergodic properties, this at least provides one framework for interpreting the dynamics of ensembles of DDE solutions.

5.2.1 DDE solution is the trace of a dynamical system

Recall (*cf.* Chapter 3) that the DDE

$$x'(t) = f\big(x(t), x(t-1)\big), \quad x(t) \in \mathbb{R}^n, \quad t \geq 0, \tag{5.6}$$

can be interpreted as a dynamical system on the phase space C of continuous functions from $[-1, 0]$ into $\mathbb{R}^n$. The corresponding semigroup $\{S_t : t \geq 0\}$ of evolution operators $S_t : C \to C$ defined by

$$(S_t\phi)(s) = x(t+s), \quad s \in [-1, 0], \tag{5.7}$$

where x is the solution of (5.6) with initial function

$$x(t) = \phi(t), \quad t \in [-1, 0]. \tag{5.8}$$

For any given initial function $\phi \in C$ there corresponds a trajectory $\{S_t\phi : t \geq 0\} \subset C$ for this dynamical system. Let $x_t \in C$ denote the phase point on this trajectory at time t. That is, x_t is the function

$$x_t(s) \equiv (S_t\phi)(s) = x(t+s), \quad s \in [-1, 0]. \tag{5.9}$$

Then $x(t)$ can be expressed as

$$x(t) = \pi(x_t), \tag{5.10}$$

where the functional $\pi : C \to \mathbb{R}^n$ is given by

$$\pi(u) = u(0). \tag{5.11}$$

This gives an interpretation of the solution variable $x(t)$ as the image under π of the phase point of the corresponding dynamical system.

In general if $\{S_t : t \geq 0\}$ is a dynamical system on Y then a function $h : \mathbb{R} \to X$ is called a *trace* of $\{S_t\}$ if there is a continuous map $\pi : Y \to X$ and a $y \in Y$ such that

$$h(t) = \pi\big(S_t(y)\big), \quad t \geq 0. \tag{5.12}$$

In other words, $h(t)$ is the continuous image under π of some trajectory of the dynamical system. A familiar example of a trace is the projection onto two dimensions of a trajectory of a three-dimensional dynamical system. See [74, p. 193] for further discussion of this concept.

Thus we can interpret a solution $x(t)$ of the DDE (5.6) as a trace of the corresponding dynamical system. Indeed, equation (5.7) together with (5.11) gives

$$x(t) = \pi(S_t\phi), \quad t \geq 0, \tag{5.13}$$

so $x(t)$ is the trace of the trajectory of $\{S_t\}$ through ϕ. It is readily verified that π is continuous if C is equipped with the sup norm. With this interpretation it is straightforward to show how various properties of the dynamical system $\{S_t\}$ are manifested as corresponding properties of solutions of the DDE. In particular we have the following:

- If $\{S_t\}$ has an invariant measure, μ, then for an ensemble of phase points $x_t \in C$ distributed according to μ, the corresponding ensemble of DDE solutions $x = \pi(x_t)$ will be distributed according to the probability measure $\pi(\mu) \equiv \mu \circ \pi^{-1}$ on $\mathbb{R}^n$, and this distribution will be invariant under the dynamics.

- If $\{S_t\}$ has an attractor $\Lambda \subset C$, then $x(t)$ lies asymptotically on the image $\pi(\Lambda) \subset \mathbb{R}^n$.

- If Λ supports an SRB measure μ, then we expect that any solution ensemble will be asymptotically distributed according to the measure $\pi(\mu)$. This provides an explanation of the convergence of ensemble histograms in the examples above.

The following sections explore these connections in greater detail.

5.2.2   Evidence of an invariant measure

In the examples above, by simulating ensembles of solutions we have found evidence for the existence of asymptotic densities for delay equations. That is, for a given density of

initial values x_0 in the DDE initial value problem (5.1), the density $\rho(x,t)$ of the ensemble of solution values x at time t evolves toward a seemingly unique density $\rho_*(x)$ as $t \to \infty$.

It is tempting to use the terminology of chapter 2 and call the limiting density ρ_* an "invariant density". This turns out to be inappropriate, since ρ_* is in fact not truly invariant in the strict sense already defined: if an ensemble of initial values x_0 is distributed with density ρ_*, the subsequent evolution of the ensemble density $\rho(x,t)$ does *not* agree with ρ_* for all time. Figure 5.4 illustrates this fact. Here we consider the Mackey-Glass equation (5.2) restricted to constant initial functions (hence $g = 0$ in (5.1)). An ensemble of 10^6 solutions has been simulated, corresponding to an ensemble of initial values x_0 distributed according to the asymptotic density ρ_* shown in Figure 5.1. From the resulting sequence of histograms, which approximate $\rho(x,t)$ at times $t = 0, 1, 2$ and 100, it is apparent that $\rho(x,t)$ initially diverges from ρ_*. Hence ρ_* is not appropriately described as an invariant density.

Nevertheless, as can be seen from the histogram representing $\rho(x, 100)$ in Figure 5.4, $\rho(x,t)$ does appear to eventually converge to ρ_* once again. As t increases beyond about 100 the ensemble histograms (not shown here) agree with ρ_*. Thus it seems appropriate to call ρ_* an "asymptotically invariant" density.

This phenomenon has a simple explanation if the infinite dimensional dynamical system corresponding to the delay equation has an invariant measure. Indeed, suppose the dynamical system $\{S_t : t \geq 0\}$ (as defined in equation (5.7)) has an invariant measure μ_*, *i.e.*,

$$\mu_* = S_t(\mu_*) \equiv \mu_* \circ S_t^{-1} \quad \forall t \geq 0. \tag{5.14}$$

If an ensemble of phase points $x_t \in C$ is distributed according to μ_*, then the corresponding ensemble of solution values $x(t) \in \mathbb{R}^n$ will be distributed according to the measure

$$\eta_* = \pi(\mu_*) \equiv \mu_* \circ \pi^{-1}, \tag{5.15}$$

since $x(t) = \pi(x_t)$ where the trace map $\pi : C \to \mathbb{R}^n$ is defined by equation (5.11). Under the evolution prescribed by the delay equation the distribution of solution values $x(t)$ does not change with time since, under evolution by S_t, η_* transforms to

$$\pi\big(S_t(\mu_*)\big) = \pi(\mu_*) = \eta_*. \tag{5.16}$$

This is just what we see with the asymptotic histograms shown in Figures 5.1–5.3, and
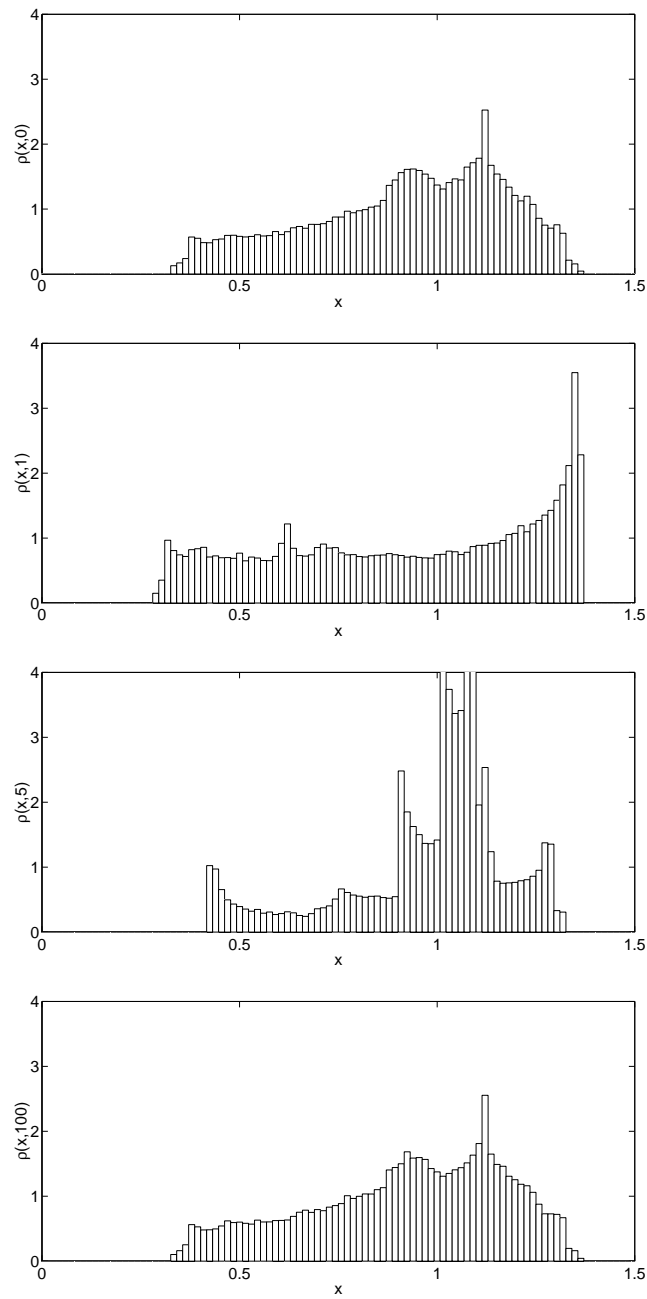
Figure 5.4: Histograms of solution values $x(t)$ for an ensemble of 10^6 solutions of the Mackey-Glass equation (5.2), corresponding to an ensemble of constant initial functions with values distributed according to the asymptotic density ρ_* depicted in Figure 5.1. Histograms are shown for times $t = 0, 1, 5, 100$.

suggests the interpretation of these histograms as approximating measures η_* that are just projections of *invariant* measures μ_* on the phase space C.

As discussed above, η_* cannot be considered an invariant measure for the DDE, since an ensemble of initial values distributed according to η_* does not necessarily remain distributed according to η_*. This can now be understood as a consequence of the fact that the trace map π is not one-to-one. That is, there can be measures on C, other than μ_*, whose images under π coincide with η_*. In general these measures will not be invariant. In particular, there is an ensemble of constant initial functions with values distributed according to η_* (corresponding to the first histogram in Figure 5.4), but this ensemble is not invariant under S_t. As this ensemble evolves, its projection under π (corresponding to the subsequent histograms in Figure 5.4) diverge from η_*.

The foregoing considerations show that by interpreting a DDE solution $x(t)$ as the trace of the corresponding dynamical system $\{S_t\}$ in C, we obtain a framework in which statistical properties of DDE solutions can be understood in terms of statistical properties of $\{S_t\}$. In particular, existence of an invariant measure μ_* for $\{S_t\}$ implies existence of a corresponding measure η_* on $\mathbb{R}^n$ that is invariant under the DDE dynamics. This suggests one explanation for the origin of the asymptotic densities shown in Figures 5.1–5.3. The following sections show how the trace map π carries over other properties of trajectories in C to corresponding properties of DDE solutions in $\mathbb{R}^n$.

5.2.3 Existence of an attractor

Theorem 5.1. *Suppose $F : \mathbb{R} \to \mathbb{R}$ maps an interval $I = (-k, k)$ into itself for some k. Let the semigroup $\{S_t : t \geq 0\}$ be defined by (5.7) for the delay equation*

$$x'(t) = -\alpha x(t) + F\big(x(t-1)\big), \quad t \geq 1 \tag{5.17}$$

with $\alpha \geq 1$. Then $\{S_t\}$ has an attractor, Λ. That is [103], there is a compact set $\Lambda \subset C$ and a neighborhood U of Λ such that

(a) For every neighborhood V of Λ, $S_t(U) \subset V$ for all sufficiently large t.

(b) $S_t(\Lambda) = \Lambda \ \forall t \geq 0$.

(c) $\Lambda = \bigcap_{t \geq 0} S_t(U)$.

Proof. (Adapted from a sketch given in [41]). By [103, Prop. 3.2] it is sufficient to show that for some open $U \subset C$, for all $t > 1$ $S_t(U)$ is relatively compact and contained in U.

To this end let

$$U = \{u \in C : |u| < k/\alpha\}. \tag{5.18}$$

Recall (Section 3.4.1) that the time-one map $S_1 : C \to C$ can be written

$$(S_1 u)(s) = u(0)e^{-\alpha(s+1)} + \int_{-1}^{s} e^{\alpha(t-s)} F\big(u(t)\big)\, dt, \quad s \in [-1,0]. \tag{5.19}$$

Thus for $u \in U$ we have

$$
\begin{aligned}
\big|(S_1 u)(s)\big| &\le \Big|u(1)e^{-\alpha(s+1)}\Big| + \Big|\int_{-1}^{s} e^{\alpha(t-s)} F\big(u(t)\big)\, dt\Big| \\
&< \frac{k}{\alpha} e^{-\alpha(s+1)} + \int_{-1}^{s} e^{\alpha(t-s)} k\, dt \\
&= \frac{k}{\alpha} e^{-\alpha(s+1)} + \frac{k}{\alpha}\big(1 - e^{-\alpha(s+1)}\big) \\
&= k/\alpha.
\end{aligned} \tag{5.20}
$$

By (5.7) we have, $\forall t \in [0,1]$,

$$(S_t \phi)(s) = \begin{cases} \phi(s+t) & \text{if } s \in [-1, -t] \\ (S_1 \phi)(s + t - 1) & \text{if } s \in (-t, 0] \end{cases} \tag{5.21}$$

so that $|(S_t u)(s)| < k/\alpha \ \forall t \in [0,1]$. The semigroup property then implies

$$S_t(U) \subseteq U \quad \forall t \ge 0. \tag{5.22}$$

Furthermore, if $\phi \in U$ then the solution $x(t)$ of (5.17) with initial function ϕ satisfies, for $t \in [0,1]$,

$$
\begin{aligned}
|x'(t)| &\le |-\alpha x(t)| + |F\big(\phi(t-1)\big)| \\
&\le \alpha |x(t)| + k,
\end{aligned} \tag{5.23}
$$

so for $u \in U$ we have

$$(S_t u)'(s) \le \alpha \frac{k}{\alpha} + k = 2k. \tag{5.24}$$

Thus for $t \ge 1$ every element of $S_t(U)$ has a bounded derivative, hence the set $S_t(U)$ is equicontinuous and therefore relatively compact (by the Arzela-Ascoli Theorem [73,

p. 57]). Thus by [103, Prop. 3.2] the set

$$\Lambda = \bigcap_{t \geq 0} S_t(U) \tag{5.25}$$

is compact and satisfies (a)–(c) above. □

In each of the examples given in Section 5.1 the delay equation satisfies the conditions of the theorem above (in particular it suffices that F be bounded), so the corresponding dynamical system has a compact attractor $\Lambda \subset C$. Thus for any initial function ϕ in some open ball $U \subset C$, the trajectory $\{S_t\phi : t \geq 0\} \subset C$ lies asymptotically on (or near) Λ. The proof gives an explicit formula for the radius of U in terms of α and the radius of the "maximal invariant interval" I such that $F(I) \subseteq I$. The full basin of attraction of Λ, $W = \bigcup_{t \geq 0} S_t^{-1}(U)$, might actually be much larger than U.

Since any initial function $\phi \in W$ has $\mathrm{dist}(S_t(\phi), \Lambda) \to 0$ as $t \to \infty$, continuity of the trace map π implies that

$$\mathrm{dist}\big(x(t), \pi(\Lambda)\big) = \mathrm{dist}\big(\pi(S_t\phi), \pi(\Lambda)\big) \to 0 \quad \text{as} \quad t \to \infty. \tag{5.26}$$

That is, the solution $x(t)$ corresponding to the initial function ϕ lies asymptotically in the image of Λ under the mapping $\pi : C \to \mathbb{R}^n$.

If S_t possesses an invariant measure μ_* describing the asymptotic statistics of its trajectories then, since trajectories in C lie asymptotically on Λ, μ_* will be concentrated on Λ. According to the previous section, there is a corresponding measure $\eta_* = \pi(\mu_*)$ on $\mathbb{R}^n$ that is invariant under the dynamics, and this measure will be concentrated on $\pi(\Lambda)$. In particular, as the following section shows, if Λ carries an SRB measure that characterizes the distribution of orbits on Λ, then the image of this measure under π describes the asymptotic statistics of typical solutions $x(t)$.

5.2.4 Evidence of an SRB measure

Each asymptotic density ρ_* shown in Figures 5.1–5.3 is constructed from an ensemble of solutions $x(t)$ of a given DDE, evaluated at a particular (large) time. If we instead sample values $\{x_n = x(nh) : n = 0, \dots, N\}$ along a *single* solution $x(t)$, where h is some fixed time increment (*e.g.* the time step for numerical integration), the histogram of these values approaches ρ_* as $N \to \infty$.

Figure 5.5 illustrates this phenomenon. Here we consider the DDE (5.5) with piecewise-linear feedback, restricted to constant initial functions (hence $g = 0$ in (5.1)). For an arbitrarily chosen initial value x_0 in the DDE problem (5.1) we have computed a single numerical solution and constructed histograms as described above. Comparison with Figure 5.3 shows good agreement between the asymptotic histogram obtained as $N \to \infty$, and the asymptotic histogram obtained by ensemble simulation. This behavior can also be seen in the DDEs of examples 5.1.1 and 5.1.2. Moreover, the asymptotic histogram thus obtained seems to be independent of the initial value x_0, with the exception of initial values on equilibrium solutions of the DDE (*e.g.* $x_0 = 0$ generates the zero solution of the Mackey-Glass equation (5.2), hence a trivial histogram concentrated at the origin).

Convergence of a histogram along a single solution requires existence of the time average

$$\lim_{N \to \infty} \frac{\#\{n = 1, \ldots, N : x_n \in A\}}{N} = \lim_{N \to \infty} \frac{1}{N} \sum_{k=1}^{N} 1_A(x_k), \qquad (5.27)$$

for each histogram bin $A \subset \mathbb{R}$. This behavior is indeed expected, for solutions corresponding to μ-almost every initial function ϕ, if the DDE's attractor supports an ergodic invariant measure μ on C (*cf.* equation (2.46), page 31). However, the set of allowable (*e.g.* constant) initial functions selected by (5.1) are not on the attractor, hence not in the support of the supposed ergodic measure μ. The fact that the time average (5.27) exists all the same suggests a property stronger than ergodicity. We conjecture that the density ρ_* characterizes the asymptotic statistics of every "typical" solution $x(t)$, with "typical" taken in the sense of "on a set of positive Lebesgue measure". This hypothesis is similar to the existence of an SRB measure (*cf.* Section 2.4.4).

If the sequence $\{x_n = x(nh) : n \geq 0\}$ does in fact have a well-defined asymptotic distribution according to a probability measure η_*, then convergence of histograms of $\{x_n\}$ to η_* can be expressed as.

$$\frac{1}{N} \sum_{n=1}^{N} \delta_{x_n}(A) \longrightarrow \eta_*(A) \quad \text{as} \quad N \to \infty, \qquad (5.28)$$

for each measurable A, where δ_{x_n} is the probability measure corresponding to a point mass at x_n:

$$\delta_x(A) = \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{otherwise.} \end{cases} \qquad (5.29)$$
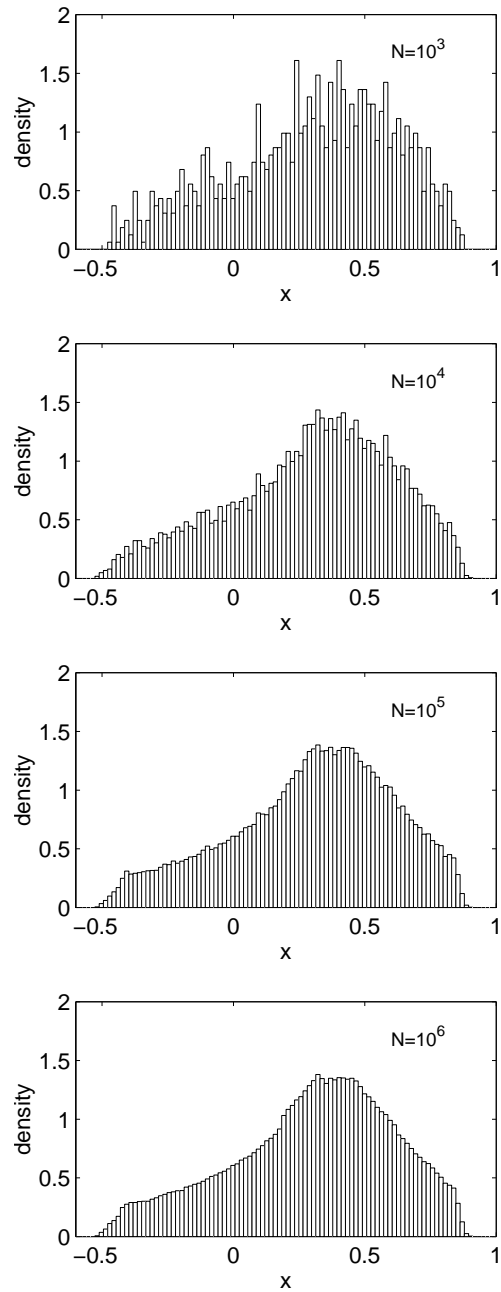
Figure 5.5: Histograms of solution values $\{x_n = x(nh) : n = 0, \ldots, N\}$ for a single numerical solution of the DDE (5.5) with fixed time step h.

The authors of [14] show that (5.28) implies

$$\frac{1}{N}\sum_{i=1}^{N}\varphi(x_n) \longrightarrow \int \varphi \, d\eta_*. \tag{5.30}$$

for any bounded continuous function $\varphi : \mathbb{R} \to \mathbb{R}$. Thus we have (supposedly) that the average of φ along any typical solution $x(t)$ is given by the average of φ with respect to η_*. Equation (5.30) is just defining condition for η_* to be an SRB measure (*cf.* Section 2.4.4 page 34), except that $\{x_n\}$ is not the orbit of a dynamical system, but rather the trace of a dynamical system.

In fact, if the dynamical system $\{S_t\}$ corresponding to the DDE does possess an SRB measure μ_*, then η_* is just the image of μ_* under the trace map π. To see this, suppose μ_* is an SRB measure for $\{S_t\}$, *i.e.*, for any bounded continuous functional $\psi : C \to \mathbb{R}$,

$$\lim_{N\to\infty}\frac{1}{N}\sum_{k=1}^{N}\psi(S^k\phi) = \int_C \psi \, d\mu_* \tag{5.31}$$

for all initial functions ϕ in a set of positive m-measure.[3] Here $S = S_h$ is the time-h map. Let x be the DDE solution corresponding to initial function ϕ. Then for the particular functional $\psi = \varphi \circ \pi$, where $\varphi : \mathbb{R}^n \to \mathbb{R}$ is bounded and continuous, and $\pi : C \to \mathbb{R}^n$ is the trace map (5.11), equation (5.31) gives

$$\begin{aligned}\lim_{N\to\infty}\frac{1}{N}\sum_{k=1}^{N}\varphi(x_n) &= \lim_{N\to\infty}\frac{1}{N}\sum_{k=1}^{N}\varphi\big(\pi(S^k(\phi))\big) \\ &= \int_C \varphi \circ \pi \, d\mu_* \\ &= \int_{\mathbb{R}^n} \varphi \, d(\mu_* \circ \pi^{-1}).\end{aligned} \tag{5.32}$$

The last line follows from measurability of π and Theorem 2.3 (page 21). Comparison with equation (5.30) gives (via the Riesz Representation Theorem [55])

$$\eta_* = \pi(\mu_*) \equiv \mu_* \circ \pi^{-1}. \tag{5.33}$$

---

[3]We suppose m is a measure on C that provides the relevant notion of "almost every". As discussed in Section 3.6 the appropriate choice of m is ambiguous, so we leave it unspecified.

