Chapter One

One-Dimensional Dynamics

The goal of this first chapter is to introduce many of the basic techniques from the theory of dynamical systems in a setting that is as simple as possible. Accordingly, all of the dynamical systems that we will encounter take place in one dimension, either on the real line or on the unit circle in the plane. For that reason, much of this chapter can be read with only a solid background in calculus.

We regard the first twelve sections of this chapter as central to the theory of dynamical systems. Here we introduce such topics as hyperbolicity, symbolic dynamics, topological conjugacy, structural stability, and chaos. These form the essential background for all that follows. Indeed, the last two chapters of this text may be regarded as extensions and refinements of the material presented in these introductory sections.

Our main thrust in this chapter is to understand what it means for a dynamical system to be chaotic. We feel that this is best understood in light of examples. Hence most of our initial effort revolves around a single family of examples, the family of quadratic maps $F_\mu(x) = \mu x(1-x)$. Later, using such tools as Sarkovskii's Theorem and the Schwarzian derivative, we will show that the seemingly specialized results for the quadratic map actually hold for a large collection of dynamical systems.

The next four sections in this chapter present material which is somewhat more technical than the preceding material. The concepts introduced – subshifts of finite type, Morse-Smale maps, the rotation number, and homoclinic bifurcation – are important in the sequel, however. The final three sections on the kneading theory and the period-doubling route to chaos should be regarded as special topics. They will not be used in what follows. However, for the reader interested in recent work on the transition to chaotic dynamics, these sections should provide an introduction to many of the topics in the current literature.

§1.1 EXAMPLES OF DYNAMICAL SYSTEMS

This brief section is intended merely as motivation for the succeeding sections. Our aim is to give a couple of simple examples of dynamical systems. These examples show how dynamical systems occur in the "real world" and how some very simple phenomena from nature yield rather complicated dynamical systems.

First, what is a dynamical system? The answer is quite simple: take a scientific calculator and input any number whatsoever. Then start striking one of the function keys over and over again. This iterative procedure is an example of a *discrete dynamical system*. For example, if we repeatedly strike the "exp" key, given an initial input x, we are computing the sequence of numbers

$$x,\ e^x\ e^{e^x},\ e^{e^{e^x}},\dots .$$

That is, we are iterating the exponential function. If this experiment is performed over and over again, it becomes apparent that any choice of initial x leads rather quickly to an "overflow" message from the calculator: that is, successive iterations of $\exp(x)$ tend to ∞. This is, in fact, the main question we will ask in the sequel: given a function f and an initial value x_0, what ultimately happens to the sequence of iterates

$$x_0,\ f(x_0),\ f(f(x_0)),\ f(f(f(x_0))),\dots .$$

As another example, consider $\sin x$. A few keystrokes on the calculator will be enough to convince the reader that any initial x_0 leads to a sequence of iterates which tends to 0. Similarly, for $\cos x$, any x_0 yields a sequence which converges fairly rapidly to .73908... (in radians, or to .99984... in degrees). The reader may begin to suspect that iteration of a given function on a given initial value always yields a sequence that converges to a fixed limit (maybe x_0, maybe 0, in any case a unique limit). Actually, nothing could be further from the truth. Very simple functions, even the simplest quadratic functions on the real line, lead to bizarre and unpredictable results when iterated. For example, program a computer or calculator to iterate the simple function $f(x) = 4x(1 - x)$. Input a random number between 0 and 1 and watch the results of the iteration. One gets dramatically different behaviors depending upon which initial x is input. Sometimes the values repeat; other times they do not. Most often they wander aimlessly about the unit interval with no discernible pattern. Now change the parameter from 4 to 3.839, i.e., iterate the function $f(x) = 3.839x(1-x)$. For a random entry between 0 and 1, one observes that the iterates of this point eventually settle down to a repeating cycle of three numbers, .149888..., .489172..., and .959299..., repeated over and over again in succession. Two comments are in order. The first example illustrates the phenomenon of chaos or unpredictability that forms one of the major themes of this book. Despite its complexity, we will see how to analyze this unpredictability completely. Second, chaos occurs in many, many dynamical systems. The second example above, which seems comparatively rather tame, also admits a set of initial x values which behave just as unpredictably as in the first example. However, due to roundoff or "experimental" error, we do not see this randomness at first glance. Nevertheless, as we shall see, it lurks in the background and has an increasingly important effect on the system as the accuracy of the computations is increased.

At this juncture, we should note that there are many other types of dynamical systems besides iterated functions. For example, differential equations are examples of *continuous*, as opposed to discrete, dynamical systems. In this book, we will not deal with these types of systems at all. These types of systems are much easier to understand once the basic behavior of discrete systems has been mastered.

Let us now consider several "applied" examples. Dynamical systems occur in all branches of science, from the differential equations of classical mechanics in physics to the difference equations of mathematical economics and biology. We will first describe a simple model from population biology which will serve as motivation for all of the succeeding chapter.

Population biologists are interested in the long-term behavior of the population of a certain species or collection of species. Given certain observed or experimentally determined parameters (number of predators, severity of climate, availability of food, etc.), the biologist sets up a mathematical model to describe the fluctuations in the population. This may take the form of a differential equation or a difference equation, depending upon whether the population is assumed to change continuously or discretely, such as when the population is measured once a year or once a generation. In either case, the population biologist is interested in what happens to an initial population of P_0 members. Does the population tend to zero as time goes on, leading to extinction of the species? Does the population become arbitrarily large, indicating eventual overcrowding? Or does the population fluctuate periodically or even randomly? Thus the problem facing the population biologist is a typical dynamical systems question: given P_0, can one predict the long-term behavior of the population?

Several simple biological models are encountered in elementary calculus courses. For example, the differential equation of exponential growth or decay is often the first differential equation a student is exposed to. In this model, we assume that the population of a single species changes at a rate that is directly proportional to the population present at the given time. This is, of course, an extremely naive model, which does not take into account obvious factors such as overcrowding, the death rate, etc. However, this model does produce an especially simple differential equation which is readily solved. If $P(t)$ denotes the population at time t, the assumptions above may be translated into

$$\frac{dP}{dt} = kP.$$

The solution to this equation is $P(t) = P_0 e^{kt}$ where $P_0 = P(0)$ is the initial population of the species. Hence, if the constant of proportionality is positive, $P(t) \to \infty$ as $t \to \infty$ leading to population explosion. If $k < 0$, then $P(t) \to 0$ as $t \to \infty$, leading to extinction.

This procedure illustrates (in an exceedingly simple situation) the typical application of dynamical systems in science. A population biologist sets up a mathematical model for which the mathematician is asked to provide some idea about the long-term behavior of the solutions.

This simple model can also be studied as a difference equation. Let us write P_n = population after n generations, where n is a natural number. The simplest growth law one can imagine is that the population in the next generation is directly proportional to that in the present generation. That is

$$P_{n+1} = k\,P_n$$

where again k is a constant.

We have

$$
\begin{aligned}
P_1 &= kP_0 \\
P_2 &= kP_1 = k^2 P_0 \\
P_3 &= kP_2 = k^3 P_0 \\
&\vdots \\
P_n &= kP_{n-1} = k^n P_0
\end{aligned}
$$

so that the ultimate fate of the population is again easy to decide. If $k > 1$, $P_n \to \infty$, whereas if $0 < k < 1$, then $P_n \to 0$.

For later use, let us recast this difference equation as a function. Let $x = P_0$ and set $f(x) = kx$. Note that, in the above terms, $f(x) = P_1$, $f(f(x)) = k^2 x = P_2$, $f(f(f(x))) = P_3$, etc. Hence the ultimate behavior of the population is intimately related to the asymptotic behavior of the iteration of the function f.

In either of the above models, one has a rather idealized situation. There are essentially only two possibilities: unchecked growth or extinction. Experience tells the population biologist that more complicated patterns arise in nature. So the biologist tries to incorporate additional constraints or parameters in the model, hoping for a better reflection of reality. One such approach again often encountered in calculus is to assume that there is some limiting value L for the population. If $P(t)$ exceeds L, the population should tend to decrease (there is overcrowding, not enough food, etc.) On the other hand, if $P(t) < L$, there is room for more of the species so $P(t)$ should increase. The simplest biological model leading to this behavior is

$$\frac{dP}{dt} = kP(L - P)$$

Note that we have simply tacked on the factor $L - P$ to the previous model.

Let us assume that $k > 0$, the case that previously led to unlimited growth. Here we note that

1. if $P = L$, $\dfrac{dP}{dt} = 0$
2. if $P > L$, $\dfrac{dP}{dt} < 0$
3. if $P < L$, $\dfrac{dP}{dt} > 0$

Thus elementary calculus shows that this model behaves according to our expectations. The population remains constant, decreases, or increases depending upon whether $P = L, P > L$, or $P < L$. In fact, one can explicitly solve the above differential equation via separation of variables and integration by partial fractions. One finds that

$$P(t) = \frac{LP_0 e^{Lkt}}{L - P_0 + P_0 e^{Lkt}}.$$

Using this formula, one may easily sketch the solutions of this system.

While this model conforms more to reality than the exponential growth model, nevertheless we see no cyclic behavior or other fluctuations in the population. One might naively expect that the corresponding difference equation behaves similarly. However, we are in for a great surprise: the analogous difference equation leads to one of the most complicated dynamical systems imaginable. To this day, the dynamics of this system are not completely understood. Moreover, this system exhibits many of the pathologies of higher dimensional systems and for this reason may be considered as one of the most basic nonlinear dynamical systems. We will return to it throughout the chapter as it provides a rich source of illustrative examples.

Let us make a simplification in our model. Let us assume that $L = 1$ is the limiting value. Obviously, we are not now talking about populations but rather percentage of population. P_n represents the percentage of the limiting population present in generation n. The population is then assumed to satisfy the following difference equation

$$P_{n+1} = kP_n(1 - P_n),$$

where again k is a positive constant. As before, we may write $x = P_0$ and $f(x) = kx(1-x)$. This, of course, is the quadratic function mentioned above. We have

$$\begin{aligned} P_1 &= f(x) \\ P_2 &= f(f(x)) \\ P_3 &= f(f(f(x))), \end{aligned}$$

and so on. Thus to determine the fate of a population for a given constant k, we must determine the asymptotic behavior of the function $kx(1 - x)$. This function, known as the logistic function, and its dynamics have been the subject of much contemporary mathematical research. In the following chapters, we will only begin to describe the complications and pathologies that arise in this simple system.

Another example of a dynamical system which arises in practical applications is Newton's method for finding the roots of a polynomial. Let

$$Q(x) = a_n x^n + a_{n-1} x^{n-1} + \ldots + a_0$$

be a polynomial. In general, it is impossible to factor Q if the degree of Q is high. Nevertheless, it is often important in applications to find a root of Q. One such procedure for doing this is the classical recursion scheme of Newton. Let x_0 be a real number. Consider the recursion

$$\begin{aligned} x_1 &= x_0 - \frac{Q(x_0)}{Q'(x_0)} \\ x_2 &= x_1 - \frac{Q(x_1)}{Q'(x_1)} \\ &\vdots \\ x_n &= x_{n-1} - \frac{Q(x_{n-1})}{Q'(x_{n-1})}. \end{aligned}$$

For most choices of the initial value x_0, it is well known from calculus that the sequence of values $x_0, x_1, x_2, \ldots$ converges to one of the roots of Q.

Given the polynomial Q, we thus see that Newton's method determines a dynamical system. Let

$$N(x) = x - \frac{Q(x)}{Q'(x)}.$$

As long as $Q'(x) \neq 0$, this function is well defined. As in our population model, Newton's method reduces to the iteration of N. Again we ask the same question: given x, what happens as we compute successively higher iterates of N at x ?

We remark that Newton's method does not always converge. For certain initial values x_0, the iterative scheme does not yield convergence to a root of Q. The structure of the set where N fails to converge is extremely interesting (especially in the complex plane) and leads to unpredictable behavior similar to the logistic function. We will take up this topic in chapter three.

§1.2 PRELIMINARIES FROM CALCULUS

In this section, we recall some elementary (and not-so-elementary) notions from single variable and multivariable calculus. In the sequel, we will also need a few notions from point-set topology, so we include them here as well. First, we fix some notation. $\mathbf{R}$ denotes the real numbers. I or J will always denote closed intervals in $\mathbf{R}$, i.e., all points x satisfying $a \le x \le b$ for some a and b. $\mathbf{R}^2$ denotes the Cartesian plane.

Let $f\colon \mathbf{R} \to \mathbf{R}$ be a function. We denote the derivative of f at x by $f'(x)$, the second derivative by $f''(x)$, and higher derivatives by $f^{(r)}(x)$. We say that f is of class C^r on I if $f^{(r)}(x)$ exists and is continuous at all $x \in I$. A function is said to be *smooth* if it is of class C^1. The function $f(x)$ is C^∞ if *all* derivatives exist and are continuous. Throughout this book, function means C^∞ function; occasionally we will use functions which are continuous but non-differentiable as examples, but in general, when we say function, we mean C^∞ function.

There are other classes of functions which are commonly studied in calculus. For example, analytic functions (i.e., those with convergent power series representations) are often encountered. For our purposes in this chapter, these types of functions are too rigid in the following sense. We wish to allow small changes in or perturbations of the functions which will change the function in a certain interval but not everywhere. This is accomplished by the use of *bump functions* which we will introduce in the Exercises. These small changes are impossible if we are restricted to analytic functions, for a small change in any of the coefficients of the power series affects the behavior of the function *everywhere*. Later, in chapter three, when we discuss complex analytic dynamical systems, we will restrict our attention solely to these types of functions.

There are some special classes of functions that often arise. The function $f(x)$ is linear if $f(x) = ax$ for some constant a; $f(x)$ is affine if $f(x) = ax + b$; $f(x)$ is piecewise linear if $f(x)$ is affine on a collection of intervals. For example, $f(x) = |x|$ is piecewise linear, the "pieces" being the positive and negative reals on each of which $f(x)$ is linear.

Definition 2.1. $f(x)$ is one-to-one if $f(x) \neq f(y)$ whenever $x \neq y$.

Clearly, increasing or decreasing functions are the only types of continuous one-to-one functions of a real variable. If $f: I \to J$ is one-to-one, then we may define the inverse of f, written $f^{-1}(x)$, by the rule $f^{-1}(x) = y$ if and only if $f(y) = x$. For example, if $f(x) = x^3$, then $f^{-1}(x) = \sqrt[3]{x}$ and if $g(x) = \tan x$, then $g^{-1}(x) = \arctan x$. Here $g: (-\pi/2, \pi/2) \to \mathbf{R}$ so $g^{-1}: \mathbf{R} \to (-\pi/2, \pi/2)$.

Definition 2.2. Let I and J be intervals and $f: I \to J$. The function f is *onto* if for any y in J there is an $x \in I$ such that $f(x) = y$. See Fig. 2.1.

Definition 2.3. Let $f: I \to J$. The function $f(x)$ is a *homeomorphism* if $f(x)$ is one-to-one, onto, and continuous, and $f^{-1}(x)$ is also continuous.

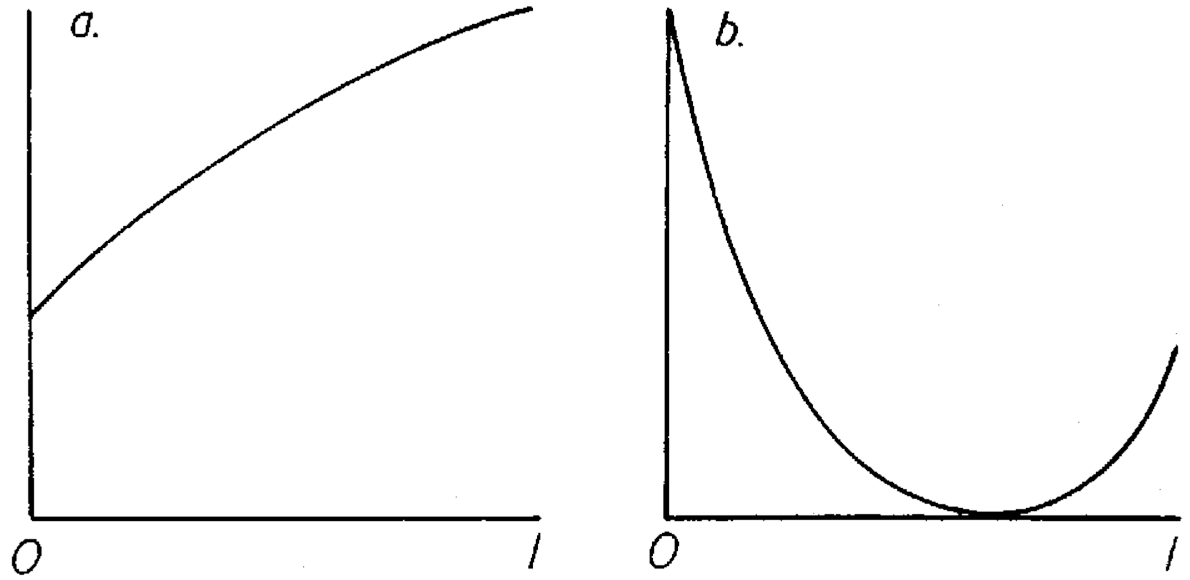

Fig. 2.1. In a. $f(x)$ is one-to-one on the interval $[0, 1]$; in b. $f(x)$ is onto the interval $[0, 1]$.

For example, $\tan x$ is a homeomorphism between $(-\pi/2, \pi/2)$ and $\mathbf{R}$. Thus we say the open interval $(-\pi/2, \pi/2)$ is *homeomorphic* to $\mathbf{R}$. Functions which are one-to-one are also said to be *injective*, while functions which are onto are also called *surjective*.

Definition 2.4. Let $f: I \to J$. The function $f(x)$ is a C^r-*diffeomorphism* if $f(x)$ is a C^r-homeomorphism such that $f^{-1}(x)$ is also C^r.

For example, it is easy to see that $\tan x$ is a C^∞ diffeomorphism from $(-\pi/2, \pi/2)$ to $\mathbf{R}$, whereas $f(x) = x^3$ is a homeomorphism which is *not* a diffeomorphism since $f^{-1}(x) = x^{1/3}$ and $(f^{-1})'(0)$ does not exist.

We will see in subsequent chapters that diffeomorphisms on the real line are extremely simple, dynamically speaking. Therefore, in this chapter, we will primarily consider non-invertible functions. In higher dimensions, diffeomorphisms become much more interesting and therefore become the focal point for dynamical systems theory.

We denote the *composition* of two functions by $f \circ g(x) = f(g(x))$. The n-fold composition of f with itself recurs over and over again in the sequel. We denote this function by $f^n(x) = \underbrace{f \circ \ldots \circ f}_{n\,times}(x)$. Note that f^n does not mean $f(x)$ raised to the n^{th} power, a function which we will *never* use, nor does it mean the n^{th} derivative of $f(x)$, which we denote by $f^{(n)}(x)$. If $f^{-1}(x)$ exists, we write $f^{-n}(x) = f^{-1} \circ \ldots \circ f^{-1}(x)$.

