Introduction to Dynamical Systems

MICHAEL BRIN
University of Maryland

GARRETT STUCK
University of Maryland

CAMBRIDGE
UNIVERSITY PRESS

PUBLISHED BY THE PRESS SYNDICATE OF THE UNIVERSITY OF CAMBRIDGE
The Pitt Building, Trumpington Street, Cambridge, United Kingdom

CAMBRIDGE UNIVERSITY PRESS
The Edinburgh Building, Cambridge CB2 2RU, UK
40 West 20th Street, New York, NY 10011-4211, USA
477 Williamstown Road, Port Melbourne, VIC 3207, Australia
Ruiz de Alarcón 13, 28014 Madrid, Spain
Dock House, The Waterfront, Cape Town 8001, South Africa

http://www.cambridge.org

First published 2002

Printed in the United States of America

Typefaces Times Ten 10/13 pt. and Avenir   *System* LaTeX 2_ε [TB]

A catalog record for this book is available from the British Library.

Library of Congress Cataloging in Publication Data

Brin, Michael.
Introduction to dynamical systems / Michael Brin, Garrett Stuck.
p. cm.
Includes bibliographical references and index.
ISBN 0-521-80841-3
1. Differentiable dynamical systems. I. Stuck, Garrett, 1961– II. Title.
QA614.8 .B75 2002
514′.74 – dc21   2002022281

ISBN 0 521 80841 3 hardback

Contents

CHAPTER ONE

Examples and Basic Concepts

Dynamical systems is the study of the long-term behavior of evolving systems. The modern theory of dynamical systems originated at the end of the 19th century with fundamental questions concerning the stability and evolution of the solar system. Attempts to answer those questions led to the development of a rich and powerful field with applications to physics, biology, meteorology, astronomy, economics, and other areas.

By analogy with celestial mechanics, the evolution of a particular state of a dynamical system is referred to as an *orbit*. A number of themes appear repeatedly in the study of dynamical systems: properties of individual orbits; periodic orbits; typical behavior of orbits; statistical properties of orbits; randomness vs. determinism; entropy; chaotic behavior; and stability under perturbation of individual orbits and patterns. We introduce some of these themes through the examples in this chapter.

We use the following notation throughout the book: $\mathbb{N}$ is the set of positive integers; $\mathbb{N}_0 = \mathbb{N} \cup \{0\}$; $\mathbb{Z}$ is the set of integers; $\mathbb{Q}$ is the set of rational numbers; $\mathbb{R}$ is the set of real numbers; $\mathbb{C}$ is the set of complex numbers; $\mathbb{R}^+$ is the set of positive real numbers; $\mathbb{R}_0^+ = \mathbb{R}^+ \cup \{0\}$.

1.1 The Notion of a Dynamical System

A *discrete-time dynamical system* consists of a non-empty set X and a map $f\colon X \to X$. For $n \in \mathbb{N}$, the nth iterate of f is the n-fold composition $f^n = f \circ \cdots \circ f$; we define f^0 to be the identity map, denoted Id. If f is invertible, then $f^{-n} = f^{-1} \circ \cdots \circ f^{-1}$($n$ times). Since $f^{n+m} = f^n \circ f^m$, these iterates form a group if f is invertible, and a semigroup otherwise.

Although we have defined dynamical systems in a completely abstract setting, where X is simply a set, in practice X usually has additional structure that is preserved by the map f. For example, (X, f) could be a measure space and a measure-preserving map; a topological space and a continuous map; a metric space and an isometry; or a smooth manifold and a differentiable map.

A *continuous-time dynamical system* consists of a space X and a one-parameter family of maps of $\{f^t\colon X \to X\}$, $t \in \mathbb{R}$ or $t \in \mathbb{R}_0^+$, that forms a one-parameter group or semigroup, i.e., $f^{t+s} = f^t \circ f^s$ and $f^0 = \mathrm{Id}$. The dynamical system is called a *flow* if the time t ranges over $\mathbb{R}$, and a *semiflow* if t ranges over $\mathbb{R}_0^+$. For a flow, the *time-t map* f^t is invertible, since $f^{-t} = (f^t)^{-1}$. Note that for a fixed t_0, the iterates $(f^{t_0})^n = f^{t_0 n}$ form a discrete-time dynamical system.

We will use the term *dynamical system* to refer to either discrete-time or continuous-time dynamical systems. Most concepts and results in dynamical systems have both discrete-time and continuous-time versions. The continuous-time version can often be deduced from the discrete-time version. In this book, we focus mainly on discrete-time dynamical systems, where the results are usually easier to formulate and prove.

To avoid having to define basic terminology in four different cases, we write the elements of a dynamical system as f^t, where t ranges over $\mathbb{Z}$, $\mathbb{N}_0$, $\mathbb{R}$, or $\mathbb{R}_0^+$, as appropriate. For $x \in X$, we define the *positive semiorbit* $\mathcal{O}_f^+(x) = \bigcup_{t \geq 0} f^t(x)$. In the invertible case, we define the *negative semiorbit* $\mathcal{O}_f^-(x) = \bigcup_{t \leq 0} f^t(x)$, and the *orbit* $\mathcal{O}_f(x) = \mathcal{O}_f^+(x) \cup \mathcal{O}_f^-(x) = \bigcup_t f^t(x)$ (we omit the subscript "f" if the context is clear). A point $x \in X$ is a *periodic point* of *period* $T > 0$ if $f^T(x) = x$. The orbit of a periodic point is called a *periodic orbit*. If $f^t(x) = x$ for all t, then x is a *fixed point*. If x is periodic, but not fixed, then the smallest positive T, such that $f^T(x) = x$, is called the *minimal period* of x. If $f^s(x)$ is periodic for some $s > 0$, we say that x is *eventually periodic*. In invertible dynamical systems, eventually periodic points are periodic.