That is, η_* is just the image of μ_* under the trace map $\pi : C \to \mathbb{R}^n$. Thus the existence of an SRB measure μ_* for $\{S_t\}$ implies the convergence of histograms, and more generally the existence of time averages (5.30), where the asymptotic the measure η_* can now be interpreted as the projection $\pi(\mu_*)$ of μ_* onto $\mathbb{R}^n$.

The supposed existence of an SRB measure for the infinite dimensional dynamical system $\{S_t\}$ is only a conjecture supported by numerical evidence. At present a proof appears to be unattainable. Indeed, justifying (5.31) is a formidable task even for finite-dimensional dynamical systems [40, 120]. For infinite dimensional systems, even the definition of SRB measure, and in particular the appropriate choice of reference measure m, is ambiguous. Nevertheless, supposing the existence of an SRB measure does provide a plausible framework that explains the apparent existence of asymptotic densities for ensembles of solutions of some DDEs. This model is helpful to the discussion in the following sections, where we consider methods of computing these asymptotic densities.

5.2.5   Higher dimensional traces

The trace map $\pi : C \to \mathbb{R}^n$ defined by (*cf.* Section 5.2.1)

$$\pi(u) = u(0) \tag{5.34}$$

is a natural way to construct finite-dimensional images of objects in C. In particular
if $\Lambda \subset C$ is an attractor for the dynamical system $\{S_t\}$ corresponding to the delay
equation (5.1) then $\pi(\Lambda)$ gives a finite-dimensional picture of Λ, in the space $\mathbb{R}^n$ of the
solution variable $x(t)$. If $\{S_t\}$ has an invariant measure μ then the measure $\pi(\mu)$ on $\mathbb{R}^n$
describes the statistics of an ensemble of solutions $x(t)$, and is also invariant under the
dynamics. Since π is many-to-one, some information is lost in these projections. Indeed,
if $x(t) \in \mathbb{R}$ then $\pi(\Lambda)$ is just an interval, and gives little information about the structure
of Λ.

   If one was to choose a different mapping $\pi : C \to \mathbb{R}^M$ with $M > n$ then intuitively
$\pi(\Lambda)$ should give a more accurate picture of Λ. Ideally we would like π to be one-to-one
on Λ, in which cased case the image $\pi(\Lambda)$ is called an *embedding* of Λ in $\mathbb{R}^M$ [59, 106].
The following theorem establishes that for M sufficiently large, almost *any* $\pi : C \to \mathbb{R}^M$
in a certain class will yield an embedding of a given finite-dimensional set in C.

Theorem 5.2 (after [59]). *Let X be a Banach space, $A \subset X$ a compact set with box-
counting dimension D. If $M > 2D$ then almost every[4] bounded linear function $\pi : X \to
\mathbb{R}^M$ is one-to-one on A.*

   To illustrate how a multi-dimensional trace map can be used to visualize the attractor
of a DDE, we consider the mapping $\pi : C \to \mathbb{R}^2$ defined by

$$\pi(u) = \big(u(-1), u(0)\big). \tag{5.35}$$

If $x_t = S_t(\phi) \in C$ is a phase point on a trajectory of $\{S_t\}$ then we have (by equation (5.9))

$$
\begin{aligned}
\pi(x_t) &= \big(x_t(-1), x_t(0)\big) \\
         &= \big(x(t-1), x(t)\big)
\end{aligned}
\tag{5.36}
$$

where $x(t)$ is the solution of the DDE with initial function ϕ. The image $\pi(\Lambda)$ can be
approximated by computing a numerical solution of the given DDE and plotting the set

---

[4]In the sense of prevalence [58], *cf.* Section 3.6.4.

of points $\{(x(t-1), x(t)) : t \geq T\}$ in the plane (where T is a sufficiently large time for transients to die out, *i.e.* for $S_t(\phi)$ to approach Λ). This amounts to plotting $x(t)$ vs. $x(t-1)$; in the literature this is occasionally done to construct phase plots of DDE solutions in the "pseudo phase space" $\mathbb{R}^2$. The results of this procedure applied to the examples in Section 5.1 are shown in Figures 5.6, 5.8 and 5.10, respectively. Even in these 2-dimensional images an intricate (presumably fractal) structure of the attractor is apparent.

The previous section gave evidence of an SRB measure μ_* supported on Λ, and showed how the histograms depicted in Figures 5.1–5.3 can be interpreted as approximations of the image of μ_* under $\pi : u \mapsto u(0)$. Just as with the attractor Λ, a more accurate image of μ_* is obtained under the 2-dimensional trace map (5.36). The measure $\pi(\mu_*)$ can be approximated by computing a typical numerical solution of the given DDE and constructing a two-dimensional histogram of the sequence of vectors $\{(x(t-1), x(t)) : t > T\}$. The results of this procedure applied to the examples of Section 5.1 are shown in Figures 5.7, 5.9 and 5.11, respectively. In each figure, part of the $(x(t-1), x(t))$-plane is divided into a grid of rectangles (the histogram bins). Each rectangle is uniformly shaded with a level of grayscale intensity that indicates the histogram height, which approximates the measure $\pi(\mu_*)$ of that rectangle.
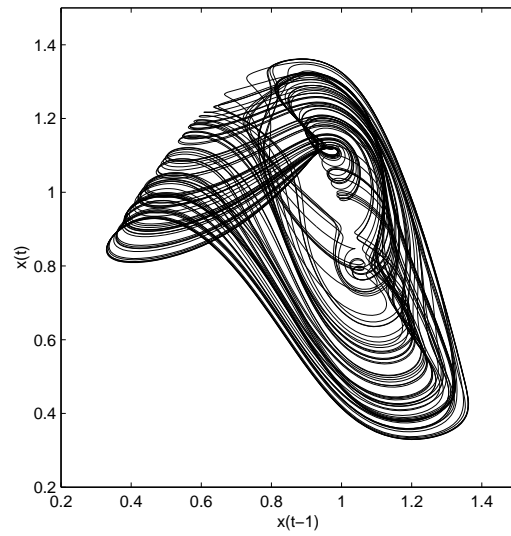
Figure 5.6: Numerical approximation of the image $\pi(\Lambda)$ of the attractor $\Lambda \subset C$ of the Mackey-Glass equation (5.2), under the trace map (5.36).
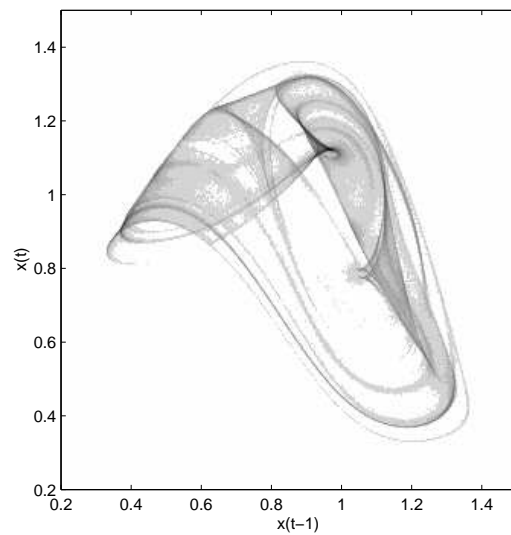
Figure 5.7: Two-dimensional histogram approximating the projection $\pi(\mu_*)$ of the supposed SRB measure μ_* for the Mackey-Glass equation (5.2), under the trace map (5.36).
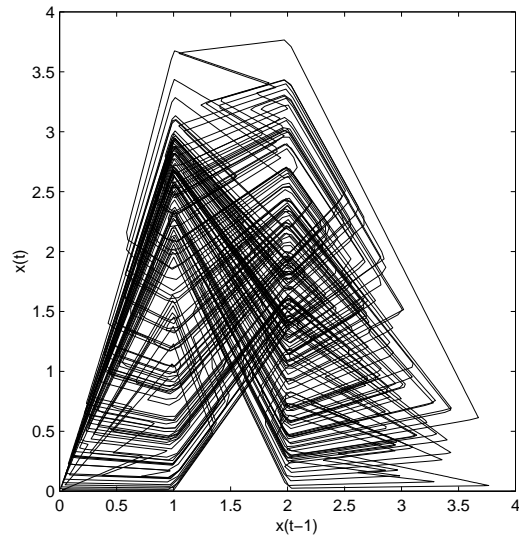
Figure 5.8: Numerical approximation of the image $\pi(\Lambda)$ of the attractor $\Lambda \subset C$ for the delay equation (5.3), under the trace map (5.36).
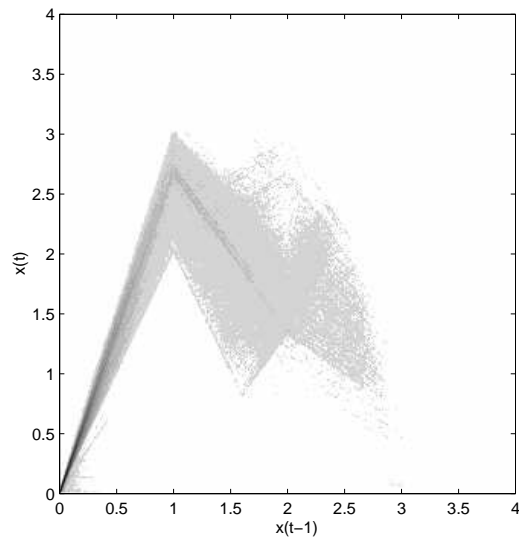
Figure 5.9: Two-dimensional histogram approximating the projection $\pi(\mu_*)$ of the supposed SRB measure μ_* for the delay equation (5.3), under the trace map (5.36).
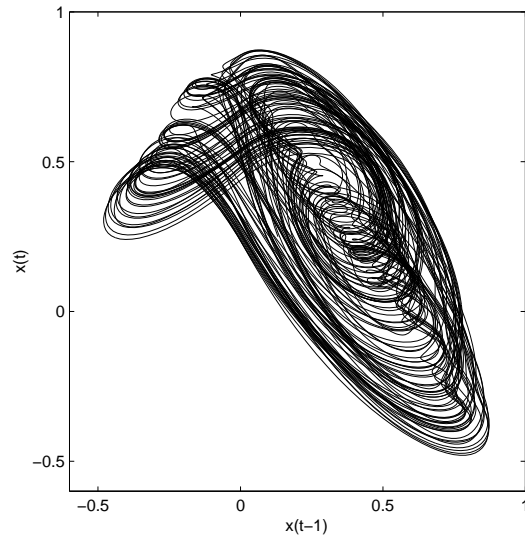
Figure 5.10: Numerical approximation of the image $\pi(\Lambda)$ of the attractor $\Lambda \subset C$ of the delay equation (5.5), under the trace map (5.36).
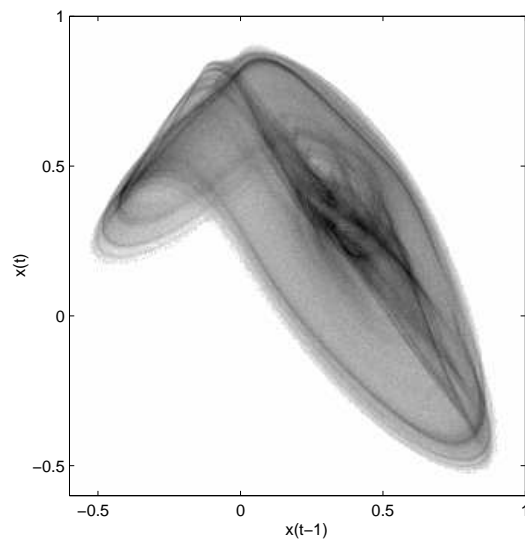
Figure 5.11: Two-dimensional histogram approximating the projection $\pi(\mu_*)$ of the supposed SRB measure μ_* for the delay equation (5.5), under the trace map (5.36).

5.3 Ulam's Method

The remainder of this chapter is concerned with methods for computing asymptotic densities for delay equations. The methods considered so far are of the "brute force" type: simulating large ensembles of solutions (Section 5.1) and computing statistics on a single long solution (Section 5.2.5). The utility of these methods is limited by their large sampling requirements, owing to the slow $O(1/\sqrt{N})$ rate of convergence of histograms as the number N of solutions is increased (see p. 79). In the following sections we seek more efficient methods.

A preliminary observation is in order, related to the remark at the beginning of Section 4.5. The obvious approach to finding asymptotic densities is to begin with an evolution equation for $\rho(x,t)$, of the form

$$\frac{d}{dt}\rho = \{\text{some operator}\}(\rho). \tag{5.37}$$

Invariance of ρ_* could then be characterized by setting the left-hand side equal to zero, resulting in the equation

$$\{\text{some operator}\}(\rho_*) = 0, \tag{5.38}$$

which hopefully could be solved, at least approximately, for ρ_*. This approach fails since, for reasons discussed in Section 4.5, $\rho(x,t)$ cannot be described by an evolution equation of the desired form (or, for that matter, by any evolution equation in terms of $\rho(x,t)$ alone). For that matter, ρ_* cannot even be considered invariant in the above sense (*cf.* Section 5.2.2). At best, we can only apply this approach to an *approximate* evolution equation for ρ. One such possibility is considered in Section 5.4.

An alternative approach, and probably the best known technique for approximating invariant measures for dynamical systems, is Ulam's method [119, 77, 37]. The following section presents the basic idea, after which we consider how the method might be adapted to delay differential equations.

5.3.1 Stochastic approximation of dynamical systems

Let a discrete-time dynamical system be defined by iterates of a map $S : X \to X$ ($X \subset \mathbb{R}^n$) and suppose S has an invariant measure μ (for continuous time systems, the following applies to a suitable discrete-time map, *e.g.* the time-one map). For a given partition $\mathcal{A} = \{A_1, \ldots, A_n\}$ of X, let $p_i(k)$ denote the probability at time k that the

system state $x(k) = S^k(x) \in A_i$.

Ulam's method approximates the evolution of the probability vector $p(k) = (p_1(k), \ldots, p_n(k))$ by a Markov chain

$$p(k + 1) = Pp(k), \tag{5.39}$$

where P is a transition matrix that models the dynamics of S. The basic idea is to ignore the details of the dynamics within each A_i, and instead consider the "coarse-grained" dynamics with respect to the partition. Thus, given that $x(k) \in A_j$, we suppose that $x(k)$ is equally likely to be anywhere in A_j, i.e., $x(k)$ is distributed according to normalized Lebesgue measure λ on A_j. Then the "transition probability" that $x(k + 1) \in A_i$ is

$$P_{ij} = \frac{\lambda(A_j \cap S^{-1}A_i)}{\lambda(A_j)}, \tag{5.40}$$

i.e., the fraction (with respect to Lebesgue measure) of A_j that is mapped by S into A_i. This defines the $n \times n$ transition matrix P for the Markov chain (5.39), a stochastic process that (hopefully) approximates the probabilistic dynamics of S, in the following sense.

Each probability vector $p = p(k)$ defines a probability measure ν on X such that $\nu(A_i) = p_i$, and in general

$$\nu(A) = \sum_{i=1}^{n} \frac{\lambda(A \cap A_i)}{\lambda(A_i)} \cdot p_i. \tag{5.41}$$

This measure has piecewise constant density

$$\rho(x) = \frac{p_i}{\lambda(A_i)} \quad \text{if } x \in A_i. \tag{5.42}$$

Thus the evolution equation (5.39) for $p(k)$ implicitly defines a sequence of piecewise constant densities, approximating a sequence of densities evolving under the action of S. To be more precise, P can be interpreted as a projection of the Perron-Frobenius operator $\tilde{P}$ corresponding to S onto the space of piecewise constant densities (with respect to the partition $\mathcal{A}$). Formally, P can be written in terms of $\tilde{P}$ as [77]

$$P(Qf) = Q(\tilde{P}f) \quad \forall f \in L^1(X), \tag{5.43}$$

where the operator Q projects f to the piecewise constant function $Qf \in L^1(X)$ given by

$$(Qf)(x) = \sum_{i=1}^{n} \frac{\int_{A_i} f \, d\lambda}{\lambda(A_i)} \cdot 1_{A_i}(x). \qquad (5.44)$$

Since P approximates the Perron-Frobenius operator $\tilde{P}$, it can be hoped that a fixed point of $\tilde{P}$ (*i.e.*, an invariant density for S) can be approximated by a fixed point of P. That is, suppose that p satisfies

$$p = Pp \qquad (5.45)$$

i.e., p is an eigenvector of P with eigenvalue 1, normalized so that $\sum p_i = 1$. The corresponding piecewise constant density defined by equation (5.42) then approximates the density of the invariant measure μ. Ulam conjectured [119] that as the partition is refined ($\lambda(A_i) \to 0$), the sequence of approximations obtained in this way should converge to the fixed point of $\tilde{P}$, *i.e.*, the true invariant density. This conjecture has in fact been proved for particular cases, such as piecewise expanding transformations on intervals [77] and on rectangles in $\mathbb{R}^n$ [38]

5.3.2 Application to delay equations

In the case of delay differential equations, we are concerned with an invariant measure μ for a dynamical system S_t on the space $C([-1, 0])$. If μ has support in some bounded set $X \subset C$ (*e.g.*, when μ is an SRB measure supported on a compact attractor Λ) then in principle Ulam's method could be applied to the time-one map $S = S_1$ and for some partition $\mathcal{A}$ of X. However, it is unclear (*cf.* Chapter 3) what measure on C is an appropriate analog of Lebesgue measure in the definition of the transition matrix P (equation (5.40)). For now, suppose we do have some such reference measure, m, such that $0 < m(A_i) < \infty$ for each i. Then, in analogy with (5.40) the matrix P with elements

$$P_{ij} = \frac{m(A_j \cap S^{-1}A_i)}{m(A_j)} \qquad (5.46)$$

defines a Markov chain $p(k+1) = Pp(k)$ that hopefully models the dynamics of S. In particular, we can hope that a fixed point p of P approximates the invariant measure μ on $\mathcal{A}$, *i.e.*, $p_i \approx \mu(A_i)$.

Suppose, in accordance with the framework of Section 5.2, that $\eta_* = \pi(\mu_*)$ is the measure on $\mathbb{R}^n$ corresponding to the asymptotic density ρ_* (where $\pi : C \to \mathbb{R}^n$ is defined

by (5.11)) and that the support of η_* is contained in some bounded $B \subset \mathrm{I\!R}^n$. Then, for a given partition $\mathcal{B} = \{B_1, \ldots, B_r\}$ of B, a careful choice of $\mathcal{A}$ permits an interpretation p as a piecewise constant approximation of ρ_* with respect to $\mathcal{B}$. The following shows how this can be done.

Note that η_* has support in the bounded set $B = \pi(X)$. Let $\mathcal{B} = \{B_1, \ldots, B_r\}$ be a partition of B on which we wish to approximate η_* by a piecewise constant density. Define a partition $\mathcal{A}$ of X by

$$A_i = \pi^{-1}(B_i), \quad i = 1, \ldots, r, \tag{5.47}$$

or more explicitly,

$$A_i = \{u \in C : u(0) \in B_i\}. \tag{5.48}$$

Then, since $\eta_* = \mu_* \circ \pi^{-1}$, we have

$$\mu_*(A_i) = \eta_*(B_i). \tag{5.49}$$

Thus if we use Ulam's method to obtain a vector $p = Pp$ of probabilities $p_i \approx \mu_*(A_i)$ then we also have $p_i \approx \eta_*(B_i)$, yielding a piecewise constant approximation

$$\rho_*(x) \approx \frac{p_i}{\lambda(B_i)} \quad \text{if } x \in B_i \tag{5.50}$$

of the density ρ_*.

The immediate difficulties in implementing this method are in evaluating the elements of the transition matrix (equation (5.46)), where we must compute the pre-images $S^{-1}(A_i)$ under the infinite dimensional map S (*e.g.*, as given explicitly in equation (5.19)) and evaluate the reference measure m of $S^{-1}(A_i)$. Both of these difficulties can be circumvented if we take as the reference measure m the invariant measure μ_*. If μ_* is an SRB measure (which seems to be the case in the examples of Section 5.1) then for any bounded continuous $\psi : C \to \mathrm{I\!R}$ we have

$$\lim_{N \to \infty} \frac{1}{N} \sum_{k=1}^{N} \psi\big(S^k(\phi)\big) = \int \psi \, d\mu_* \tag{5.51}$$

for every initial function ϕ in some set of positive m-measure. In particular, for $\psi = 1_A$

we have[5]

$$\frac{1}{N} \sum_{k=1}^{N} 1_A(S^k(\phi)) = \frac{1}{N} \#\{k \in 1, \dots, N : S^k(\phi) \in A\}$$

$$\xrightarrow{N \to \infty} \int 1_A \, d\mu_* = \int_A d\mu_* = \mu_*(A). \tag{5.52}$$

Then for sufficiently large N and for any typical initial function ϕ, we can approximate $\mu_*(A_i)$ by

$$\mu_*(A_i) \approx \frac{1}{N} \#\{k \in 1, \dots, N : S^k(\phi) \in A_i\}$$

$$= \frac{1}{N} \#\{k \in 1, \dots, N : (S^k \phi)(0) \in B_i\} \tag{5.53}$$

$$= \frac{1}{N} \#\{k \in 1, \dots, N : x_k \in B_i\},$$

where $x_k = \big(S^k(\phi)\big)(0) = x(k)$ and $x(t)$ is the solution of the DDE with initial function ϕ. Thus, to construct the transition matrix P, we compute (*e.g.*, numerically) a long sequence $\{x_k = x(k) : k = 1, \dots, N\}$ along a single solution $x(t)$ of the DDE, and approximate P_{ij} in equation (5.46) by (see also *e.g.* [15])

$$P_{ij} = \frac{\#\{k : x_k \in B_j \text{ and } x_{k+1} \in B_i\}}{\#\{k : x_k \in B_j\}}, \tag{5.54}$$

assuming $\mathcal{B}$ has been chosen so that $\{x_k\} \cap B_j \neq \emptyset \ \forall j$.

The transition matrix P defines a Markov chain that approximates the asymptotic probabilistic dynamics of the given DDE. In particular, we hope that a fixed point p of P will provide a piecewise constant approximation of the asymptotic density ρ_*, via (5.50). Indeed, when applied to the examples of Section 5.1 this method identically reproduces the asymptotic density found by computing a histogram of $\{x_k\}$ (*i.e.* as in Section 5.2.4, Figure 5.5).

It turns out that this must be the case. Let $p = (p_1, \dots, p_r)$ represent the normalized

---

[5]There is a technical difficulty here: 1_A is not continuous so (5.51) does not strictly apply with $\psi = 1_A$. However, we can approximate 1_A from below by continuous functions $\{\psi_n\}$ with $\psi_n \to 1_A$ μ_*-almost everywhere, so $\int \psi_n \, d\mu_* \to \int 1_A \, d\mu_*$ by the Lebesgue dominated convergence theorem [74, p. 22]. See [71, p. 134] for further details.

histogram of $\{x_k\}$ taken with respect to the partition $\mathcal{B}$, *i.e.*,

$$p_i = \frac{1}{N}\#\{k \in 1, \ldots, N : x_k \in B_i\} \tag{5.55}$$

A simple calculation shows that

$$
\begin{aligned}
(Pp)_i &= \sum_{j=1}^{r} P_{ij}p_j \\
&= \sum_{j=1}^{r} \frac{\#\{k : x_k \in B_j \text{ and } x_{k+1} \in B_i\}}{\#\{k : x_k \in B_j\}} \cdot \frac{\#\{k : x_k \in B_j\}}{N} \\
&= \sum_{j=1}^{r} \frac{\#\{k : x_k \in B_j \text{ and } x_{k+1} \in B_i\}}{N} \\
&= \frac{\#\{k : x_{k+1} \in B_i\}}{N} \\
&= \frac{\#\{k : x_k \in B_i\} \pm 1}{N} \\
&= p_i \pm \frac{1}{N} \;\; \longrightarrow \;\; p_i \text{ as } N \to \infty.
\end{aligned}
\tag{5.56}
$$

Thus p is (almost) a fixed point of P, and Ulam's method as applied here simply reproduces the results of computing a histogram along a single solution of the given DDE. Essentially, the construction is circular and no new information about the asymptotic density is gained.