Perhaps the most important feature from elementary calculus that we will use is the Chain Rule:

Proposition 2.5. *If f and g are functions, then*

$$(f \circ g)'(x) = f'(g(x))g'(x).$$

In particular, if $h(x) = f^n(x)$, then

$$h'(x) = f'(f^{n-1}(x)) \cdot f'(f^{n-2}(x)) \cdot \ldots \cdot f'(x).$$

Another important notion from elementary calculus is the Mean Value Theorem:

Theorem 2.6. *Suppose $f\colon [a,b] \to \mathbf{R}$ is C^1. Then there exists $c \in [a,b]$ such that*

$$f(b) - f(a) = f'(c)(b-a).$$

Fig. 2.2 illustrates the content of the Mean Value Theorem. The third important result from calculus is the Intermediate Value Theorem:

Theorem 2.7. *Suppose $f\colon [a,b] \to \mathbf{R}$ is continuous. Suppose that $f(a) = u$ and $f(b) = v$. Then for any z between u and v, there exists c, $a \le c \le b$, such that $f(c) = z$.*

One of the most abstract and seemingly useless theorems from multivariable calculus is the Implicit Function Theorem. Most beginning students

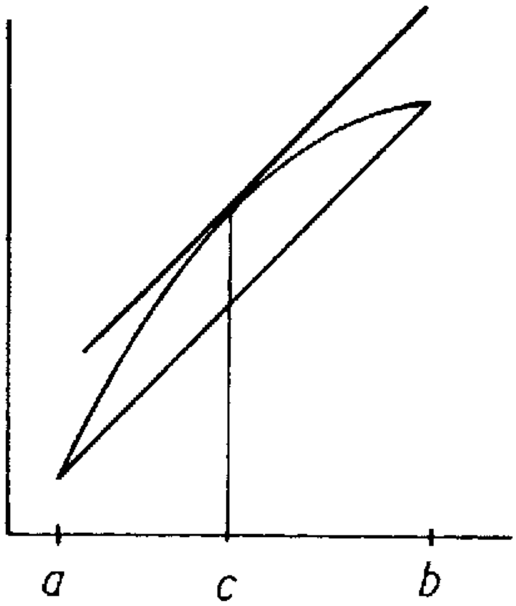

Fig. 2.2. The Mean Value Theorem.

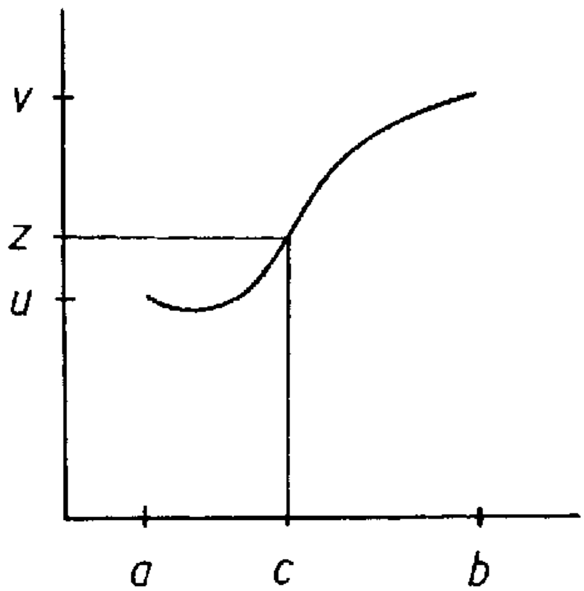

Fig. 2.3. The Intermediate Value Theorem.

have no appreciation of the power of this Theorem when they encounter it in their first analysis course. We hope that the geometric results in bifurcation theory that we will encounter later will help dispel any misconceptions about the usefulness of this theorem.

Theorem 2.8. *Suppose* $G: \mathbf{R}^2 \to \mathbf{R}^1$ *is a* C^1*-function (i.e., both partial derivatives of* G *exist and are continuous.) Suppose further that*

1. $G(x_0, y_0) = 0$
2. $\dfrac{\partial G}{\partial y}(x_0, y_0) \neq 0.$

Then there exist open intervals I about x_0 and J about y_0 and a C^1-function $p: I \to J$ satisfying

1. $p(x_0) = y_0$
2. $G(x, p(x)) = 0$ *for all* $x \in I$.

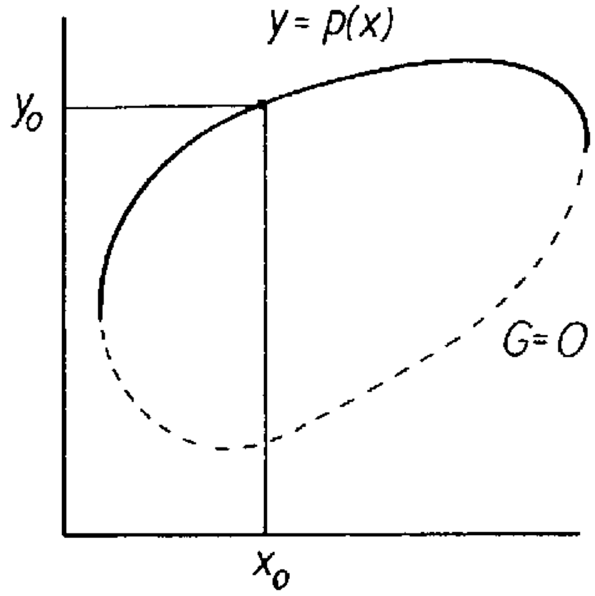

Fig. 2.4. The Implicit Function Theorem.

Rather than prove the Implicit Function Theorem, we give several examples of how to apply it. While these examples are obviously concocted, they nevertheless are typical, as we shall see later.

Example 2.9. Let $G(x, y) = x^2 + y^2 - 1$. The level sets of G are clearly circles, and $G = 0$ defines the unit circle in the plane.

Suppose $G(x_0, y_0) = 0$ and $y_0 > 0$, i.e., (x_0, y_0) is a point on the upper semicircle. Clearly,

$$\frac{\partial G}{\partial y}(x_0, y_0) = 2y_0 \neq 0$$

so the Implicit Function Theorem applies. The result is a function $p(x)$ which satisfies $G(x, p(x)) = 0$ for all x sufficiently close to x_0. What is $p(x)$? In this case, we can construct $p(x)$ explicitly. Clearly, $p(x) = \sqrt{1 - x^2}$, which is C^∞ as long as $x \neq \pm 1$ (when $y = 0$). We have $G(x, \sqrt{1 - x^2}) = 0$ for $|x| < 1$, as the Implicit Function Theorem guarantees. If $y_0 < 0$, then we must choose $p(x) = -\sqrt{1 - x^2}$.

It is important to realize that, in practice, one cannot very often solve for the function $p(x)$ as we did here. Nevertheless, the Implicit Function

Theorem guarantees its existence (whether or not we can explicitly write it down), and that is often exactly what we need.

Example 2.10. $G(x,y) = x^5y^4 - xy^5 - yx^2 + 1$ satisfies $G(1,1) = 0$ and

$$\frac{\partial G}{\partial y}(1,1) = -2.$$

Hence there is a function $p(x)$ defined in some interval about $x = 1$ and which satisfies $G(x, p(x)) = 0$. Solving $G(x,y) = 0$ for $y = p(x)$ is impossible, however.

Fixed points for functions are points x which satisfy $f(x) = x$. These points will play a dominant role in the theory of dynamical systems. The following easy application of the Intermediate Value Theorem gives an important criterion for the existence of a fixed point. See Fig. 2.5.

Proposition 2.11. *Let $I = [a,b]$ be an interval and let $f\colon I \to I$ be continuous. Then f has at least one fixed point in I.*

Proof. Let $g(x) = f(x) - x$. Clearly, $g(x)$ is continuous on I. Suppose $f(a) > a$ and $f(b) < b$ (otherwise, one of a or b is fixed). We thus have $g(a) > 0$ and $g(b) < 0$, so the Intermediate Value Theorem gives the existence of c between a and b for which $g(c) = 0$. Therefore, $f(c) = c$ and we are done. q.e.d.

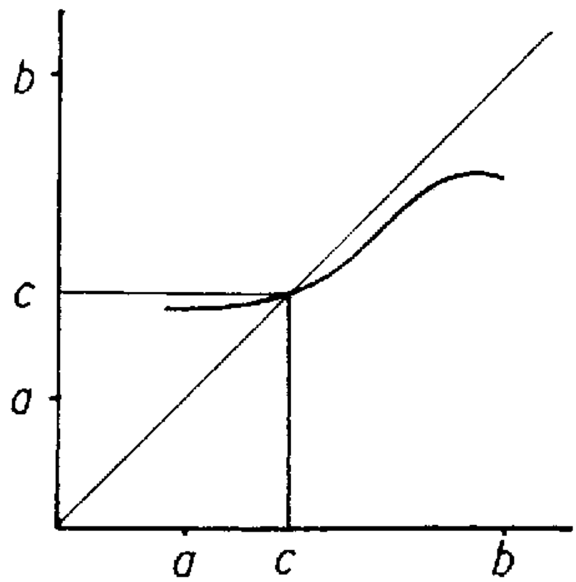

Fig. 2.5. $f\colon I \to I$ has at least one fixed point.

This theorem is a special case of a much more general theorem called the Brouwer Fixed Point Theorem, which gives a similar sufficient condition for the existence of fixed points in higher dimensions. One can actually do better with a little differentiability. The following result is a special case of the Contraction Mapping Theorem.

Proposition 2.12. *Let $f: I \to I$ and assume that $|f'(x)| < 1$ for all x in I. Then there exists a unique fixed point for f in I. Moreover*

$$|f(x) - f(y)| < |x - y|$$

for all $x, y \in I, x \neq y$.

Proof. Proposition 2.11 guarantees at least one fixed point for f, so we suppose that both x and y are fixed points, $x \neq y$. By the Mean Value Theorem, there is a c between x and y such that

$$f'(c) = \frac{f(y) - f(x)}{y - x} = 1.$$

But this contradicts our assumption that $|f'(c)| < 1$ for all c in I. Hence $x = y$.

To establish the second assertion of the Proposition, we again use the Mean Value Theorem to assert that for any $x, y \in I, x \neq y$

$$|f(y) - f(x)| = |f'(c)||y - x| < |y - x|$$

as required.

q.e.d.

We close this section with a few notions from general topology. In general, these notions are beyond the scope of elementary calculus courses. However, many of them occur in the simplest possible setting on the real line, and this is precisely the setting in which we will work.

Definition 2.13. Let $S \subset \mathbf{R}$. A point $x \in \mathbf{R}$ is a limit point of S if there is a sequence of distinct points $x_n \in S$ converging to x. S is a closed set if it contains all of its limit points.

Clearly, closed intervals of the form $a \leq x \leq b$ are closed sets. Any finite union of closed sets is also closed. Infinite unions of closed sets, however, need not be closed, as the following example shows.

Example 2.14. Let $I_n = [\frac{1}{n}, 1]$. Then

$$\bigcup_{n=1}^{\infty} I_n = (0, 1]$$

which is not closed, since 0 is a limit point of S which is not in S.

Intersections of closed sets yield closed sets, however (the empty set is, by definition, a closed set.) Moreover, if I_n is a closed, non-empty, and bounded interval for each n and $I_{n+1} \subset I_n$, then $\cap_{n=1}^{\infty} I_n$ is a closed, *non-empty* set. The crucial word here is, of course, non-empty.

Definition 2.15. Let $S \subset \mathbf{R}$. S is an open set if, for any $x \in S$, there is an $\epsilon > 0$ such that all points t in the open interval $x - \epsilon < t < x + \epsilon$ are contained in S.

It is clear that the complement of a closed set is open and vice versa. Unlike closed sets, infinite unions of open intervals are open sets in $\mathbf{R}$. However, infinite intersections of open intervals are not open sets. For example, if $J_n = (-\frac{1}{n}, \frac{1}{n})$, then $\cap_{n=1}^{\infty} J_n = \{0\}$ which is closed.

For any set S, we denote the closure of S by $\overline{S}$. $\overline{S}$ consists of all points in S together with all limit points of S. For example, if S is the open interval $(0, 1)$, then $\overline{S}$ is the closed interval $[0, 1]$. Clearly, if S is closed, then $\overline{S} = S$.

Definition 2.16. A subset U of S is *dense* in S if $\overline{U} = S$.

For example, any open set S is dense in its closure $\overline{S}$. A more interesting example is the set of rational numbers Q, which is dense in $\mathbf{R}$. Similarly, the irrationals are dense in $\mathbf{R}$. We caution the reader against thinking that dense subsets are necessarily large. Even open and dense sets may be quite small in the sense of total length. Here is an example in the unit interval I given by $0 \le x \le 1$. Since the rationals form a countable set in I, we may list them in some order. One such ordering is

$$0, 1, \frac{1}{2}, \frac{1}{3}, \frac{2}{3}, \frac{1}{4}, \frac{3}{4}, \frac{1}{5}, \frac{2}{5}, \frac{3}{5}, \frac{4}{5}, \frac{1}{6}, \ldots .$$

Now let $\epsilon > 0$ be small. Consider the open interval of length ϵ^n about the n^{th} rational in this list. The union of all of these intervals is an open set in I which is clearly dense since it contains all of the rationals in I. However, the total length of this set is quite small. Indeed, the length is given by

$$\sum_{n=1}^{\infty} \epsilon^n = \frac{\epsilon}{1 - \epsilon}.$$

This example shows clearly the difference between the topological approach to dynamics that we will adopt in the sequel and the measure theoretic approach. In a topological sense, an open, dense subset is considered "large." These sets may or may not be large in a measure theoretic sense, i.e., in the sense of total length.

Exercises

1. Decide whether each of the following functions are one-to-one, onto, homeomorphisms, or diffeomorphisms on their domains of definition.

a. $f(x) = x^{5/3}$
b. $f(x) = x^{4/3}$
c. $f(x) = 3x + 5$
d. $f(x) = e^x$
e. $f(x) = 1/x$
f. $f(x) = 1/x^2$

2. Identify which of the following subsets of $\mathbf{R}$ are closed, open, or neither.

a. $\{x | x \text{ is an integer }\}$
b. $\{x | x \text{ is a rational number }\}$
c. $\{x | x = \frac{1}{n} \text{ for some natural number } n\}$
d. $\{x | \sin(\frac{1}{x}) = 0\}$
e. $\{x | x\sin(\frac{1}{x}) = 0\}$
f. $\{x | \sin(\frac{1}{x}) > 0\}$

3. Prove that the set of rational numbers of the form $p/2^n$ for $p, n \in \mathbf{Z}$ is dense in $\mathbf{R}$.

The goal of the next few exercises is to construct special functions which will be useful later when we perturb or change slightly a given function. These functions are called "bump functions." Define

$$B(x) = \begin{cases} \exp(-1/x^2) & \text{if } x > 0 \\ 0 & \text{if } x \leq 0 \end{cases}$$

4. Sketch the graph of $B(x)$.

5. Prove that $B'(0) = 0$.

6. Inductively prove that $B^{(n)}(0) = 0$ for all n. Conclude that $B(x)$ is a C^∞ function.

7. Modify $B(x)$ to construct a C^∞ function $C(x)$ which satisfies

a. $C(x) = 0$ if $x \le 0$.

b. $C(x) = 1$ if $x \ge 1$.

c. $C'(x) > 0$ if $0 < x < 1$.

8. Modify $C(x)$ to construct a C^∞ *bump function* $D(x)$ on the interval $[a, b]$, i.e., $D(x)$ satisfies

a. $D(x) = 1$ for $a \le x \le b$.

b. $D(x) = 0$ for $x < \alpha$ and $x > \beta$ where $\alpha < a$ and $\beta > b$.

c. $D'(x) \neq 0$ on the intervals (α, a) and (b, β).

9. Use a bump function to construct a diffeomorphism $f: [a, b] \to [c, d]$ which satisfies $f'(a) = f'(b) = 1$ and $f(a) = c$, $f(b) = d$.

§1.3 ELEMENTARY DEFINITIONS

The basic goal of the theory of dynamical systems is to understand the eventual or asymptotic behavior of an iterative process. If this process is a differential equation whose independent variable is time, then the theory attempts to predict the ultimate behavior of solutions of the equation in either the distant future ($t \to \infty$) or the distant past ($t \to -\infty$). If the process is a discrete process such as the iteration of a function, then the theory hopes to understand the eventual behavior of the points $x, f(x), f^2(x), \ldots, f^n(x)$ as n becomes large. That is, dynamical systems asks the somewhat nonmathematical sounding question: where do points go and what do they do when they get there? In this chapter, we will attempt to answer this question at least partially for one of the simplest classes of dynamical systems, functions of a single real variable. Functions which determine dynamical systems are also called *mappings*, or *maps*, for short. This terminology connotes the geometric process of taking one point to another. As much of the sequel will in fact be geometric, we will use all of these terms synonymously.

Definition 3.1. The forward orbit of x is the set of points $x, f(x), f^2(x), \ldots$ and is denoted by $O^+(x)$. If f is a homeomorphism, we may define the full orbit of x, $O(x)$, as the set of points $f^n(x)$ for $n \in \mathbf{Z}$, and the backward orbit of x, $O^-(x)$, as the set of points $x, f^{-1}(x), f^{-2}(x), \ldots$.

Thus our basic goal is to understand all orbits of a map. Orbits and forward orbits of points can be quite complicated sets, even for very simple nonlinear mappings. However, there are some orbits which are especially simple and which will play a central role in the study of the entire system.

Definition 3.2. The point x is a fixed point for f if $f(x) = x$. The point x is a periodic point of period n if $f^n(x) = x$. The least positive n for which $f^n(x) = x$ is called the prime period of x. We denote the set of periodic points of (not necessarily prime) period n by $\mathrm{Per}_n(f)$, and the set of fixed points by $\mathrm{Fix}\,(f)$. The set of all iterates of a periodic point form a periodic orbit.

Maps may have many fixed points. For example, the identity map $id(x) = x$ fixes all points in $\mathbf{R}$, whereas the map $f(x) = -x$ fixes the origin, while all other points have period 2. These, however, are atypical dynamical systems; maps with intervals of fixed or periodic points are rare in a sense which will be made precise later. Most of the dynamical systems we will encounter will have isolated periodic points.

Example 3.3. The map $f(x) = x^3$ has $0, 1$, and -1 as fixed points and no other periodic points. The map $P(x) = x^2 - 1$ has fixed points at $(1 \pm \sqrt{5})/2$, while the points 0 and -1 lie on a periodic orbit of period 2.

Example 3.4. Let S^1 denote the unit circle in the plane. We denote a point in S^1 by its angle θ measured in radians in the standard manner. Hence a point is determined by any angle of the form $\theta + 2k\pi$ for an integer k. Now let $f(\theta) = 2\theta$. (Note that $f(\theta + 2\pi) = f(\theta)$ on the circle so this map is well defined.) Now $f^n(\theta) = 2^n\theta$, so that θ is periodic of period n if and only if $2^n\theta = \theta + 2k\pi$ for some integer k, i.e., if and only if $\theta = 2k\pi/(2^n - 1)$ where $0 \le k \le 2^n$ is an integer. Hence the periodic points of period n for f are the $(2^n - 1)^{th}$ roots of unity. It follows that the set of periodic points are dense in S^1. See Exercise 10.

Definition 3.5. A point x is eventually periodic of period n if x is not periodic but there exists $m > 0$ such that $f^{n+i}(x) = f^i(x)$ for all $i \ge m$. That is, $f^i(x)$ is periodic for $i \ge m$.

Example 3.6. Let $f(x) = x^2$. Then $f(1) = 1$ is fixed, while $f(-1) = 1$ is eventually fixed.

Example 3.7. Let $f(\theta) = 2\theta$ on the circle. Note that $f(0) = 0$ is fixed. If $\theta = 2k\pi/2^n$ then $f^n(\theta) = 2k\pi$ so that θ is eventually fixed. It follows that eventually fixed points are also dense in S^1. See Exercise 11.