For a subset $A \subset X$ and $t > 0$, let $f^t(A)$ be the image of A under f^t, and let $f^{-t}(A)$ be the preimage under f^t, i.e., $f^{-t}(A) = (f^t)^{-1}(A) = \{x \in X\colon f^t(x) \in A\}$. Note that $f^{-t}(f^t(A))$ contains A, but, for a non-invertible dynamical system, is generally not equal to A. A subset $A \subset X$ is *f-invariant* if $f^t(A) \subset A$ for all t; *forward f-invariant* if $f^t(A) \subset A$ for all $t \geq 0$; and *backward f-invariant* if $f^{-t}(A) \subset A$ for all $t \geq 0$.

In order to classify dynamical systems, we need a notion of equivalence. Let $f^t\colon X \to X$ and $g^t\colon Y \to Y$ be dynamical systems. A *semiconjugacy* from (Y, g) to (X, f) (or, briefly, from g to f) is a surjective map $\pi\colon Y \to X$ satisfying $f^t \circ \pi = \pi \circ g^t$, for all t. We express this formula schematically by saying that the following diagram commutes:

$$\begin{array}{ccc} Y & \xrightarrow{g} & Y \\ {\scriptstyle\pi}\downarrow & & \downarrow{\scriptstyle\pi} \\ X & \xrightarrow{f} & X \end{array}$$

An invertible semiconjugacy is called a *conjugacy*. If there is a conjugacy from one dynamical system to another, the two systems are said to be *conjugate*; conjugacy is an equivalence relation. To study a particular dynamical system, we often look for a conjugacy or semiconjugacy with a better-understood model. To classify dynamical systems, we study equivalence classes determined by conjugacies preserving some specified structure. Note that for some classes of dynamical systems (e.g., measure-preserving transformations) the word *isomorphism* is used instead of "conjugacy."

If there is a semiconjugacy π from g to f, then (X, f) is a *factor* of (Y, g), and (Y, g) is an *extension* of (X, f). The map $\pi: Y \to X$ is also called a *factor map* or *projection*. The simplest example of an extension is the *direct product*

$$(f_1 \times f_2)^t: X_1 \times X_2 \to X_1 \times X_2$$

of two dynamical systems $f_i^t: X_i \to X_i$, $i = 1, 2$, where $(f_1 \times f_2)^t(x_1, x_2) = (f_1^t(x_1), f_2^t(x_2))$. Projection of $X_1 \times X_2$ onto X_1 or X_2 is a semiconjugacy, so (X_1, f_1) and (X_2, f_2) are factors of $(X_1 \times X_2, f_1 \times f_2)$.

An extension (Y, g) of (X, f) with factor map $\pi: Y \to X$ is called a *skew product* over (X, f) if $Y = X \times F$, and π is the projection onto the first factor or, more generally, if Y is a fiber bundle over X with projection π.

Exercise 1.1.1. Show that the complement of a forward invariant set is backward invariant, and vice versa. Show that if f is bijective, then an invariant set A satisfies $f^t(A) = A$ for all t. Show that this is false, in general, if f is not bijective.

Exercise 1.1.2. Suppose (X, f) is a factor of (Y, g) by a semiconjugacy $\pi: Y \to X$. Show that if $y \in Y$ is a periodic point, then $\pi(y) \in X$ is periodic. Give an example to show that the preimage of a periodic point does not necessarily contain a periodic point.

1.2 Circle Rotations

Consider the unit circle $S^1 = [0, 1]/\sim$, where $\sim$ indicates that 0 and 1 are identified. Addition mod 1 makes S^1 an abelian group. The natural distance on [0, 1] induces a distance on S^1; specifically,

$$d(x, y) = \min(|x - y|, 1 - |x - y|).$$

Lebesgue measure on [0, 1] gives a natural measure λ on S^1, also called Lebesgue measure λ.

We can also describe the circle as the set $S^1 = \{z \in \mathbb{C} : |z| = 1\}$, with complex multiplication as the group operation. The two notations are related by $z = e^{2\pi i x}$, which is an isometry if we divide arc length on the multiplicative circle by 2π. We will generally use the additive notation for the circle.

For $\alpha \in \mathbb{R}$, let R_α be the rotation of S^1 by angle $2\pi\alpha$, i.e.,

$$R_\alpha x = x + \alpha \mod 1.$$

The collection $\{R_\alpha : \alpha \in [0, 1)\}$ is a commutative group with composition as group operation, $R_\alpha \circ R_\beta = R_\gamma$, where $\gamma = \alpha + \beta \mod 1$. Note that R_α is an isometry: It preserves the distance d. It also preserves Lebesgue measure λ, i.e., the Lebesgue measure of a set is the same as the Lebesgue measure of its preimage.