For lack of any other reasonable reference measure m on C that we can evaluate (or approximate), our formulation of Ulam's method for DDEs does not provide an independent estimate of the asymptotic density. However, the Markov chain $p \mapsto Pp$ defined by (5.54) is interesting in its own right, as a simple model of the asymptotic probabilistic dynamics of the given DDE. An intuitive way to represent such a Markov chain is to graph the matrix of transition probabilities P. Figures 5.12–5.14 give examples of such plots, with P computed as in equation (5.54), for each of the delay equations considered in the examples of Section 5.1. For each figure the support of the asymptotic density in $\mathbb{R}$ has been partitioned into 100 intervals B_i, $i = 1, \ldots, 100$, of equal length. Each rectangle $B_j \times B_i$ in the plane is shaded uniformly with grayscale level indicating the probability P_{ij} of transition from $x(t) \in B_j$ to $x(t+1) \in B_i$. Thus darker rectangles indicate likely transitions; white regions indicate transitions that never occur, at least asymptotically.
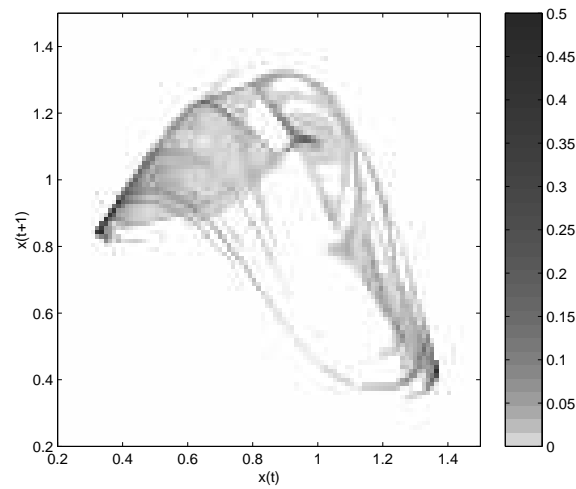
Figure 5.12: Graphical representation of the matrix P of transition probabilities (defined by equation (5.54)) for the Markov chain approximating the asymptotic dynamics of the Mackey-Glass equation (5.2).

Figures 5.12–5.14 are remarkably similar to Figures 5.7, 5.9 and 5.11. There is in fact an intimate connection between these figures, owing to the fact that the numerator in equation (5.54) is equivalent to a two-dimensional histogram of the sequence of vectors $\{(x(t), x(t-1)) : t = 0, 1, \ldots\}$, with bins $\{B_j \times B_i : i, j = 1, \ldots, N\}$. The entries of the matrix P are therefore identical to the heights of the corresponding two-dimensional histograms in Figures 5.7, 5.9 and 5.11, except that each column of P is normalized so that $\sum_{i=1}^{r} P_{ij} = 1$.
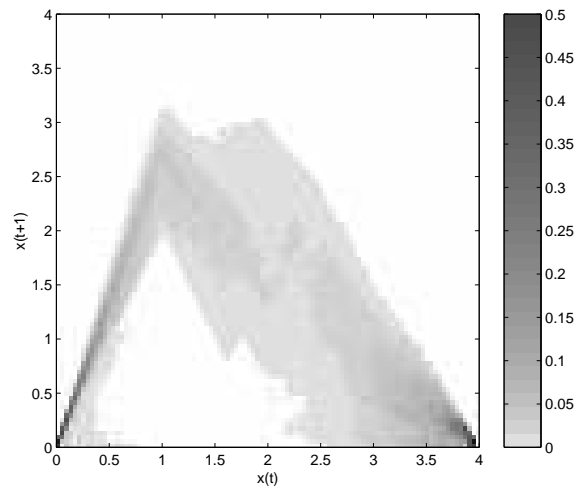
Figure 5.13: Graphical representation of the matrix P of transition probabilities (defined by equation (5.54)) for the Markov chain approximating the asymptotic dynamics of the delay equation (5.3).
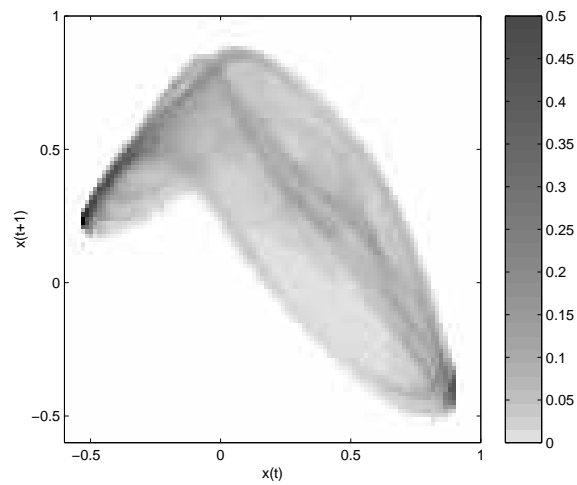
Figure 5.14: Graphical representation of the matrix P of transition probabilities (defined by equation (5.54)) for the Markov chain approximating the asymptotic dynamics of the delay equation (5.5).

5.4 Fixed Points of Approximate Markov Operators

Ulam's method and its generalizations are the only techniques we are aware of for approximating invariant measures of general dynamical systems. However, Lepri *et al.* [75] present a method specifically aimed at approximating invariant measures for discrete-time systems with delayed dynamics, based on a similar method for coupled map lattices [69]. This seems like a promising approach to asymptotic densities for delay differential equations. In this section we consider the application of the method in [75] to a discretized approximation of a particular class of DDEs.

5.4.1 Approximate Markov operator

Consider the DDE[6]

$$x'(t) = -\alpha x(t) + f\big(x(t-1)\big), \quad t \geq 0, \quad x(t) \in \mathbb{R}, \tag{5.57}$$

with initial function

$$x(t) = \phi(t), \quad t \in [-1, 0]. \tag{5.58}$$

Euler discretization of (5.57) with time step $h = 1/N$ yields

$$\frac{x(t+h) - x(t)}{h} \approx -\alpha x(t) + f\big(x(t-1)\big), \tag{5.59}$$

which gives the explicit formula

$$x_{n+1} = (1 - \alpha h)x_n + hf(x_{n-N}) \tag{5.60}$$

for the approximate solution $x_n = x(t_n)$ at the "mesh points" $t_n = nh$, $n = 0, 1, 2, \ldots$. Together with initial values $x_i = \phi(t_i)$, $i = -N, \ldots, 0$, this formula can be iterated to construct a sequence $\{x_n : n = 0, 1, 2, \ldots\}$ that approximates the solution of (5.57)–(5.58).

The problem we consider here is to estimate the asymptotic density $\rho_*(x)$ (supposing one exists) of an ensemble of systems evolving under (5.60). This is just the problem considered in [75] for more general discrete-time systems with delayed dynamics. Their approach is easily adapted to the particular system (5.60) as follows.

---

[6]The ideas of this section generalize in a straightforward way to DDEs in $\mathbb{R}^n$, but this requires more complex notation. To simplify the presentation we will restrict our attention to DDEs in one dimension only.

With

$$y_n = x_{n-N} \tag{5.61}$$

equation (5.60) becomes

$$x_{n+1} = T(x_n, y_n), \tag{5.62}$$

where

$$T : (x, y) \mapsto (1 - \alpha h)x + hf(y). \tag{5.63}$$

If $f : \mathbb{R} \to \mathbb{R}$ is measurable then T is a measurable, nonsingular transformation from $\mathbb{R}^2$ into $\mathbb{R}$. Let $\rho_2(x, y; n)$ be the density at time n of the pair (x_n, y_n), for an ensemble of sequences governed by (5.60). Then in analogy with the definition of the Perron-Frobenius operator, the ensemble of values $x_{n+1} = T(x_n, y_n)$ will be distributed with one-dimensional density $\rho_1(x; n + 1)$ satisfying

$$\int_A \rho_1(x; n + 1) \, dx = \int_{T^{-1}(A)} \rho_2(x, y; n) \, dx \, dy \qquad \forall \text{ Borel } A \subset \mathbb{R}. \tag{5.64}$$

For $A = (-\infty, s)$ we have

$$T^{-1}(A) = \left\{ (x, y) \in \mathbb{R}^2 : x < \frac{s - hf(y)}{1 - \alpha h} \right\}, \tag{5.65}$$

so that

$$\int_{-\infty}^{s} \rho_1(x; n + 1) \, dx = \int_{-\infty}^{\infty} \int_{-\infty}^{\frac{s - hf(y)}{1 - \alpha h}} \rho_2(x, y; n) \, dx \, dy. \tag{5.66}$$

Differentiating with respect to s yields the explicit formula

$$\rho_1(x; n + 1) = \frac{1}{1 - \alpha h} \int_{-\infty}^{\infty} \rho_2\left(\frac{x - hf(y)}{1 - \alpha h}, y; n\right) dy. \tag{5.67}$$

If the densities $\rho_{1*}(x)$ and $\rho_{2*}(x, y)$ are invariant under the process defined by (5.60) then we can drop the dependence on n to yield

$$\rho_{1*}(x) = \frac{1}{1 - \alpha h} \int_{-\infty}^{\infty} \rho_{2*}\left(\frac{x - hf(y)}{1 - \alpha h}, y\right) dy. \tag{5.68}$$

As it stands, this equation cannot be used on its own to determine ρ_{1*}, since it requires prior knowledge of the 2-dimensional density ρ_{2*}. With a similar approach it is possible write an analogous equation defining ρ_{2*}, but this in turn requires knowledge of the 3-

dimensional density ρ_{3*} of the triple (x_n, x_{n-N}, x_{n-2N}). In general, the evolution equation for the density ρ_k of the k-tuple $(x_n, x_{n-N}, \ldots, x_{n-(k-1)N})$ requires knowledge of $\rho_{(k+1)}$, so that we obtain an open recursion relation for the corresponding invariant densities ρ_{k*}.

To close this open recursion relation so it can be solved for ρ_{1*}, one can make an approximation whereby for some k, ρ_{k*} can be expressed in terms of the $\{\rho_{j*} : j \leq k\}$. The simplest such approximation is the factorization

$$\rho_{2*}(x, y) = \rho_{1*}(x)\rho_{1*}(y). \tag{5.69}$$

This amounts to assuming that x and y are independent, *i.e.*, uncorrelated. It is a "self-consistent" approximation, in that x and y are both supposed to be distributed according to ρ_{1*}. This is exactly what we expect if ρ_{1*} is an invariant density: if x_n is distributed according to ρ_{1*} for all sufficiently large n, then so must be $y_n = x_{n-N}$. This approximation closes equation (5.68), which becomes

$$\rho_{1*}(x) = \frac{1}{1 - \alpha h} \int_{-\infty}^{\infty} \rho_{1*}\left(\frac{x - hf(y)}{1 - \alpha h}\right) \rho_{1*}(y)\, dy \equiv (Q\rho_{1*})(x). \tag{5.70}$$

By its construction Q maps densities to densities, but it is not a Markov operator since it is nonlinear. Nevertheless, in [69] the operator analogous to Q is called a "self-consistent Perron-Frobenius operator". Q can be interpreted intuitively as the operator that effects the evolution of densities under the action of (5.60) with the assumption that at each time step, x_{n-N} is a random variable independent of x_n and distributed with the same density as x_n.

Since Q approximates, in some sense, the probabilistic dynamics of the Euler discretization of the given DDE, it is hoped that a density that is invariant under Q (*i.e.*, a fixed point of Q) will approximate the asymptotic density ρ_*, *e.g.* as observed in Figures 5.1–5.3. One approach to approximating the solution of the operator equation $\rho_* = Q\rho_*$ is fixed point iteration: if a sequence of densities $\{u_{n+1} = Qu_n\}$ can be found that converges in L^1, then the limit $\rho_* = \lim_{n\to\infty} u_n$ furnishes a solution of (5.70).[7] In practice this iteration is carried out numerically.

---

[7]Actually this requires continuity of Q, which seems to require restricting Q to L^∞. So far we have not found a satisfactory proof.

5.4.2   Numerical implementation

The integral in (5.70) resembles a convolution. In fact by the change of variables

$$v(x) = \frac{1}{1 - \alpha h} \, u\Big(\frac{x}{1 - \alpha h} \Big) \tag{5.71}$$

$$w(x) = \sum_{s \in f^{-1}\{x/h\}} \frac{u(s)}{h|f'(s)|}, \tag{5.72}$$

$Q(u)$ can be written as a convolution integral,

$$(Qu)(x) = \int_{-\infty}^{\infty} v(x - z)w(z)dz. \tag{5.73}$$

This integral can be approximated numerically as follows.

Let the densities u, $Q(u)$, and w be approximated by the corresponding vectors of values they assume on a uniform grid $\{x_1, \ldots, x_M\}$,

$$x_i = x_1 + (i - 1)\Delta x, \quad i = 1, \ldots, M, \tag{5.74}$$

and designate a vector of weights $\{\alpha_j\}$ appropriate for numerical quadrature on this grid (*e.g.*, by Simpson's Rule [94, p. 134]). Then $(Qu)(x_i)$ can be approximated by

$$
\begin{aligned}
(Qu)_i = (Qu)(x_i) &\approx \sum_{j=1}^{M} v(x_i - x_j)w(x_j)\alpha_j \\
&= \sum_{j=1}^{M} v\big((i - j)\Delta x\big)w(x_j)\alpha_j \\
&= \sum_{j=1}^{M} v(y_{i-j+M})w(x_j)\alpha_j \\
&= \sum_{j=1}^{M} v_{i+M-j}w_j\alpha_j,
\end{aligned}
\tag{5.75}
$$

where

$$y_k = (k - M)\Delta x, \quad k = 1, \ldots, 2M - 1, \tag{5.76}$$

and the vectors

$$v_k = v(y_k)$$
$$w_j = w(x_j). \tag{5.77}$$

are evaluated according to (5.71), using interpolation of the $u_i = u(x_i)$. The final line of (5.75) is a discrete convolution, representing a moving average of length-M windows of $\{v_k\}$ with respect to the vector of weights $\{w_j \alpha_j\}$. Using standard techniques [94, ch. 12], the $(Qu)_i$ can then be evaluated efficiently using a Fast Fourier Transform.

5.4.3 Case study

The delay equation (5.57) with piecewise linear feedback term

$$f(x) = g(x)/\varepsilon$$
$$g(x) = 1 - 1.9|x| \tag{5.78}$$

and $\alpha = 1/\varepsilon$ was found in Example 5.1.3 to exhibit an asymptotic density ρ_* when $\varepsilon = 0.3$ (*cf.* Figure 5.3). However, when the method described above is applied to this equation, it does not yield an approximation of ρ_*. Instead, for any initial density u, iterating $u \mapsto Qu$ results in convergence toward a point mass concentrated at $x_* = 1/2.9$, which is readily seen to be an unstable fixed point of the map $x \mapsto g(x)$.

Given the success of this method for estimating invariant densities for other systems with delayed dynamics [75] this result is surprising, but a partial explanation can be advanced as follows. The assumption implicit in the factorization (5.69) essentially removes any explicit delay from the dynamics of the discretization (5.60), since the delayed coordinate $y_n = x_{n-N}$ is always assumed to have the same density as x_n. In other words, the delayed coordinate is being modeled by a stochastic variable distributed like x_n. Iteration of $u \mapsto Qu$ gives a probabilistic description of the map $(x, y) \mapsto T(x, y)$ where at every iteration y is assumed to have the same distribution as x. This is sort of (but not quite) like evolving a density under the one-dimensional map $x \mapsto T(x, x)$. From equation (5.63) we have

$$T(x, x) = x + \frac{h}{\varepsilon}\big(g(x) - x\big), \tag{5.79}$$

so that $x \mapsto T(x, x)$ has the same fixed points as g. In particular the stability of the fixed

point $x_* = 1/2.9$ is determined by

$$
\begin{aligned}
\left| \frac{d}{dx_*} T(x_*, x_*) \right| &= \left| 1 + \frac{h}{\varepsilon} \big(g'(x_*) - 1 \big) \right| \\
&= \left| 1 + \frac{h}{\varepsilon} (-1.9 - 1) \right| \\
&= \left| 1 - 2.9 \frac{h}{\varepsilon} \right| < 1 \quad \text{for } h \ll \varepsilon.
\end{aligned}
\tag{5.80}
$$

Thus the fixed point $x_* = 1/2.9$ of T is stable, hence the apparent convergence of densities to a point mass δ_{x_*}.

A similar phenomenon occurs when fixed-point iteration is attempted to solve (5.70) in the case of either the Mackey-Glass equation (5.2) or the delay equation (5.3) with piecewise constant feedback. In neither case does the method yield an approximation of the asymptotic density found by ensemble simulation (*cf.* Figures 5.1 and 5.2), but yields rather a point mass concentrated as a fixed point of $T(x, x)$.

Evidently the instability of the dynamics of the DDE (5.57) is delay-induced. Indeed, with zero delay and f defined as in (5.78) the DDE becomes

$$
x' = F(x) \equiv -\frac{1}{\varepsilon} x + \frac{1}{\varepsilon} \big(1 - 1.9|x| \big),
\tag{5.81}
$$

which has a single stable fixed point $x_* = 1/2.9$. It is easily shown that the situation is similar with the other DDEs considered in examples 5.1.1 and 5.1.2: with zero delay the asymptotic dynamics are trivial. Instability is essential to the existence of a nontrivial invariant density. Since the assumption in (5.69) effectively removes the delay, the resulting condition (5.70) reasonably does not provide an approximation of the observed asymptotic density ρ_*. The examples considered in [75] exhibit chaotic behavior even for zero delay; this helps explain why they did not encounter the difficulties we find here.

5.4.4   Second-order method

A less restrictive assumption than (5.69), specifically one that retains the essential delay in the dynamics, might yield an effective method of approximating ρ_*. One possibility, suggested in [75] as a more accurate variant on the original method, is to make a different approximation that truncates the recursion relationship for the ρ_{k*} at some $k > 2$. The following gives a sketch of how this might be done.

With $y_n = x_{n-N}$ and $z_n = x_{n-2N}$, the discretization (5.60) becomes

$$
\begin{aligned}
(x_{n+1}, y_{n+1}) &= \big((1 - \alpha h)x_n + hf(y_n), (1 - \alpha h)y_n + hf(z_n)\big) \\
&= \big(T(x_n, y_n), T(y_n, z_n)\big) \\
&\equiv \tilde{T}(x_n, y_n, z_n).
\end{aligned}
\tag{5.82}
$$

Then $\tilde{T}$ is a measurable, nonsingular transformation from $\mathbb{R}^3$ into $\mathbb{R}^2$ .Let $\rho_3(x, y, z; n)$ be the density of the triple (x_n, y_n, z_n), for an ensemble of sequences governed by (5.60). The ensemble of pairs (x_{n+1}, y_{n+1}) will be distributed with 2-dimensional density $\rho_2(x, y; n+1)$ satisfying

$$
\int_A \rho_2(x, y; n+1)\, dx\, dy = \int_{\tilde{T}^{-1}(A)} \rho_3(x, y, z; n)\, dx\, dy\, dz \qquad \forall \text{ Borel } A \subset \mathbb{R}^2.
\tag{5.83}
$$

Since

$$
\tilde{T}(x, y, z) \in (-\infty, r] \times (-\infty, s] \implies
\begin{cases}
x < (r - hf(y))/(1 - \alpha h) \\
y < (s - hf(z))/(1 - \alpha h),
\end{cases}
\tag{5.84}
$$

equation (5.83) yields (with $A = (-\infty, r] \times (-\infty, s]$)

$$
\int_{-\infty}^{s} \int_{-\infty}^{r} \rho_2(x, y; n+1)\, dx\, dy = \int_{-\infty}^{\infty} \int_{-\infty}^{\frac{s-hf(z)}{1-\alpha h}} \int_{-\infty}^{\frac{r-hf(y)}{1-\alpha h}} \rho_3(x, y, z; n)\, dx\, dy\, dz.
\tag{5.85}
$$

Differentiating with respect to r and s yields the explicit formula

$$
\begin{aligned}
\rho_2(x, y; n+1) &= \frac{1}{(1 - \alpha h)^2} \int_{-\infty}^{\infty} \rho_3\Big(\frac{x - hf(\frac{y-hf(z)}{1-\alpha h})}{1 - \alpha h}, \frac{y - hf(z)}{1 - \alpha h}, z; n\Big)\, dz \\
&= \frac{1}{(1 - \alpha h)^2} \int_{-\infty}^{\infty} \rho_3\Big(G\big(x, G(y, z)\big), G(y, z), z; n\Big)\, dz
\end{aligned}
\tag{5.86}
$$

where

$$
G(y, z) = (y - hf(z))/(1 - \alpha h).
\tag{5.87}
$$

If the densities $\rho_{2*}(x, y)$ and $\rho_{3*}(x, y, z)$ are invariant under the dynamics then we can drop the dependence on n. To close this equation requires an approximation whereby ρ_{3*}

can be written in terms of ρ_{2*}. One possibility, suggested in [75], is the factorization

$$\rho_{3*}(x, y, z) = \frac{\rho_{2*}(x, y)\, \rho_{2*}(y, z)}{\rho_{1*}(y)}, \tag{5.88}$$

where

$$\rho_{1*}(y) = \int \rho_{2*}(x, y)\, dx. \tag{5.89}$$

Thus both (x_n, x_{n-N}) and (x_{n-N}, x_{n-2N}) are distributed with the same density ρ_{2*}; this is consistent with ρ_{2*} being invariant under the dynamics. Substituting these relationships into (5.86) yields a nonlinear operator equation

$$\rho_{2*} = Q\rho_{2*} \tag{5.90}$$

for the two-point density ρ_{2*}. Solving this equation by fixed point iteration approximates the evolution of a two-point density for the system (5.60), where at each iteration the pair (x_{n-N}, x_{n-2N}) is assumed to have the same density as (x_n, x_{n-N}). Since this method retains the inherent delay in the dynamics, it is possible that (5.90) will have a solution ρ_{2*}, whereby equation (5.89) gives an approximation of the invariant density $\rho_* = \rho_{1*}$. However, the computational complexity is much greater than in the previous method, and to date we have not developed an implementation.

5.4.5   Continuous-time formulation

In the previous sections we considered a discretized version of a given delay differential equation, to which the method developed in [75] could be directly applied. However, it seems more natural to avoid the discretization step altogether. It is in fact possible to formulate a continuous-time approach analogous to the discrete time method in [75], as illustrated below.