We remark that eventually periodic points cannot occur if the map is a homeomorphism.

Definition 3.8. Let p be periodic of period n. A point x is forward asymptotic to p if $\lim_{i\to\infty} f^{in}(x) = p$. The stable set of p, denoted by $W^s(p)$, consists of all points forward asymptotic to p.

If p is non-periodic, we may still define forward asymptotic points by requiring $|f^i(x) - f^i(p)| \to 0$ as $i \to \infty$. Also, if f is invertible, we may consider *backward asymptotic* points by letting $i \to -\infty$ in the above definition. The set of points backwards asymptotic to p is called the *unstable set* of p and is denoted by $W^u(p)$.

Example 3.9. Let $f(x) = x^3$. Then $W^s(0)$ is the open interval $-1 < x < 1$. $W^u(1)$ is the positive real axis, whereas $W^u(-1)$ is the negative real axis.

Definition 3.10. A point x is a critical point of f if $f'(x) = 0$. The critical point is non-degenerate if $f''(x) \neq 0$. The critical point is degenerate if $f''(x) = 0$.

For example $f(x) = x^2$ has a non-degenerate critical point at 0, but $f(x) = x^n$ for $n > 2$ has a degenerate critical point at 0. Note that degenerate critical points may be maxima, minima, or saddle points (as in the case of $f(x) = x^3$). But non-degenerate critical points must be either maxima or minima. Critical points cannot occur for diffeomorphisms, but their existence for non-invertible maps is one reason why these kinds of maps are more complicated.

The goal of dynamical systems is to understand the nature of all orbits, and to identify the set of orbits which are periodic, eventually periodic, asymptotic, etc. Generally, this is an impossible task. For example, if $f(x)$ is a quadratic polynomial, then finding explicitly the periodic points of period n necessitates solving the equation $f^n(x) = x$, which is a polynomial equation of degree 2^n. A computer does not help matters much, for numerical computations of periodic points are often misleading. Round-off errors tend to accumulate and make many periodic points invisible to the computer. Therefore we are left with only qualitative or geometric techniques to understand the dynamics of a given system. This means that we should look for a geometric picture of the behavior of all orbits of a system. This geometric picture is provided by the phase portrait which we now discuss.

The graph of a function on the reals provides information about its first iterate, but gives very little information about subsequent iterates. To understand higher iterates, we could attempt to sketch each of their graphs, but this is a cumbersome procedure. There is a much more efficient, geometric method for describing the orbits of a dynamical system, the *phase portrait*. This is a picture, on the real line itself, as opposed to the plane, of all orbits of a system. For example, to indicate that all non-zero orbits of $f(x) = -x$ have period 2, we could sketch the phase portrait as in Fig. 3.1.a. This figure also depicts the phase portraits of some other simple maps.

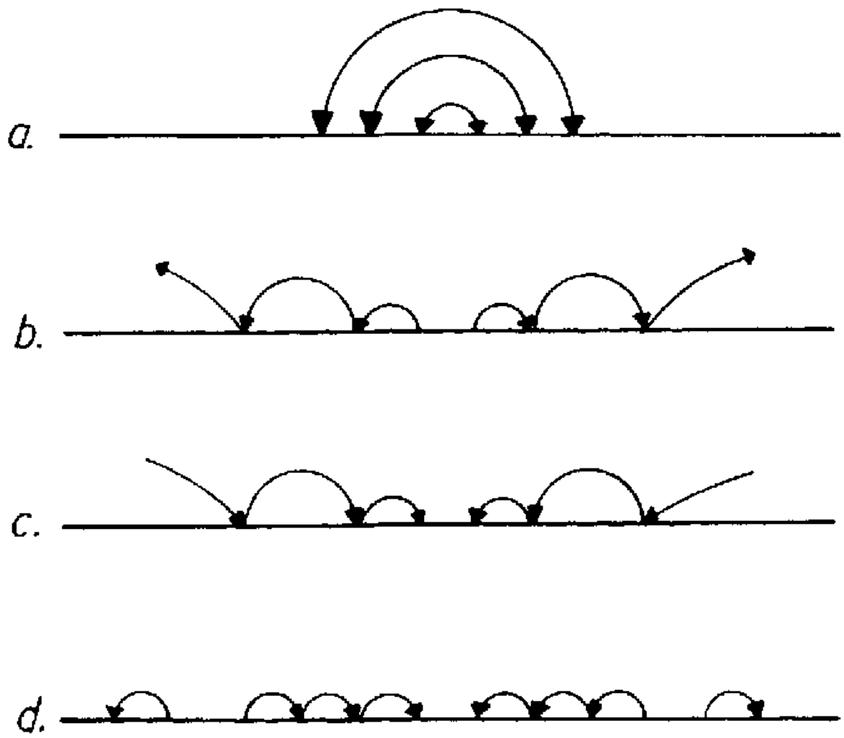

Fig. 3.1. The phase portraits of
a. $f(x) = -x$, b. $f(x) = 2x$,
c. $f(x) = \frac{1}{2}x$, d. $f(x) = x^3$.

The graph of $f(x)$ does of course contain information about the first iteration of f. We may use it to gain insight into higher iterations and hence the phase portrait via the following procedure which we call *graphical analysis*. Identify the diagonal $\Delta = \{(x, x) | x \in \mathbf{R}\}$ with $\mathbf{R}$ in the obvious way. A vertical line from (p, p) to the graph of f meets the graph at $(p, f(p))$. Then a horizontal line from $(p, f(p))$ to Δ meets the diagonal at $(f(p), f(p))$. Hence a vertical line to the graph followed by a horizontal line back to Δ yields the image of the point p under f on the diagonal. We may thus visualize the phase portrait of a map as taking place on the diagonal rather than

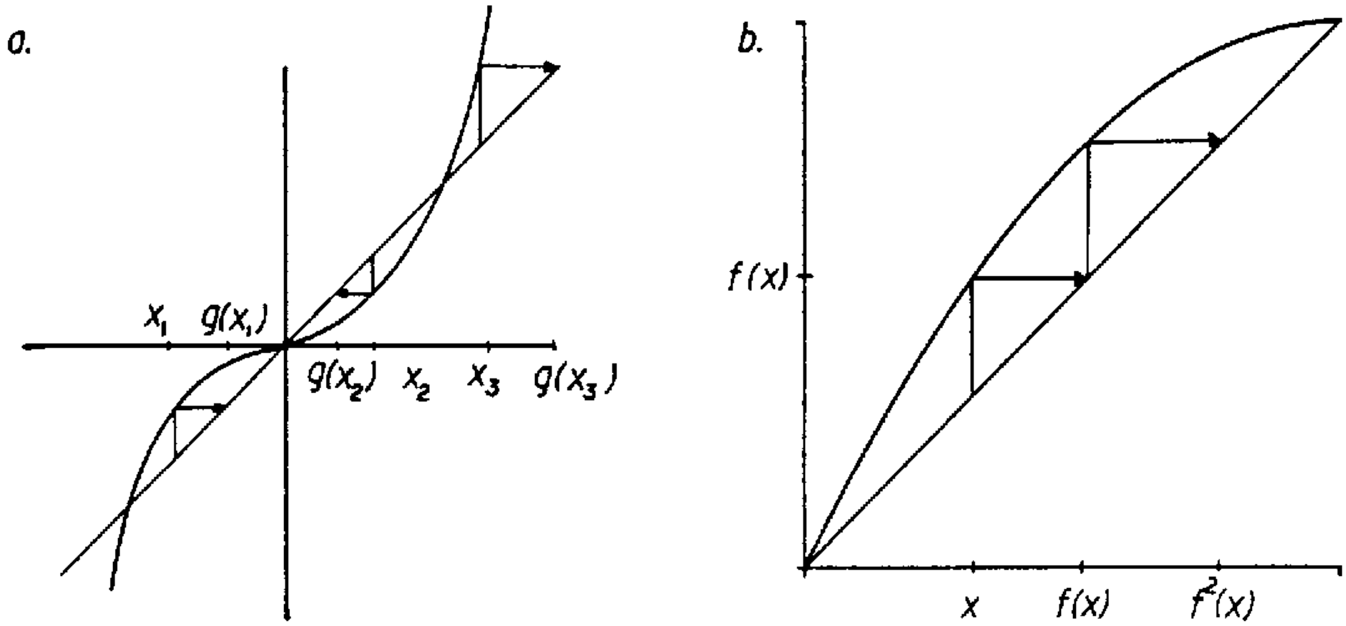

Fig. 3.2. Graphical analyses of
a. $g(x) = x^3$ and b. $f(x) = 2x - x^2$.

on the x-axis. Then an orbit is given by repeatedly drawing line segments vertically from Δ to the graph and then horizontally from the graph to Δ. Fig. 3.2 illustrates this procedure for $g(x) = x^3$ and $f(x) = 2x - x^2$.

Diffeomorphisms of the circle form an interesting class of maps which are somewhat different from maps of $\mathbf{R}$. The following example is typical.

Example 3.11. Let $f(\theta) = \theta + \epsilon \sin(2\theta)$ for $0 < \epsilon < 1/2$. Note that f has fixed points at $0, \pi/2, \pi$, and $3\pi/2$. We compute $f'(0) = f'(\pi) = 1 + 2\epsilon > 1$ whereas $f'(\pi/2) = f'(3\pi/2) = 1 - 2\epsilon < 1$. Hence 0 and π are repelling fixed points and $\pi/2$ and $3\pi/2$ are attracting. More generally, $f(\theta) = \theta + \epsilon \sin(N\theta)$ has N attracting and N repelling fixed points arranged alternately around the circle as long as $0 < \epsilon < 1/N$.

The phase portraits of these maps may be sketched as in Fig. 3.3. Another important class of circle maps are the translation maps.

Example 3.12. Translations of the circle. Let $\lambda \in \mathbf{R}$ and $T_\lambda(\theta) = \theta + 2\pi\lambda$. The maps T_λ behave quite differently depending upon the rationality or irrationality of λ. If $\lambda = p/q$, where p and q are integers, then $T_\lambda^q(\theta) = \theta + 2\pi p = \theta$ so that all points are fixed by T_λ^q. When λ is irrational, the situation is quite different. The following result is known as Jacobi's Theorem.

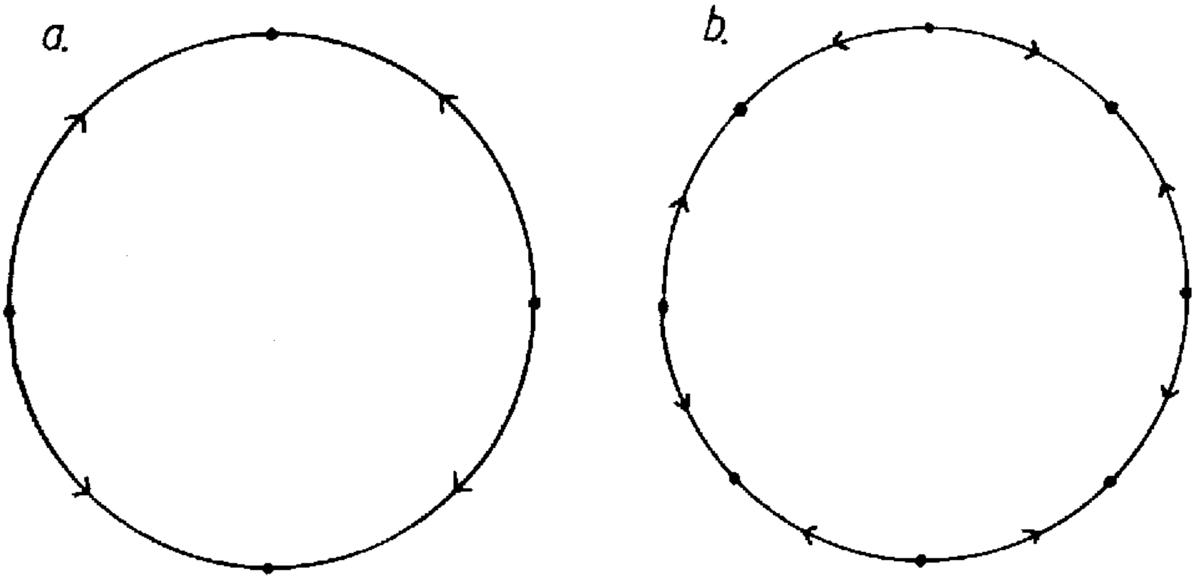

Fig. 3.3. The phase portraits of
a. $f(\theta) = \theta + \epsilon \sin(2\theta)$ and
b. $f(\theta) = \theta + \epsilon \sin(4\theta)$.

Theorem 3.13. *Each orbit T_λ is dense in S^1 if λ is irrational.*

Proof. Let $\theta \in S^1$. The points on the orbit of θ are distinct for if $T_\lambda^n(\theta) = T_\lambda^m(\theta)$ we would have $(n-m)\lambda \in \mathbf{Z}$, so that $n = m$. Any infinite set of points on the circle must have a limit point. Thus, given any $\epsilon > 0$, there must be integers n and m for which $|T_\lambda^n(\theta) - T_\lambda^m(\theta)| < \epsilon$. Let $k = n - m$. Then $|T_\lambda^k(\theta) - \theta| < \epsilon$.

Now T_λ preserves lengths in S^1. Consequently, T_λ^k maps the arc connecting θ to $T_\lambda^k(\theta)$ to the arc connecting $T_\lambda^k(\theta)$ and $T_\lambda^{2k}(\theta)$ which has length less than ϵ. In particular it follows that the points $\theta, T_\lambda^k(\theta), T_\lambda^{2k}(\theta), \ldots$ partition S^1 into arcs of length less than ϵ. Since ϵ was arbitrary, this completes the proof.

q.e.d.

Exercises

1. Use a calculator to iterate each of the following functions (using an arbitrary initial value) and explain these results.

a. $C(x) = \cos(x)$
b. $S(x) = \sin(x)$
c. $E(x) = e^x$
d. $F(x) = \frac{1}{e}e^x$
e. $A(x) = \arctan(x)$

2. Using the graph of the function, identify the fixed points for each of the maps in the previous Exercise.

3. List all periodic points for each of the following maps. Then use the graph of $f(x)$ to sketch the phase portrait of $f(x)$ on the indicated interval.

a. $f(x) = -\frac{1}{2}x, \quad -\infty < x < \infty$
b. $f(x) = -3x, \quad -\infty < x < \infty$
c. $f(x) = x - x^2, \quad 0 \leq x \leq 1$
d. $f(x) = \frac{\pi}{2}\sin x, \quad 0 \leq x \leq \pi$
e. $f(x) = -x^3, \quad -\infty < x < \infty$
f. $f(x) = \frac{1}{2}(x^3 + x), \quad -1 \leq x \leq 1$

4. Identify the stable sets of each of the fixed points for the maps in the previous Exercise.

5. For each of the following functions, list all critical points and decide whether each is degenerate or non-degenerate.

a. $f(x) = x^3 - x$
b. $S(x) = \sin(x)$
c. $f(x) = x^4 - 2x^2$
d. $g(x) = x^3 + x^4$

6. Describe the phase portrait of the map of the circle given by

$$f(\theta) = \theta + \frac{2\pi}{n} + \epsilon\sin(n\theta)$$

for $0 < \epsilon < 1/n$.

7. Prove that a homeomorphism of $\mathbf{R}$ can have no periodic points with prime period greater than 2. Give an example of a homeomorphism that has a periodic point of period 2.

8. Prove that a homeomorphism cannot have eventually periodic points.

9. Let $S\colon S^1 \to S^1$ be given by $S(\theta) = \theta + \omega + \epsilon\sin(\theta)$ where ω and ϵ are constants. Prove that S is a homeomorphism of the circle if $|\epsilon| < 1$.

10. Let $f(\theta) = 2\theta$ be the map of S^1 discussed in Example 3.4. Prove that periodic points of f are dense in S^1.

11. Prove that eventually fixed points for the map in Exercise 10 are also dense in S^1.

§1.4 HYPERBOLICITY

Simple maps like $id(x) = x$ and $f(x) = -x$ are, unfortunately, atypical among dynamical systems. There are many reasons why this is so, but perhaps the most unusual feature of these maps is the fact that all points are periodic under iteration of these maps. Most maps do not have this type of behavior. Periodic points tend to be more spread out on the line. In this section we will introduce one of the main themes of this book, hyperbolicity. Maps with hyperbolic periodic points are the ones that occur typically in many dynamical systems and, moreover, they provide the simplest types of periodic behavior to analyze.

Definition 4.1. Let p be a periodic point of prime period n. The point p is hyperbolic if $|(f^n)'(p)| \neq 1$. The number $(f^n)'(p)$ is called the multiplier of the periodic point.

Example 4.2. Consider the diffeomorphism $f(x) = \frac{1}{2}(x^3 + x)$. There are 3 fixed points: $x = 0, 1$, and -1. Note that $f'(0) = 1/2$ and $f'(\pm 1) = 2$. Hence each fixed point is hyperbolic. The graph and phase portrait of $f(x)$ are depicted in Fig. 4.1.

Example 4.3. Let $f(x) = -\frac{1}{2}(x^3 + x)$. 0 is a hyperbolic fixed point, with $f'(0) = -\frac{1}{2}$. The points ± 1 now lie on a periodic orbit of period 2. We compute $(f^2)'(\pm 1) = f'(1) \cdot f'(-1) = 4$ by the chain rule. Hence this periodic point is hyperbolic, and the phase portrait is depicted in Fig. 4.2. Note that points in the interval $(-1, 1)$ spiral toward 0 and away from ± 1.

We observe that, in the above two examples, we have $|f'(0)| < 1$ and that points close to 0 are forward asymptotic to 0. This situation occurs often:

Proposition 4.4. *Let p be a hyperbolic fixed point with $|f'(p)| < 1$. Then there is an open interval U about p such that if $x \in U$, then*

$$\lim_{n \to \infty} f^n(x) = p.$$

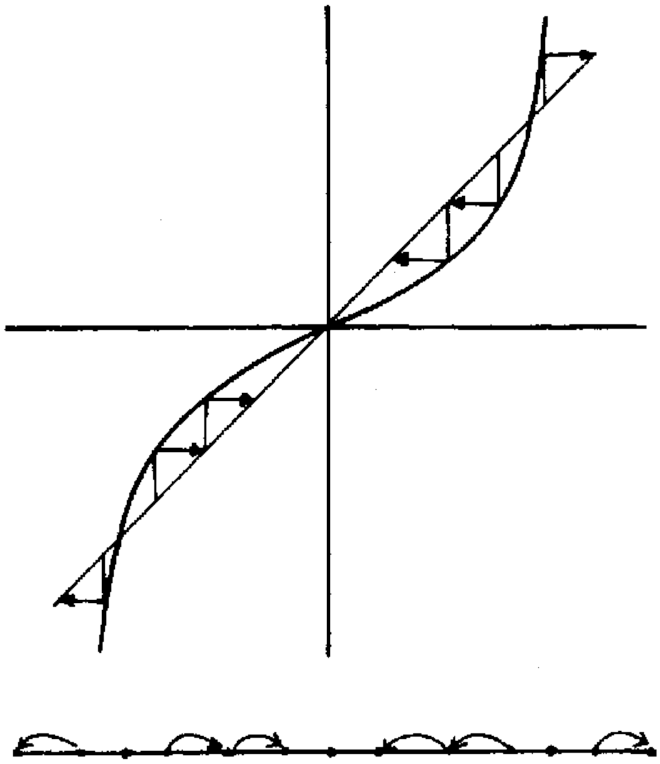

Fig. 4.1. The graph and phase portraits of
$f(x) = \frac{1}{2}(x^3 + x)$.