If $\alpha = p/q$ is rational, then $R_\alpha^q = \mathrm{Id}$, so every orbit is periodic. On the other hand, if α is irrational, then every positive semiorbit is dense in S^1. Indeed, the pigeon-hole principle implies that, for any $\epsilon > 0$, there are $m, n < 1/\epsilon$ such that $m < n$ and $d(R_\alpha^m, R_\alpha^n) < \epsilon$. Thus R^{n-m} is rotation by an angle less than ϵ, so every positive semiorbit is *ϵ-dense* in S^1 (i.e., comes within distance ϵ of every point in S^1). Since ϵ is arbitrary, every positive semiorbit is dense.

For α irrational, density of every orbit of R_α implies that S^1 is the only R_α-invariant closed non-empty subset. A dynamical system with no proper closed non-empty invariant subsets is called *minimal*. In Chapter 4, we show that any measurable R_α-invariant subset of S^1 has either measure zero or full measure. A measurable dynamical system with this property is called *ergodic*.

Circle rotations are examples of an important class of dynamical systems arising as group translations. Given a group G and an element $h \in G$, define maps $L_h : G \to G$ and $R_h : G \to G$ by

$$L_h g = hg \quad \text{and} \quad R_h g = gh.$$

These maps are called *left* and *right translation* by h. If G is commutative, $L_h = R_h$.

A *topological group* is a topological space G with a group structure such that group multiplication $(g, h) \mapsto gh$, and the inverse $g \mapsto g^{-1}$ are continuous maps. A continuous homomorphism of a topological group to itself is called an *endomorphism*; an invertible endomorphism is an *automorphism*. Many important examples of dynamical systems arise as translations or endomorphisms of topological groups.

Exercise 1.2.1. Show that for any $k \in \mathbb{Z}$, there is a continuous semiconjugacy from R_α to $R_{k\alpha}$.

Exercise 1.2.2. Prove that for any finite sequence of decimal digits there is an integer $n > 0$ such that the decimal representation of 2^n starts with that sequence of digits.

Exercise 1.2.3. Let G be a topological group. Prove that for each $g \in G$, the closure $H(g)$ of the set $\{g^n\}_{n=-\infty}^{\infty}$ is a commutative subgroup of G. Thus, if G has a minimal left translation, then G is abelian.

***Exercise 1.2.4.** Show that R_α and R_β are conjugate by a homeomorphism if and only if $\alpha = \pm\beta \bmod 1$.

1.3 Expanding Endomorphisms of the Circle

For $m \in \mathbb{Z}$, $|m| > 1$, define the *times-m* map $E_m \colon S^1 \to S^1$ by

$$E_m x = mx \bmod 1.$$

This map is a non-invertible group endomorphism of S^1. Every point has m preimages. In contrast to a circle rotation, E_m expands arc length and distances between nearby points by a factor of m: If $d(x, y) \le 1/(2m)$, then $d(E_m x, E_m y) = md(x, y)$. A map (of a metric space) that expands distances between nearby points by a factor of at least $\mu > 1$ is called *expanding*.

The map E_m preserves Lebesgue measure λ on S^1 in the following sense: if $A \subset S^1$ is measurable, then $\lambda(E_m^{-1}(A)) = \lambda(A)$ (Exercise 1.3.1). Note, however, that for a sufficiently small interval I, $\lambda(E_m(I)) = m\lambda(I)$. We will show later that E_m is ergodic (Proposition 4.4.2).

Fix a positive integer $m > 1$. We will now construct a semiconjugacy from another natural dynamical system to E_m. Let $\Sigma = \{0, \dots, m-1\}^{\mathbb{N}}$ be the set of sequences of elements in $\{0, \dots, m-1\}$. The *shift* $\sigma \colon \Sigma \to \Sigma$ discards the first element of a sequence and shifts the remaining elements one place to the left:

$$\sigma((x_1, x_2, x_3, \dots)) = (x_2, x_3, x_4, \dots).$$

A base-m expansion of $x \in [0, 1]$ is a sequence $(x_i)_{i \in \mathbb{N}} \in \Sigma$ such that $x = \Sigma_{i=1}^{\infty} x_i / m^i$. In analogy with decimal notation, we write $x = 0.x_1 x_2 x_3 \dots$.

Base-m expansions are not always unique: A fraction whose denominator is a power of m is represented both by a sequence with trailing $m-1$s and a sequence with trailing zeros. For example, in base 5, we have $0.144\ldots = 0.200\ldots = 2/5$.

Define a map

$$\phi\colon \Sigma \to [0,1], \qquad \phi((x_i)_{i\in\mathbb{N}}) = \sum_{i=1}^{\infty} \frac{x_i}{m^i}.$$

We can consider ϕ as a map into S^1 by identifying 0 and 1. This map is surjective, and one-to-one except on the countable set of sequences with trailing zeros or $m-1$'s. If $x = 0.x_1x_2x_3\ldots \in [0,1)$, then $E_m x = 0.x_2x_3\ldots$. Thus, $\phi \circ \sigma = E_m \circ \phi$, so ϕ is a semiconjugacy from σ to E_m.

We can use the semiconjugacy of E_m with the shift σ to deduce properties of E_m. For example, a sequence $(x_i) \in \Sigma$ is a periodic point for σ with period k if and only if it is a periodic sequence with period k, i.e., $x_{k+i} = x_i$ for all i. It follows that the number of periodic points of σ of period k is m^k. More generally, (x_i) is eventually periodic for σ if and only if the sequence (x_i) is eventually periodic. A point $x \in S^1 = [0,1]/\!\sim$ is periodic for E_m with period k if and only if x has a base-m expansion $x = 0.x_1x_2\ldots$ that is periodic with period k. Therefore, the number of periodic points of E_m of period k is $m^k - 1$ (since 0 and 1 are identified).