Consider the delay equation

$$x'(t) = f\big(x(t), x(t - 1)\big), \quad x(t) \in \mathbb{R}^n, \quad t \geq 0, \tag{5.91}$$

written as

$$x'(t) = f\big(x(t), y(t)\big) \tag{5.92}$$

where $y(t) = x(t - 1)$. For any given value of $y = y(t)$ this is just an ordinary differential

equation prescribing the flow of $x(t)$ along a vector field $x \mapsto f(x, y)$, in which y acts as a fixed parameter. Of course this interpretation is valid only instantaneously, *i.e.*, at a particular time, since $y(t)$ itself changes in time. If $\rho_1(x; t)$ is the density at time t of an ensemble of solutions $x(t)$ that all share the *same* value of $y(t)$, then the transportation of ρ_1 along this flow is described by a continuity equation (*cf.* Section 2.3.2),

$$\frac{\partial \rho_1(x; t)}{\partial t} = -\nabla_x \cdot (\rho_1(x; t) f(x, y)), \tag{5.93}$$

where $\nabla_x = (\partial/\partial x_1, \ldots, \partial/\partial x_n)$.

Suppose now that we have an ensemble of solutions without any restriction on $y(t)$, and that the ensemble of pairs $(x(t), y(t))$ is distributed with 2-dimensional density $\rho_2(x, y; t)$. It is helpful to think of this ensemble as being partitioned into sub-ensembles according to the value of $y(t)$. Then as the ensemble evolves under the action of (5.91), each sub-ensemble contributes an increment to ρ_1 according to (5.93). Proceeding heuristically, the total increment to ρ_1 can be found by summing equation (5.93) over these sub-ensembles, *i.e.*,

$$\frac{\partial \rho_1(x; t)}{\partial t} = -\int \nabla_x \cdot (\rho_2(x, y; t) f(x, y)) \, dy. \tag{5.94}$$

A less heuristic derivation of this result is as follows. Suppose that vectors $(x, y) \in \mathbb{R}^{2n}$ are distributed with density $\rho_2(x, y; t)$, and that x evolves according to equation (5.92). Let

$$\rho_1(x; t) = \int \rho(x, y; t) \, dy \tag{5.95}$$

denote the "collapsed" density of the ensemble of values x. For the purposes of determining the instantaneous increment of $\rho_1(x; t)$ under the action of (5.92), the dynamics of $y(t)$ are a second-order effect and can be ignored. Thus we can make the simplifying assumption that $y'(t) = 0$, and write

$$\frac{d}{dt} \begin{bmatrix} x(t) \\ y(t) \end{bmatrix} = \begin{bmatrix} f(x(t), y(t)) \\ 0 \end{bmatrix} \equiv H(x(t), y(t)), \tag{5.96}$$

which gives the flow of $(x(t), y(t)) \in \mathbb{R}^{2n}$ along the vector field $H : \mathbb{R}^{2n} \to \mathbb{R}^{2n}$. Under transportation by this flow, the density $\rho_2(x, y; t)$ evolves according to the continuity

equation

$$
\begin{aligned}
\frac{\partial \rho_2(x, y; t)}{\partial t} &= -\nabla \cdot \big(\rho_2(x, y; t) H(x, y) \big) \\
&= -\nabla_x \cdot \big(\rho_2(x, y; t) f(x, y) \big) - \nabla_y \cdot \big(\rho_2(x, y; t) \cdot 0 \big) \\
&= -\nabla_x \cdot \big(\rho_2(x, y; t) f(x, y) \big),
\end{aligned}
\tag{5.97}
$$

where $\nabla_y = (\partial/\partial y_1, \dots, \partial/\partial y_n)$. Then we have

$$
\begin{aligned}
\frac{\partial}{\partial t} \rho_1(x; t) &= \int \frac{\partial}{\partial t} \rho_2(x, y; t) \, dy \\
&= -\int \nabla_x \cdot \big(\rho_2(x, y; t) f(x, y) \big) \, dy,
\end{aligned}
\tag{5.98}
$$

in agreement with (5.94).

If the one- and two-point densities $\rho_{1*}(x)$ and $\rho_{2*}(x, y)$ are invariant under the dynamics then equation (5.98) gives

$$
\int \nabla_x \cdot \big(\rho_{2*}(x, y) f(x, y) \big) \, dy = 0.
\tag{5.99}
$$

Assuming again that ρ_{2*} can be factored as in (5.69), this yields the condition

$$
\int \rho_{1*}(y) \, \nabla_x \cdot \big(\rho_{1*}(x) f(x, y) \big) \, dy = 0,
\tag{5.100}
$$

which is analogous to the discrete-time condition (5.70). It is unclear whether this relationship uniquely determines (within a constant multiple) a unique approximate invariant density ρ_{1*}.

However, here again we have resorted to an assumption (equation (5.69)) that effectively removes the explicit delay from the dynamics. As discussed in the previous section, this assumption leads to trivial asymptotic dynamics. Without a more sophisticated approach that retains the essential delay, it seems unwarranted to pursue these ideas further.

5.5 Conclusions

For a variety of delay differential equations, numerically computed solution ensembles appear to converge to an asymptotic distribution, described by an asymptotic measure η on $\mathbb{R}^n$. This phenomenon can be understood in terms of ergodic properties of the associated infinite dimensional dynamical system $\{S_t\}$ on C. In the examples considered, $\{S_t\}$ is known to possess a compact attractor Λ: any ensemble of trajectories starting in the basin of attraction of Λ will, asymptotically, be distributed on Λ. The numerical evidence supports the existence of a natural invariant probability measure (*i.e.*, an SRB measure) μ_* supported on Λ. This serves to explain the convergence of solution ensembles to a particular asymptotic distribution η, as well as the fact that averages along "typical" individual solutions of the DDE coincide with spatial averages or expectations with respect to this same measure.

The practical and theoretical importance of invariant measures, and SRB measures especially, makes the computation of invariant measures for DDEs a desirable goal. However, an effective solution to this problem remains elusive. Previously published methods of estimating invariant measures for dynamical systems do not adapt well to delay equations.

Ulam's method—the most widely known technique for estimating invariant measures—can be formulated for DDEs in such a way that it yields an approximation of the asymptotic measure η. This approximation turns out to be identical to the histogram of a time series generated by a typical solution of the given DDE. Thus, at least in our formulation, Ulam's method *per se* is not a useful approach to DDEs but merely points to the fact that if the desired invariant measure is an SRB measure, then it can be estimated by computing a histogram along a single long-time solution. This is, in fact, a far more efficient method than the ensemble simulation approach, in which on the order of 10^6 individual solutions must be computed.

An alternative approach to estimating invariant measures for DDEs is the "self-consistent Perron-Frobenius operator" method of [75]. A suitable discretization of a given DDE yields a discrete-time system of the type to which this method applies. Somewhat surprisingly, however, a straightforward application of the method fails to generate the desired approximate invariant density.

Chapter 6

Transient Chaos

6.1   Introduction

Studies of chaotic dynamical systems have focused mainly on persistent, or attracting chaos—*i.e.*, on systems that possess a chaotic attractor. The phenomenon of *transient* chaos has aroused less interest despite its ubiquity [70]. Systems exhibiting transient chaos have the distinguishing feature that their evolutions are very irregular (chaotic) during a transient period, but eventually become periodic. This behavior is seen in many physical systems, including fluid dynamics [2, 27] and chaotic scattering [21], as well as in mathematical dynamical systems such as the Hénon map [49, 56], the Lorenz system [78, 124], and the forced damped pendulum [9].

There appear to be a number of underlying universal features of transient chaos, despite the diversity of its manifestations; see [111] for a review. The central notion is the existence in phase space of an unstable invariant set on which the dynamics are chaotic (*e.g.*, in the sense of Li and Yorke [76]). Such a set is called a *strange repeller* or *chaotic saddle*, since the instability is typically of saddle type. Except for its instability, this set plays a role similar to that of a strange attractor: the dynamics on the repeller are closely related to the irregularity of nearby trajectories.

The generally accepted model of the phase space dynamics underlying transient chaos is as follows. A typical initial phase point is attracted, under the system dynamics, along the stable manifold of the chaotic saddle. The trajectory subsequently wanders in a neighborhood of the saddle for some time, during which it exhibits the dynamics associated with the saddle. Eventually it exits along the saddle's unstable manifold, and arrives asymptotically at one of the system's attractors (typically a periodic orbit or equilibrium point, but possibly a chaotic attractor). Because the saddle is the dynamical invariant that determines the behavior of trajectories during their transient phase, its structure and the dynamics on it are the objects of primary interest in the analysis of transient chaos.

To date there has been no published account of transient chaos in delay differential equations.[1] However, a number of published results would suggest that transient chaos occurs in some DDEs. The results that motivated the present study were the observations in [81] and [1][2] of fractal basins of attraction (a hallmark of transient chaos [111]) in delay

---

[1][70] references unpublished work by P. Grassberger and I. Procaccia.

[2]In the DDE studied in [1] fractal basins are present even in the absence of a delay. Our primary interest here is in systems where the chaotic dynamics are "delay induced", *i.e.*, arise only in the presence

equations. Transverse homoclinic orbits, which imply the existence of a chaotic saddle and Yorke-type chaos [51], have been proved to exist in some DDEs [5, 7, 121]. Hale and Sternberg [54] found numerical evidence of transverse homoclinic orbits in the Mackey-Glass equation [86]. These results have all been presented amid discussions of attracting chaos, whereas transient chaos in DDEs has not been specifically investigated.

Aside from the importance of delay equations in describing natural and industrial processes, transient chaos in DDEs has special relevance to the study of infinite dimensional dynamical systems. Because numerical integration of DDEs is relatively easy as compared, for example, with partial differential equations, Farmer [42] pointed out that DDEs make convenient prototypical models for the study of attractors of infinite dimensional systems. In the same spirit, DDEs could serve as simple models for the study of *transient* chaos in infinite dimensional systems.

The following section provides numerical evidence of transient chaos in delay equations of the form

$$x'(t) = -\alpha x(t) + F\big(x(t-1)\big). \tag{6.1}$$

We extend the results of [81] and show how the existence of fractal basins of attraction can be used to find solutions with long-lived chaotic transients in a first-order DDE having only periodic attractors. In addition (and in contradiction with the negative result of [81]) we find parameter sets of the Mackey-Glass equation that yield fractal basins of attraction, and we illustrate the existence of long-lived chaotic transients for this system.

Numerical analysis of transient chaos (*e.g.*, approximation of the saddle, its dimension, Lyapunov exponents, entropy, *etc.*) requires a method for computing trajectories on (or very near) the chaotic saddle. Various methods have been proposed and applied to finite-dimensional systems [9, 70, 90, 110]. There are no published accounts of attempts to apply these methods to infinite dimensional systems such as DDEs. In section 6.4 we develop an adaptation of the "stagger-and-step" method [110] and apply it to the DDEs for which we have found evidence of transient chaos. Having constructed a numerical approximation of the saddle, we illustrate graphical methods for visualizing the saddle, and characterize its geometry quantitatively using standard methods for estimating ergodic parameters such as Lyapunov exponents and fractal dimensions.

---

of an intrinsic delay in the dynamics.

6.2   Evidence of Transient Chaos

6.2.1   Fractal basins of attraction

The (necessarily open) set of all initial phase points eventually asymptotic to a given attractor is called that attractor's *basin of attraction*. In a system with multistability, *i.e.* one that possesses more than one attractor, the points that do not lie in any basin of attraction constitute the *basin boundary*. The basin boundary is necessarily invariant under the system dynamics. If this set is fractal (*i.e.* has non-integer capacity dimension) then the dynamics in a neighborhood of the boundary exhibits sensitivity to initial conditions [87], hence transient chaos. This happens, for example, if there is a "horseshoe" [71, 109] in the dynamics (*e.g.* if there is a transverse homoclinic orbit [51]); in this case the basin boundary has a Cantor-like structure. However, multistability and existence of a fractal basin boundary are not necessary for transient chaos: *e.g.* the Hénon map [56] for some parameter values exhibits transient chaos but has only a single attractor (at infinity) [91].

Unstable invariant sets such as unstable fixed points and chaotic saddles and their stable manifolds must lie within the basin boundary, since they are not in any basin of attraction. The basin boundary can consist entirely of the stable manifolds of unstable invariant sets, but this need not be the case [87].

Multistability and fractal basin boundaries for delay differential equations are reported in [1, 81]. However, the connection of this observation to the possible existence of transient chaos has not been investigated. In the following we reproduce the fractal basin boundaries for the DDE considered in [81], and also provide evidence of fractal basins in the much-studied Mackey-Glass equation [86].

DDE with piecewise-constant feedback

In [81], Losson *et al.* studied the delay equation

$$x'(t) = -\alpha x(t) + F\big(x(t-1)\big),$$

$$F(x) = \begin{cases} c & \text{if } x \in [x_1, x_2] \\ 0 & \text{otherwise.} \end{cases} \tag{6.2}$$
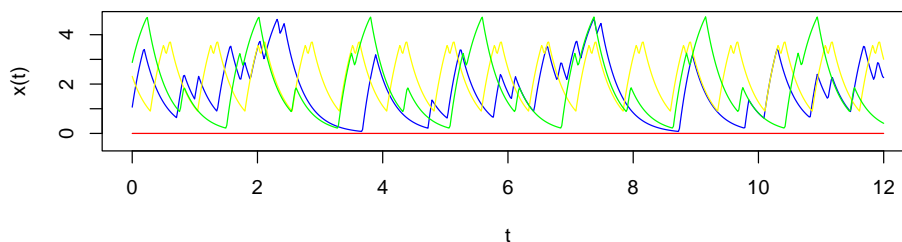
Figure 6.1: Four coexisting attracting periodic solutions (including the trivial solution) of the delay equation (6.2) with parameters $\alpha = 3.25$, $c = 20.5$, $x_1 = 1$ and $x_2 = 2$.

For parameters $\alpha = 3.25$, $c = 20.5$, $x_1 = 1$ and $x_2 = 2$, they found[3] three coexisting attracting periodic solutions. Figure 6.1 illustrates these solutions, together with the zero solution which is also attracting, found by numerically integrating[4] equation (6.2) to large t with different initial functions.

Recall that the phase space of equation (6.2), considered as a dynamical system $\{S_t : t \geq 0\}$ (*cf.* Chapter 3), is the space C of continuous real-valued functions on the interval $[-1, 0]$. The phase point $x_t \in C$ at time t is the solution history,

$$x_t(s) = x(t + s), \quad s \in [-1, 0]. \tag{6.3}$$

Each of the solutions shown in Figure 6.1 corresponds to a periodic orbit in $\Gamma \subset C$. Each such Γ has a corresponding basin of attraction, which is the set of initial functions in C that are asymptotic to Γ under the action of S_t as $t \to \infty$. Thus basins of attraction for DDEs are subsets of the infinite dimensional space C. Consequently, visualizing the basins of attraction and their boundaries is problematic.

One possibility for visualizing the basins of attraction of the DDE (6.2) is to visualize "cross-sections" through C. Consider for example the subspace of C spanned by the

---

[3]There seems to be an error in [81], which gives $\alpha = 3.75$.

[4]Using the Fortran code DKLAG6 [25].

functions $s \mapsto 1$ and $s \mapsto s$, that is, functions of the form

$$\phi(s) = A + Bs, \quad s \in [-1, 0]. \tag{6.4}$$

These constitute a two-dimensional subspace Σ of C, parametrized by coordinates $(A, B) \in \mathbb{R}^2$. For a given attracting periodic orbit Γ of (6.2), the intersection of its basin of attraction with Σ can be approximated numerically and visualized by the following method. For a given point in the (A, B)-plane, numerically integrate (6.2) with the corresponding initial function ϕ of the form (6.4). If the resulting orbit is asymptotic to Γ then ϕ is in the basin of attraction of Γ, so plot the point (A, B). Repeating this procedure for a grid of points in the (A, B)-plane yields a picture approximating part of the basin of attraction's intersection with Σ.

If a different color is associated with each of the various basins of attraction then all four basins of attraction can be visualized on a single graph, as in Figure 6.2. This figure shows the results of the procedure above, carried out for the delay equation (6.2). Here we have made a slight change from equation (6.4) and taken initial functions on the interval $[0, 1]$ of the form

$$x(t) = A + Bt, \quad t \in [0, 1], \tag{6.5}$$

or equivalently

$$\phi(s) = A + B(s + 1), \quad s \in [-1, 0]. \tag{6.6}$$

Figure 6.2 corresponds to Figure 11 of [81], except for a different choice of axes. As already pointed out in [81], the resulting image suggests that the basin boundary is a fractal set with Cantor-like structure. This is supported by evidence given in [54] of the presence of a transverse homoclinic orbit, and implies the existence of transiently chaotic solutions. Indeed in section 6.2.2 we are able, with the help of Figure 6.2, to find solutions of (6.2) that exhibit long chaotic transients.

To illustrate that there is nothing very special about the subspace Σ spanned by functions of the form (6.4), Figure 6.3 shows basins of attraction for the DDE (6.2) for initial functions of the form

$$x(t) = A \cos(2\pi t) + B \sin(2\pi t), \quad t \in [0, 1]. \tag{6.7}$$

This figure exhibits a fractal structure similar to that seen in Figure 6.2.

Figure 6.2: Basins of attraction for the delay equation (6.2) for initial functions of the form $t \mapsto A + Bt$, $t \in [0, 1]$. The basin colors are those of the corresponding solutions shown in Figure 6.1. The second image shows an enlargement, by a factor of 10, of part of the first image.
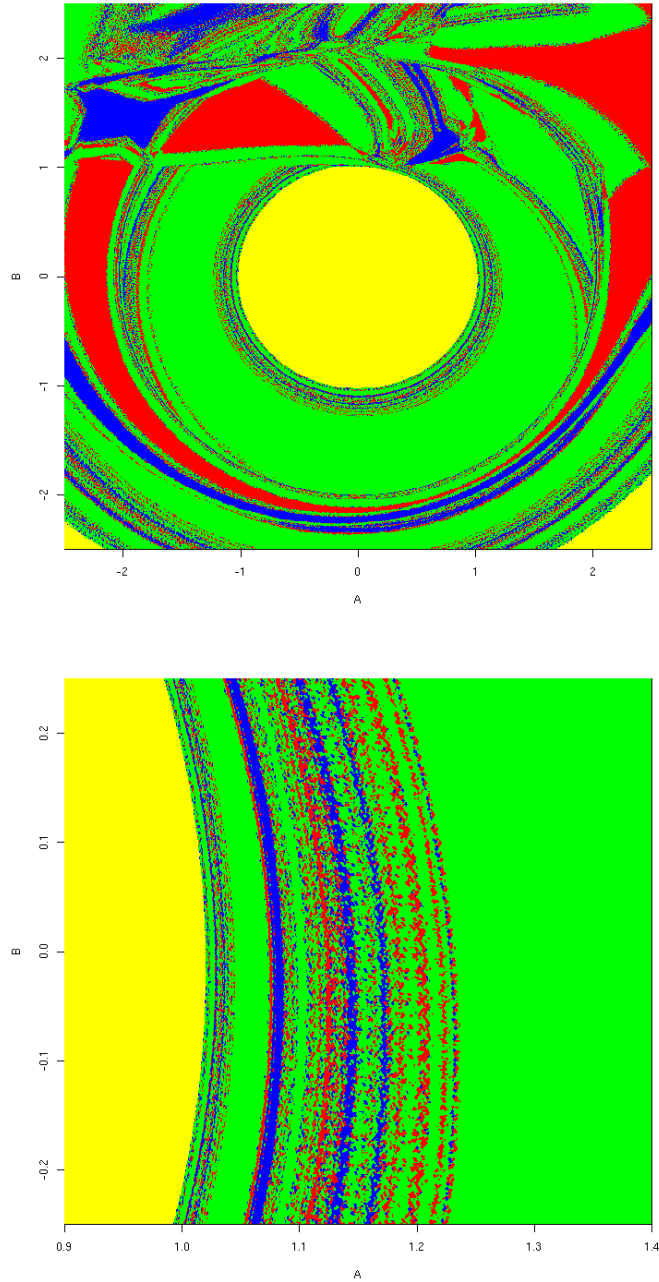
Figure 6.3: As Figure 6.2, except with initial functions of the form $t \mapsto A\sin(2\pi t) + B\cos(2\pi t)$. The second image shows an enlargement, by a factor of 10, of part of the first image (the granularity is a result of limited numerical accuracy).
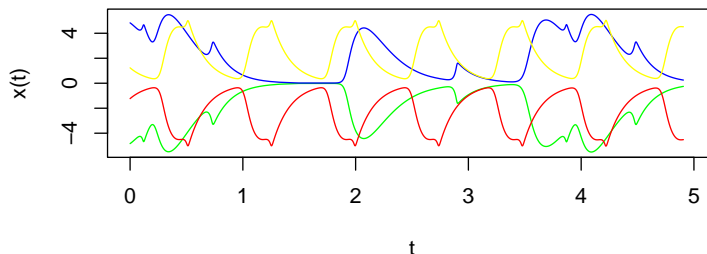
Figure 6.4: Four coexisting attracting periodic solutions of the Mackey-Glass equation (6.8) with parameters $\alpha = 1/0.1625$, $\beta = 12/0.1625$.

Mackey-Glass equation

The Mackey-Glass delay differential equation [86],

$$x'(t) = -\alpha x(t) + \beta \frac{x(t-1)}{1 + x(t-1)^{10}}, \tag{6.8}$$

was originally introduced to model oscillations in neutrophil populations. It has subsequently been the subject of much study because of the variety of dynamical phenomena it exhibits.

Losson *et al.* [81] found multistability (coexistence of two attracting periodic orbits) in the DDE (6.8), but for the parameter sets they considered they found basins of attraction with only simple, non-fractal boundaries. We have made a more thorough search of parameter space, sampling the rectangle $[0.5, 10] \times [0, 200]$ in (α, β)-space at 20×20 resolution. For each sample, 100 different numerical solutions of (6.8) were computed. In this way we found numerous parameter values for which equation (6.8) has higher-order multi-stability and basins of attraction with apparently fractal boundaries. For example, at $\alpha = 1/0.1625$, $\beta = 12/0.1625$, there are the four coexisting attracting periodic solutions shown in Figure 6.4. These solutions occur in symmetric positive and negative pairs, due to the invariance of equation (6.8) under the transformation $x(t) \mapsto -x(t)$.