Proof. Since f is C^1, there is $\epsilon > 0$ such that $|f'(x)| < A < 1$ for $x \in [p - \epsilon, p + \epsilon]$. By the Mean Value Theorem

$$|f(x) - p| = |f(x) - f(p)| \leq A|x - p| < |x - p| \leq \epsilon.$$

Hence $f(x)$ is contained in $[p - \epsilon, p + \epsilon]$ and, in fact, is closer to p than x is. Via the same argument

$$|f^n(x) - p| \leq A^n|x - p|$$

so that $f^n(x) \to p$ as $n \to \infty$.

q.e.d.

Remarks.

1. It follows that the interval $[p - \epsilon, p + \epsilon]$ is contained in the stable set associated to p, $W^s(p)$.

2. A similar result is true for hyperbolic periodic points of period n. In this case, we get an open interval U about p which is mapped inside itself by f^n. Of course, the assumption in this case is that $|(f^n)'(p)| < 1$.

Definition 4.5. Let p be a hyperbolic periodic point of period n with $|(f^n)'(p)| < 1$. The point p is called an attracting periodic point (an attractor) or a sink.

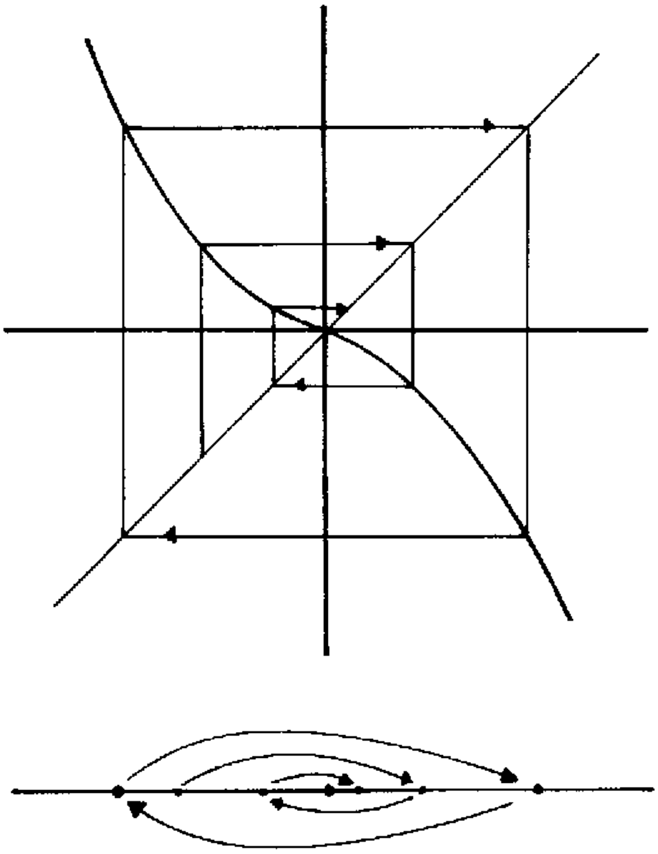

Fig. 4.2. The graph and phase portraits of $f(x) = -\frac{1}{2}(x^3 + x)$.

Attracting periodic points of period n thus have neighborhoods which are mapped inside themselves by f^n. Such a neighborhood is called the *local stable set* and is denoted by W^s_{loc}. We may actually distinguish three different types of attracting fixed points, namely those where $f'(p) = 0$, $0 < f'(p) < 1$, and $-1 < f'(p) < 0$. The behavior near these types of fixed points is illustrated in Fig. 4.3.

The behavior of a map near periodic points where the derivative is larger than one in absolute value is quite different from that of sinks.

Proposition 4.6. *Let p be a hyperbolic fixed point with $|f'(p)| > 1$. Then there is an open interval U of p such that, if $x \in U$, $x \neq p$, then there exists $k > 0$ such that $f^k(x) \notin U$.*

The proof is similar to the proof of the preceding proposition and is therefore left as an exercise. Graphically, the result is quite clear; see Fig. 4.4.

Definition 4.7. A fixed point p with $|f'(p)| > 1$ is called a repelling fixed point (a repellor) or source. The neighborhood described in the Proposition is called the local unstable set and denoted W^u_{loc}.

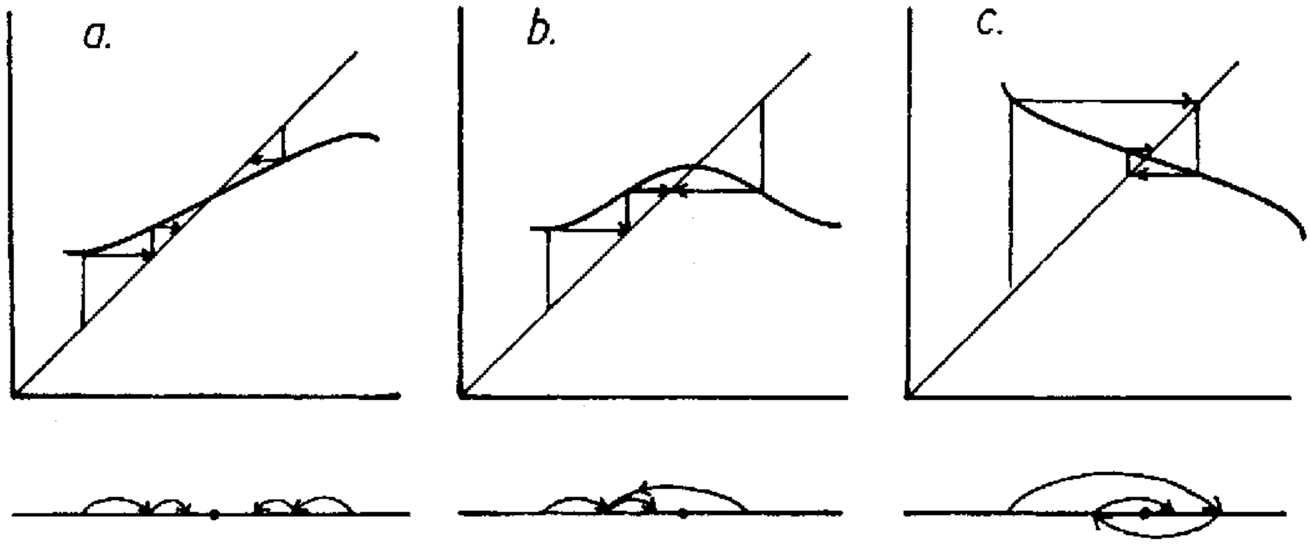

Fig. 4.3. The phase portraits near an attracting fixed point p. in case a. $0 < f'(p) < 1$, b. $f'(p) = 0$, c. $-1 < f'(p) < 0$.

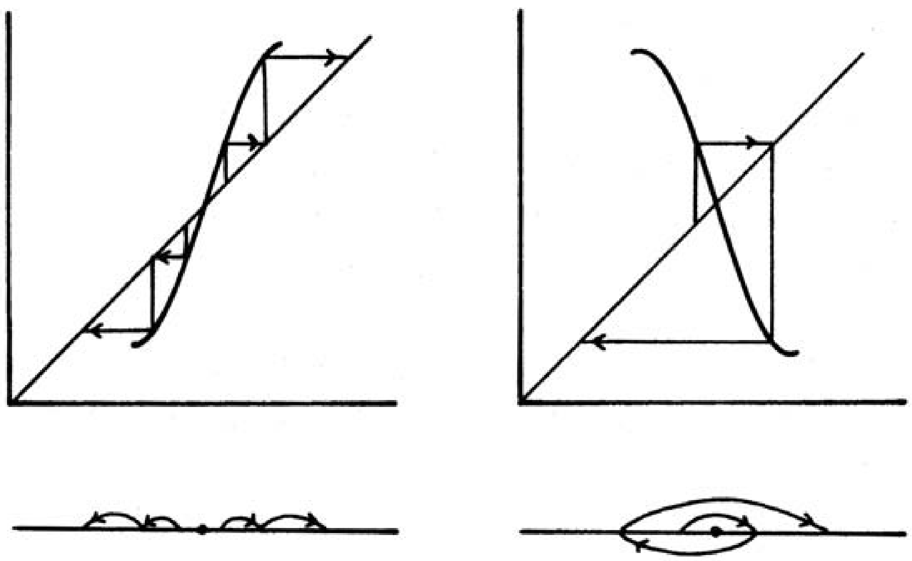

Fig. 4.4. The phase portraits near a repelling fixed point.

We remark that periodic points of period n exhibit similar behavior when $|(f^n)'(p)| > 1$.

Hyperbolic periodic points therefore have local behavior which is governed by the derivative at the periodic point. This is not true when the point is indifferent or non-hyperbolic, as the following example shows.

Example 4.8. Each of the maps in Fig. 4.5 satisfy $f(0) = 0$ and $f'(0) = 1$, but each have vastly different phase portraits near 0. In a., the map $f(x) = x + x^3$ has a *weakly* repelling fixed point at 0. In b., the map $f(x) = x - x^3$

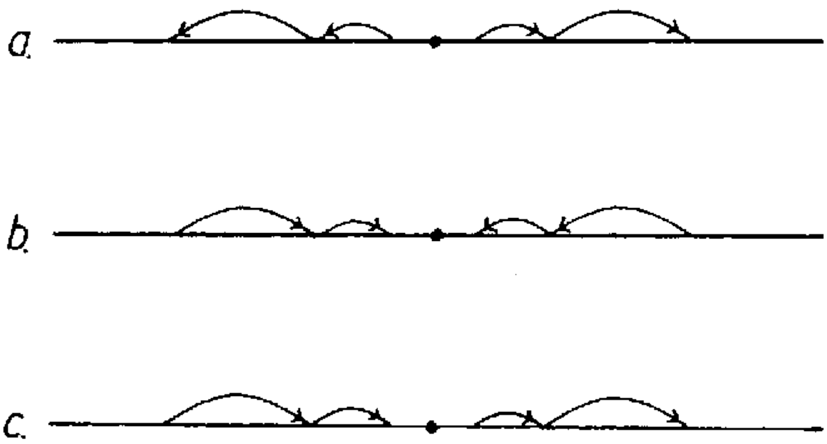

Fig. 4.5. The phase portraits of a. $f(x) = x + x^3$, b. $f(x) = x - x^3$, c. $f(x) = x + x^2$.

has a *weakly* attracting fixed point at 0. In c., the map $f(x) = x + x^2$ is weakly repelling from the right but weakly attracting from the left.

Most maps have only hyperbolic periodic points, as we shall see later. However, non-hyperbolic periodic points often occur in families of maps. When this happens, the periodic point structure often undergoes a *bifurcation*. We will deal with bifurcation theory more extensively later, but for now we give several examples.

Example 4.9. Consider the family of quadratic functions $Q_c(x) = x^2 + c$, where c is a parameter. The graphs of Q_c assume three different positions relative to the diagonal depending upon whether $c > 1/4$, $c = 1/4$, or $c < 1/4$. See Fig. 4.6. Note that Q_c has no fixed points for $c > 1/4$. When $c = 1/4$, Q_c has a unique non-hyperbolic fixed point at $x = 1/2$. And when $c < 1/4$, Q_c has a pair of fixed points, one attracting and one repelling. Thus the phase portrait of Q_c changes as c decreases through $1/4$. This change is an example of a bifurcation.

Example 4.10. Let $F_\mu(x) = \mu x(1-x)$ with $\mu > 1$. F_μ has two fixed points: one at 0 and the other at $p_\mu = (\mu - 1)/\mu$. Note that $F'_\mu(0) = \mu$ and $F'_\mu(p_\mu) = 2 - \mu$. Hence 0 is a repelling fixed point for $\mu > 1$ and p_μ is attracting for $1 < \mu < 3$. When $\mu = 3$, $F'_\mu(p_\mu) = -1$. We sketch the graphs of F^2_μ for μ near 3. See Fig. 4.7. Note that 2 new fixed points for F^2_μ appear as μ increases through 3. These are new periodic points of period 2. Another bifurcation has occurred: this time we have a change in $\mathrm{Per}_2(F_\mu)$.

This quadratic family actually exhibits many of the phenomena that are crucial in the general theory. The next section is devoted entirely to this function.

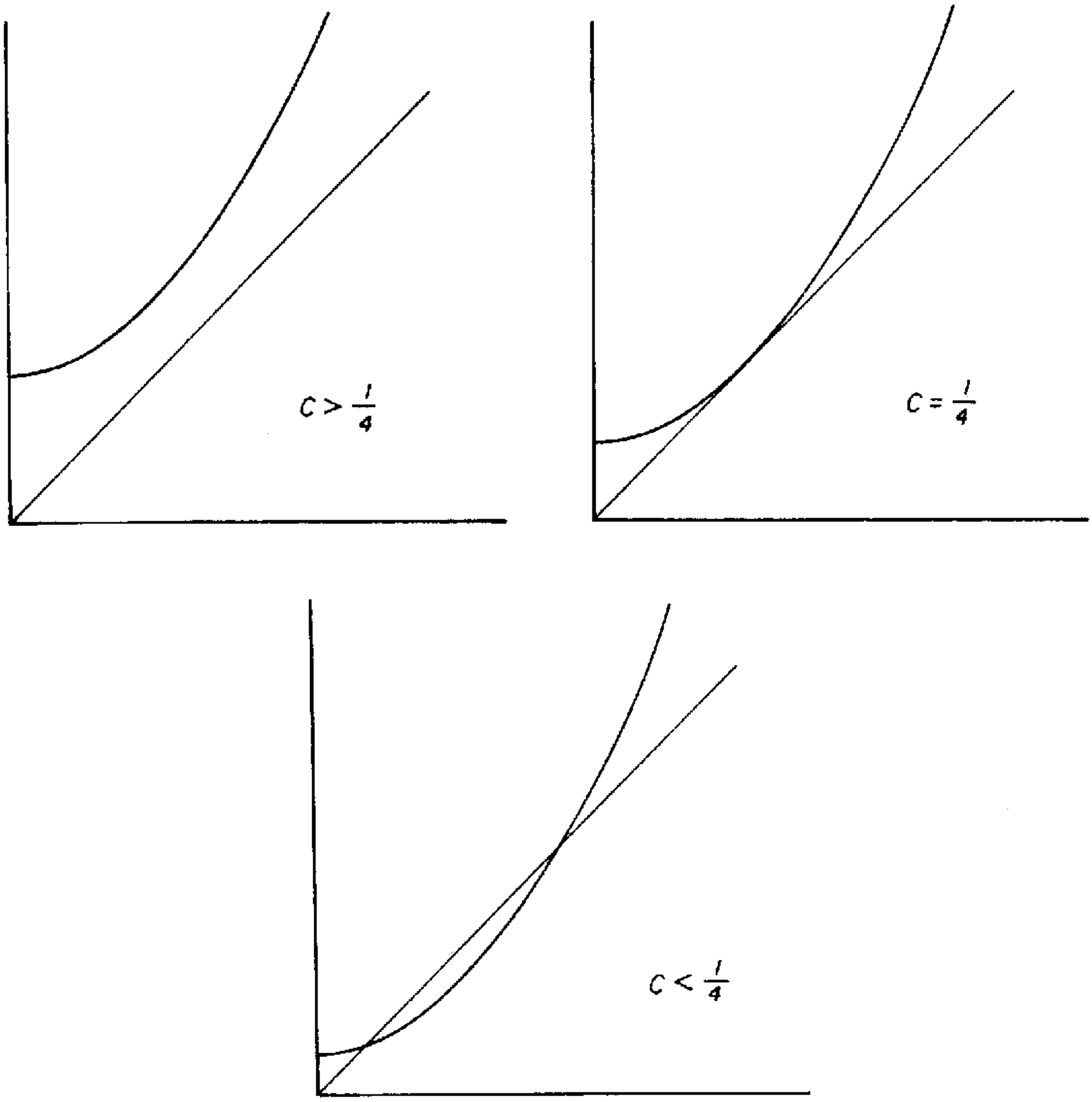

Fig. 4.6. The graphs of $Q_c(x) = x^2 + c$ for $c > 1/4$, $c = 1/4$, and $c < 1/4$.

Exercises

1. Find all periodic points for each of the following maps and classify them as attracting, repelling, or neither. Sketch the phase portraits.

a. $f(x) = x - x^2$

b. $f(x) = 2(x - x^2)$

c. $f(x) = x^3 - \frac{1}{9}x$

d. $f(x) = x^3 - x$

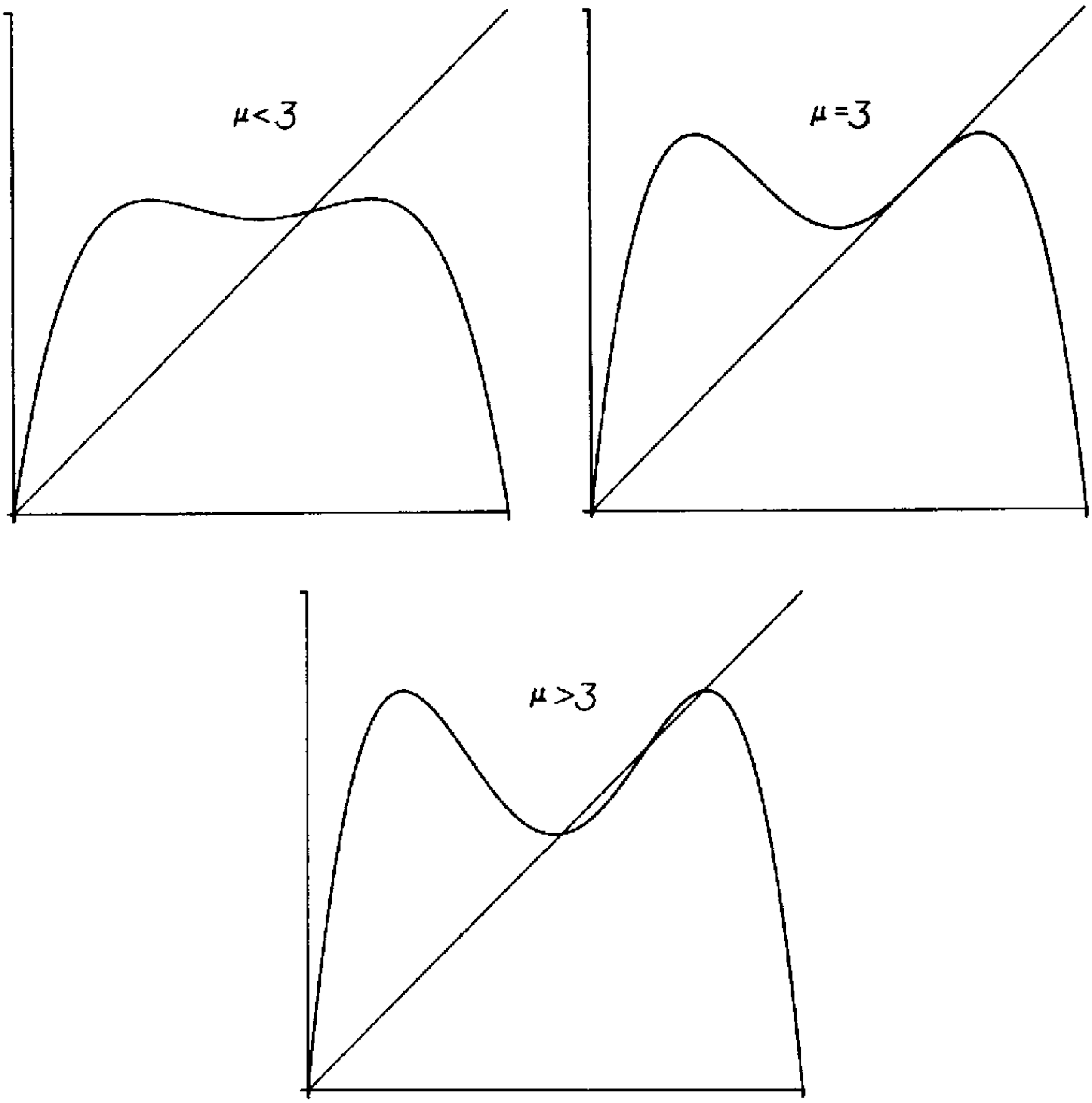

Fig. 4.7. The graphs of $F_\mu^2(x)$ where $F_\mu(x) = \mu x(1-x)$ for $\mu < 3, \mu = 3$, and $\mu > 3$.

e. $S(x) = \frac{1}{2}\sin(x)$
f. $S(x) = \sin(x)$
g. $E(x) = e^{x-1}$
h. $E(x) = e^x$
i. $A(x) = \arctan x$
j. $A(x) = \frac{\pi}{4}\arctan x$

k. $A(x) = -\frac{\pi}{4} \arctan x$

2. Discuss the bifurcations which occur in the following families of maps for the indicated parameter value

a. $S_\lambda(x) = \lambda \sin x, \quad \lambda = 1$
b. $E_\lambda(x) = \lambda e^x, \quad \lambda = 1/e$
c. $E_\lambda(x) = \lambda e^x, \quad \lambda = -e$
d. $Q_c(x) = x^2 + c, \quad c = -3/4$
e. $F_\mu(x) = \mu x(1-x), \quad \mu = 1$
f. $A_\lambda(x) = \lambda \arctan x, \quad \lambda = 1$
g. $A_\lambda(x) = \lambda \arctan x, \quad \lambda = -1$

3. Suppose f is a diffeomorphism. Prove that all hyperbolic periodic points are isolated.

4. Show via an example that hyperbolic periodic points need not be isolated.

5. Find an example of a C^1 diffeomorphism with a non-hyperbolic fixed point which is an accumulation point of other hyperbolic fixed points.