Let $\mathcal{F}_m = \bigcup_{k=1}^{\infty}\{0,\ldots,m-1\}^k$ be the set of all finite sequences of elements of the set $\{0,\ldots,m-1\}$. A subset $A \subset [0,1]$ is dense if and only if every finite sequence $w \in \mathcal{F}_m$ occurs at the beginning of the base-m expansion of some element of A. It follows that the set of periodic points is dense in S^1. The orbit of a point $x = 0.x_1x_2\ldots$ is dense in S^1 if and only if every finite sequence from $\mathcal{F}_m$ appears in the sequence (x_i). Since $\mathcal{F}_m$ is countable, we can construct such a point by concatenating all elements of $\mathcal{F}_m$.

Although ϕ is not one-to-one, we can construct a right inverse to ϕ. Consider the partition of $S^1 = [0,1]/\!\sim$ into intervals

$$P_k = [k/m, (k+1)/m), \qquad 0 \le k \le m-1.$$

For $x \in [0,1]$, define $\psi_i(x) = k$ if $E_m^i x \in P_k$. The map $\psi\colon S^1 \to \Sigma$, given by $x \mapsto (\psi_i(x))_{i=0}^{\infty}$, is a right inverse for ϕ, i.e., $\phi \circ \psi = \mathrm{Id}\colon S^1 \to S^1$. In particular, $x \in S^1$ is uniquely determined by the sequence $(\psi_i(x))$.

The use of partitions to code points by sequences is the principal motivation for *symbolic dynamics*, the study of shifts on sequence spaces, which is the subject of the next section and Chapter 3.

Exercise 1.3.1. Prove that $\lambda(E_m^{-1}([a,b])) = \lambda([a,b])$ for any interval $[a,b] \subset [0,1]$.

Exercise 1.3.2. Prove that $E_k \circ E_l = E_l \circ E_k = E_{kl}$. When is $E_k \circ R_\alpha = R_\alpha \circ E_k$?

Exercise 1.3.3. Show that the set of points with dense orbits is uncountable.

Exercise 1.3.4. Prove that the set

$$C = \{x \in [0,1]: E_3^k x \notin (1/3, 2/3)\ \forall\, k \in \mathbb{N}_0\}$$

is the standard middle-thirds Cantor set.

***Exercise 1.3.5.** Show that the set of points with dense orbits under E_m has Lebesgue measure 1.

1.4 Shifts and Subshifts

In this section, we generalize the notion of shift space introduced in the previous section. For an integer $m > 1$ set $\mathcal{A}_m = \{1, \ldots, m\}$. We refer to $\mathcal{A}_m$ as an *alphabet* and its elements as *symbols*. A finite sequence of symbols is called a *word*. Let $\Sigma_m = \mathcal{A}_m^{\mathbb{Z}}$ be the set of infinite two-sided sequences of symbols in $\mathcal{A}_m$, and $\Sigma_m^+ = \mathcal{A}_m^{\mathbb{N}}$ be the set of infinite one-sided sequences. We say that a sequence $x = (x_i)$ contains the word $w = w_1 w_2 \ldots w_k$ (or that w *occurs* in x) if there is some j such that $w_i = x_{j+i}$ for $i = 1, \ldots, k$.

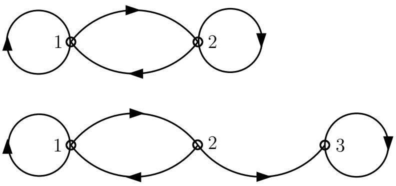
Given a one-sided or two-sided sequence $x = (x_i)$, let $\sigma(x) = (\sigma(x)_i)$ be the sequence obtained by shifting x one step to the left, i.e., $\sigma(x)_i = x_{i+1}$. This defines a self-map of both Σ_m and Σ_m^+ called the *shift*. The pair (Σ_m, σ) is called the *full two-sided shift*; (Σ_m^+, σ) is the *full one-sided shift*. The two-sided shift is invertible. For a one-sided sequence, the leftmost symbol disappears, so the one-sided shift is non-invertible, and every point has m preimages. Both shifts have m^n periodic points of period n.

The shift spaces Σ_m and Σ_m^+ are compact topological spaces in the product topology. This topology has a basis consisting of *cylinders*

$$C^{n_1,\ldots,n_k}_{j_1,\ldots,j_k} = \{x = (x_l): x_{n_i} = j_i,\ i = 1, \ldots, k\},$$

where $n_1 < n_2 < \cdots < n_k$ are indices in $\mathbb{Z}$ or $\mathbb{N}$, and $j_i \in \mathcal{A}_m$. Since the preimage of a cylinder is a cylinder, σ is continuous on Σ_m^+ and is a homeomorphism of Σ_m. The metric

$$d(x, x') = 2^{-l}, \qquad \text{where} \quad l = \min\{|i|: x_i \neq x'_i\}$$



$$\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$$

$$\begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 0 & 1 \end{pmatrix}$$

Figure 1.1. Examples of directed graphs with labeled vertices and the corresponding adjacency matrices.

generates the product topology on Σ_m and Σ_m^+ (Exercise 1.4.3). In Σ_m, the open ball $B(x, 2^{-l})$ is the symmetric cylinder $C^{-l,-l+1,\dots,l}_{x_{-l},x_{-l+1},\dots,x_l}$, and in Σ_m^+, $B(x, 2^{-l}) = C^{1,\dots,l}_{x_1,\dots,x_l}$. The shift is expanding on Σ_m^+; if $d(x, x') < 1/2$, then $d(\sigma(x), \sigma(x')) = 2d(x, x')$.