Figure 6.5 shows the basins of attraction for equation (6.8), or rather, part of the intersection of these basins with the subspace of C spanned by functions of the form (6.4). Also shown is a sequence of magnifications, spanning four orders of magnitude, that

suggest the basin boundaries are fractal sets with Cantor-like structure. The reflection symmetry in the first figure is due to invariance of the DDE under the transformation $x(t) \mapsto -x(t)$, or equivalently $(A, B) \mapsto (-A, -B)$.
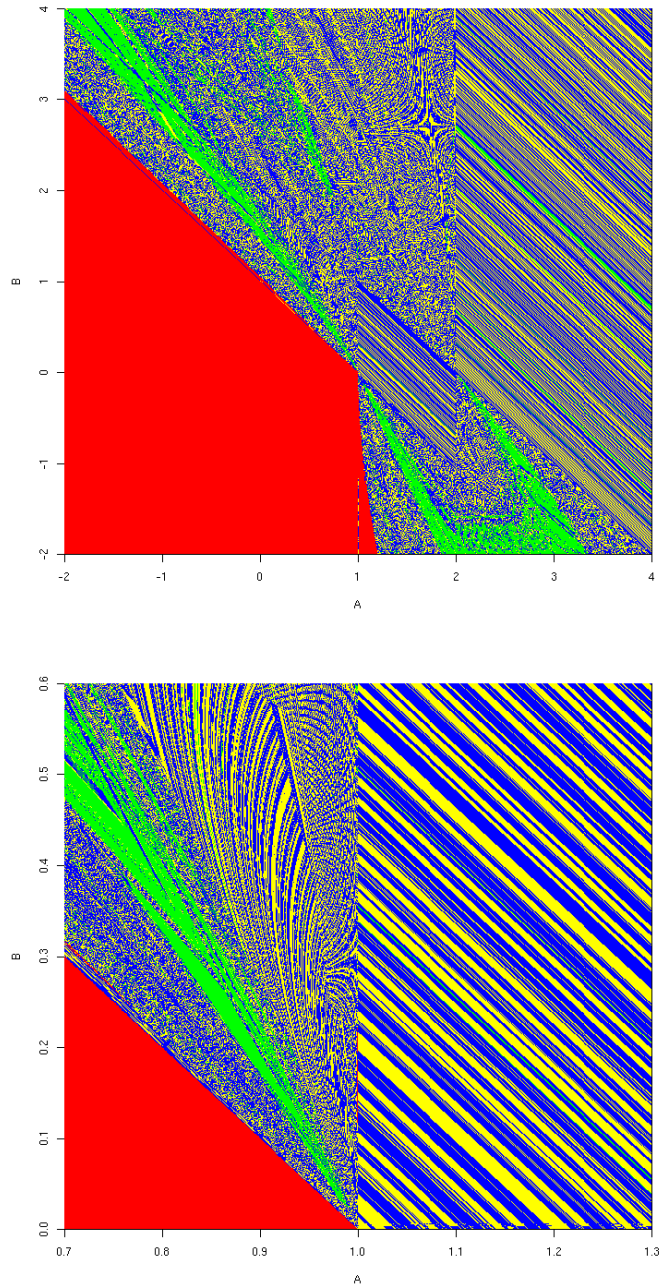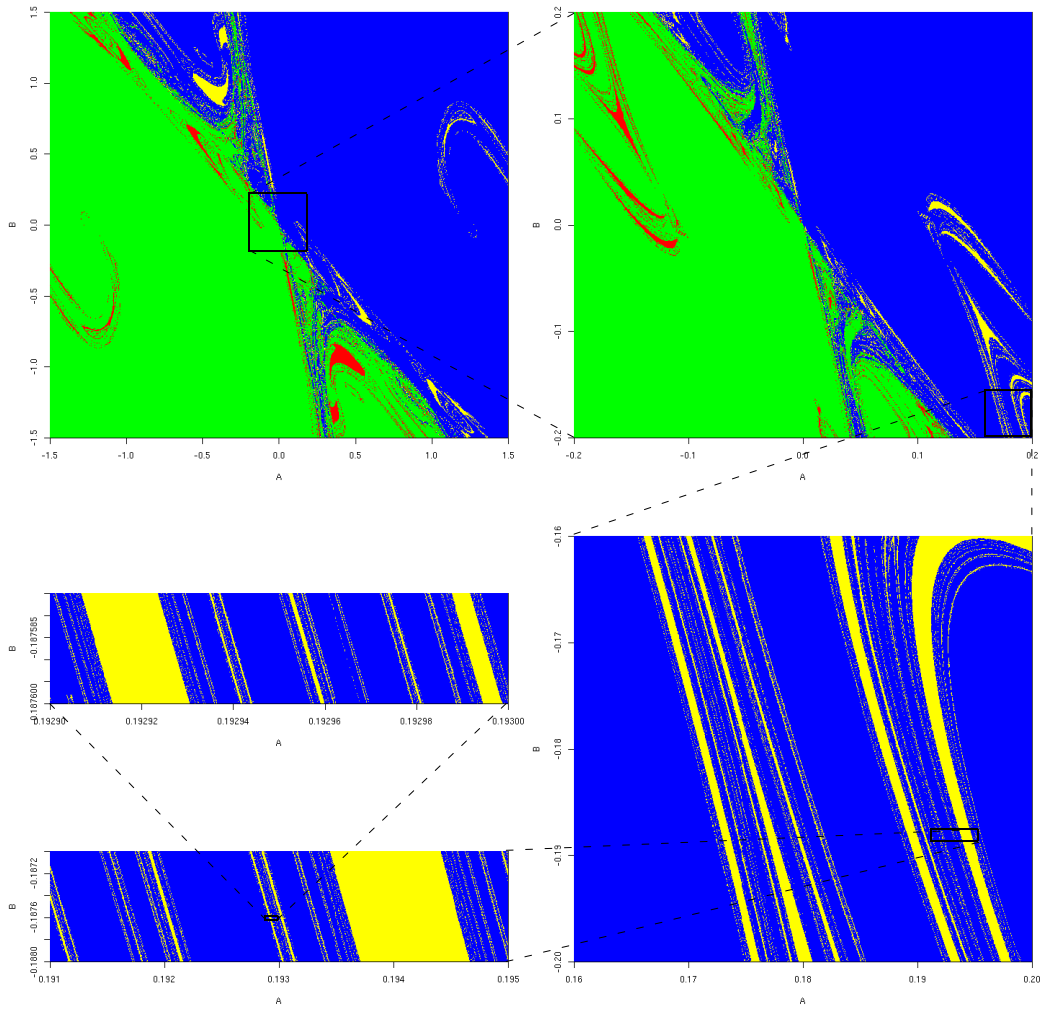
Figure 6.5: Basins of attraction of the Mackey-Glass equation (6.8) for initial functions of the form $t \mapsto A + Bt$, $t \in [0, 1]$. The basin colors are those of the corresponding solutions shown in Figure 6.4. A sequence of magnifications over 4 orders of magnitude is shown to highlight the fractal structure of the basin boundaries.
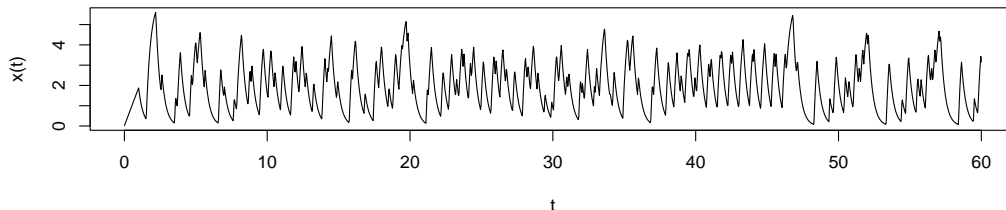
Figure 6.6: A transient chaotic solution of the delay equation (6.2), corresponding to the initial function $x(t) = 0.0131614 + 1.870858t$, $t \in [0, 1]$.

6.2.2   Chaotic transients

If transient chaos does occur in the delay equations (6.2) and (6.8) then it should be possible to find solutions with long chaotic transients. If there is a chaotic saddle, the basin boundary will contain the saddle and its stable manifold. Thus chaotic transients should be found for initial functions close to the basin boundary. Such initial functions, of the form $x(t) = A + Bt$, $t \in [0, 1]$, can simply be read off Figures 6.2 and 6.5 by choosing a point (A, B) near a boundary between basins of attraction. This point can be refined, using a bisection algorithm, to obtain a point (A, B) and corresponding initial function ϕ arbitrarily close (within numerical precision) to the basin boundary. Integration of the DDE forward from this initial function is then expected to yield a solution with a long chaotic transient.

Figure 6.6 shows a numerical solution of equation (6.2) corresponding to an initial function found in this way. This solution does indeed appear to exhibit aperiodic behavior for a considerable duration (about 50 time units) before settling down to one of the attracting periodic solutions. Figure 6.7 shows a solution of the Mackey-Glass equation 6.8 obtained in the same manner, also with a long chaotic transient that settles down to an attracting periodic orbit after about 50 time units.

The observation of long chaotic transients, together with the existence of seemingly fractal basins of attraction, suggests the presence of chaotic saddles in the delay equations (6.2) and (6.8). It is of interest to approximate these saddles numerically. This problem is considered in the following sections.
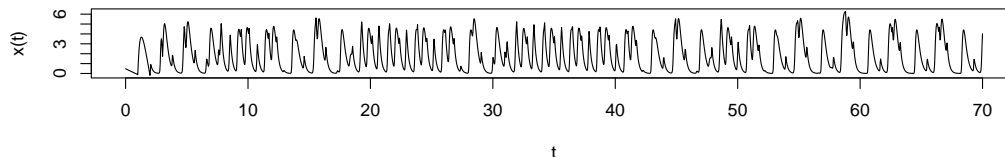
Figure 6.7: A transient chaotic solution of the Mackey-Glass equation (6.8), corresponding to the initial function $x(t) = 0.482417 - 0.599163t$, $t \in [0, 1]$.

6.3 Numerical Analysis

Analysis of transient chaos—that is, finding a numerical approximation of the saddle and characterizing it quantitatively, *e.g.* in terms of ergodic properties such as Lyapunov exponents and fractal dimensions—requires a method for computing arbitrarily long orbits on or very near the chaotic saddle. Because the saddle is unstable, this is not straightforward: rather than being attracted to the saddle, any numerical trajectory starting near the saddle, no matter how close, will eventually depart from it.

A variety of algorithms have been developed for computing numerical trajectories very near a chaotic saddle, including the "straddle-orbit"[9], "PIM" (Proper Interior Maximum) [90], and "stagger-and-step" [110] methods. Each of these algorithms uses the following strategy. Starting with an initial phase point near the saddle (or its stable manifold), evolve this point forward under the system dynamics, occasionally applying a perturbation of size less than ε to keep the trajectory within a small neighborhood of the saddle. By construction the resulting trajectory $\{x_n\}$ satisfies

$$|x_{n+1} - S(x_n)| < \varepsilon, \tag{6.9}$$

so $\{x_n\}$ is an ε-pseudo-orbit of S. It is presumed that if ε is sufficiently small then $\{x_n\}$ approximates a true orbit on the saddle (*e.g.* by the Shadowing Lemma, *cf.* Section 2.6). The important differences between the various algorithms, which we outline below, lie in the methods used to find appropriate perturbations.

Throughout the following, let $\{S^n : n \in \mathbb{Z}_+\}$ be a discrete-time dynamical system defined by iterates of a transformation $S : X \to X$.

6.3.1   Straddle orbit method

The goal of the straddle orbit method is to construct a perturbed orbit $\{x_n\}$ that follows the basin boundary. Between iterates of S, the straddle orbit method [9] employs a bisection algorithm to perturb an orbit to within a small neighborhood of the basin boundary. The algorithm is as follows.

1. Choose phase points $x_A, x_B \in X$ that lie in different basins.

2. Let x_C be the midpoint of the segment $x_A x_B$; determine which basin x_C is in, *e.g.* by iteration of S.

3. If x_C is in the same basin as x_A, let $x_A := x_C$, else let $x_B := x_C$.

4. If $|x_A - x_B| \geq \varepsilon$, return to 2.

5. Let $x_A := S(x_A)$, $x_B := S(x_B)$.

6. If $|x_A - x_B| \geq \varepsilon$, return to 2, else return to 5.

The perturbation phase of the algorithm is carried out in steps 2–4, which use a bisection algorithm to isolate the basin boundary between points x_A and x_B separated by a distance less than ε. Points x_A and x_B are both evolved under iterates of S, and the bisection algorithm is repeated whenever x_A and x_B differ by more than ε.

The straddle orbit $\{x_n\}$ is the sequence of points x_A (alternatively x_B) obtained from step 5. By construction, the straddle orbit remains within distance ε of the basin boundary. It is presumed that for ε sufficiently small the straddle orbit approximates an orbit on the boundary. If the boundary consists of the stable manifold of the chaotic saddle then the straddle orbit should follow this stable manifold and, after an initial transient phase, should remain with ε of the saddle itself.

6.3.2   PIM method

The goal of the PIM method [90, 91] is to construct a perturbed orbit $\{x_n\}$ that remains indefinitely within some neighborhood R of the chaotic saddle. This is done by perturbing the orbit so as to increase the escape time $T(x_n)$, which is the number of iterates of S required to take x_n out of R.

Suppose the dynamical system $S : X \to X$ has a chaotic saddle Λ. Let $R \subset X$ be a *transient region*, such that $\Lambda \subset R$ and R contains no attractor. For $x \in R$, define the *escape time* $T(x)$ by

$$T(x) = \min\{n > 0 : S^n(x) \notin R\}$$
$$(= \infty \text{ if } S^n(x) \in R \; \forall n > 0).$$

(6.10)

In the following, an ordered set (x_a, x_b, x_c) of points $x_a, x_b, x_c \in X$ is said to be a PIM triple if x_b lies on the segment $x_a x_c$ and $T(x_b) > \max\big(T(x_a), T(x_c)\big)$.

---

1. Choose a PIM triple (x_a, x_b, x_c) such that x_a and x_c lie in different basins.

2. Choose N equally spaced points on the segment $x_a x_c$. From these choose a PIM triple $(\bar{x}_a, \bar{x}_b, \bar{x}_c)$ such that the segment $\bar{x}_a \bar{x}_c$ is a proper subset of the segment $x_a x_b$.

3. Let $x_a := \bar{x}_a$, $x_c := \bar{x}_c$.

4. If $|x_a - x_c| \geq \varepsilon$, return to 2.

5. Let $x_a := S(x_a)$, $x_c := S(x_c)$.

6. If $|x_a - x_c| \geq \varepsilon$, return to 2, else return to 5.

The PIM orbit $\{x_n\}$ is the sequence of points x_a (alternatively x_b) obtained from step 5. By construction the PIM orbit $\{x_n\}$ satisfies $T(x_n) > 0$ for all n, so $\{x_n\}$ remains within the transient region R for all time.

6.3.3 Stagger-and-step method

The stagger-and-step method [110] is also based on seeking perturbations that increase the escape time of the orbit, but perturbations are not restricted to a particular line segment. Let the escape time $T(x)$ be defined as in the PIM method above, and let $T_* > 0$ (the minimum allowed escape time). Then a stagger-and-step trajectory $\{x_n\}$ results from iterating the following algorithm.

1. Choose $x_0 \in X$ with $T(x_0) > T_*$.

2. If $T(x_n) \leq T_*$, find a random perturbation $r \in X$, $|r| < \varepsilon$, such that $T(x_n + r) > T_*$, and let $x_n := x_n + r$.

3. Let $x_{n+1} := S(x_n)$.

4. Return to 2.

6.4   Application to Delay Equations

6.4.1   Approximate discrete-time map

The algorithms described above are formulated in the context of discrete-time dynamical systems. For a continuous-time system they can be applied to an appropriate discretized version of the dynamics, for example iterates of the time-one map. In the case of a delay equation (6.1) the time-one map is a transformation $S = S_1 : C \to C$. In a numerical simulation this transformation cannot be represented exactly, and we must resort to a finite dimensional approximation. For this purpose the phase point $x_t \in C$ can be represented by the vector $\mathbf{u}(t) = \big(u_0(t), \ldots, u_N(t)\big)$ of values

$$u_i(t) = x_t(s_i) = x(t + s_i), \tag{6.11}$$

that x_t takes on a uniform grid

$$s_i = -1 + i/N, \quad i = 0, \ldots, N, \tag{6.12}$$

typically with N of the order 10^2 or 10^3. The time-one map $S(\phi)$ can be carried out by constructing a numerical solution x with initial function ϕ to time $t = 1$, and evaluating the vector of solution values $u_i = x(1 + s_i)$, $i = 0, \ldots, N$.

To be more precise, in numerical integration of the DDE (6.1) with time step $0 < h \ll 1$, $\mathbf{u}(t + h)$ is approximated by

$$\mathbf{u}(t) \mapsto \mathbf{u}(t + h) \approx T\big(\mathbf{u}(t)\big) \tag{6.13}$$

for some transformation $T : \mathbb{R}^{N+1} \to \mathbb{R}^{N+1}$, the details of which depend on the choice

of integration scheme. With $h = 1/N$ we have

$$
\begin{aligned}
\mathbf{u}(t+1) &= \mathbf{u}(t+Nh) \\
&\approx T^N\big(\mathbf{u}(t)\big),
\end{aligned}
\tag{6.14}
$$

so that

$$
\tilde{S} = T^N
\tag{6.15}
$$

approximates the time-one map S. The transformation $\tilde{S} : \mathbb{R}^{N+1} \to \mathbb{R}^{N+1}$ is readily implemented using any of the various codes available for numerical integration of DDEs.[5]

6.4.2 Computing escape times

Implementing the PIM and stagger-and-step methods requires a method of computing the escape time function T. This in turn requires a practical method of describing the transient region R. One method is to define R to be the set

$$
R = \{x \in X : \text{dist}(x, A_i) > \delta \; \forall i\},
\tag{6.16}
$$

for some $\delta > 0$, where the A_i are the attractors of the system. Presumably these are already known or approximated, so that $\text{dist}(x, A_i)$ can be computed. This approach was used in [90] for low-dimensional systems.

Although it has not received mention elsewhere, it seems clear that the A_i must also include periodic orbits whose stability is of saddle type. Otherwise, the presence of such an orbit implies the existence of trajectories with arbitrarily large escape times, but which do not lie near the chaotic saddle (*e.g.*, trajectories that follow the stable manifold of the unstable periodic orbit). If a particular such orbit A is not counted among the A_i then the PIM and stagger-and-step methods fail, with the trajectory $\{x_n\}$ converging to A.[6] In this case the resulting trajectory $\{x_n\}$ automatically yields an approximation of A, which can then be included among the A_i and the method applied again.

---

[5]Throughout this chapter, numerical integration is performed using the Fortran code DKLAG6 [25] interfaced with the R language [60].

[6]As a side-effect that might be exploited, the PIM and stagger-and-step methods appear to be novel ways to find and approximate saddle type periodic orbits.
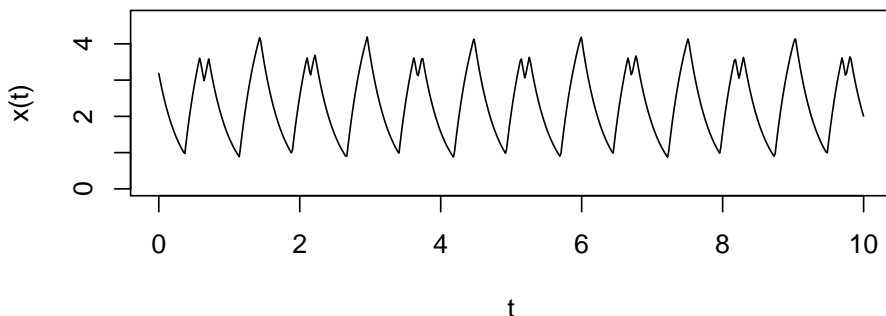
Figure 6.8: An unstable (saddle type) periodic solution of the delay equation (6.2).

6.4.3   Failure of existing algorithms

In principle, any of the methods in [9, 90, 110] could be used to approximate the chaotic saddle (if one exists) for the finite dimensional map $\tilde{S} : \mathbb{R}^{N+1} \rightarrow \mathbb{R}^{N+1}$ defined in Section 6.4.1. We have implemented all of these methods and applied them to the delay equations (6.2) and (6.8) for which we have found evidence of transient chaos. We find that each of these methods does in fact generate an aperiodic trajectory of some duration, but that eventually this trajectory converges to an unstable (saddle type) periodic orbit, regardless of the method used. The unstable periodic solutions found in this way for equations (6.2) and (6.8) are shown in Figures 6.8 and 6.9, respectively. This mode of failure is interesting since it has not been reported before.

There is a simple plausible explanation for the failure of the straddle orbit method. For each of the DDEs under consideration there appears to be a saddle type periodic orbit. The stable manifold of any such orbit is contained within the basin boundary. A straddle orbit, which by construction lies close to the basin boundary, is eventually perturbed onto (or near) this stable manifold. The orbit thereafter follows this stable manifold rather than the stable manifold of the chaotic saddle, and consequently the straddle orbit is asymptotic to the periodic orbit.

In both the PIM and stagger-and-step methods the unstable periodic orbit can be avoided by excluding it from the transient region R, as described in the previous section.
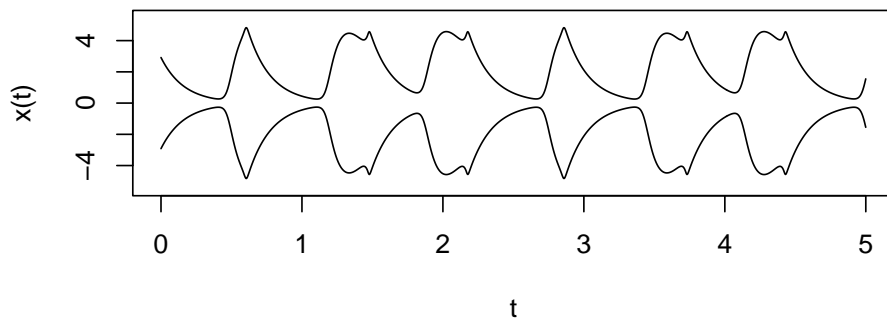
Figure 6.9: Unstable (saddle type) periodic solutions of the Mackey-Glass equation (6.8).

This prevents these methods from converging to the periodic orbit, since an orbit that enters a neighborhood of an unstable periodic orbit is deemed to have left the transient region. However, with this provision both algorithms eventually stall at a point where they are unable to find a perturbation that yields an increased escape time. The reason for this failure seems to be the mechanism illustrated in Figure 6.10. Here the orbit through the phase point x exits the transient region R by entering a neighborhood Ω of an unstable periodic orbit, represented by the point P. The orbit through the perturbed phase point $\tilde{x}$ narrowly misses Ω, so that $T(\tilde{x}) > T(x)$. Since it increases the escape time of the trajectory, $\tilde{x}$ is taken as a "successful" perturbation in either the PIM or stagger-and-step method. However, this perturbation is spurious since $\tilde{x}$ is carried by successive iterations of S into a region where further successful perturbations, spurious or otherwise, do not exist. Both algorithms come to a halt when the trajectory through $\tilde{x}$ reaches a point where all phase points within distance ε exit R via Ω, and all have the same escape time.

It has been pointed out [110] that the PIM method is expected to fail if the chaotic saddle has more than one unstable direction. This would suggest an alternative explanation for the failure of the method here. However, the results of Section 6.5.2 provide evidence that the chaotic saddle has only one unstable direction, for both of the DDEs (6.2) and (6.8). Therefore the failure of the method does not seem to be caused by the presence of multiple unstable directions.
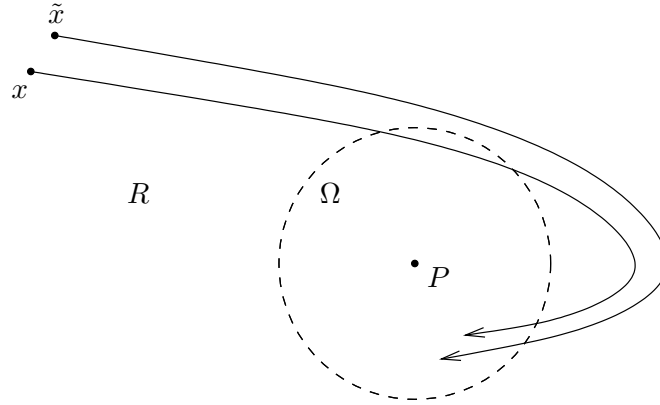
Figure 6.10: Mechanism of failure of the PIM and stagger-and-step methods. The phase point x exits the transient region R via a neighborhood Ω of an unstable periodic orbit P. The perturbed phase point $\tilde{x}$ just misses Ω, so $T(\tilde{x}) > T(x)$.