6. Discuss the dynamics of the family $f_\alpha(x) = x^3 - \alpha x$ for $-\infty < \alpha \leq 1$. Find all parameter values where bifurcations occur. Describe how the phase portrait of f_α changes at these points.

7. Consider the linear maps $f_k(x) = kx$. Show that there are four open sets of parameters for which the phase portraits of f_k are similar. The exceptional cases are $k = 0, \ \pm 1$.

§1.5 AN EXAMPLE: THE QUADRATIC FAMILY

In this section, we will continue the discussion of the quadratic family $F_\mu(x) = \mu x(1-x)$. Actually, we will return to this example repeatedly throughout the remainder of this chapter, since it illustrates many of the most important phenomena that occur in dynamical systems.

Proposition 5.1.

1. $F_\mu(0) = F_\mu(1) = 0$ *and* $F_\mu(p_\mu) = p_\mu$, *where* $p_\mu = \dfrac{\mu - 1}{\mu}$.
2. $0 < p_\mu < 1$ *if* $\mu > 1$.

The proof of this Proposition is straightforward. From now on we will concentrate on the case $\mu > 1$. The following Proposition shows that most points behave rather tamely under iteration of F_μ: all points which do not lie in the interval $[0, 1]$ tend to $-\infty$.

Proposition 5.2. *Suppose* $\mu > 1$. *If* $x < 0$, *then* $F_\mu^n(x) \to -\infty$ *as* $n \to \infty$. *Similarly, if* $x > 1$, *then* $F_\mu^n(x) \to -\infty$ *as* $n \to \infty$.

Proof. If $x < 0$, then $\mu x(1-x) < x$ so $F_\mu(x) < x$. Hence $F_\mu^n(x)$ is a decreasing sequence of points. This sequence cannot converge to p, for then we would have $F_\mu^{n+1}(x) \to F_\mu(p) < p$, whereas $F_\mu^n(x) \to p$. Hence $F_\mu^n(p) \to -\infty$ as required. If $x > 1$, then $F_\mu(x) < 0$ so $F_\mu^n(x) \to -\infty$ as well.

q.e.d.

Graphical analysis yields the above results easily, as shown in Fig. 5.1. As a consequence of this Proposition, all of the interesting dynamics of the quadratic family occur in the unit interval $I = \{x \,|\, 0 \leq x \leq 1\}$. For low values of μ, the dynamics of F_μ are not too complicated.

Proposition 5.3. *Let* $1 < \mu < 3$.

1. F_μ *has an attracting fixed point at* $p_\mu = (\mu - 1)/\mu$ *and a repelling fixed point at* 0.
2. *If* $0 < x < 1$, *then*

$$\lim_{n\to\infty} F_\mu^n(x) = p_\mu.$$

Proof. Part 1 was proved in Example 4.10 at the end of the last section. For part 2, we first deal with the case $1 < \mu < 2$. Suppose x lies in the interval $(0, 1/2]$. Then graphical analysis immediately shows that

$$|F_\mu(x) - p_\mu| < |x - p_\mu|$$

if $x \neq p_\mu$. See Fig. 5.2. Consequently, $F_\mu^n(x) \to p_\mu$ as $n \to \infty$. If, on the other hand, x lies in the interval $(1/2, 1)$, then $F_\mu(x)$ lies in $(0, 1/2)$, so that the previous argument implies

$$F_\mu^n(x) = F_\mu^{n-1}(F_\mu(x)) \to p_\mu$$

as $n \to \infty$.

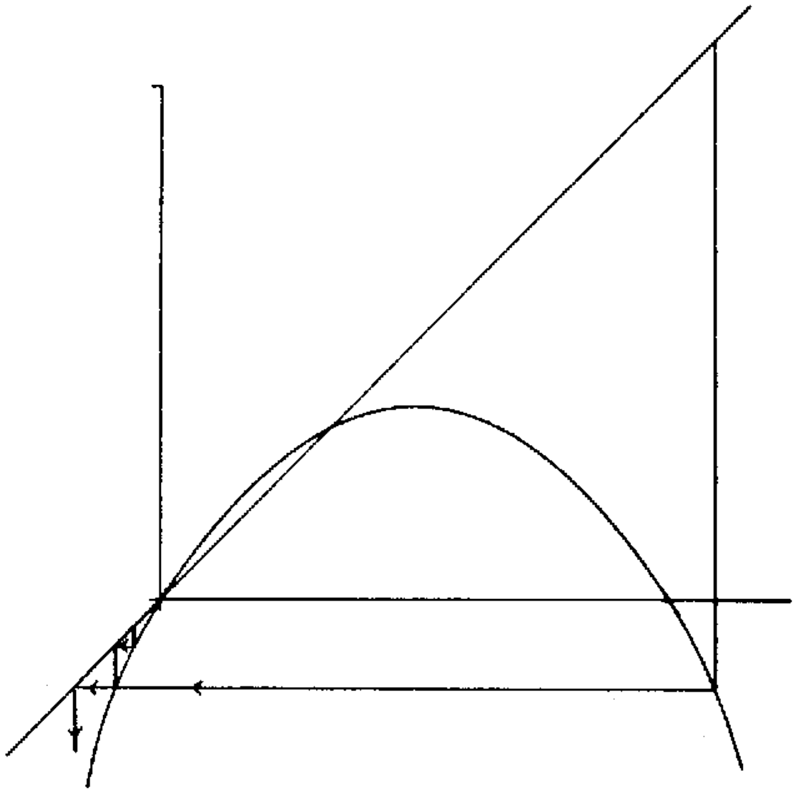

Fig. 5.1. Graphical analysis of $F_\mu(x) = \mu x(1-x)$ when $\mu > 1$.

The case when $2 < \mu < 3$ is more difficult. Graphical analysis shows what is different in this case. See Fig. 5.2. Note that $1/2 < p_\mu < 1$. Let $\hat{p}_\mu$ denote the unique point in the interval $(0, 1/2)$ that is mapped onto p_μ by F_μ. Then the reader may easily check that F_μ^2 maps the interval $[\hat{p}_\mu, p_\mu]$ inside $[1/2, p_\mu]$. It follows that $F_\mu^n(x) \to p_\mu$ as $n \to \infty$ for all $x \in [\hat{p}_\mu, p_\mu]$. Now suppose $x < \hat{p}_\mu$. Again graphical analysis shows that there exists $k > 0$ such that $F_\mu^k(x) \in [\hat{p}_\mu, p_\mu]$. Thus $F_\mu^{k+n}(x) \to p_\mu$ as $n \to \infty$ in this case as well. Finally, as before, F_μ maps the interval $(p_\mu, 1)$ onto $(0, p_\mu)$, so the result follows here as well. Since $(0,1) = (0, \hat{p}_\mu) \cup [\hat{p}_\mu, p_\mu] \cup (p_\mu, 1)$, we are finished. We leave the intermediate case $\mu = 2$ to the reader. See Exercise 1.

q.e.d.

Hence for $1 < \mu < 3$, F_μ has only two fixed points and all other points in I are asymptotic to p_μ. Thus the dynamics of F_μ are completely understood for μ in this range. The phase portraits of F_μ are depicted in Fig. 5.3.

As we showed in Example 4.10 in the previous section, as μ passes through 3, the dynamics of F_μ become slightly more complicated: a new periodic point of period 2 is born. This is the beginning of a long story: as μ continues to increase the dynamics of F_μ become increasingly more complicated until the phase portrait of F_μ is dramatically different from the above picture. This is a scenario that we will investigate in much more detail later.

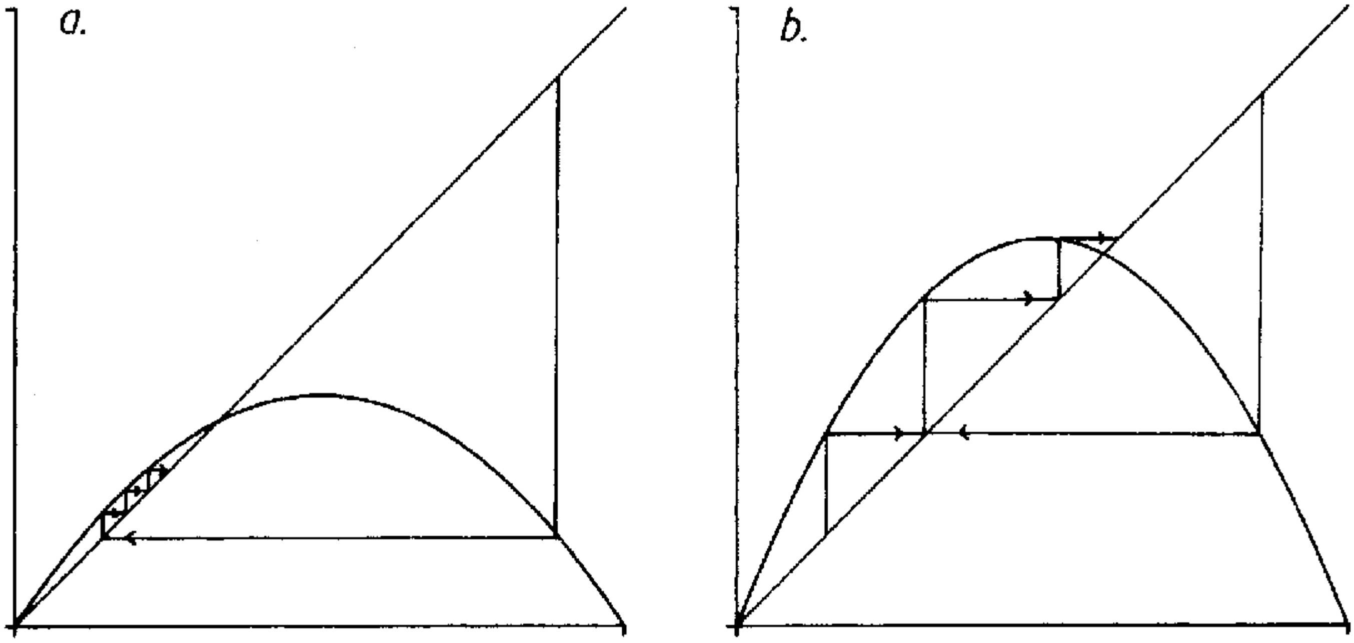

Fig. 5.2. Graphical analysis of $F_\mu(x) = \mu x(1-x)$ when a. $1 < \mu < 2$, and b. $2 < \mu < 3$.

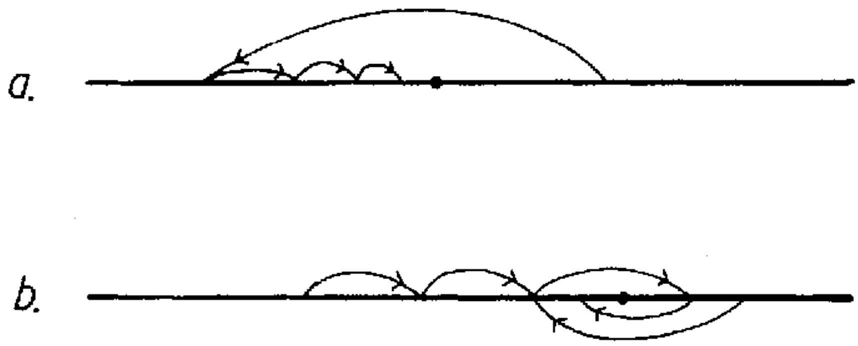

Fig. 5.3. The phase portraits for $F_\mu(x) = \mu x(1-x)$ when a. $1 < \mu < 2$, and b. $2 < \mu < 3$.

We now turn to the case when $\mu > 4$. For the remainder of this section, we will drop the subscript μ and write F instead of F_μ. As above, all of the interesting dynamics of F occur in the unit interval I. Note that, since $\mu > 4$, the maximum value $\mu/4$ of F is larger than one. Hence certain points leave I after one iteration of F. Denote the set of such points by A_0. Clearly, A_0 is an open interval centered at $\frac{1}{2}$ and has the property that, if $x \in A_0$, then $F(x) > 1$, so $F^2(x) < 0$ and $F^n(x) \to -\infty$. A_0 is the set of points which immediately escape from I. All other points in I remain in I after one iteration of F.

Let $A_1 = \{x \in I \mid F(x) \in A_0\}$. If $x \in A_1$, then $F^2(x) > 1$, $F^3(x) < 0$, and so, as before, $F^n(x) \to -\infty$. Inductively, let $A_n = \{x \in I | F^n(x) \in A_0\}$.

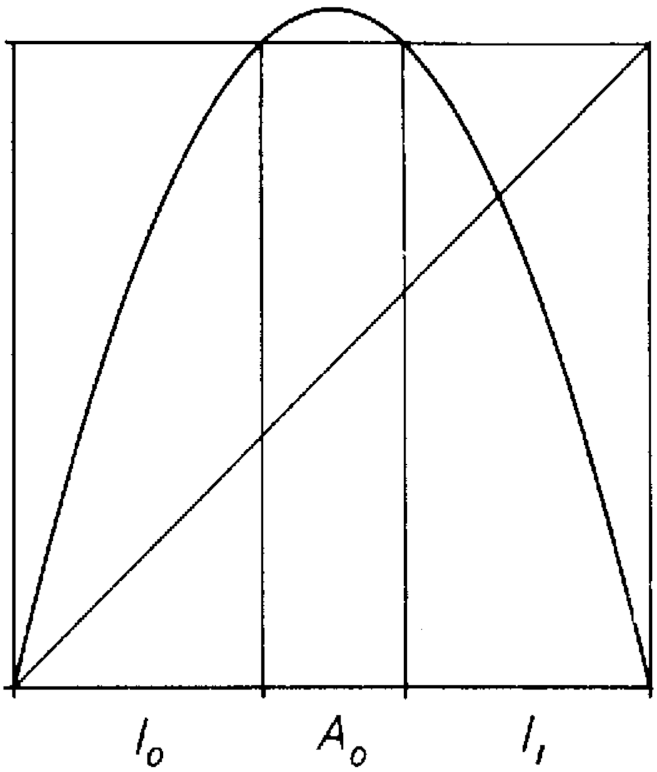

Fig. 5.4.

That is, $A_n = \{x \in I | F^i(x) \in I$ for $i \leq n$ but $F^{n+1}(x) \notin I\}$, so that A_n consists of all points which escape from I at the $n+1^{st}$ iteration. As above, if x lies in A_n, it follows that the orbit of x tends eventually to $-\infty$. Since we therefore know the ultimate fate of any point which lies in the A_n, it therefore remains only to analyze the behavior of those points which never escape from I, i.e., the set of points which lie in

$$I - \left(\bigcup_{n=0}^{\infty} A_n\right).$$

Let us denote this set by Λ. Our first question is: what precisely is this set of points? To understand Λ, we describe more carefully its recursive construction.

Since A_0 is an open interval centered at $1/2$, $I - A_0$ consists of two closed intervals, I_0 on the left and I_1 on the right. See Fig. 5.4.

Note that F maps both I_0 and I_1 monotonically onto I; F is increasing on I_0 and decreasing on I_1. Since $F(I_0) = F(I_1) = I$, there are a pair of open intervals, one in I_0 and one in I_1, which are mapped into A_0 by F. Therefore this pair of intervals is precisely the set A_1.

Now consider $I - (A_0 \cup A_1)$. This set consists of 4 closed intervals and F maps each of them monotonically onto either I_0 or I_1. Consequently F^2 maps each of them onto I. We therefore see that each of the four intervals in

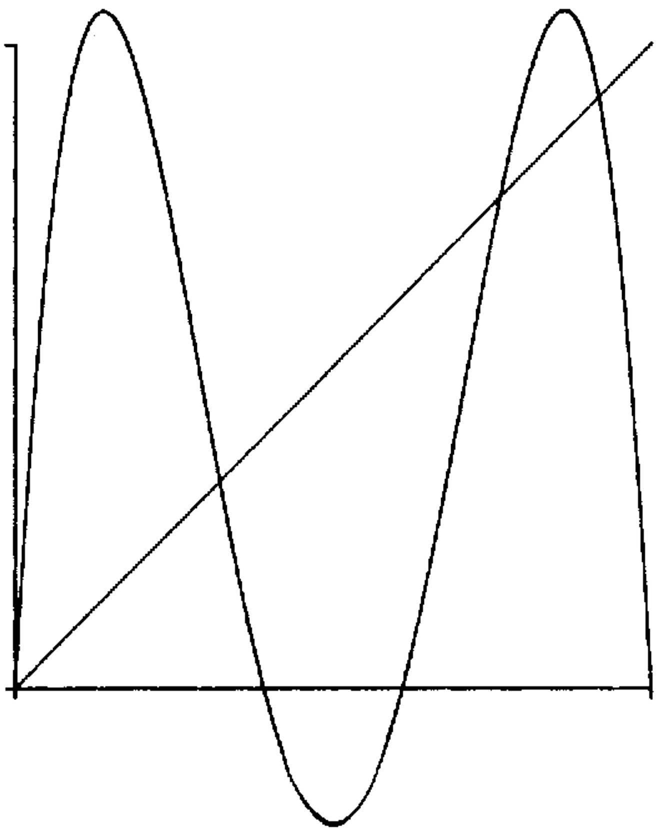

Fig. 5.5. The graph of F^2.