In the product topology, periodic points are dense, and there are dense orbits (Exercise 1.4.5).

Now we describe a natural class of closed shift-invariant subsets of the full shift spaces. These *subshifts* can be described in terms of *adjacency matrices* or their associated *directed graphs*. An adjacency matrix $A = (a_{ij})$ is an $m \times m$ matrix whose entries are zeros and ones. Associated to A is a directed graph Γ_A with m vertices such that a_{ij} is the number of edges from the ith vertex to the jth vertex. Conversely, if Γ is a finite directed graph with vertices $v_1, \dots, v_m$, then Γ determines an adjacency matrix B, and $\Gamma = \Gamma_B$. Figure 1.1 shows two adjacency matrices and the associated graphs.

Given an $m \times m$ adjacency matrix $A = (a_{ij})$, we say that a word or infinite sequence x (in the alphabet $\mathcal{A}_m$) is *allowed* if $a_{x_i x_{i+1}} > 0$ for every i; equivalently, if there is a directed edge from x_i to x_{i+1} for every i. A word or sequence that is not allowed is said to be *forbidden*. Let $\Sigma_A \subset \Sigma_m$ be the set of allowed two-sided sequences (x_i), and $\Sigma_A^+ \subset \Sigma_m^+$ be the set of allowed one-sided sequences. We can view a sequence $(x_i) \in \Sigma_A$ (or Σ_A^+) as an infinite walk along directed edges in the graph Γ_A, where x_i is the index of the vertex visited at time i. The sets Σ_A and Σ_A^+ are closed shift-invariant subsets of Σ_m and Σ_m^+, and inherit the subspace topology. The pairs (Σ_A, σ) and (Σ_A^+, σ) are called the two-sided and one-sided *vertex shifts* determined by A.

A point $(x_i) \in \Sigma_A$ (or Σ_A^+) is periodic of period n if and only if $x_{i+n} = x_i$ for every i. The number of periodic points of period n (in Σ_A or Σ_A^+) is equal to the trace of A^n (Exercise 1.4.2).

Exercise 1.4.1. Let A be a matrix of zeros and ones. A vertex v_i can be *reached* (in n steps) from a vertex v_j if there is a path (consisting of n edges) from v_i to v_j along directed edges of Γ_A. What properties of A correspond to the following properties of Γ_A?

(a) Any vertex can be reached from some other vertex.
(b) There are no terminal vertices, i.e., there is at least one directed edge starting at each vertex.
(c) Any vertex can be reached in one step from any other vertex .
(d) Any vertex can be reached from any other vertex in exactly n steps.

Exercise 1.4.2. Let A be an $m \times m$ matrix of zeros and ones. Prove that:
(a) the number of fixed points in Σ_A (or Σ_A^+) is the trace of A;
(b) the number of allowed words of length $n+1$ beginning with the symbol i and ending with j is the i, jth entry of A^n; and
(c) the number of periodic points of period n in Σ_A (or Σ_A^+) is the trace of A^n.

Exercise 1.4.3. Verify that the metrics on Σ_m and Σ_m^+ generate the product topology.

Exercise 1.4.4. Show that the semiconjugacy $\phi \colon \Sigma \to [0, 1]$ of §1.3 is continuous with respect to the product topology on Σ.

Exercise 1.4.5. Assume that all entries of some power of A are positive. Show that in the product topology on Σ_A and Σ_A^+, periodic points are dense, and there are dense orbits.

1.5 Quadratic Maps

The expanding maps of the circle introduced in §1.3 are *linear maps* in the sense that they come from linear maps of the real line. The simplest nonlinear dynamical systems in dimension one are the quadratic maps

$$q_\mu(x) = \mu x(1-x), \quad \mu > 0.$$

Figure 1.2 shows the graph of q_3 and successive images $x_i = q_3^i(x_0)$ of a point x_0.

If $\mu > 1$ and $x \notin [0, 1]$, then $q_\mu^n(x) \to -\infty$ as $n \to \infty$. For this reason, we focus our attention on the interval $[0, 1]$. For $\mu \in [0, 4]$, the interval $[0, 1]$ is forward invariant under q_μ. For $\mu > 4$, the interval $(1/2 - \sqrt{1/4 - 1/\mu}$, $1/2 + \sqrt{1/4 - 1/\mu})$ maps outside $[0, 1]$; we show in Chapter 7 that the set of points Λ_μ whose forward orbits stay in $[0, 1]$ is a Cantor set, and (Λ_μ, q_μ) is equivalent to the full one-sided shift on two symbols.