6.4.4   Modified stagger-and-step method

The following slight modification of the stagger-and-step method circumvents the difficulty described above, and proves effective for the delay equations considered here. Before applying it to DDEs, we present the modified algorithm in its general form.

The idea behind the method is to make a more aggressive search for a "successful stagger", *i.e.* a perturbation that moves x_n onto a nearby trajectory that remain within a neighborhood of the saddle for a longer period of time $T(x_n)$. In the original stagger-and-step method a perturbation is sought only when $T(x_n) = T_*$. For $T(x_n) > T_*$, x_n is simply iterated forward under S until $T(x_n) = T_*$. In our modification, up to $N > 0$ random perturbations are selected in an attempt to find a successful stagger at *each* iteration.

For a given $T_* > 0$ (the minimum allowed escape time) and small $\varepsilon > 0$, a trajectory $\{x_n\}$ results from iteration of the following algorithm.

1. Choose $x_0 \in X$ with $T(x_0) > T_*$.

2. Set $j := 1$.

3. Choose a random perturbation $r \in X$, $|r| < \varepsilon$.

4. If $T(x_n + r) > T(x_n)$ then set $x_n := x_n + r$ and go to 6.

5. If $T(x_n) \leq T_*$ or $j < N$ then set $j := j + 1$ and return to 3.

6. Let $x_{n+1} := S(x_n)$.

7. Return to 2.

Thus each iteration consists of a possible successful stagger of the current phase point, followed by an iteration of S. Because $T(x_n) \geq T_* > 0$ for all n, we have $x_n \in R$ for all time; that is, the trajectory never leaves the transient region. If the trajectory is aperiodic, it is taken (after removal of an initial transient where it follows the stable manifold of the chaotic saddle) as an approximation of the chaotic saddle.

The parameters ε, T_* and N are adjustable. As discussed in [110], the success of the stagger-and-step method can depend on a careful choice of T_*; this observation applies also to the modified algorithm. If T_* is outside a range of suitable values, a very large and possibly infinite number of "stagger attempts" are necessary before a successful stagger (one that increases $T(x_n)$ above T_*) is found.

Parameter N is the maximum number of "stagger attempts" at each iteration if $T(x_n) > T_*$. The original stagger-and-step method is similar but not identical to the case $N = 1$. In this case only one stagger attempt is made for phase points with $T(x_n) > T_*$. Usually this attempt fails. After each such failure $T(x_n)$ decreases by one, until $T(x_n) = T_*$ where an exhaustive search for a successful stagger is made.

In applying the method to delay equations we find, regardless of the choice of T_*, that very frequently a successful stagger cannot be found when $T(x_n) = T_*$, and the algorithm "gets stuck". This phenomenon appears to be related to the instability of trajectories in a neighborhood of the saddle: under iteration of $\tilde{S}$, the potential successful staggers within an ε-ball at x_n quickly diverge to a distance greater than ε from the current phase point. If this divergence occurs in fewer than the number of iterations required for $T(x_n)$ to fall to T_*, a successful stagger will fail to exist when $T(x_n) = T_*$ is reached. With

our modification, taking $N > 1$ ($N = 5$ has worked well in practice) greatly reduces the frequency of this outcome, since a more thorough search for a successful stagger is made at each iteration.

Even with this approach it occasionally happens that a phase point is reached where $T(x_n) = T_*$ and where it is impossible to find a successful stagger. In such cases an effective remedy is to revert to a previous iteration, make a thorough search until a successful stagger is found, and continue the algorithm from this point.

6.5   Numerical Analysis Results

6.5.1   Visualizing the saddle and its invariant measure

The stagger-step algorithm described above, when applied to the approximate discrete-time map $\tilde{S}$ as defined for the delay equations (6.2) and (6.8), yields a numerical trajectory $\{\mathbf{u}(n) \in \mathbb{R}^{N+1} : n = 0, 1, 2, \ldots\}$ that approximates a trajectory in C near the presumed chaotic saddle. The corresponding solution of the DDE can be constructed from equation (6.11). Figures 6.11 and 6.12 show chaotic solutions, computed in this manner, for the delay equations (6.2) and (6.8), respectively.

With the trajectory $\{\mathbf{u}(n)\}$ in hand, it is possible to visualize and otherwise characterize the geometry of the saddle. The phase space dimension of this trajectory is finite but large, so that visualizing it (and hence the chaotic saddle it approximates) is difficult. As discussed in Sections 5.2.1 and 5.2.5, one way to visualize an object in the infinite dimensional phase space C is to plot its image under a suitable "trace map"

$$\pi : C \to \mathbb{R}^M. \tag{6.17}$$

For delay equations like those considered here, a common practice is to use the map

$$\pi : u \mapsto \big(u(-1), u(0)\big), \tag{6.18}$$

i.e. to plot $x(t)$ vs. $x(t-1)$. With respect to the finite-dimensional vector $\mathbf{u}$ that approximates u, π acts according to

$$\pi : \mathbf{u} \mapsto (u_0, u_N). \tag{6.19}$$

Figure 6.13 shows a two-dimensional image, computed in the manner described above, of the chaotic saddle for the delay equation (6.2). Similarly, Figure 6.15 presents an
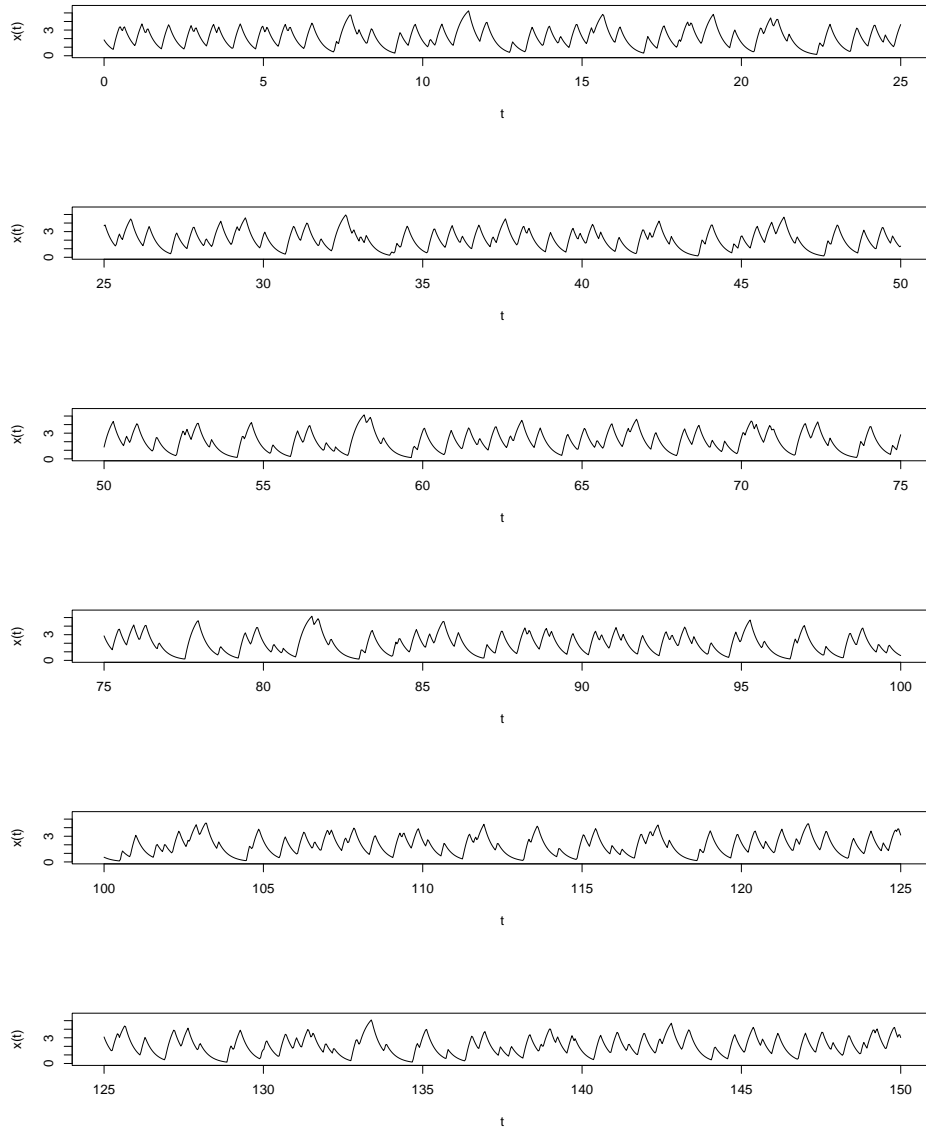
Figure 6.11: A segment of a numerical solution near the chaotic saddle of the delay equation (6.2), computed using the modified stagger-and-step algorithm.
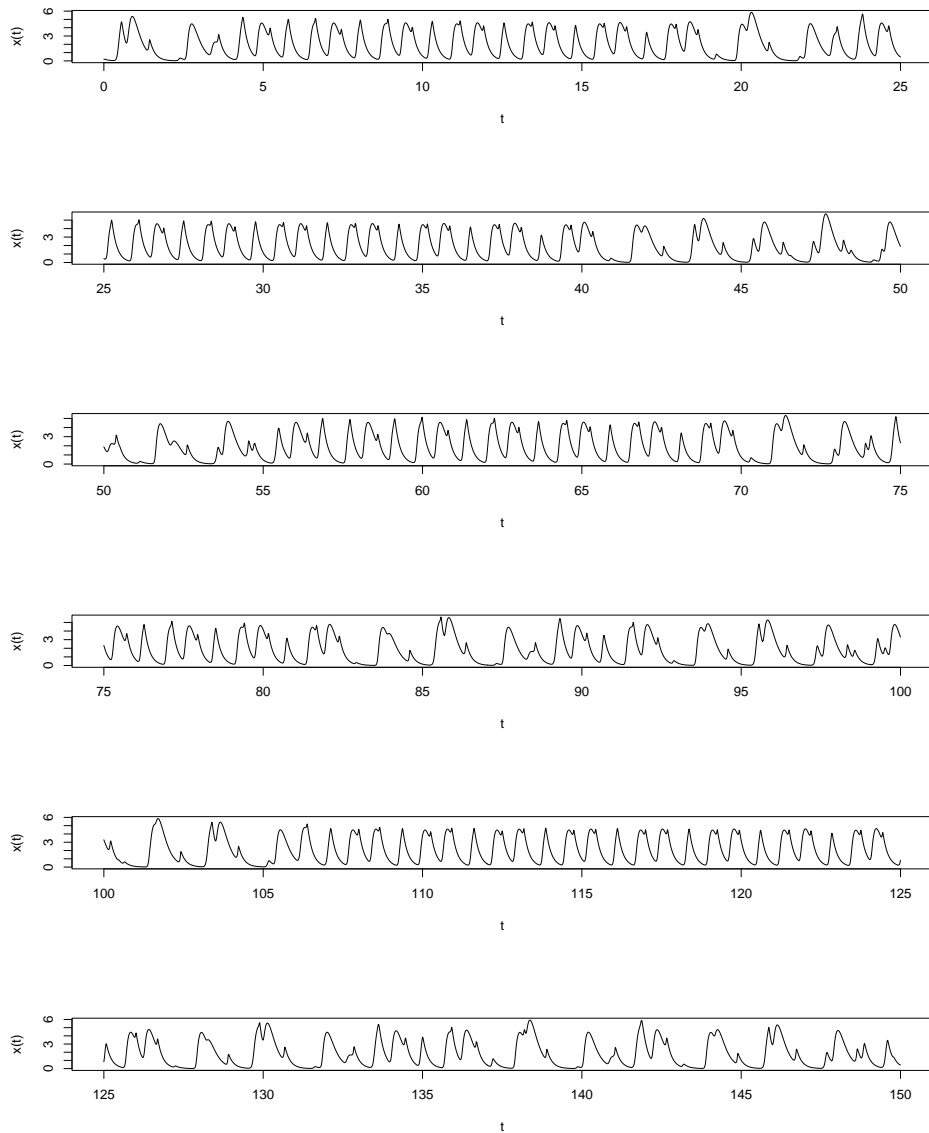
Figure 6.12: A segment of a numerical solution near the chaotic saddle of the Mackey-Glass equation (6.8), computed using the modified stagger-and-step algorithm.

image of the chaotic saddle of the Mackey-Glass equation (6.8). These images provide a "flattened" view of a trajectory on the saddle. Of course, the trajectory itself lives in the phase space C (or, at least, the numerical approximation of the trajectory lives in $\mathbb{R}^{N+1}$).

The asymptotic statistics of a trajectory on the saddle induces an invariant measure. That is, for a given initial phase point $u \in C$ on the saddle, the measure μ on C defined by

$$\mu(A) = \lim_{M \to \infty} \frac{1}{M} \sum_{n=1}^{M} 1_A\big(S^n(u)\big), \tag{6.20}$$

provided the limit exists, is invariant under the dynamics S (*cf.* Section 2.4.2). Using a stagger-step trajectory $\{\mathbf{u}(n) : n = 1, 2, \ldots, M\}$ of large length M, this measure can be approximated by the measure

$$\tilde{\mu}(A) = \frac{1}{M} \sum_{n=1}^{M} 1_A\big(\mathbf{u}(n)\big). \tag{6.21}$$

on $\mathbb{R}^{N+1}$.

As discussed in Section 5.2.5, it is possible to visualize $\tilde{\mu}$ by computing its two-dimensional image under the trace map π (equation (6.19)). This amounts to simply computing a two-dimensional histogram (*i.e.*, density) of pairs (u_0, u_N) along the numerical trajectory $\{\mathbf{u}(n)\}$. The resulting images of the invariant measures for the delay equations (6.2) and (6.8) are shown in Figures 6.14 and 6.16, respectively.

In principle there may be an uncountable number of distinct invariant measures on the saddle. In practice, however, any stagger-step trajectory appears to yield the same approximate invariant measure, for both of the delay equations considered here. This suggests the existence of a unique "natural" invariant measure for these systems, analogous to SRB measure in that it is the invariant measure naturally selected by numerical simulations and (presumably) physical experiments. See [34, 35] and references therein for discussion of how the notion of natural invariant measure should be defined in the context of transient chaos.

In terms of the solution $x(t)$, the asymptotic statistics on the chaotic saddle are described by the measure ν on $\mathbb{R}$ given by

$$\nu(A) = \lim_{T \to \infty} \frac{1}{T} \int_0^T 1_A\big(x(t)\big)\, dt, \tag{6.22}$$
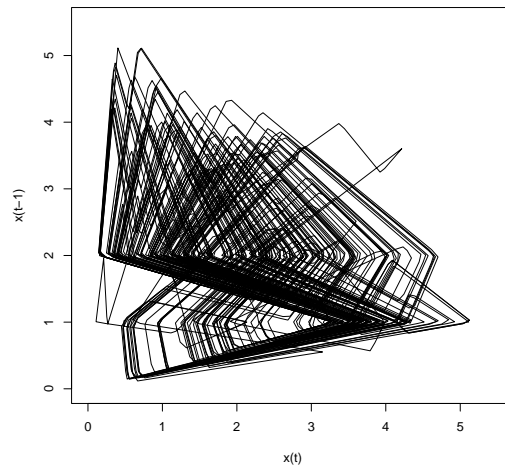
Figure 6.13: Two-dimensional projection, under the trace map (6.18), of part of a trajectory on the chaotic saddle of the delay equation (6.2).
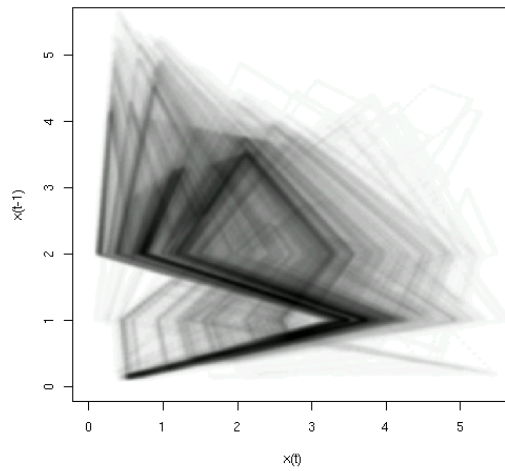
Figure 6.14: Two-dimensional projection, under the trace map (6.18), of the natural invariant measure on the chaotic saddle of the delay equation (6.2).
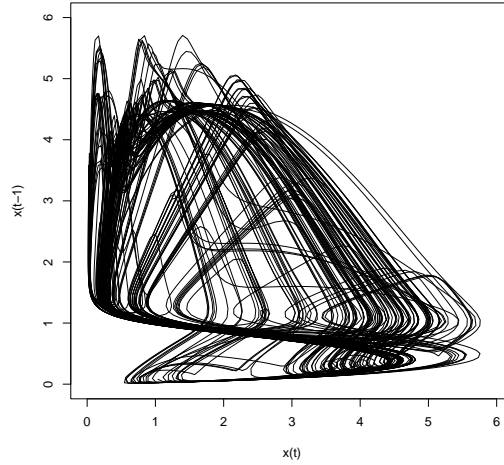
Figure 6.15: Two-dimensional projection, under the trace map (6.18), of part of a trajectory on the chaotic saddle of the Mackey-Glass equation (6.8).
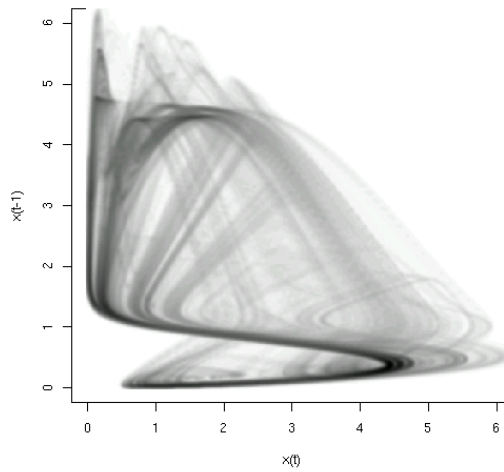
Figure 6.16: Two-dimensional projection, under the trace map (6.18), of the natural invariant measure on the chaotic saddle of the Mackey-Glass equation (6.8). Density is indicated by grayscale intensity.
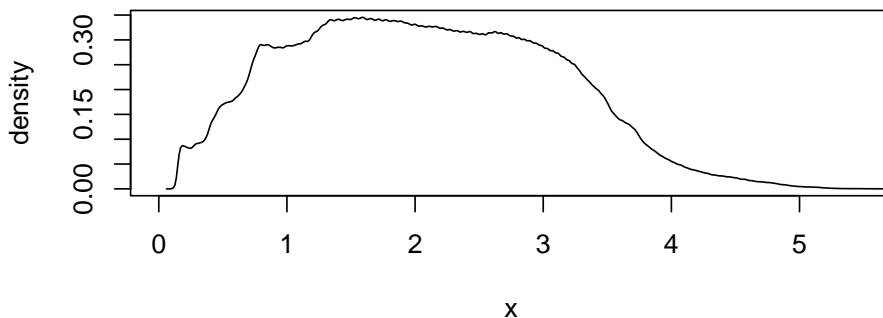
Figure 6.17: One-dimensional invariant density for the chaotic saddle of the delay equation (6.2).

or in terms of the discrete-time map S,

$$\nu(A) = \lim_{M \to \infty} \frac{1}{M} \sum_{n=1}^{M} 1_A((S^n u)(1)). \tag{6.23}$$

This is just the one-dimensional projection, under the trace map $\pi : u \mapsto u(0)$, of the natural invariant measure μ. The measure ν has the practical significance of describing where a solution on the saddle spends most of its time. It is more intuitively understood in terms of its density (the so-called "invariant density"), which can be approximated by a histogram of values $x(t)$ along a long solution near the saddle. Invariant densities found in this way for the delay equations (6.2) and (6.8) are shown in Figures 6.17 and 6.18, respectively.

6.5.2   Quantitative characterization: Ergodic parameters

It is considered *de rigeur*, whenever a new chaotic invariant set is found, to characterize its geometry and the dynamics on it by computing its fractal dimensions and Lyapunov exponents, and possibly other ergodic parameters (*cf.* Section 2.5). For the chaotic saddles of the delay equations considered here, for which we are able to compute arbitrarily long numerical trajectories $\{\mathbf{u}(n)\}$ approximating the saddle, these quantities can be found by
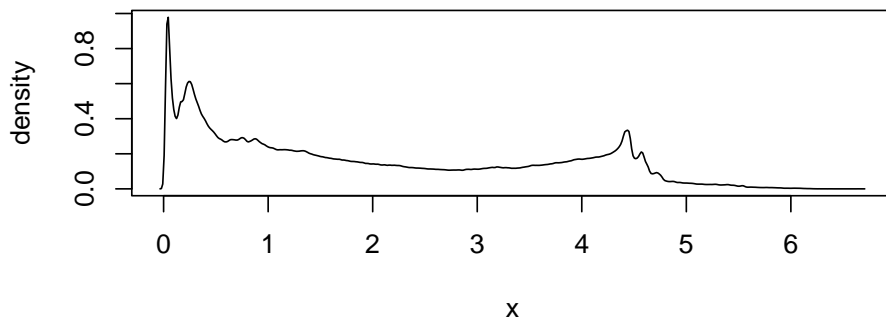
Figure 6.18: One-dimensional invariant density for the chaotic saddle of the Mackey-Glass equation (6.8).

straightforward computation using the algorithms referenced in Section 2.5.

Table 6.1 summarizes the results of computations of the largest five Lyapunov exponents λ_i [12, 13], the Lyapunov dimension [43]

$$d_L = j + \frac{\sum_{i=1}^{j} \lambda_i}{|\lambda_{j+1}|} \quad \text{where} \quad j = \max_{k} \sum_{i=1}^{k} \lambda_i > 0, \tag{6.24}$$

and the correlation dimension [48] of the chaotic saddles for each of the delay equations (6.2) and (6.8). Both DDEs have a zero Lyapunov exponent. This is to be expected for continuous-time systems in general, with the zero Lyapunov exponent corresponding to neutral expansion along the direction tangent to the flow [52].

Note that both DDEs exhibit just a single positive Lyapunov exponent. The existence of a positive exponent confirms that there is sensitivity to initial conditions in a neighborhood of the saddle. The fact that there is only one single positive exponent suggests that the saddle has only one unstable direction. Thus the failure of the PIM method for these DDEs cannot be due to the presence of multiple unstable directions.

System	Parameters	Lyapunov spectrum (bits/time)					Lyapunov dimension	Correlation dimension
		λ_1	λ_2	λ_3	λ_4	λ_5		
$F(x) = \begin{cases} c & \text{if } x \in [x_1, x_2] \\ 0 & \text{otherwise} \end{cases}$	$\begin{cases} \alpha = 3.25 \\ c = 20.5 \\ x_1 = 1 \\ x_2 = 2 \end{cases}$	0.54	0.00	-1.5	-8.2	-12	2.36	1.96
$F(x) = \beta \dfrac{x}{1 + x^{10}}$	$\begin{cases} \alpha = 1/0.1625 \\ \beta = 12/0.1625 \end{cases}$	0.60	0.00	-0.50	-3.1	-3.8	3.03	2.24

Table 6.1: Lyapunov exponents and fractal dimensions for chaotic saddles of delay differential equations having the form $x'(t) = -\alpha x(t) + F\big(x(t-1)\big)$.