$I - (A_0 \cup A_1)$ contains an open subinterval which is mapped by F^2 onto A_0. Therefore, points in these intervals escape from I upon the third iteration of F. This is the set we called A_2. For later use, we observe that F^2 is alternately increasing and decreasing on these four intervals. It follows that the graph of F^2 must therefore have two humps as shown in Fig. 5.5.

Continuing in this manner we note two facts. First, A_n consists of 2^n disjoint open intervals. Hence $I - (A_0 \cup \ldots \cup A_n)$ consists of 2^{n+1} closed intervals since

$$1 + 2 + 2^2 + \ldots + 2^n = 2^{n+1} - 1.$$

Secondly, F^{n+1} maps each of these closed intervals monotonically onto I. In fact, the graph of F^{n+1} is alternately increasing and decreasing on these intervals. Thus the graph of F^{n+1} has exactly 2^n humps on I, and it follows that the graph of F^n crosses the line $y = x$ at least 2^n times. This implies that F^n has at least 2^n fixed points or, equivalently, $\mathrm{Per}_n(F)$ consists of 2^n points in I. Clearly, the structure of Λ is much more complicated when $\mu > 4$ than the earlier case $\mu < 3$.

The construction of Λ is reminiscent of the construction of the Cantor Middle Thirds set: Λ is obtained by successively removing open intervals from the "middles" of a set of closed intervals.

Definition 5.4. A set Λ is a Cantor set if it is a closed, totally disconnected, and perfect subset of I. A set is totally disconnected if it contains no intervals; a set is perfect if every point in it is an accumulation point or limit point of other points in the set.

Example 5.5. *The Cantor Middle-Thirds Set.* This is the classical example of a Cantor set. Start with I but remove the open "middle third," i.e., the interval $(\frac{1}{3}, \frac{2}{3})$. Next, remove from what remains the two middle thirds again, i.e., the pair of intervals $(\frac{1}{9}, \frac{2}{9})$ and $(\frac{7}{9}, \frac{8}{9})$. Continue removing middle thirds in this fashion; note that 2^n open intervals are removed at the n^{th} stage of this process. Thus, this procedure is entirely analogous to our construction above. Exercise 7 shows that the Cantor Middle-Thirds set is indeed a Cantor set as defined in 5.4.

Remark. The Cantor Middle-Thirds set is an example of a *fractal*. Intuitively, a fractal is a set which is self-similar under magnification. In the Cantor Middle-Thirds set, suppose we look only at those points which lie in the left-hand interval $[0, \frac{1}{3}]$. Under a microscope which magnifies this interval by a factor of three, the "piece" of the Cantor set in $[0, \frac{1}{3}]$ looks exactly like the original set. More precisely, the linear map $L(x) = 3x$ maps the portion of the Cantor set in $[0, \frac{1}{3}]$ homeomorphically onto the entire set. See Exercise 10. This process does not stop at the first level: one may magnify any piece of the Cantor set at the n^{th} stage of the construction by a factor of 3^n and obtain the original set. See Exercise 11.

To guarantee that our set Λ is a Cantor set, we need an additional hypothesis on μ. Let us assume that μ is large enough so that $|F'(x)| > 1$ for all $x \in I_0 \cup I_1$. The reader may check that $\mu > 2 + \sqrt{5}$ suffices. Hence, for these values of μ, there exists $\lambda > 1$ such that $|F'(x)| > \lambda$ for all $x \in \Lambda$. By the chain rule, it follows that $|(F^n)'(x)| > \lambda^n$ as well. We claim that Λ contains no intervals. Indeed, if this were so, we could choose $x, y \in \Lambda$, $x \neq y$, with the closed interval $[x, y] \subset \Lambda$. But then, $|(F^n)'(\alpha)| > \lambda^n$ for all $\alpha \in [x, y]$. Choose n so that $\lambda^n|y - x| > 1$. By the Mean Value Theorem, it then follows that $|F^n(y) - F^n(x)| \geq \lambda^n|y - x| > 1$, which implies that at least one of $F^n(y)$ or $F^n(x)$ lies outside of I. This is a contradiction, and so Λ is totally disconnected.

Since Λ is a nested intersection of closed intervals, Λ is closed. We now prove that Λ is perfect. First note that any endpoint of an A_k is in Λ: indeed, such points are eventually mapped to the fixed point at 0, and so they stay in I under iteration. Now if $p \in \Lambda$ were isolated, every nearby point must leave I under iteration of F. Such points must belong to some A_k. Either there is a sequence of endpoints of the A_k converging to p, or else all points in a deleted neighborhood of p are mapped out of I by some power of F. In the former case, we are done as the endpoints of the A_k map to 0 and hence are in Λ. In the latter, we may assume that F^n maps p to 0 and all other points in a neighborhood of p into the negative real axis. But then F^n has a maximum at p so that $(F^n)'(p) = 0$. By the chain rule, we must have $F'(F^i(p)) = 0$ for some $i < n$. Hence $F^i(p) = 1/2$. But then $F^{i+1}(p) \notin I$ and so $F^n(p) \to -\infty$, contradicting the fact that $F^n(p) = 0$.

Hence we have proved

Theorem 5.6. *If $\mu > 2 + \sqrt{5}$, then Λ is a Cantor set.*

Remark. The theorem is true for $\mu > 4$, but the proof is more delicate.

We have now succeeded in understanding the gross behavior of orbits of F_μ when $\mu > 4$. Either a point tends to $-\infty$ under iteration of F_μ, or else its entire orbit lies in Λ. Hence we understand the orbit of a point under F_μ perfectly well as long as the point does not lie in Λ. In the next section, we will complete the analysis of the dynamics of F_μ by analyzing the dynamics of F_μ on Λ.

When $\mu > 2 + \sqrt{5}$, we have shown that $|F_\mu'(x)| > 1$ on $I_0 \cup I_1$. This implies that $|F_\mu'(x)| > 1$ on Λ. This is a condition similar to the hyperbolicity condition of §3, except that we require $|F_\mu'(x)| \neq 1$ on a whole set, not just at a periodic point. This motivates the definition of a hyperbolic set:

Definition 5.7. A set $\Gamma \subset \mathbf{R}$ is a repelling (resp. attracting) hyperbolic set for f if Γ is closed, bounded and invariant under f and there exists an $N > 0$ such that $|(f^n)'(x)| > 1$ (resp. < 1) for all $n \geq N$ and all $x \in \Gamma$.

The Cantor set Λ for the quadratic map when $\mu > 2 + \sqrt{5}$ is of course a repelling hyperbolic set with $N = 1$.

Exercises

1. Prove that $F_2(x) = 2x(1-x)$ satisfies: if $0 < x < 1$, then $F_2^n(x) \to 1/2$ as $n \to \infty$.

2. Sketch the graph of $F_4^n(x)$ on the unit interval, where $F_4(x) = 4x(1-x)$. Conclude that F_4 has at least 2^n periodic points of period n.

3. Sketch the graph of the tent map

$$T_2(x) = \begin{cases} 2x & 0 \leq x \leq 1/2 \\ 2 - 2x & \frac{1}{2} \leq x \leq 1 \end{cases}$$

on the unit interval. Use the graph of T_2^n to conclude that T_2 has exactly 2^n periodic points of period n.

4. Prove that the set of all periodic points of $T(x)$ are dense in $[0, 1]$.

5. Sketch the graph of the baker map

$$B(x) = \begin{cases} 2x & 0 \le x < 1/2 \\ 2x - 1 & 1/2 \le x < 1 \end{cases}.$$

How many periodic points of period n does B have?

6. The following exercises deal with the family of functions $F(x) = x^3 - \lambda x$ for $\lambda > 0$.

a. Find all periodic points and classify them when $0 < \lambda < 1$.
b. Prove that, if $|x|$ is sufficiently large, then $|f^n(x)| \to \infty$.
c. Prove that if λ is sufficiently large, then the set of points which do not tend to infinity is a Cantor set.

7. Prove that the Cantor Middle-Thirds set described in Example 5.5 is closed, nonempty, perfect, and totally disconnected.

8. Show that, at the n^{th} stage of the construction of the Cantor Middle-Thirds set, the sum of the lengths of the remaining intervals is

$$1 - \frac{1}{3}\Big(\sum_{i=1}^{n}\Big(\frac{2}{3}\Big)^{i-1}\Big).$$

Conclude that the sum of the lengths of these intervals tends to 0 as $n \to \infty$.

9. Construct a Middle-Fifths Cantor set in which the middle fifth of each remaining subinterval of the unit interval is removed. What can be said about the sum of the lengths of the remaining intervals in this case?

10. Let Γ be the Cantor Middle-Thirds set. Prove that the linear map $L(x) = 3x$ maps $\Gamma \cap [0, \frac{1}{3}]$ homeomorphically onto Γ.

11. Generalize Exercise 10 to show that the portion of Γ contained in an interval remaining at the n^{th} stage of the construction of Γ is homeomorphic to Γ.

§1.6 SYMBOLIC DYNAMICS

Our goal in this section is to give a model for the rich dynamical structure of the quadratic map on the Cantor set Λ discussed in the previous section. To do this we will set up a model mapping which is completely equivalent to F. At first, this model may seem artificial and unintuitive. But, as we go along, it will become clear that such symbolic models describe the dynamics of F completely and also in the simplest possible way.

We need a "space" on which our model map will act. The points in this space will be infinite sequences of 0's and 1's. We don't worry about convergence of these sequences; rather, the difficult notion here is to imagine such an infinite sequence as representing a single "point" in space.

Definition 6.1. $\Sigma_2 = \{\mathbf{s} = (s_0 s_1 s_2 \ldots) | s_j = 0 \text{ or } 1\}$.

Σ_2 is called the *sequence space* on the two symbols 0 and 1. More generally, we can consider the space Σ_n consisting of infinite sequences of integers between 0 and $n-1$. Elements of Σ_2 are infinite strings of integers, like (000...) or (0101...). We may make Σ_2 into a metric space as follows. For two sequences $\mathbf{s} = (s_0 s_1 s_2 \ldots)$ and $\mathbf{t} = (t_0 t_1 t_2 \ldots)$, define the distance between them by

$$d[\mathbf{s}, \mathbf{t}] = \sum_{i=0}^{\infty} \frac{|s_i - t_i|}{2^i}$$

Since $|s_i - t_i|$ is either 0 or 1, this infinite series is dominated by the geometric series

$$\sum_{i=0}^{\infty} \frac{1}{2^i} = 2$$

and therefore it converges.

For example, if $\mathbf{s} = (000...)$ and $\mathbf{t} = (111...)$, then $d[\mathbf{s}, \mathbf{t}] = 2$. If $\mathbf{r} = (1010...)$, then

$$d[\mathbf{s}, \mathbf{r}] = \sum_{i=0}^{\infty} \frac{1}{2^{2i}} = \frac{1}{1 - \frac{1}{4}} = \frac{4}{3}.$$

Proposition 6.2. *d is a metric on Σ_2.*

Proof. Clearly, $d[\mathbf{s}, \mathbf{t}] \geq 0$ for any $\mathbf{s}, \mathbf{t} \in \Sigma_2$, and $d[\mathbf{s}, \mathbf{t}] = 0$ iff $s_i = t_i$ for all i. Since $|s_i - t_i| = |t_i - s_i|$, it follows that $d[\mathbf{s}, \mathbf{t}] = d[\mathbf{t}, \mathbf{s}]$. Finally, if $\mathbf{r}, \mathbf{s}$, and $\mathbf{t} \in \Sigma_2$, then $|r_i - s_i| + |s_i - t_i| \geq |r_i - t_i|$ from which we deduce that $d[\mathbf{r}, \mathbf{s}] + d[\mathbf{s}, \mathbf{t}] \geq d[\mathbf{r}, \mathbf{t}]$.

q.e.d.

The metric d allows us to decide which subsets of Σ_2 are open and which are closed, as well as which sequences are close to each other.

Proposition 6.3. *Let* $\mathbf{s}, \mathbf{t} \in \Sigma_2$ *and suppose* $s_i = t_i$ *for* $i = 0, 1, \ldots, n$. *Then* $d[\mathbf{s}, \mathbf{t}] \leq 1/2^n$. *Conversely, if* $d[\mathbf{s}, \mathbf{t}] < 1/2^n$, *then* $s_i = t_i$ *for* $i \leq n$.

Proof. If $s_i = t_i$ for $i \leq n$, then

$$\begin{aligned} d[\mathbf{s}, \mathbf{t}] &= \sum_{i=0}^{n} \frac{|s_i - s_i|}{2^i} + \sum_{i=n+1}^{\infty} \frac{|s_i - t_i|}{2^i} \\ &\leq \sum_{i=n+1}^{\infty} \frac{1}{2^i} = \frac{1}{2^n}. \end{aligned}$$

On the other hand, if $s_j \neq t_j$ for some $j \leq n$, then we must have

$$d[\mathbf{s}, \mathbf{t}] \geq \frac{1}{2^j} \geq \frac{1}{2^n}$$

consequently, if $d[\mathbf{s}, \mathbf{t}] < 1/2^n$, then $s_i = t_i$ for $i \leq n$.

q.e.d.

The importance of this result is that we can decide quickly whether or not two sequences are close to each other. Intuitively, this result says that two sequences in Σ_2 are close provided their first few entries agree. We now define the most important ingredient in symbolic dynamics, the shift map on Σ_2.

Definition 6.4. The shift map $\sigma\colon \Sigma_2 \to \Sigma_2$ is given by $\sigma(s_0 s_1 s_2 \ldots) = (s_1 s_2 s_3 \ldots)$.

The shift map simply "forgets" the first entry in a sequence, and shifts all other entries one place to the left. Clearly, σ is a two-to-one map of Σ_2, as s_0 may be either 0 or 1. Moreover, in the metric defined above, σ is a continuous map.

Proposition 6.5. $\sigma\colon \Sigma_2 \to \Sigma_2$ *is continuous.*

Proof. Let $\epsilon > 0$ and $\mathbf{s} = s_0 s_1 s_2 \ldots$. Pick n such that $1/2^n < \epsilon$. Let $\delta = 1/2^{n+1}$. If $\mathbf{t} = t_0 t_1 t_2 \ldots$ satisfies $d[\mathbf{s}, \mathbf{t}] < \delta$, then by Proposition 6.3 we have $s_i = t_i$ for $i \leq n + 1$. Hence the i^{th} entries of $\sigma(\mathbf{s})$ and $\sigma(\mathbf{t})$ agree for $i \leq n$. Therefore $d[\sigma(\mathbf{s}), \sigma(\mathbf{t})] \leq 1/2^n < \epsilon$.

q.e.d.

In the next section, we will show that the shift map is an exact model for the quadratic map F_μ when $\mu > 4$. Here we will simply show that the dynamics of σ can be understood completely. For example, periodic points correspond exactly to repeating sequences, i.e., sequences of the form $\mathbf{s} = (s_0 \dots s_{n-1}, s_0 \dots s_{n-1}, s_0 \dots s_{n-1} \dots)$. Hence there are 2^n periodic points of period n for σ, each generated by one of the 2^n finite sequence of 0's and 1's of length n.

Eventually periodic points are equally abundant and easy to recognize. For example, any sequence of the form $(s_0 \dots s_n 1111 \dots)$ is eventually fixed, while any eventually repeating sequence is eventually periodic for σ.

Another interesting fact about σ is that periodic points form a dense subset of Σ_2. Recall that a subset is dense in Σ_2 provided its closure is the entire space Σ_2. To prove that $\mathrm{Per}(\sigma)$ is dense, we must produce a sequence of periodic points τ_n which converge to an arbitrary point $\mathbf{s} = (s_0 s_1 s_2 \dots)$ in Σ_2. We define the sequence $\tau_n = (s_0 \dots s_n, s_0 \dots s_n, \dots)$, i.e., τ_n is the repeating sequence whose entries agree with $\mathbf{s}$ up to the n^{th} entry. By Proposition 6.3, $d[\tau_n, \mathbf{s}] \leq 1/2^n$, so that we have $\tau_n \to \mathbf{s}$.

Of course, not all points in Σ_2 are periodic or eventually periodic. Any non-repeating sequence can never be periodic. In fact, the non-periodic sequences greatly outnumber the periodic sequences in Σ_2. Moreover, there are non-periodic orbits in Σ_2 which wind densely about Σ_2, i.e., the closure of the orbit is Σ_2 itself. Another way to say this is there are points in Σ_2 whose orbit comes arbitrarily close to any given sequence in Σ_2. To see this, consider

$$\mathbf{s}^* = (\underbrace{0\ 1}_{1 blocks} \,|\, \underbrace{00\ 01\ 10\ 11}_{2 blocks} \,|\, \underbrace{000\ 001 \cdots}_{3 blocks} \,|\, \underbrace{\cdots}_{4 blocks}).$$

$\mathbf{s}^*$ is constructed by successively listing all blocks of 0's and 1's of length n, then length $n+1$, etc. Clearly, some iterate of σ applied to $\mathbf{s}^*$ yields a sequence which agrees with any given sequence in an arbitrarily large number of places. Mappings which have dense orbits are called *topologically transitive*.

Let us list these properties of σ:

Proposition 6.6.

1. *The cardinality of* $\mathrm{Per}_n(\sigma)$ *is* 2^n.
2. $\mathrm{Per}(\sigma)$ *is dense in* Σ_2.
3. *There exists a dense orbit for* σ *in* Σ_2.

In the next section, we will show that the shift map on Σ_2 is in fact the "same" map as f on Λ.

Symbolic dynamics is one of the main themes of this book. It will appear in various guises throughout, including later in this chapter when we introduce subshifts of finite type and also the kneading theory to describe the dynamics of F_μ when $\mu < 4$.

Exercises

1. Let

$$\mathbf{s} = (001\,001\,001\ldots)$$
$$\mathbf{t} = (01\,01\,01\ldots)$$
$$\mathbf{r} = (10\,10\,10\ldots).$$

Compute:
   a. $d[\mathbf{s}, \mathbf{t}]$
   b. $d[\mathbf{t}, \mathbf{r}]$
   c. $d[\mathbf{s}, \mathbf{r}]$.

2. Identify all sequences in Σ_2 which are periodic points of period 3 for σ. Which sequences lie on the same orbit under σ?

3. Rework Exercise 2 for periods four and five.

4. Let Σ' consist of all sequences in Σ_2 satisfying: if $s_j = 0$ then $s_{j+1} = 1$. In other words, Σ' consists of only those sequences in Σ_2 which never have two consecutive zeros.
   a. Show that σ preserves Σ' and that Σ' is a closed subset of Σ.
   b. Show that periodic points of σ are dense in Σ'.
   c. Show that there is a dense orbit in Σ'.
   d. How many fixed points are there for $\sigma, \sigma^2, \sigma^3$ in Σ' ?
   e. Find a recursive formula for the number of fixed points of σ^n in terms of the number of fixed points of σ^{n-1} and σ^{n-2}.

5. Let Σ_N consist of all sequences of natural numbers $1, 2, \ldots, N$. There is a natural shift on Σ_N.
   a. How many periodic points does σ have in Σ_N ?
   b. Show that σ has a dense orbit in Σ_N.