Let X be a locally compact metric space and $f \colon X \to X$ a continuous map. A fixed point p of f is *attracting* if it has a neighborhood U such that $\bar{U}$ is compact, $f(\bar{U}) \subset U$, and $\bigcap_{n\geq 0} f^n(U) = \{p\}$. A fixed point p is *repelling*

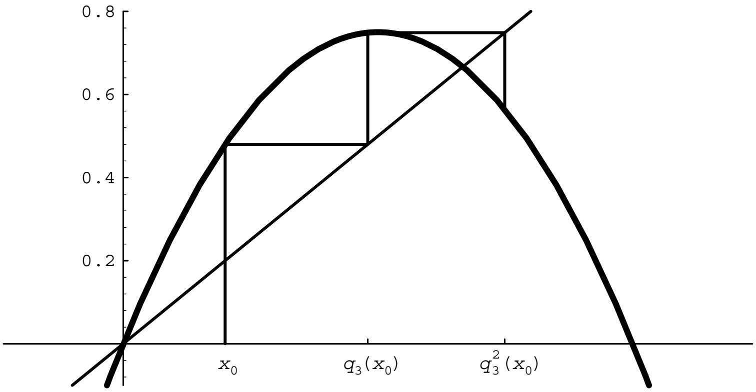


Figure 1.2. Quadratic map of q_3.

if it has a neighborhood U such that $\bar{U} \subset f(U)$, and $\bigcap_{n\geq 0} f^{-n}(U) = \{p\}$. Note that if f is invertible, then p is attracting for f if and only if it is repelling for f^{-1}, and vice versa. A fixed point p is called *isolated* if there is a neighborhood of p that contains no other fixed points.

If x is a periodic point of f of period n, then we say that f is an *attracting* (*repelling*) *periodic point* if x is an attracting (repelling) fixed point of f^n. We also say that the periodic orbit $\mathcal{O}(x)$ is attracting or repelling, respectively.

The fixed points of q_μ are 0 and $1 - 1/\mu$. Note that $q'_\mu(0) = \mu$ and that $q'_\mu(1 - 1/\mu) = 2 - \mu$. Thus, 0 is attracting for $\mu < 1$ and repelling for $\mu > 1$, and $1 - 1/\mu$ is attracting for $\mu \in (1, 3)$ and repelling for $\mu \notin [1, 3]$ (Exercise 1.5.4).

The maps q_μ, $\mu > 4$, have interesting and complicated dynamical behavior. In particular, periodic points abound. For example,

$$q_\mu([1/\mu, 1/2]) \supset [1 - 1/\mu, 1],$$
$$q_\mu([1 - 1/\mu, 1]) \supset [0, 1 - 1/\mu] \supset [1/\mu, 1/2].$$

Hence, $q_\mu^2([1/\mu, 1/2]) \supset [1/\mu, 1/2]$, so the Intermediate Value Theorem implies that q_μ^2 has a fixed point $p_2 \in [1/\mu, 1/2]$. Thus, p_2 and $q_\mu(p_2)$ are non-fixed periodic points of period 2. This approach to showing existence of periodic points applies to many one-dimensional maps. We exploit this technique in Chapter 7 to prove the Sharkovsky Theorem (Theorem 7.3.1), which asserts, for example, that for continuous self-maps of the interval the existence of an orbit of period three implies the existence of periodic orbits of all orders.

Exercise 1.5.1. Show that for any $x \notin [0, 1]$, $q_\mu^n(x) \to -\infty$ as $n \to \infty$.

Exercise 1.5.2. Show that a repelling fixed point is an isolated fixed point.

Exercise 1.5.3. Suppose p is an attracting fixed point for f. Show that there is a neighborhood U of p such that the forward orbit of every point in U converges to p.

Exercise 1.5.4. Let $f: \mathbb{R} \to \mathbb{R}$ be a C^1 map, and p be a fixed point. Show that if $|f'(p)| < 1$, then p is attracting, and if $|f'(p)| > 1$, then p is repelling.

Exercise 1.5.5. Are 0 and $1 - 1/\mu$ attracting or repelling for $\mu = 1$? for $\mu = 3$?

Exercise 1.5.6. Show the existence of a non-fixed periodic point of q_μ of period 3, for $\mu > 4$.

Exercise 1.5.7. Is the period-2 orbit $\{p_2, q_\mu(p_2)\}$ attracting or repelling for $\mu > 4$?

1.6 The Gauss Transformation

Let $[x]$ denote the greatest integer less than or equal to x, for $x \in \mathbb{R}$. The map φ: $[0, 1] \to [0, 1]$ defined by

$$\varphi(x) = \begin{cases} 1/x - [1/x] & \text{if} \quad x \in (0, 1], \\ 0 & \text{if} \quad x = 0 \end{cases}$$

was studied by C. Gauss, and is now called the *Gauss transformation*. Note that φ maps each interval $(1/(n+1), 1/n]$ continuously and monotonically onto $[0, 1)$; it is discontinuous at $1/n$ for all $n \in \mathbb{N}$. Figure 1.3 shows the graph of φ.

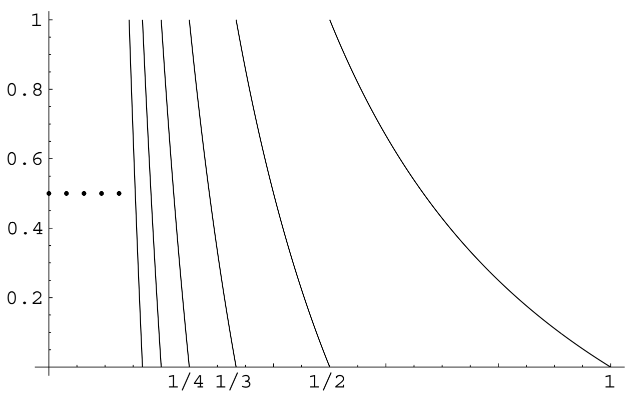


Figure 1.3. Gauss transformation.