6.6 Conclusions

Transient chaos in delay differential equations, although anticipated by numerous published results [1, 5, 7, 54, 81, 121], has not specifically been investigated before. Multistability with fractal basins of attraction, a key signature of transient chaos, has been reported previously for the delay equation (6.2). We have also found multistability and fractal basins in the Mackey-Glass equation (6.8). Existence of fractal basins suggests the existence of transiently chaotic trajectories and the presence of a chaotic saddle [87]. Indeed, by computing basins of attraction for these two delay equations we have been able to illustrate, numerically, the existence of solutions with long chaotic transients.

The published methods for approximating chaotic saddles fail when applied to the delay equations considered here. The PIM method [90] in particular is expected to fail if the saddle has more than one unstable direction [110]. However, our computations of Lyapunov exponents point to the presence of only a single unstable direction, so this explanation is inadequate. Instead, the failure of the known methods stems from the presence of saddle type unstable periodic orbits, to which each of the methods eventually converge. This mode of failure has not been observed before.

A slightly modified version of the stagger-and-step method (Section 6.4.4) avoids unstable periodic orbits, and appears to be an effective method for approximating chaotic saddles for delay equations. Using this method we are able to compute chaotic numerical solutions of arbitrarily long duration, for both of equations (6.2) and (6.8).

Despite the fact that the saddle is embedded in an infinite dimensional phase space (or at least a finite- but high-dimensional phase space used for numerical approximation), it is possible to go some way toward visualizing it by graphing its projections onto two dimensions. We have done this for the delay equations considered here, projecting the numerical approximation of the chaotic saddle onto two dimensions by applying a particular "trace map" commonly used for visualizing the dynamics of delay equations.

The distribution of orbits on the chaotic saddle can be characterized by an invariant measure. This too can be approximated numerically from a stagger-and-step trajectory for a given DDE, and projected onto two dimensions for the purpose of visualization.

Chaotic invariant sets are typically characterized in terms of their ergodic parameters such as Lyapunov exponents and dimensions. These can be found by the standard algorithms, involving straightforward computations on the numerical trajectory resulting from the stagger-and-step algorithm. For each of the delay equations (6.2) and (6.8) we

have applied these techniques to the numerical trajectory approximating the saddle, and thereby estimated its Lyapunov spectrum, Lyapunov dimension, and correlation dimension, the results of which are summarized in Table 6.1. Two interesting results follow. Firstly, despite the infinite dimensionality of the phase space of the delay equation, the saddle itself has quite low dimension, of the order $2 \sim 3$. The same observation has been made with regard to chaotic attractors of delay equations [42]. Second, for both of the DDEs considered the saddle has only one positive Lyapunov exponent, hence only one unstable direction. This lends support to our hypothesis that the previously published algorithms for approximating the saddle fail for some reason other than the existence of multiple unstable directions.

Chapter 7

Conclusion

7.1   Summary of Conclusions

The probabilistic approach to evolutionary delay differential equations suffers from the chief difficulty, encountered in various guises throughout this thesis, that the phase space of a delay equation is infinite dimensional. This difficulty might explain the absence in the literature of a thorough discussion of how such an approach might be developed. In Chapter 3 we have attempted to bridge this gap.

The phase space that arises naturally in the formulation of delay equations as dynamical systems is the space C of continuous functions from the interval $[-1, 0]$ into $\mathbb{R}^n$. Any solution of a given DDE can be identified with the evolution of a corresponding phase point in C. Within this context Chapter 3 develops the basic framework for the application of ergodic concepts to delay equations, and explores the implications of this framework. Its main conclusions, mostly negative, are consequences of the infinite dimensionality of C.

An ergodic approach to delay equations entails an adequate theory of probability on the phase space C, for which measure theoretic probability provides a sufficiently

abstract setting. However, there are peculiarities of probability in infinite dimensions that make an ergodic approach to DDEs problematic. The main analytical tool of applied ergodic theory is the Perron-Frobenius operator [74], which prescribes the evolution of probability densities under the action of a dynamical system. The Perron-Frobenius operator formalism has been very successful in the analysis of finite dimensional systems. However, infinite dimensional systems cannot be expected to have well-defined densities, owing to singularity of the evolution operator (Section 3.6.2). Consequently one cannot define a Perron-Frobenius operator corresponding to a delay equation. The absence of a well-defined Perron-Frobenius operator precludes the application of the major part of the ergodic theoretic toolbox, *e.g.* in [74].

An ergodic approach to delay equations will also require a theory of integration with respect to measures on C. For example, the Birkhoff Ergodic Theorem (*cf.* Section 2.4.3) allows one to express a time average in terms of an integral (*i.e.* spatial average or expectation) with respect to an ergodic measure. The evaluation of such integrals requires an adequate theory of integration on function space. This theory is lacking, except in the special case of Wiener measure. Because Wiener measure is invariant under the quantum field equations, the theory of integration with respect to Wiener measure has been well developed in the physics literature. However, we cannot expect invariance of Wiener measure under delay differential equations, except perhaps in special cases. The lack of a general theory of integration on function space is a serious barrier to developing an ergodic theory of delay equations.

In the application of ergodic concepts to physical systems, the notion of SRB measure plays a central role. An SRB measure characterizes the asymptotic statistics of almost every orbit of a dynamical system. For finite dimensional systems Lebesgue measure provides a natural and essentially unique translation-invariant notion of "almost every". However, in infinite dimensions there is no measure analogous to Lebesgue measure. In particular, there is no non-trivial translation-invariant measure on C to provide the requisite notion of "almost every". Consequently the definition of SRB measure for delay equations is ambiguous. One way to resolve this ambiguity is to substitute the notion of prevalence [58] in place of "Lebesgue almost every" in the definition of SRB measure. This is consistent with the present definition of SRB measure for finite dimensional systems and therefore provides a natural extension to the infinite dimensional case.

The difficulties associated with the infinite dimensionality of DDEs can be avoided either by finite dimensional approximation or by removing the requirement that DDEs

be treated in a dynamical systems context (thereby precluding any discussion of ergodic concepts, since the notion of an evolution semigroup underlies all of ergodic theory).

Despite the infinite dimensionality of the phase space, a delay equation nevertheless prescribes the evolution of some finite dimensional quantity $x(t) \in \mathbb{R}^n$. In Chapter 4 we have investigated probabilistic approaches to this evolution problem. In order that a DDE prescribes a finite dimensional evolutionary process, it is necessary to restrict the set of allowable initial functions to some finite dimensional subset of C, *e.g.* the subspace of constant functions. If the set of allowable initial functions is n-dimensional then one can define a family of solution maps $S_t : \mathbb{R}^n \to \mathbb{R}^n$. For some simple DDEs for which the solution map can be found analytically, one can derive an explicit formula for the Perron-Frobenius operator corresponding to S_t and thereby analytically solve the density evolution problem (Section 4.2). For more complicated equations this method is impractical, due both to the difficulty of finding S_t analytically and to the non-invertibility of S_t once it has been found. In particular this method fails to provide an analytical approach to the evolution of densities for chaotic delay equations with interesting statistical properties.

In the absence of a generally applicable analytical method, it is desirable to have an effective computational approach to the evolution of densities for DDEs. Of the numerical methods considered, the simplest is the "brute force" method of simulating large ensembles of solutions and compiling histograms to approximate densities (Section 4.3). Because this method relies on adequate statistical sampling to obtain accurate results it is computationally intensive, to the point of being impractical for many applications. Nevertheless, due to general results on the reliability of statistics computed from numerical simulations [11], one has reason to hope that ensemble simulation is a robust means of estimating ensemble densities.

As an alternative to ensemble simulation we have developed a numerical method for computing the evolution of densities for DDEs, based on piecewise linear approximation of the solution map (Section 4.4). This approximate solution map is easily inverted, and leads to an approximate version of the analytical approach considered in Section 4.2. This method is much more efficient for computing the evolution of densities. However, its effectiveness is limited for DDEs with chaotic dynamics, for which the solution map for large times can be extremely complex and difficult to approximate. Thus for chaotic DDEs this numerical approach is not suitable for evolving densities to arbitrarily large times.

The method of steps, frequently used to find analytical solutions of delay equations, can be used to express a delay equation as a sequence of ordinary differential equations. With appropriate modifications to this method, a DDE can be expressed as a system of simultaneous ODEs. One can then formulate the density evolution problem for a DDE in terms of a corresponding ODE system, and use methods for ODEs to write an evolution equation for the density (Section 4.5). For simple DDEs this equation can be solved analytically by the method of characteristics, reproducing the analytical results obtained in Section 4.2. For more complicated equations an analytical solution is not possible, but the method retains some intuitive appeal because it provides a model for the evolution of densities for DDEs, in terms of the transportation of a line mass by a flow in $\mathbb{R}^N$. A numerical implementation of this model provides geometrical insight into the results of our other numerical approaches to density evolution (Section 4.5.4). However, the model has limited utility when densities are evolved to large time, both because the dimension N increases with time, and, for chaotic DDEs, the solution map becomes very complicated and the resulting distribution of the line mass in $\mathbb{R}^N$ becomes difficult to approximate and in any case loses its intuitive appeal.

For a variety of delay differential equations, numerically computed solution ensembles appear to converge to a unique asymptotic distribution, characterized by an asymptotic probability measure η_*, with density ρ_*, on $\mathbb{R}^n$ (Section 5.1). The same density is observed if one constructs a histogram of solution values along a *single* solution of the DDE. This phenomenon can be understood in terms of (hypothetical) ergodic properties of the associated infinite dimensional dynamical system $\{S_t\}$ on C (Section 5.2). In the examples considered, $\{S_t\}$ is known to possess a compact attractor Λ: any ensemble of trajectories starting in the basin of attraction of Λ will, asymptotically, be distributed on Λ. The numerical evidence supports the existence of a natural invariant probability measure (*i.e.*, an SRB measure) μ_* supported on Λ. The asymptotic measure η_* can be interpreted as the image of μ_* under a suitable projection from C into $\mathbb{R}^n$.

The practical and theoretical importance of invariant measures, and SRB measures especially, makes the computation of asymptotic measures for DDEs a desirable goal. The simplest approach to this problem is to evolve an initial density forward in time until the asymptotic statistics become apparent. However, of the methods discussed above for computing the evolution of densities for DDEs, none is well suited to evolving densities to large times except the "brute force" ensemble simulation method. The increase in time of both the complexity of the solution map and the dimension of the system are funda-

mental obstacles to evolving densities to large times. In Chapter 5 we have investigated alternative approaches to estimating asymptotic densities.

For some DDEs we have hypothesized the existence of an SRB measure, since a histogram of a single solution reproduces the asymptotic density found by ensemble simulation. For DDEs with this property the asymptotic density can be approximated more efficiently by simulating a single long solution, rather than the many (*e.g.* 10^6) different solutions required for ensemble simulation (Section 5.2.4). While this approach can be used to quickly estimate the asymptotic density, it falls into the class of "brute force" methods that provide no insight into underlying the mechanism.

Ulam's method [119]—the most widely known technique for estimating invariant measures—can be formulated for DDEs so that it yields an approximation of the asymptotic measure η_* (Section 5.3). However, this approximation turns out to be identical (by definition) to the histogram of a time series generated by a typical solution of the given DDE, because our construction is somewhat circular. Thus, at least in our formulation, Ulam's method does not provide an independent estimate of the invariant density.

The "self-consistent Perron-Frobenius operator" method of [69, 75] is a promising approach to estimating asymptotic densities for DDEs. A suitable discretization of a given DDE yields a discrete-time system to which this method can be applied. However, a straightforward implementation (Section 5.4) fails to generate the desired approximate invariant density. Instead, the method converges to an unstable fixed point of the DDE. This failure can be explained, in part, by the fact that in the delay equations we consider, the instability (hence chaotic behavior and nontrivial statistical properties) is delay-induced, *i.e.* is not present in the absence of a non-zero delay. In the systems considered in [75], chaotic behavior is present even in the case of zero delay. Adapting the approach of [75] to systems with delay-induced instability remains an open problem.

A solution to the problem of estimating asymptotic densities for DDEs remains elusive. Previously published methods of estimating invariant measures for dynamical systems do not adapt well to delay equations. At present the only effective methods are based on computing statistics on numerical solutions.

Transient chaos in delay differential equations, although anticipated by a number of publications, has not been investigated before. We have found multistability with fractal basins of attraction, a key signature of transient chaos, in a delay equation with piecewise-constant nonlinearity (also reported in [81]) as well as the Mackey-Glass equation (Section 6.2). Existence of fractal basins suggests the existence of transiently chaotic

trajectories and the presence of a chaotic saddle. Indeed, by computing basins of attraction for these two delay equations we have been able to illustrate, numerically, the existence of solutions with long chaotic transients.

Previously published numerical methods for approximating unstable chaotic sets (*i.e.* chaotic saddles) all fail when applied to the delay equations considered here. The PIM method [90] in particular is expected to fail if the saddle has more than one unstable direction. However, our calculations of Lyapunov exponents point to the presence of only a single unstable direction, so this explanation is inadequate. Instead, the failure of the known algorithms appears to stem from the presence of saddle type unstable periodic orbits, to which each of the algorithms eventually converge. This mode of failure has not been observed before.

We have developed a modified version of the stagger-and-step method (Section 6.4.4), aimed at avoiding unstable periodic orbits. With this method, aperiodic numerical trajectories of arbitrarily long duration can be found for both of the delay equations considered in our examples. It is presumed that such a trajectory approximates the chaotic saddle.

Despite the fact that the saddle is embedded in an infinite dimensional phase space (or at least a finite- but high-dimensional phase space used for numerical approximation), it is possible to go some way toward visualizing the saddle by graphing its projections onto two dimensions. We have done this for the delay equations considered here, projecting the numerical approximation of the chaotic saddle onto two dimensions by applying a particular "trace map" commonly used for visualizing the dynamics of delay equations (Section 6.5.1).

The distribution of orbits on the chaotic saddle can be characterized by an invariant measure. This too can be approximated numerically from a stagger-and-step trajectory, and projected onto two dimensions for the purpose of visualization.

Chaotic invariant sets are typically characterized in terms of their ergodic parameters such as Lyapunov exponents and dimensions. These can be found using standard algorithms, involving straightforward computations on the numerical trajectory resulting from the stagger-and-step algorithm. For each of the delay equations considered we have applied these techniques and thereby estimated the Lyapunov spectrum, Lyapunov dimension, and correlation dimension of the respective chaotic saddles (Section 6.5.2). Despite the infinite dimensionality of the phase space of the delay equation, the saddle itself has quite low dimension, of the order $2 \sim 3$. The same observation has been made with regard to chaotic *attractors* of delay equations [42]. For both of the DDEs con-

sidered, the saddle has only one positive Lyapunov exponent, hence only one unstable direction. This lends support to our hypothesis that the previously published algorithms for approximating the saddle fail for some reason other than the existence of multiple unstable directions.

7.2 Directions for Further Research

A number of the investigations undertaken in this thesis suggest avenues for further research. Some of the more promising directions are outlined below, along with some new ideas that have not been sufficiently developed to warrant inclusion in the foregoing.

In Section 3.6 we showed that the notion of physical or SRB measure is problematic for infinite dimensional systems such as DDEs, owing to the lack (or ambiguity) of an appropriate notion of "almost every" in infinite dimensions. Prevalence [58] is one such notion that appears not to have been investigated in the context of physical measures. Its translation invariance makes prevalence a good candidate for what might be considered a "natural" sense of almost every, which is what we desire in the identification of a natural or physical invariant measure. Since prevalence is equivalent to "Lebesgue almost every" in finite dimensional spaces, using prevalence in the definition of SRB measure would be consistent with the already-accepted but inadequate definition. Prevalence of a given set also happens to be relatively easy to prove, thanks to a number of results established in [58]. Especially as SRB measure is expected to play an important role in an ergodic theoretic understanding of such infinite dimensional phenomena as *e.g.* turbulence in fluids, further investigation of the definition of SRB measure for infinite dimensional systems, and the potential role of the notion of prevalence in particular, is warranted.

Many results in ergodic theory are expressed in terms of integrals over the phase space. For this reason it is argued above that an ergodic theoretic understanding of delay equations will require a theory of integration on the function space C. Such a theory has been developed if the measure of integration is Wiener measure, but we have argued that this is inadequate for an understanding of delay equations because, in general, Wiener measure is not expected to be invariant under a given DDE. This position is dissatisfying, if only because the physics literature has amassed such a wealth of theory and tools (*e.g.* Feynman diagrams) for integrating functionals with respect to Wiener measure. The possibility of exploiting these tools in the context of delay equations has not been adequately explored. One avenue, for example, might follow the approach

of [17, 99, 100, 101], where exactness is proved for a class of partial differential equations by establishing a conjugacy with a certain dynamical system for which Wiener measure is invariant. It would be interesting to seek a class of delay equations for which a similar argument could be made. This would provide the first rigorous result on strong ergodic properties of a delay equation as a dynamical system in C.

In Section 4.4 we develop a numerical method for approximating the evolution of densities, based on piecewise-linear approximation of the solution map for a given DDE. The method itself is not limited to delay equations, and could be applied to any system for which a solution map can be approximated. It would be interesting, for example, to apply this approach to estimating invariant densities for ordinary differential equations.

Ulam's method may yet turn out to be an effective approach to estimating asymptotic densities for delay equations. The Euler discretization with time step h, considered in Section 5.4, defines a discrete-time dynamical system $S : \mathbb{R}^N \to \mathbb{R}^N$. For some delay equations S is known to have strong ergodic properties [80], making it a good candidate for the application of Ulam's method. That is, Ulam's method might be used to approximate the N-dimensional natural S-invariant measure. This would require a partition $\mathcal{A}$ of $\mathbb{R}^N$ and the definition of a transition probability matrix relative to $\mathcal{A}$ (equation (5.40), with Lebesgue measure λ on $\mathbb{R}^N$). At first this seems impractical, since even a relatively coarse partition would contain on the order of 100^N cells (with N large, say $N \gtrsim 100$, so that the discretization accurately models the DDE). However, attractors for delay equations have generally been found to be low-dimensional (e.g., $D \approx 3$) [42], and therefore should be resolvable with a carefully chosen partition containing on the order of 100^3 cells. This is within the limits of practical computation. An "automatic refinement" method for choosing an optimal partition, coupled with efficient construction of the transition matrix P, has recently been implemented in the software package GAIO [32, 33, 67]. Some exciting recent developments in this area show that numerical simulations can be used to prove rigorous results about dynamical systems [116, 117] and infinite dimensional systems in particular [29]. It seems reasonable that this software could be applied to the estimation of invariant measures for delay equations.

In Section 5.4 we attempted to develop a method for estimating asymptotic densities for DDEs, based on the ideas in [69, 75]. The results of this investigation were disappointing and somewhat surprising. Using the same ideas, the authors of [69] and [75] have successfully estimated "collapsed" (i.e., one- and two-dimensional) invariant densities for some other high-dimensional systems. We have hypothesized that this discrepancy is

due to the fact that for the systems considered in [75], chaotic behavior occurs even in the absence of an explicit delay in the dynamics, whereas for the DDEs we consider the instability is inherently delay-induced. This explanation is not entirely satisfactory. It seems that some variation on this theme should be effective for delay differential equations. In particular, the second-order method outlined in Section 5.4.4 warrants further investigation.

In Chapter 6, previously published algorithms for approximating chaotic saddles failed when applied to the delay equations we have considered. This failure is apparently due to the presence of saddle type periodic orbits, which capture the numerical trajectory that would otherwise approximate the saddle. This mechanism should not be specific to delay equations. It would be interesting to investigate the behavior of these algorithms for simpler (*e.g.* finite dimensional) systems that also possess saddle-type periodic orbits, both to test our hypotheses regarding this mode of failure and to investigate more effective remedies.

Our modification of the stagger-and-step method [110] successfully avoids the difficulty that otherwise occurs in the presence of saddle type periodic orbits. An obvious extension of this work would be an attempt at a similar modification of the PIM method. Our numerical results suggest that the chaotic saddle has only one unstable direction, at least for the DDEs we have considered. Thus in principle the PIM method [90] should be applicable, but for the difficulty presented by the presence of saddle type periodic orbits. A successful modification of this method would be a valuable tool not just for delay equations but for the numerical analysis of transient chaos in general.

A theme that arises in various contexts throughout this thesis is that the analysis of DDEs is complicated by their infinite dimensionality. Chaos in infinite-dimensional systems is not a new subject: the field has a rich literature, with a focus mainly on chaotic partial differential equations (PDEs). However, little has been said about PDEs in an ergodic theoretic context; consequently this thesis provides only scant discussion of the PDE literature. Nevertheless, a number of fruitful avenues for further research are suggested by the problem of adapting and applying to DDEs the techniques that have been developed for PDEs.

One such collection of techniques is motivated by the possible existence of *inertial manifolds* for DDEs. An inertial manifold (IM), a subset of phase space, is a smooth, finite-dimensional invariant manifold that exponentially attracts all trajectories in a certain neighborhood (see *e.g.* [112, 113, 114] and references therein). Inertial manifolds

have been shown to exist for a broad class of dissipative dynamical systems including some PDEs. If an inertial manifold $\mathcal{M}$ exists then any global attractor will be contained in $\mathcal{M}$; thus the subsystem obtained by restricting the original dynamical system to $\mathcal{M}$ faithfully reproduces the asymptotic dynamics. This subsystem is itself equivalent to a certain *finite* system of ordinary differential equations. This drastic simplification makes it possible to analyze the asymptotic dynamics using methods for finite dimensional systems. This approach has led, for example, to rigorous bounds on the dimension of the attractor for some PDEs [63]. Techniques have also been developed for approximating inertial manifolds numerically, and results have been obtained on the persistence of inertial manifolds under perturbation and numerical approximation [31, 62, 64, 65, 66].

We are unaware of any investigations of inertial manifolds for chaotic DDEs (in the case of a DDE we are interested in finite-dimensional invariant manifolds in the function space C). However, the observations in [42] and elsewhere of low-dimensional attractors for some DDEs suggest the presence of an inertial manifold. In a preliminary investigation in this direction, we have considered the special case of seeking a flat inertial manifold, *i.e.* an attracting invariant linear subspace of C. If such an invariant subspace exists it should be readily observable, *e.g.* by employing a Gram-Schmidt procedure to show (numerically) the existence of a finite basis for points on the attractor. We have carried out this procedure for the chaotic DDEs considered in Section 5.1, with results indicating that there is no finite basis for the attractor, hence no flat inertial manifold. The possible existence of curved inertial manifolds warrants further investigation. The finite dimensional reduction such an invariant manifold would afford might be instrumental in circumventing some of the difficulties encountered in this thesis.

As mentioned in Section 3.2, delay differential equations are a special case of a more general class of retarded functional differential equations (RFDEs). Throughout this thesis we have restricted our attention to delay equations having the particular form of equation (3.1) (*e.g.* with a single, fixed delay). At times we have been able to exploit the relative simplicity of the special form of this equation (*e.g.* Section 4.5) to obtain a desired result. However, many of the results of this thesis do not rely on the special form of this equation. An interesting and obvious avenue for further research is to extend the present work, where possible, to more general RFDEs.

Bibliography

[1] J. M. Aguirregabiria and J. R. Etxebarria, *Fractal basin boundaries of a delay-differential equation*, Phys. Lett. A **122** (1987), no. 5, 241–244.

[2] G. Ahlers and R. W. Walden, *Turbulence near onset of convection*, Phys. Rev. Lett. **44** (1980), no. 7, 445–448.