6. Let $\mathbf{s} \in \Sigma_2$. Define the stable set of $\mathbf{s}$, $W^s(\mathbf{s})$, to be the set of sequences $\mathbf{t}$ such that $d[\sigma^i(\mathbf{s}), \sigma^i(\mathbf{t})] \to 0$ as $i \to \infty$. Identify all of the sequences in $W^s(\mathbf{s})$.

§1.7 TOPOLOGICAL CONJUGACY

The goal of this section is to relate the shift map discussed in the previous section to the quadratic map $F_\mu(x) = \mu x(1-x)$ when μ is sufficiently large. Recall that all points in $\mathbf{R}$ tend to $-\infty$ under iteration of F_μ with the exception of those points in the Cantor set Λ. In order to complete the description of the dynamics of F_μ, we must then understand the restriction of F_μ to Λ.

Recall first that $\Lambda \subset I_0 \cup I_1$. If $x \in \Lambda$, then all points on the orbit of x lie in Λ and hence in one of these two intervals. We can thus get a rough idea of the behavior of the orbit by noting in which of these intervals the various iterates of x fall. Accordingly, we make the following definition.

Definition 7.1. The *itinerary of x* is a sequence $S(x) = s_0 s_1 s_2 \ldots$ where $s_j = 0$ if $F_\mu^j(x) \in I_0$, $s_j = 1$ if $F_\mu^j(x) \in I_1$.

Thus the itinerary of x is an infinite sequence of 0's and 1's. That is, $S(x)$ is a point in the sequence space Σ_2. We think of S as a map from Λ to Σ_2. This map has several interesting properties.

Theorem 7.2. *If $\mu > 2 + \sqrt{5}$, then $S : \Lambda \to \Sigma_2$ is a homeomorphism.*

Proof. We first show that S is one-to-one. Let $x, y \in \Lambda$ and suppose $S(x) = S(y)$. Then, for each n, $F_\mu^n(x)$ and $F_\mu^n(y)$ lie on the same side of $1/2$. This implies that F_μ is monotonic on the interval between $F_\mu^n(x)$ and $F_\mu^n(y)$. Consequently, all points in this interval remain in $I_0 \cup I_1$. This contradicts the fact that Λ is totally disconnected.

To see that S is onto, we first introduce the following notation. Let $J \subset I$ be a closed interval. Let

$$F_\mu^{-n}(J) = \{x \in I | F_\mu^n(x) \in J\}.$$

In particular, $F_\mu^{-1}(J)$ denotes the preimage of J. Observe that, if $J \subset I$ is a closed interval, then $F_\mu^{-1}(J)$ consists of two subintervals, one in I_0 and one in I_1. See Fig. 7.1.

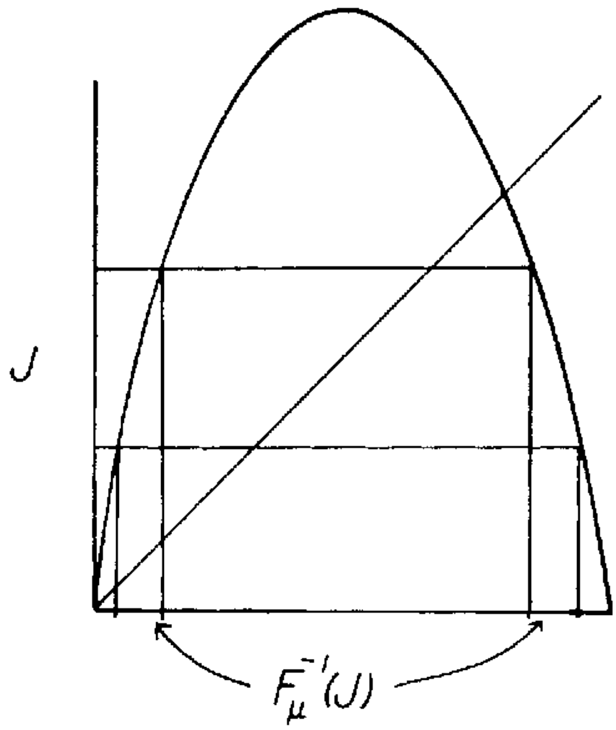

Fig. 7.1. The preimage of a closed interval J is a pair of closed intervals, one in I_0 and one in I_1.

Now let $\mathbf{s} = s_0s_1s_2\ldots$. We must produce $x \in \Lambda$ with $S(x) = \mathbf{s}$. To that end we define

$$\begin{aligned} I_{s_0s_1\ldots s_n} &= \{x \in I | x \in I_{s_0},\ F_\mu(x) \in I_{s_1}, \ldots, F_\mu^n(x) \in I_{s_n}\} \\ &= I_{s_0} \cap F_\mu^{-1}(I_{s_1}) \cap \ldots \cap F_\mu^{-n}(I_{s_n}). \end{aligned}$$

We claim that the $I_{s_0\ldots s_n}$ form a nested sequence of nonempty closed intervals as $n \to \infty$. Note that

$$I_{s_0s_1\ldots s_n} = I_{s_0} \cap F_\mu^{-1}(I_{s_1\ldots s_n}).$$

By induction, we may assume that $I_{s_1\ldots s_n}$ is a nonempty subinterval, so that, by the observation above, $F_\mu^{-1}(I_{s_1\ldots s_n})$ consists of two closed intervals, one in I_0 and one in I_1. Hence $I_{s_0} \cap F_\mu^{-1}(I_{s_1\ldots s_n})$ is a single closed interval.

These intervals are nested because

$$I_{s_0\ldots s_n} = I_{s_0\ldots s_{n-1}} \cap F_\mu^{-n}(I_{s_n}) \subset I_{s_0\ldots s_{n-1}}\ .$$

Therefore we conclude that

$$\bigcap_{n\geq 0} I_{s_0s_1\ldots s_n}$$

is nonempty. Note that if $x \in \cap_{n\geq 0} I_{s_0s_1\ldots s_n}$, then $x \in I_{s_0}$, $F_\mu(x) \in I_{s_1}$, etc. Hence $S(x) = (s_0s_1\ldots)$. This proves that S is onto.

Note that $\cap_{n\geq 0} I_{s_0s_1\ldots s_n}$ consists of a unique point. This follows immediately from the fact that S is one-to-one. In particular, we have that diam $I_{s_0s_1\ldots s_n} \to 0$ as $n \to \infty$.

Finally, to prove continuity of S, we choose $x \in \Lambda$ and suppose that $S(x) = s_0s_1s_2\ldots$. Let $\epsilon > 0$. Pick n so that $1/2^n < \epsilon$. Consider the closed subintervals $I_{t_0t_1\ldots t_n}$ defined above for all possible combinations $t_0t_1\ldots t_n$. These subintervals are all disjoint, and Λ is contained in their union. There are 2^{n+1} such subintervals, and $I_{s_0s_1\ldots s_n}$ is one of them. Hence we may choose δ such that $|x-y| < \delta$ and $y \in \Lambda$ implies that $y \in I_{s_0s_1\ldots s_n}$. Therefore, $S(y)$ agrees with $S(x)$ in the first $n+1$ terms. Hence, by Proposition 6.3,

$$d[S(x), S(y)] < \frac{1}{2^n} < \epsilon.$$

This proves the continuity of S. It is easy to check that S^{-1} is also continuous. Thus, S is a homeomorphism.

q.e.d.

This theorem shows that, as sets, Λ and Σ_2 are the same. More importantly, the coding S also gives an equivalence between the dynamics of F_μ on Λ and σ on Σ_2. This is the content of the following theorem.

Theorem 7.3. $S \circ F_\mu = \sigma \circ S$.

Proof. A point x in Λ may be defined uniquely by the nested sequence of intervals

$$\bigcap_{n\geq 0} I_{s_0s_1\ldots s_n\ldots}$$

determined by the itinerary $S(x)$. Now

$$I_{s_0\ldots s_n} = I_{s_0} \cap F_\mu^{-1}(I_{s_1}) \cap \ldots \cap F_\mu^{-n}(I_{s_n})$$

so that $F_\mu(I_{s_0\ldots s_n})$ may be written

$$I_{s_1} \cap F_\mu^{-1}(I_{s_2}) \cap \ldots F_\mu^{-n+1}(I_{s_n}) = I_{s_1\ldots s_n} \ ,$$

since $F_\mu(I_{s_0}) = I$. Hence

$$\begin{aligned} SF_\mu(x) &= SF_\mu\Big(\cap_{n=0}^{\infty} I_{s_0s_1\ldots s_n}\Big) \\ &= S\Big(\cap_{n=1}^{\infty} I_{s_1\ldots s_n}\Big) \\ &= s_1s_2\ldots = \sigma S(x). \end{aligned}$$

q.e.d.

Definition 7.4. Let $f: A \to A$ and $g: B \to B$ be two maps. f and g are said to be topologically conjugate if there exists a homeomorphism $h: A \to B$ such that, $h \circ f = g \circ h$. The homeomorphism h is called a topological conjugacy.

Mappings which are topologically conjugate are completely equivalent in terms of their dynamics. For example, if f is topologically conjugate to g via h, and p is a fixed point for f, then $h(p)$ is fixed for g. Indeed, $h(p) = hf(p) = gh(p)$. Similarly, h gives a one-to-one correspondence between $\mathrm{Per}_n(f)$ and $\mathrm{Per}_n(g)$. One may also check that eventually periodic and asymptotic orbits for f go over via h to similar orbits for g, and that f is topologically transitive if and only if g is. In particular, since F_μ on Λ is topologically conjugate to the shift, we have now proved that the quadratic map enjoys the striking properties we uncovered so easily for σ in the last section. These may be summarized as follows.

Theorem 7.5. *Let $F_\mu(x) = \mu x(1-x)$ with $\mu > 2 + \sqrt{5}$. Then*
1. *The cardinality of $\mathrm{Per}_n(F_\mu)$ is 2^n.*
2. *$\mathrm{Per}(F_\mu)$ is dense in Λ.*
3. *F_μ has a dense orbit in Λ.*

This Theorem shows the power of symbolic dynamics and topological conjugacy. Actually computing the 2^n periodic points of period n for F_μ is a hopeless task. But topological conjugacy guarantees that these orbits are there, and, moreover, symbolic dynamics gives a rough measure of the complexity of the orbits in Λ. Thus these two notions provide justification for our statement that the shift map is an accurate model for the quadratic map.

Exercises

1. Let $Q_c(x) = x^2 + c$. Prove that if $c < 1/4$, there is a unique $\mu > 1$ such that Q_c is topologically conjugate to $F_\mu(x) = \mu x(1-x)$ via a map of the form $h(x) = \alpha x + \beta$.

2. A point p is a *non-wandering* point for f, if, for any open interval J containing p, there exists $x \in J$ and $n > 0$ such that $f^n(x) \in J$. Note that we do not require that p itself return to J. Let $\Omega(f)$ denote the set of non-wandering points for f.

a. Prove that $\Omega(f)$ is a closed set.
b. If F_μ is the quadratic map with $\mu > 2 + \sqrt{5}$, show that $\Omega(F_\mu) = \Lambda$.
c. Identify $\Omega(F_\mu)$ for each μ satisfying $0 < \mu \leq 3$.

3. A point p is *recurrent* for f if, for any open interval J about p, there exists $n > 0$ such that $f^n(p) \in J$. Clearly, all periodic points are recurrent.

a. Give an example of a non-periodic recurrent point for F_μ when $\mu > 2 + \sqrt{5}$.
b. Give an example of a non-wandering point for F_μ which is not recurrent.

4. *Order of the periodic points.* Let Γ_n denote the set of repeating sequences of period n in Σ_2. Identify such a sequence with a finite string $s_1, \ldots, s_n$ of 0's and 1's in the natural way. Under the topological conjugacy, each element of Γ_n corresponds to a unique point in I for a given value of $\mu > 2 + \sqrt{5}$.

a. Prove that the order of these points in I is independent of $\mu > 2+\sqrt{5}$. Let $N(s_1, \ldots, s_n)$ denote the integer between 0 and $2^n - 1$ corresponding to this order, numbering from left to right, i.e., $N(0, \ldots, 0) = 0$. Let $B(s_1, \ldots, s_n)$ denote N in binary form. That is, $B(s_1, \ldots, s_n) = (a_1, \ldots, a_n)$ where $a_j = 0$ or 1 and

$$N(s_1, \ldots, s_n) = a_1 \cdot 2^{n-1} + a_2 \cdot 2^{n-2} + \ldots + a_n \cdot 2^0.$$

b. Use induction to prove that B is given by the following formula:

$$a_j = \sum_{i=1}^{j} s_j \bmod 2.$$

For example, the fixed point $1, 1, 1 \in \Gamma_3$ occupies position 5 on the real line since

$$a_1 = s_1 = 1$$
$$a_2 = s_1 + s_2 = 0 \bmod 2$$
$$a_3 = s_1 + s_2 + s_3 = 1 \bmod 2.$$

c. List all points in Γ_n for $n = 2, 3, 4$ according to this ordering.
d. Describe an algorithm for ordering the points in Γ_n knowing the ordering of Γ_{n-1}.

§1.8 CHAOS

The quadratic map exhibits in stunning fashion a phenomenon which is only partially understood: the chaotic behavior of orbits of a dynamical system. There are many possible definitions of chaos, ranging from measure theoretic notions of randomness in ergodic theory to the topological approach we will adopt here.

Definition 8.1. $f: J \to J$ is said to be topologically transitive if for any pair of open sets $U, V \subset J$ there exists $k > 0$ such that $f^k(U) \cap V \neq \emptyset$.

Intuitively, a topologically transitive map has points which eventually move under iteration from one arbitrarily small neighborhood to any other. Consequently, the dynamical system cannot be decomposed into two disjoint open sets which are invariant under the map. Note that if a map possesses a dense orbit, then it is clearly topologically transitive. The converse is also true (for compact subsets of $\mathbf{R}$ or S^1), but we will not prove it here since the proof depends on the Baire Category Theorem.

Definition 8.2. $f: J \to J$ has sensitive dependence on initial conditions if there exists $\delta > 0$ such that, for any $x \in J$ and any neighborhood N of x, there exists $y \in N$ and $n \geq 0$ such that $|f^n(x) - f^n(y)| > \delta$.

Intuitively, a map possesses sensitive dependence on initial conditions if there exist points arbitrarily close to x which eventually separate from x by at least δ under iteration of f. We emphasize that not *all* points near x need eventually separate from x under iteration, but there must be at least one such point in every neighborhood of x. If a map possesses sensitive dependence on initial conditions, then for all practical purposes, the dynamics of the map defy numerical computation. Small errors in computation which are introduced by round-off may become magnified upon iteration. The results of numerical computation of an orbit, no matter how accurate, may bear no resemblance whatsoever with the real orbit.

Example 8.3. The quadratic map $\mu x(1-x)$ with $\mu > 2 + \sqrt{5}$ possesses sensitive dependence on initial conditions on Λ. To see this, choose δ less than the diameter of A_0, where A_0 is the gap between I_0 and I_1. Let $x, y \in \Lambda$. If $x \neq y$, then $S(x) \neq S(y)$, so the itineraries of x and y must differ in at least one spot, say the n^{th}. But this means that $F_\mu^n(x)$ and $F_\mu^n(y)$ lie on opposite sides of A_0, so that

$$|F_\mu^n(x) - F_\mu^n(y)| > \delta.$$

Example 8.4. An irrational rotation of the circle is topologically transitive but not sensitive to initial conditions, since all points remain the same distance apart after iteration.

We turn now to one of the main themes of this book, the notion of a chaotic dynamical system. There are many possible definitions of chaos in a dynamical system, some stronger and some weaker than ours. We choose this particular definition because it applies to a large number of important examples and because, in many cases, it is easy to verify.

Definition 8.5. Let V be a set. $f\colon V \to V$ is said to be chaotic on V if

1. f has sensitive dependence on initial conditions.
2. f is topologically transitive.
3. periodic points are dense in V.

To summarize, a chaotic map possesses three ingredients: unpredictability, indecomposability, and an element of regularity. A chaotic system is unpredictable because of the sensitive dependence on initial conditions. It cannot be broken down or decomposed into two subsystems (two invariant open subsets) which do not interact under f because of topological transitivity. And, in the midst of this random behavior, we nevertheless have an element of regularity, namely the periodic points which are dense.

Example 8.6. $f\colon S^1 \to S^1$ given by $f(\theta) = 2\theta$ is chaotic. As we have seen, the angular distance between two points is doubled upon iteration of f. Hence f is sensitive to initial conditions. Topological transitivity also follows from this observation since any small arc in S^1 is eventually expanded by some f^k to cover all of S^1 and, in particular, any other arc in S^1. The density of periodic points was established in §1.3. We remark that this map possesses a strong form of sensitive dependence called expansiveness.

Definition 8.7. $f\colon J \to J$ is expansive if there exists $\nu > 0$ such that, for any $x, y \in J$, $x \neq y$, there exists n such that $|f^n(x) - f^n(y)| > \nu$.

Expansiveness differs from sensitive dependence in that *all* nearby points eventually separate by at least ν.

Example 8.8. The quadratic maps $F_\mu(x) = \mu x(1 - x)$ are chaotic on Λ when $\mu > 2 + \sqrt{5}$.

This example differs markedly from the previous example in that the chaos is confined to a small subset of I, namely the Cantor set Λ. A much larger chaotic region for a quadratic map is given by the following example.

Example 8.9. $F_4(x) = 4x(1-x)$ is chaotic on the interval $I = [0,1]$.

Proof. Let $g(\theta) = 2\theta$ be the map on S^1 discussed in Example 8.6. Define $h_1 \colon S^1 \to [-1,1]$ by $h_1(\theta) = \cos\theta$. That is, h_1 is just projection from S^1 to the x-axis. Let $q(x) = 2x^2 - 1$. Then we have

$$\begin{aligned} h_1 \circ g(\theta) &= \cos(2\theta) \\ &= 2\cos^2\theta - 1 \\ &= q \circ h_1(\theta) \end{aligned}$$

so that h_1 conjugates g with q. Now q is also topologically conjugate to F_4. Indeed, if $h_2(t) = \frac{1}{2}(1-t)$, then we have $F_4 \circ h_2 = h_2 \circ q$. Hence we have the following diagram

$$\begin{array}{ccc} S^1 & \xrightarrow{g} & S^1 \\ h_1 \Big\downarrow & & \Big\downarrow h_1 \\ [-1,1] & \xrightarrow{q} & [-1,1] \\ h_2 \Big\downarrow & & \Big\downarrow h_2 \\ [0,1] & \xrightarrow{F_4} & [0,1]. \end{array}$$

It follows immediately that F_4 is topologically transitive, for if U and V are two open intervals in I, we may choose open arcs $\hat{U}$ and $\hat{V}$ in S^1 which project onto U and V under $h_2 \circ h_1$. Since there exists k such that $g^k(\hat{U}) \cap \hat{V} \neq \emptyset$, we therefore have $F_4^k(U) \cap V \neq \emptyset$.

To prove sensitive dependence, we note that any neighborhood U of $x \in I$ "lifts" to $\hat{U}$ in S^1. There exists n such that $g^n(\hat{U})$ covers S^1, so $F_4^n(U)$ covers I as well. Hence there are points in U which move at least $\delta = 1/2$ away from x. Finally, density of periodic points for g implies that there is a g-periodic point in $\hat{U}$. The projection of this point in U is clearly F_4-periodic.

The technique introduced in this Example can also be used to produce other examples of maps which are chaotic on an interval. The so-called Tchebycheff polynomials are important classical examples which feature this type of behavior. See Exercises 1-3.