Gauss discovered a natural invariant measure μ for φ. The Gauss measure of an interval $A=(a,b)$ is

$$\mu(A)=\frac{1}{\log 2}\int_a^b \frac{dx}{1+x}=(\log 2)^{-1}\log\frac{1+b}{1+a}.$$

This measure is φ-invariant in the sense that $\mu(\varphi^{-1}(A))=\mu(A)$ for any interval $A=(a,b)$. To prove invariance, note that the preimage of (a,b) consists of infinitely many intervals: In the interval $(1/(n+1),1/n)$, the preimage is $(1/(n+b),1/(n+a))$. Thus,

$$\begin{aligned}\mu(\varphi^{-1}((a,b)))&=\mu\left(\bigcup_{n=1}^{\infty}\left(\frac{1}{n+b},\frac{1}{n+a}\right)\right)\\&=\frac{1}{\log 2}\sum_{n=1}^{\infty}\log\left(\frac{n+a+1}{n+a}\cdot\frac{n+b}{n+b+1}\right)=\mu((a,b)).\end{aligned}$$

Note that in general $\mu(\varphi(A))\neq\mu(A)$.

The Gauss transformation is closely related to continued fractions. The expression

$$[a_1,a_2,\ldots,a_n]=\cfrac{1}{a_1+\cfrac{1}{a_2+\cdots\cfrac{1}{a_n}}},\qquad a_1,\ldots,a_n\in\mathbb{N},$$

is called a *finite continued fraction*. For $x\in(0,1]$, we have $x=1/([\frac{1}{x}]+\varphi(x))$. More generally, if $\varphi^{n-1}(x)\neq 0$, set $a_i=[1/\varphi^{i-1}(x)]\geq 1$ for $i\leq n$. Then,

$$x=\cfrac{1}{a_1+\cfrac{1}{a_2+\cfrac{1}{\cdots+\cfrac{1}{a_n+\varphi^n(x)}}}}$$

Note that x is rational if and only if $\varphi^m(x)=0$ for some $m\in\mathbb{N}$ (Exercise 1.6.2). Thus any rational number is uniquely represented by a finite continued fraction.

For an irrational number $x\in(0,1)$, the sequence of finite continued fractions

$$[a_1,a_2,\ldots,a_n]=\cfrac{1}{a_1+\cfrac{1}{a_2+\cfrac{1}{\cdots+\cfrac{1}{a_n}}}}$$

converges to x (where $a_i = [1/\varphi^{i-1}(x)]$) (Exercise 1.6.4). This is expressed concisely with the infinite continued fraction notation

$$x = [a_1, a_2, \ldots] = \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cdots}}.$$

Conversely, given a sequence $(b_i)_{i\in\mathbb{N}}$, $b_i \in \mathbb{N}$, the sequence $[b_1, b_2, \ldots, b_n]$ converges, as $n \to \infty$, to a number $y \in [0, 1]$, and the representation $y = [b_1, b_2, \ldots]$ is unique (Exercise 1.6.4). Hence $\varphi(y) = [b_2, b_3, \ldots]$, because $b_n = [1/\varphi^{n-1}(y)]$.

We summarize this discussion by saying that the continued fraction representation conjugates the Gauss transformation and the shift on the space of finite or infinite integer-valued sequences $(b_i)_{i=1}^{\omega}$, $\omega \in \mathbb{N} \cup \{\infty\}$, $b_i \in \mathbb{N}$. (By convention, the shift of a finite sequence is obtained by deleting the first term; the empty sequence represents 0.) As an immediate consequence, we obtain a description of the eventually periodic points of φ (see Exercise 1.6.3).

Exercise 1.6.1. What are the fixed points of the Gauss transformation?

Exercise 1.6.2. Show that $x \in [0, 1]$ is rational if and only if $\varphi^m(x) = 0$ for some $m \in \mathbb{N}$.

Exercise 1.6.3. Show that:

(a) a number with periodic continued fraction expansion satisfies a quadratic equation with integer coefficients; and

(b) a number with eventually periodic continued fraction expansion satisfies a quadratic equation with integer coefficients.

The converse of the second statement is also true, but is more difficult to prove [Arc70], [HW79].

***Exercise 1.6.4.** Show that given any infinite sequence $b_k \in \mathbb{N}$, $k = 1, 2, \ldots$, the sequence $[b_1, \ldots, b_n]$ of finite continued fractions converges. Show that for any $x \in \mathbb{R}$, the continued fraction $[a_1, a_2, \ldots]$, $a_i = [1/\phi^{i-1}(x)]$, converges to x, and that this continued fraction representation is unique.