[3] Uwe an der Heiden, *Delays in physiological systems*, J. Math. Biol. **8** (1979), 345–364.

[4] ———, *Stochastic properties of simple differential-delay equations*, Delay Equations, Approximation and Application (G. Meinardus and G. Nurnberger, eds.), International Series of Numerical Mathematics, vol. 74, Birkhauser, 1985, pp. 147–164.

[5] Uwe an der Heiden and Michael C. Mackey, *The dynamics of production and destruction: Analytic insight into complex behavior*, J. Math. Biol. **16** (1982), 75–101.

[6] Uwe an der Heiden, Michael C. Mackey, and Hans-Otto Walther, *Complex oscillations in a simple deterministic neuronal network*, Mathematical aspects of physiology (F. Hoppensteadt, ed.), American Mathematical Society, Providence RI, 1981, pp. 355–360.

[7] Uwe an der Heiden and Hans-Otto Walther, *Existence of chaos in control systems with delayed feedback*, J. Differential Equations **47** (1983), 273–295.

[8] D. C. Baird, *Experimentation: An introduction to measurement theory and experiment design*, 2nd ed., Prentice Hall, Englewood Cliffs NJ, 1988.

[9] P. M. Battelino, C. Grebogi, E. Ott, J. A. Yorke, and E. D. Yorke, *Multiple coexisting attractors, basin boundaries and basic sets*, Physica D **32** (1988), 296.

[10] A. Bellen and S. Maset, *Numerical solution of constant coefficient linear delay differential equations as abstract Cauchy problems*, Numerische Mathematik **84** (2000), 351–374.

[11] G. Benettin, M. Casartelli, L. Galgani, A. Giorgilli, and J.-M. Strelcyn, *On the reliability of numerical studies of stochasticity. Part I: Existence of time averages.*, Nuovo Cimento **44** (1978), no. 1, 183–196.

[12] Giancarlo Benettin, Luigi Galgani, Antonio Giorgilli, and Jean-Marie Strelcyn, *Lyapunov characteristic exponents for smooth dynamical systems; a method for computing all of them. Part 1: Theory*, Meccanica **15** (1980), 9–20.

[13] ———, *Lyapunov characteristic exponents for smooth dynamical systems; a method for computing all of them. Part 2: Numerical application*, Meccanica **15** (1980), 21–30.

[14] Michael Blank and Leonid Bunimovich, *Multicomponent dynamical systems: SRB measures and phase transitions*, Nonlinearity **16** (2003), 387–401.

[15] Erik M. Bollt, *Controlling chaos and the inverse Perron-Frobenius problem: Global stabilization of arbitrary invariant measures*, International Journal of Bifurcation and Chaos **10** (2000), no. 5, 1033–1050.

[16] R. Bowen and D. Ruelle, *The ergodic theory of Axiom A flows*, Invent. Math. **79** (1975), 181–202.

[17] Pavol Brunovsky and Jozef Komornik, *Ergodicity and exactness of the shift on $C[0, \infty)$ and the semiflow of a first-order partial differential equation*, Journal of Mathematical Analysis and Applications **104** (1984), 235–245.

[18] L. A. Buminovich, S. G. Dani, R. L. Dobrushin, M. V. Jakobson, I. P. Kornfeld, N. B. Maslova, Ya. B. Pesin, Ya. G. Sinai, J. Smillie, Y. M. Sukhov, and A. M. Vershik, *Dynamical systems, ergodic theory and applications*, 2nd ed., Encyclopedia of Mathematical Sciences, vol. 100, Springer, New York, 2000.

[19] Marek Capiński, *Hopf equation for some nonlinear differential delay equation and invariant measures for corresponding dynamical system*, Universitatis Iagellonicae Acta Mathematica **XXVIII** (1991), 171–175.

[20] George F. Carrier and Carl E. Pearson, *Partial differential equations*, 2nd ed., Academic Press, San Diego, 1988.

[21] Chaos **3** (1993), no. 4, Focus issue on scattering.

[22] Goong Chen, Sze-Bi Hsu, and Jianxin Zhou, *Chaotic vibrations of the one-dimensional wave equation due to a self-excitation boundary condition. I. Controlled hysteresis.*, Trans. Amer. Math. Soc. **350** (1998), no. 11, 4265–4311.

[23] B. D. Coleman and G. H. Renninger, *Theory of the response of the limulus retina to periodic excitation*, J. Math. Biol. **3** (1976), 103–120.

[24] P. Collet and J.-P. Eckmann, *Iterated maps of the interval as dynamical systems*, Birkhauser, Cambridge, 1980.

[25] S. P. Corwin, D. Sarafyan, and S. Thompson, *DKLAG6: A code based on continuously embedded sixth-order Runge-Kutta methods for the solution of state-dependent functional differential equations*, Appl. Numer. Math. **24** (1997), 317–330.

[26] R. F. Curtain and A. J. Pritchard, *Infinite dimensional linear systems theory*, Springer-Verlag, 1978.

[27] A. G. Darbyshire and T. Mullin, *Transition to turbulence in constant-mass-flux pipe-flow*, J. Fluid Mech. **289** (1995), 83–114.

[28] F. Darwin (ed.), *The life and letters of Charles Darwin*, Murray, London, 1887.

[29] S. Day, O. Junge, and K. Mischaikow, *A rigorous numerical method for the global analysis of infinite dimensional discrete dynamical systems*, preprint (2003).

[30] S. A. Deadwyler and R. E. Hampson, *Ensemble activity and behavior: What's the code?*, Science **270** (1995), 1316–1318.

[31] A. Debussche and R. Témam, *Convergent families of approximate inertial manifolds*, J. Math. Pures Appl. **73** (1994), no. 5, 489–522.

[32] Michael Dellnitz and Oliver Junge, *An adaptive subdivision technique for the approximation of attractors and invariant measures*, Computing and visualization science **1** (1998), 63–68.

[33] _____, *On the approximation of complicated dynamical behavior*, SIAM Journal on Numerical Analysis **36** (1999), no. 2, 491–515.

[34] Mukeshwar Dhamala and Ying-Cheng Lai, *Unstable orbits and the natural measure on nonhyperbolic chaotic saddles*, Phys. Rev. E **60** (1999), no. 5, 6176–6179.

[35] Mukeshwar Dhamala, Ying-Cheng Lai, and Eric J. Kostelich, *Analyses of transient chaotic time series*, Phys. Rev. E **64** (2001), no. 5.

[36] O. Diekmann, S. A. van Gils, S. M. Verduyn Lunel, and H.-O. Walther, *Delay equations: Functional-, complex-, and nonlinear analysis*, Applied Mathematical Sciences, vol. 110, Springer-Verlag, New York, 1995.

[37] Jiu Ding and Tien-Yien Li, *Markov finite approximation of Frobenius-Perron operator*, Nonlinear analysis: Theory, Methods & Applications **17** (1991), no. 8, 759–772.

[38] Jiu Ding and Aihui Zhou, *Finite approximations of Frobenius-Perron operators: A solution of Ulam's conjecture to multi-dimensional transformations*, Physica D **92** (1996), 61–68.

[39] R. D. Driver, *Ordinary and delay differential equations*, Springer-Verlag, New York—Heidelberg—Berlin, 1977.

[40] J.-P. Eckmann and D. Ruelle, *Ergodic theory of chaos and strange attractors*, Rev. Mod. Phys. **57** (1985), no. 3, 617–656.

[41] Sergey V. Ershov, *Asymptotic theory of multidimensional chaos*, J. Stat. Phys. **69** (1991), no. 3/4, 781–812.

[42] J. D. Farmer, *Chaotic attractors of infinite-dimensional systems*, Physica D **4** (1982), 366–392.

[43] P. Frederickson, J. Kaplan, E. Yorke, and J. Yorke, *The Lyapunov dimension of strange attractors*, J. Differential Equations **49** (1983), 185.

[44] Avner Friedman, *Foundations of modern analysis*, Holt, Rinehart and Winston, Inc., New York, 1970.

[45] Jean-Michel Ghidaglia, *Finite-dimensional behavior for weakly damped driven Schrödinger equations*, Ann. Inst. H. Poincaré Anal. Non Linéaire **5** (1988), no. 4, 365–405.

[46] H. M. Gibbs, *Optical bistability: Controlling light with light*, Academic, Orlando, 1985.

[47] Leon Glass and Michael C. Mackey, *Pathological conditions resulting from instabilities in physiological control systems*, Annals of the New York Academy of Science **316** (1979), 214–235.

[48] P. Grassberger and I. Procaccia, *Measuring the strangeness of strange attractors*, Physica D **9** (1983), 189–208.

[49] C. Grebogi, E. Ott, and J. A. Yorke, *Crises, sudden changes in chaotic attractors, and transient chaos*, Physica D **7** (1983), 181–200.

[50] R. Grimshaw and X. Tian, *k-soliton structures, period-doubling and spatial chaos in the periodically forced Korteweg-de Vries equation*, Differential Equations and Dynamical Systems **4** (1996), no. 2, 225–251.

[51] John Guckenheimer and Philip Holmes, *Nonlinear oscillations, dynamical systems, and bifurcations of vector fields*, Applied Mathematical Sciences, vol. 42, Springer-Verlag, New York—Berlin—Heidelberg—Tokyo, 1983.

[52] H. Haken, *At least one Lyapunov exponent vanishes if the trajectory of an attractor does not contain a fixed point*, Physics Letters **94** (1983), 71.

[53] Jack K. Hale and Sjoerd M. Verduyn Lunel, *Introduction to functional differential equations*, Applied Mathematical Sciences, vol. 99, Springer-Verlag, New York, 1993.

[54] Jack K. Hale and Natalia Sternberg, *Onset of chaos in differential delay equations*, Journal of Computational Physics **77** (1988), 221–239.

[55] P. R. Halmös, *Measure theory*, Springer-Verlag, New York, 1974.

[56] M. Hénon, *A two-dimensional mapping with a strange attractor*, Comm. Math. Phys. **50** (1976), 69–77.

[57] Philip Holmes and Jerrold Marsden, *A partial differential equation with infinitely many periodic orbits: Chaotic oscillations of a forced beam*, Arch. Rational Mech. Anal. **76** (1981), no. 2, 135–165.

[58] B. Hunt, T. Sauer, and J. Yorke, *Prevalence: A translation-invariant "almost every" on infinite-dimensional spaces*, Bull. Amer. Math. Soc. **27** (1992), 217–238.

[59] Brian R. Hunt and Vadim Yu Kaloshin, *Regularity of embeddings of infinite-dimensional fractal sets into finite-dimensional spaces*, Nonlinearity **12** (1999), 1263–1275.

[60] Ross Ihaka and Robert Gentleman, *R: A language for data analysis and graphics*, Journal of Computational and Graphical Statistics **5** (1996), no. 3, 299–314.

[61] Kensuke Ikeda and Kenji Matsumoto, *High-dimensional chaotic behavior in systems with delayed feedback*, Physica D **29** (1987), 223–235.

[62] M. S. Jolly, R. Rosa, and R. Témam, *Accurate computations on inertial manifolds*, SIAM J. Sci. Comput. **22** (2000), no. 6, 2216–2238.

[63] ———, *Evaluating the dimension of an inertial manifold for the Kuramoto-Sivashinsky equation*, Adv. Differential Equations **5** (2000), no. 1–3, 31–66.

[64] Don A. Jones and Andrew M. Stuart, *Attractive invariant manifolds under approximation: Inertial manifolds*, J. Differential Equations **123** (1995), no. 2, 588–637.

[65] Don A. Jones, Andrew M. Stuart, and Edriss S. Titi, *Persistence of invariant sets for dissipative evolution equations*, Journal of Mathematical Analysis and Applications **219** (1998), no. 2, 479–502.

[66] Don A. Jones and Edriss S. Titi, *Approximations of inertial manifolds for dissipative nonlinear equations*, J. Differential Equations **127** (1996), no. 1, 54–86.

[67] O. Junge, *An adaptive subdivision technique for the approximation of attractors and invariant measures: Proof of convergence*, Dynamical Systems **16** (2001), no. 3, 213–222.

[68] Mark Kac, *Integration in function spaces and some of its applications*, Accademia Nazionale dei Lincei, Scuola Normale Superiore, Pisa, 1980.

[69] Kunihiko Kaneko, *Self-consistent Perron-Frobenius operator for spatiotemporal chaos*, Phys. Lett. A **139** (1989), no. 1–2, 47–52.

[70] H. Kantz and P. Grassberger, *Repellers, semi-attractors, and long-lived chaotic transients*, Physica D **17** (1985), 75–86.

[71] A. Katok and B. Hasselblatt, *Introduction to the modern theory of dynamical systems*, Encyclopedia of Mathematics and its Applications, vol. 54, Cambridge University Press, Cambridge, 1995.

[72] Yuri A. Kuznetsov, *Elements of applied bifurcation theory*, Applied Mathematical Sciences, vol. 112, Springer-Verlag, New York, 1995.

[73] Serge Lang, *Real and functional analysis*, 3rd ed., Graduate Texts in Mathematics, vol. 142, Springer-Verlag, New York—Berlin—Heidelberg, 1993.

[74] Andrzej Lasota and Michael C. Mackey, *Chaos, fractals, and noise: Stochastic aspects of dynamics*, Applied Mathematical Sciences, vol. 97, Springer-Verlag, New York, 1994.

[75] S. Lepri, G. Giacomelli, A. Politi, and F. T. Arecchi, *High-dimensional chaos in delayed dynamical systems*, Physica D **70** (1993), 235–249.

[76] T. Y. Li and J. A. Yorke, *Period three implies chaos*, Amer. Math. Monthly **82** (1975), 985–992.

[77] Tien-Yien Li, *Finite approximation for the Frobenius-Perron operator: A solution to Ulam's conjecture*, Journal of Approximation Theory **17** (1976), 177–186.

[78] E. N. Lorenz, *Deterministic non-periodic flow*, J. Atmos. Sci. **20** (1963), 130–141.

[79] Jérôme Losson and Michael C. Mackey, *A Hopf-like equation and perturbation theory for differential delay equations*, J. Stat. Phys. **69** (1992), no. 5/6, 1025–1046.

[80] ———, *Coupled map lattices as models of deterministic and stochastic differential delay equations*, Phys. Rev. E **52** (1995), no. 1, 115–128.

[81] Jérôme Losson, Michael C. Mackey, and André Longtin, *Solution multistability in first-order nonlinear differential delay equations*, Chaos **3** (1993), no. 2, 167–176.

[82] Benjamin P. Luce, *Homoclinic explosions in the complex Ginzburg-Landau equation*, Physica D **84** (1995), no. 3–4, 553–581.

[83] Michael C. Mackey, *Commodity price fluctuations: Price dependent delays and non-linearities as explanatory factors*, J. Econ. Theory **48** (1989), no. 2, 497–509.

[84] ———, *Time's arrow: The origins of thermodynamic behaviour*, Springer-Verlag, New York, 1992.

[85] ———, *Density evolution in systems with delayed dynamics*, Private communication, March 1994.

[86] Michael C. Mackey and Leon Glass, *Oscillation and chaos in physiological control systems*, Science **197** (1977), 287–289.

[87] Steven W. McDonald, Celso Grebogi, Edward Ott, and James A. Yorke, *Fractal basin boundaries*, Physica D **17** (1985), 125–153.

[88] John G. Milton and Michael C. Mackey, *Neural ensemble coding and statistical periodicity: Speculations on the operation of the mind's eye*, Journal of Physiology— Paris **94** (2000), no. 5–6, 489–503.

[89] I. Ncube, S. A. Campbell, and E. R. Vrscay, *Stationary densities and the stochastic approximation of a certain class of random algorithms*, Differential Equations and Dynamical Systems (preprint).

[90] Helena E. Nusse and James A. Yorke, *A procedure for finding numerical trajectories on chaotic saddles*, Physica D **36** (1989), 137–156.

[91] ———, *Analysis of a procedure for finding numerical trajectories close to chaotic saddle hyperbolic sets*, Ergodic Theory and Dynamical Systems **11** (1991), 189–208.

[92] V. I. Oseledec, *A multiplicative ergodic theorem: Lyapunov characteristic exponents for dynamical systems*, Trans. Moscow Math. Soc. **19** (1968), 197.

[93] Mark Pollicott and Michiko Yuri, *Dynamical systems and ergodic theory*, London Mathematical Society Student Texts, vol. 40, Cambridge University Press, 1998.

[94] William H. Press, Saul A. Teukolsky, William T. Vetterling, and Brian P. Flannery, *Numerical recipes in C*, 2nd ed., Cambridge University Press, 1992.

[95] Ilya Prigogine, *From being to becoming: Time and complexity in the physical sciences*, W. H. Freeman and Company, San Francisco, 1980.

[96] _____, *The end of certainty: Time, chaos, and the new laws of nature*, Free Press, New York, 1996.

[97] H. L. Royden, *Real analysis*, MacMillan, London, 1968.

[98] Walter Rudin, *Principles of mathematical analysis*, 3rd ed., McGraw-Hill, New York, 1976.

[99] Ryszard Rudnicki, *Invariant measures for the flow of a first order partial differential equation*, Ergodic Theory and Dynamical Systems **5** (1985), 437–443.

[100] _____, *An abstract Wiener measure invariant under a partial differential equation*, Bulletin of the Polish Academy of Sciences: Mathematics **35** (1987), no. 5–6, 289–295.

[101] _____, *Strong ergodic properties of a first-order partial differential equation*, Journal of Mathematical Analysis and Applications **133** (1988), 14–26.

[102] David Ruelle, *A measure associated with Axiom A attractors*, Am. J. Math. **98** (1976), 619.

[103] _____, *Small random perturbations of dynamical systems and the definition of attractors*, Commun. Math. Phys. **82** (1981), 137–151.

[104] _____, *Chaotic evolution and strange attractors: The statistical analysis of time series for deterministic nonlinear systems*, Cambridge University Press, New York, 1989.

[105] L. H. Ryder, *Quantum field theory*, Cambridge University Press, Cambridge, 1985.

[106] Tim Sauer, James A. Yorke, and Martin Casdagli, *Embedology*, J. Stat. Phys. **65** (1991), no. 3/4, 579–616.

[107] L. F. Shampine and S. Thompson, *Solving DDEs in MATLAB*, Appl. Numer. Math. **37** (2001), no. 4, 441–458.

[108] Robert Shaw, *Strange attractors, chaotic behavior, and information flow*, Zeitschrift fur Naturforschung **36a** (1981), 80–112.

[109] S. Smale, *Differentiable dynamical systems*, Bull. Amer. Math. Soc. **73** (1967), 747–817.

[110] David Sweet, Helena E. Nusse, and James A. Yorke, *Stagger-and-step method: Detecting and computing chaotic saddles in higher dimensions*, Phys. Rev. Lett. **86** (2001), no. 11, 2261–2264.

[111] Tamás Tél, *Transient chaos*, Experimental study and characterization of chaos (Hao Bai-Lin, ed.), Directions in Chaos, vol. 3, World Scientific, Singapore—New Jersey—London—Hong Kong, 1990, pp. 149–211.

[112] Roger Témam, *Inertial manifolds for nonlinear evolutionary equations*, J. Differential Equations **72** (1988), no. 2, 309–353.

[113] ———, *Inertial manifolds*, Mathematical Intelligencer **12** (1990), no. 4, 68–74.

[114] ———, *Infinite-dimensional dynamical systems in mechanics and physics*, 2nd ed., Applied Mathematical Sciences, vol. 68, Springer-Verlag, New York, 1997.

[115] X. Tian and R. H. J. Grimshaw, *Low-dimensional chaos in a perturbed Korteweg-de Vries equation*, Internat. J. Bifur. Chaos Appl. Sci. Engrg. **5** (1995), no. 4, 1221–1233.

[116] Warwick Tucker, *The Lorenz attractor exists*, Comptes Rendus de l'Academie des Sciences, Serie I – Mathematique **328** (1999), no. 12, 1197–1202.

[117] ———, *A rigorous ODE solver and Smale's 14th problem*, Foundations of Computational Mathematics **2** (2002), no. 1, 53–117.

[118] A. Uçar, *A prototype model for chaos studies*, International Journal of Engineering Science **40** (2002), 251–258.

[119] S. Ulam, *Problems in modern mathematics*, Interscience, New York, 1960.

[120] Marcela Viana, *Dynamical systems: Moving into the next century*, Mathematics Unlimited: 2001 and Beyond (Björn Engquist and Wilfried Schmidt, eds.), Springer, New York, 2001, p. 1167.

[121] Hans-Otto Walther, *Homoclinic solution and chaos in $\dot{x}(t) = f(x(t-1))$*, Nonlinear Analysis **5** (1981), 775–788.

[122] Yasuo Yamasaki, *Measures on infinite dimensional spaces*, Series in Pure Mathematics, vol. 5, World Scientific, Singapore—Philadelphia, 1985.

[123] Chao Yang Yin and Wei Ding, *An estimate of the Hausdorff dimension of a uniform attractor for a class of nonautonomous reaction-diffusion equations with a quasiperiodic external force*, J. Lanzhou Univ. Nat. Sci. **33** (1997), no. 4, 14–19.

[124] J. A. Yorke and E. D. Yorke, *Metastable chaos: Transition to sustained chaotic behavior in the Lorentz model*, J. Stat. Phys. **21** (1979), 263.

[125] Erich Zauderer, *Partial differential equations of applied mathematics*, Wiley, New York, 1983.

[126] A. Zubrzycki, *Spatial inhomogeneity of periodic and chaotic attractors of a driven damped nonlinear Schrödinger equation*, Acta Phys. Polon. A **88** (1995), no. 6, 1067–1072.

Nomenclature

$\ll$ $\nu \ll \mu$ means measure ν is absolutely continuous with respect to μ, page 22

1_A indicator function of a set A, page 20

$\mathcal{A}$ σ-algebra, page 16

$C([a,b])$ space of continuous functions from $[a,b]$ into $\mathbb{R}^n$, with sup norm, page 43

C space of continuous functions from $[-\tau, 0]$ into $\mathbb{R}^n$, with sup norm, page 44

C space of continuous functions from $[-1, 0]$ into $\mathbb{R}^n$, with sup norm, page 47

$D(X)$ set of densities on a phase space X, a subset of $L^1(X)$, page 23

ϕ initial function for a delay differential equation, page 43

Λ Attractor of a dynamical system, page 123

λ Lebesgue measure on $\mathbb{R}^n$, page 18

L^p L^p space of functions, page 22

$\mu(A)$ measure of a set A, page 17

μ_f probability measure with density functional f, page 22

P Perron-Frobenius operator, page 24

P_t Perron-Frobenius operator corresponding to S_t, page 24

$\mathbb{R}_+$ the set $\{x \in \mathbb{R} : x \geq 0\}$, page 15

$\sigma(\mathcal{B})$ σ-algebra generated by a class $\mathcal{B}$ of sets, page 17

$S^{-1}(A)$ pre-image of a set A under a transformation S, page 19

S_t dynamical system or semigroup of transformations, page 15

t time variable ($\in \mathbb{R}_+$ or $\mathbb{Z}_+$) for a dynamical system, page 14

X phase space, page 15

$(X, \mathcal{A})$ measurable space, page 16

$(X, \mathcal{A}, \mu)$ measure space or probability space, page 17

$x(t)$ solution of a delay differential equation, page 43

x_t phase point in X at time t for a dynamical system, page 14

$\mathbb{Z}_+$ the set $\{x \in \mathbb{Z} : x \geq 0\} = \{0, 1, 2, \ldots\}$, page 15