1.7 Hyperbolic Toral Automorphisms

Consider the linear map of $\mathbb{R}^2$ given by the matrix

$$A = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}.$$

The eigenvalues are $\lambda = (3 + \sqrt{5})/2 > 1$ and $1/\lambda$. The map expands by a factor of λ in the direction of the eigenvector $v_\lambda = ((1 + \sqrt{5})/2, 1)$, and contracts

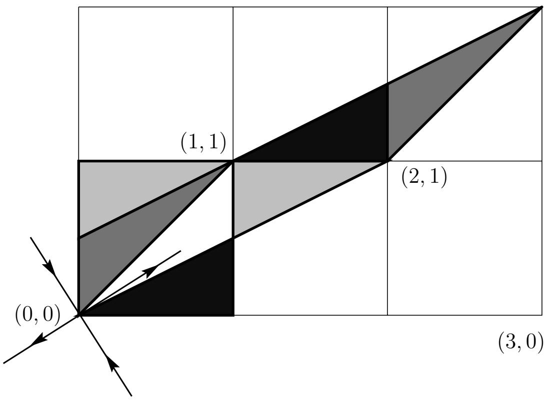


Figure 1.4. The image of the torus under A.

by $1/\lambda$ in the direction of $v_{1/\lambda} = ((1-\sqrt{5})/2, 1)$. The eigenvectors are perpendicular because A is symmetric.

Since A has integer entries, it preserves the integer lattice $\mathbb{Z}^2 \subset \mathbb{R}^2$ and induces a map (which we also call A) of the *torus* $\mathbb{T}^2 = \mathbb{R}^2/\mathbb{Z}^2$. The torus can be viewed as the unit square $[0, 1] \times [0, 1]$ with opposite sides identified: $(x_1, 0) \sim (x_1, 1)$ and $(0, x_2) \sim (1, x_2)$, $x_1, x_2 \in [0, 1]$. The map A is given in coordinates by

$$A\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} (2x_1 + x_2) \bmod 1 \\ (x_1 + x_2) \bmod 1 \end{pmatrix}$$

(see Figure 1.4). Note that $\mathbb{T}^2$ is a commutative group and A is an automorphism, since A^{-1} is also an integer matrix.

The periodic points of $A\colon \mathbb{T}^2 \to \mathbb{T}^2$ are the points with rational coordinates (Exercise 1.7.1).

The lines in $\mathbb{R}^2$ parallel to the eigenvector v_λ project to a family W^u of parallel lines on $\mathbb{T}^2$. For $x \in \mathbb{T}^2$, the line $W^u(x)$ through x is called the *unstable manifold* of x. The family W^u partitions $\mathbb{T}^2$ and is called the *unstable foliation* of A. This foliation is invariant in the sense that $A(W^u(x)) = W^u(Ax)$. Moreover, A expands each line in W^u by a factor of λ. Similarly, the *stable foliation* W^s is obtained by projecting the family of lines in $\mathbb{R}^2$ parallel to $v_{1/\lambda}$. This foliation is also invariant under A, and A contracts each *stable manifold* $W^s(x)$ by $1/\lambda$. Since the slopes of v_λ and $v_{1/\lambda}$ are irrational, each of the stable and unstable manifolds is dense in $\mathbb{T}^2$ (Exercise 1.11.1).

In a similar way, any $n \times n$ integer matrix B induces a group endomorphism of the n-torus $\mathbb{T}^n = \mathbb{R}^n/\mathbb{Z}^n = [0, 1]^n/\sim$. The map is invertible (an automorphism) if and only if B^{-1} is an integer matrix, which happens if and only if $|\det B| = 1$ (Exercise 1.7.2). If B is invertible and the eigenvalues do not lie on the unit circle, then $B: \mathbb{T}^n \to \mathbb{T}^n$ has expanding and contracting subspaces of complementary dimensions and is called a *hyperbolic toral automorphism*. The stable and unstable manifolds of a hyperbolic toral automorphism are dense in $\mathbb{T}^n$ (§5.10). This is easy to show in dimension two (Exercise 1.7.3 and Exercise 1.11.1).

Hyperbolic toral automorphisms are prototypes of the more general class of *hyperbolic dynamical systems*. These systems have uniform expansion and contraction in complementary directions at every point. We discuss them in detail in Chapter 5.

Exercise 1.7.1. Consider the automorphism of $\mathbb{T}^2$ corresponding to a nonsingular 2×2 integer matrix whose eigenvalues are not roots of 1.

(a) Prove that every point with rational coordinates is eventually periodic.

(b) Prove that every eventually periodic point has rational coordinates.

Exercise 1.7.2. Prove that the inverse of an $n \times n$ integer matrix B is also an integer matrix if and only if $|\det B| = 1$.

Exercise 1.7.3. Show that the eigenvalues of a two-dimensional hyperbolic toral automorphism are irrational (so the stable and unstable manifolds are dense by Exercise 1.11.1).

Exercise 1.7.4. Show that the number of fixed points of a hyperbolic toral automorphism A is $\det(A - I)$ (hence the number of periodic points of period n is $\det(A^n - I)$).

1.8 The Horseshoe

Consider a region $D \subset \mathbb{R}^2$ consisting of two semicircular regions D_1 and D_5 together with a unit square $R = D_2 \cup D_3 \cup D_4$ (see Figure 1.5).

Let $f: D \to D$ be a differentiable map that stretches and bends D into a horseshoe as shown in Figure 1.5. Assume also that f stretches $D_2 \cup D_4$ uniformly in the horizontal direction by a factor of $\mu > 2$ and contracts

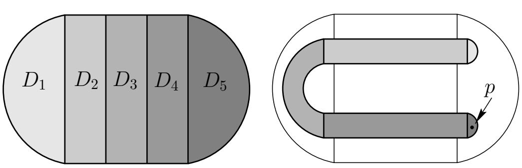


Figure 1.5. The horseshoe map.