We remark that the map h_1 above is not a homeomorphism since it is a two-to-one at most points. Thus we have *not* shown that $g(\theta) = 2\theta$ on S^1

and F_4 are topologically conjugate. Rather, we say that these two maps are *semi-conjugate*.

Exercises

1. Use the method of Example 8.9 to prove that $F(x) = 4x^3 - 3x$ is chaotic on the interval $[-1, 1]$. (Hint: consider $g(\theta) = 3\theta$ on S^1.)

2. Prove that $F(x) = 8x^4 - 8x^2 + 1$ is chaotic on $[-1, 1]$.

3. The polynomial which is given as in Exercises 1 and 2 by projection of $g(\theta) = \cos n\theta$ onto the interval $[-1, 1]$ is called the n^{th} Tchebycheff polynomial, when properly normalized. Show that these polynomials satisfy the differential equation

$$(1 - x^2)y'' - xy' + \alpha^2 y = 0$$

where α is a positive integer.

4. Prove that $T(x) = \tan x$ is chaotic on the entire real line, despite the fact that there are a dense set of points at which an iterate of T fails to be defined.

5. Prove that the baker map

$$B(x) = \begin{cases} 2x & 0 \le x \le \frac{1}{2} \\ 2x - 1 & \frac{1}{2} \le x \le 1 \end{cases}$$

is chaotic on $[0, 1]$.

The following exercises apply to the tent map

$$T_2(x) = \begin{cases} 2x & 0 \le x \le 1/2 \\ 2(1 - x) & 1/2 \le x \le 1 \end{cases}.$$

Note that the maximum of T_2 is 1 and occurs at $x = \frac{1}{2}$. To describe the dynamics of T_2 via symbolic dynamics, we thus need to modify Σ_2 somewhat since there is an ambiguity in the sequence associated to any rational number of the form $p/2^k$ where p is an integer. For example, $1/2$ may be described by either (11000...) or (01000...). To remedy this, we identify any two sequences of the form $(s_0 \ldots s_{k-1} {*} 1000\ldots)$, where $* = 0$ or 1. For example, the sequences (1101000...) and (1111000...) are to be thought of as representing the same point. Let Σ_2' denote Σ_2 with these identifications.

6. Prove that $S\colon I \to \Sigma_2'$ is one-to-one, where $S(x)$ is defined as in §1.7.

7. Prove that $\sigma \circ S = S \circ T_2$.

8. Prove that T_2 has exactly 2^n periodic points of period n.

9. Prove that T_2 is chaotic on I.

10. Prove that T_2 is topologically conjugate to the quadratic map $F_4(x) = 4x(1-x)$.

11. Construct a piecewise linear map on $[0,1]$ which is topologically conjugate to $F(x) = 4x^3 - 3x$ on $[-1,1]$.

§1.9 STRUCTURAL STABILITY

A very important notion in the study of dynamical systems is the stability or persistence of the system under small changes or perturbations. This is the concept of structural stability which we introduce in this section. Briefly, a map f is structurally stable if every "nearby" map is topologically conjugate to f and so has essentially the same dynamics. Clearly, we need to be precise about what nearby means, but the basic idea is simple. If, no matter how we perturb f or change f slightly, we get an equivalent dynamical system, then the dynamical structure of f is stable. Here, equivalent means topologically conjugate. If f and g are topologically conjugate, we will write $f \sim g$.

The notion of structural stability is extremely important in applications. Suppose our dynamical system is the solution of a differential equation or otherwise comes from a real world physical system. Ordinarily, the system itself will be only a model of real world phenomena: certain assumptions will have been made, and certain approximations and experimental errors will be present. Hence the dynamical system itself, albeit a completely accurate solution of the physical model, will nevertheless be only an approximation to reality since the model itself suffers this flaw. Now, if the dynamical system in question is not structurally stable, then the small errors and approximations made in the model have a chance of dramatically changing the structure of the real solution to the system. That is, our "solution" could be radically wrong or unstable. If, on the other hand, the dynamical system in question is structurally stable, then the small errors introduced by approximations and experimental errors may not matter at all: the solution to the model system may be equivalent or topologically conjugate to the actual solution.

This does not mean that the only interesting physical systems are the structurally stable ones. Indeed, most dynamical systems that arise in classical mechanics are not structurally stable. There are also simple examples of systems such as the Lorenz system from meteorology that are "far" from being structurally stable. These systems cannot even be approximated in a sense to be made precise below by stable systems. Nevertheless, the concept of structural stability is an important one in applications of the theory of dynamical systems.

To begin the discussion of structural stability, we need to make precise the notion of "nearness" of two functions.

Definition 9.1. Let f and g be two maps. The C^0-distance between f and g, written $d_0(f,g)$, is given by

$$d_0(f,g) = \sup_{x \in \mathbf{R}} |f(x) - g(x)|.$$

The C^r-distance $d_r(f,g)$ is given by

$$d_r(f,g) = \sup_{x \in \mathbf{R}} (|f(x) - g(x)|,\ |f'(x) - g'(x)|, \ldots, |f^{(r)}(x) - g^{(r)}(x)|).$$

Intuitively, two maps are C^r-close provided they as well as their first r derivatives differ by only a small amount. We may also consider the C^r-distance between two maps on an interval $J \subset \mathbf{R}$ by suitably restricting x and y. We caution the reader that d_r does not give a useful metric on the set of all functions. Indeed, since the real line is unbounded, two maps can easily be infinitely far apart. Moreover, even if we assume this difficulty away, the resulting topology on the set of functions is nasty. Hence we will use the C^r-distance only as a measure of the proximity of two functions and not as a global metric on all maps.

Example 9.2. $f(x) = 2x$ and $g(x) = (2+\epsilon)x$ have C^0-distance infinity. But $f(x) = 2x$ and $g(x) = 2x + \epsilon$ are C^r-ϵ apart for all r. Let $J = [0, 10]$. Then $f(x) = 2x$ and $g(x) = (2+\epsilon)x$ are C^0-10ϵ apart (and, in fact, C^r-10ϵ apart) on the interval J.

We will be primarily concerned with functions that are C^1-close or, at most, C^2-close. Fig. 9.1 illustrates the difference graphically between C^0-close, C^1-close, and C^2-close.

We now define C^r-structural stability.

Definition 9.3. Let $f\colon J \to J$. f is said to be C^r-structurally stable on J, if there exists $\epsilon > 0$ such that whenever $d_r(f,g) < \epsilon$ for $g\colon J \to J$, it follows that f is topologically conjugate to g.

Example 9.4. Let $L(x) = \frac{1}{2}x$. Then L is C^1-structurally stable on $\mathbf{R}$. To see this, we must exhibit an $\epsilon > 0$ such that, if $d_1(L,g) < \epsilon$, then L and g are topologically conjugate. We claim that any $\epsilon < 1/2$ works. For if

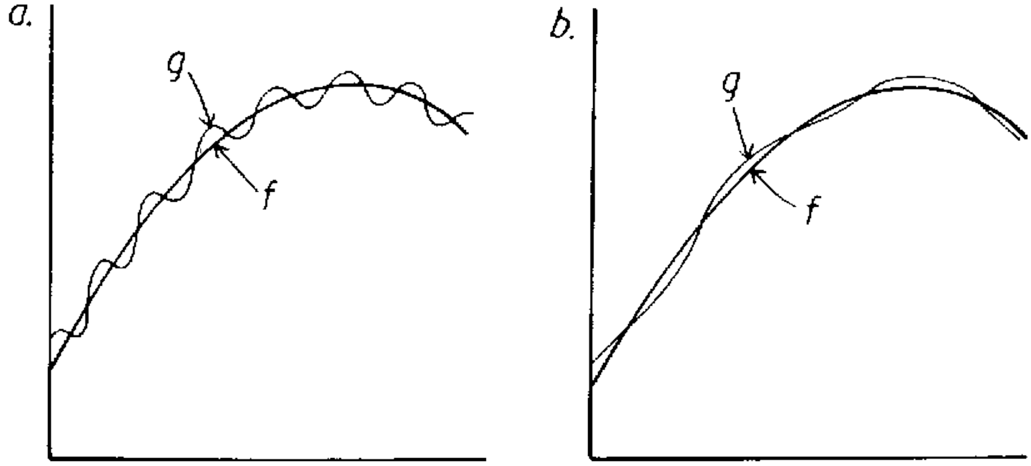

Fig. 9.1. In Fig. a, f and g are C^0-close but not C^1-close. In Fig. b, f and g are C^1-close but not C^2-close.

$d_1(L, g) < \epsilon$, then we must have $0 < g'(x) < 1$ for all $x \in \mathbf{R}$. In particular, $g(x)$ is everywhere increasing. Note also that $g(x)$ has a unique attracting fixed point p in $\mathbf{R}$ and that all points in $\mathbf{R}$ tend to p under iteration. That g has a unique fixed point follows from the Mean Value Theorem: between any two fixed points of g must lie a point with derivative $= 1$, which cannot happen. Alternatively, since $|g'(x)| < 1$, g is a global contraction.

This shows that L and g have the same dynamics, the basic idea behind structural stability. To be strictly precise, however, we must exhibit a topological conjugacy between L and g. To do this, we introduce the notion of a *fundamental domain*. This is best done by example. Consider the pair of intervals $5 < |x| \leq 10$. Note that the L-orbit of any point in $\mathbf{R}$ (with the exception of 0) enters this set exactly once. For g, we may find a similar fundamental domain: indeed, it is easy to check that the intervals $g(10) < x \leq 10$ and $-10 \leq x < g(-10)$ have the same property (Exercise 1.)

We now construct a conjugacy h such that $h \circ L = g \circ h$. First define $h\colon [5, 10] \to [g(10), 10]$ and $h\colon [-10, -5] \to [-10, g(-10)]$ to be linear, i.e., with a straight line graph. We require that h be increasing so that $h(\pm 10) = \pm 10$. (We remark that any other increasing homeomorphism works just as well.) We complete the definition of h as follows. Let $x \neq 0$. There is an $n \in \mathbf{Z}$ such that $L^n(x)$ belongs to the fundamental domain for L. Hence $h \circ L^n(x)$ is well-defined. We then set $h(x) = g^{-n} \circ h \circ L^n(x)$. Note that $h(x)$ is also well-defined, since g is a homeomorphism and so g^{-n} makes sense. Clearly, we have $g^n \circ h(x) = h \circ L^n(x)$. Moreover, if we apply the same construction to $L(x)$, we find that $g \circ h(x) = h \circ L(x)$, as required. Finally, define $h(0) =$ fixed point of g. It is easy to check that h as defined is a homeomorphism.

Intuitively, a fundamental domain is visited exactly once by each orbit, except, of course, the fixed point. Hence we may define a conjugacy in virtually any way we please on the fundamental domain, and then extend in the only way possible by iterating the map. The only question is then whether or not we can extend the conjugacy to points whose orbits never enter the fundamental domain.

We now return to the quadratic map $F_\mu(x) = \mu x(1-x)$. As we saw in §1.5, all points tend to $-\infty$ for this map with the exception of those in a set Λ on which F_μ is topologically conjugate to the shift. We claim that, if μ is large enough, then F_μ is C^2-structurally stable. This may be proved by another fundamental domain argument.

This is more complicated, but let us sketch the details. We first assume that $\mu > 2+\sqrt{5}$, so that $|F'_\mu(x)| > 1$ on $I_0 \cup I_1$. We will produce an $\epsilon > 0$ such that if g is C^2-ϵ close to F_μ, then g has the same dynamics as F_μ. Let us first choose ϵ_1 small enough so that if g is C^2- ϵ_1 close to F_μ, then $g'' < 0$, i.e., so that the graph of g is concave down. This is clearly possible since $F''_\mu \equiv -2\mu$. Next choose $\epsilon_2 < \epsilon_1$ small enough so that if g is C^1-ϵ_2 close to F_μ, then g has two fixed points, α and β, which satisfy

1. $\alpha < \beta$,
2. $g'(\alpha) > 1$,
3. $g'(\beta) < -1$.

The fact that ϵ_2 may be chosen so that g has at most two fixed points follows from the concavity of the graph of g. The fact that g has at least two fixed points can be guaranteed by making g C^0-close to F_μ. Finally, the conditions on g' at the fixed points are controlled by the C^1-distance of g from F_μ.

Note that g has a unique critical point c and that there exist points α', β' with $g(\alpha') = \alpha$, $g(\beta') = \beta$. The points α and α' play the same role as 0 and 1 do for F_μ.

We may finally choose $\epsilon < \epsilon_2$ such that, if g is C^1-ϵ close to F_μ, then $g^{-1}(\alpha')$ consists of a pair of points, a_0 and a_1, and moreover, if $x \in [\alpha, a_0] \cup [a_1, \alpha']$, then $|g'(x)| > 1$. Thus, if g is C^2-ϵ close (and therefore C^0- and C^1 -ϵ close) to F_μ, then the graph of g has all of the qualitative properties of the graph of F_μ on the interval $[\alpha, \alpha']$. See Fig. 9.2.

More importantly, F_μ and g have the same dynamics. It follows immediately that if $x < \alpha$, then $g^n(x) \to -\infty$. Similarly, if $x > \alpha'$ or if $x \in (a_0, a_1)$, then $g^n(x) \to -\infty$ as well. A similar inductive procedure on the inverse images of (a_0, a_1) as in §1.5 shows that all points except those in a Cantor set Λ_g tend to $-\infty$ eventually under iteration of g. On Λ_g, g is again topologically conjugate to the shift automorphism via arguments as in §1.5. We leave the details to the reader.

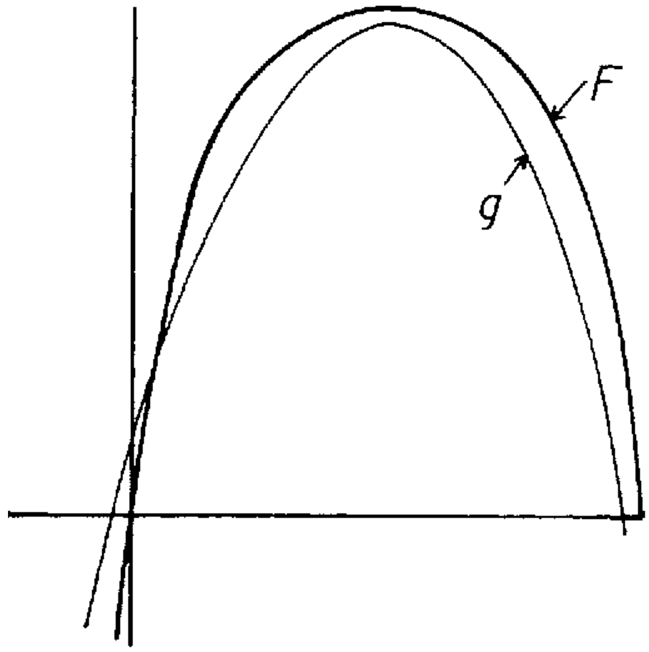

Fig. 9.2. The graphs of F_μ and g are C^2-ϵ close.

To prove that $F_\mu \sim g$, we must construct fundamental domains for both F_μ and g in order to define the conjugacy. This can be accomplished as follows. First choose $x_0 < \min(g^2(c),\ F_\mu^2(c))$. The intervals $(F_\mu(x_0), x_0)$ and $(g(x_0), x_0)$ are easily seen to be fundamental domains for F_μ on $\mathbf{R}^-$ and g on $(-\infty, \alpha)$. The conjugacy may then be defined arbitrarily on $(F_\mu(x_0), x_0)$ and extended by $h \circ F_\mu = g \circ h$ to all of $\mathbf{R}^-$.

We then extend h to the interval $(1, \infty)$ and finally to each A_n in the natural way. We remark that care must be taken on A_0 since F_μ is two-to-one on this interval. Once h is defined on all of $\mathbf{R} - \Lambda$, we extend to Λ in the only way possible to make h a homeomorphism. We leave the tedious details to the reader.

Alternatively, one can use the fact that both g on Λ_g and F_μ on Λ are topologically conjugate to the shift, hence to each other on these sets. The conjugacy may be extended off these sets by the above fundamental domain argument. In either event we have proved

Theorem 9.5. *The quadratic map $F_\mu(x) = \mu x(1-x)$ is C^2 structurally stable if $\mu > 2 + \sqrt{5}$.*

Perhaps more important than the question of when a given map is structurally stable is the converse question: when is it *not* structurally stable? One of the major ways a map can fail to be structurally stable occurs when there is a lack of hyperbolicity.

Example 9.6. Let $F_0(x) = x - x^2$. Note that $F_0(0) = 0$ and $F_0'(0) = 1$, so 0 is a non-hyperbolic fixed point. Consider $F_\epsilon(x) = x - x^2 + \epsilon$. Clearly, $F_\epsilon(x)$

is C^r-ϵ close to F_0. When $\epsilon > 0$, F_ϵ is easily seen to have two fixed points, but when $\epsilon < 0$, F_ϵ has none. Consequently, the F_ϵ do not have the same dynamics as F_0 and therefore F_0 is not structurally stable.

Example 9.7. Let $T_\lambda(x) = x^3 - \lambda x$. For $-1 < \lambda \leq 1$, T_λ has three fixed points: at 0 and at $\pm\sqrt{\lambda+1}$. All points between $\pm\sqrt{\lambda+1}$ tend to the attracting fixed point at 0. When $\lambda > 1$, this is no longer true. There exists x in the interval $[-\sqrt{\lambda+1},\ \sqrt{\lambda+1}]$ such that $T_\lambda(x) = -x$, i.e., the graph of T_λ crosses the line $y = -x$. Since $T_\lambda(-x) = -T_\lambda(x)$, we also have $T_\lambda(-x) = x$, so that x is a periodic point of period 2. Hence the dynamics of T_λ is different for $\lambda \leq 1$ and $\lambda > 1$, so that T_1 is not structurally stable. We remark that T_{-1} is also not structurally stable. See Exercise 2. Note that $T_1'(0) = -1$, so that the fixed point is again non-hyperbolic when structural stability fails to hold.

A hyperbolic fixed point for f is C^1 structurally stable locally. By this we mean there is a neighborhood of the fixed point and an $\epsilon > 0$ such that, if a map g is C^1-ϵ close to f on this neighborhood, then f is topologically conjugate to g on this neighborhood. This fact is established in a series of exercises below. Along the way, we establish the one-dimensional version of Sternberg's Theorem (sometimes called Hartman's Theorem):

Theorem 9.8. *Let p be a hyperbolic fixed point for f and suppose $f'(p) = \lambda$ with $|\lambda| \neq 0, 1$. Then there are neighborhoods U of p and V of $0 \in \mathbf{R}$ and a homeomorphism $h\colon U \to \mathbf{R}$ which conjugates f on U to the linear map $L(x) = \lambda x$ on V.*

Thus a map near a hyperbolic fixed point is always locally topologically conjugate to its derivative. This allows us to explain why we only require that the conjugacy map in the definition of topological conjugacy be a homeomorphism, not a diffeomorphism. Suppose $f(p) = p$ and $f'(p) = \lambda$. Let h be a diffeomorphism. Then $g = h \circ f \circ h^{-1}$ has a fixed point at $h(p)$, but we have

$$
\begin{aligned}
g'(h(p)) &= h'(f(p)) \cdot f'(p) \cdot (h^{-1})'(h(p)) \\
&= h'(p) \cdot \lambda \cdot \frac{1}{h'(p)} \\
&= \lambda.
\end{aligned}
$$

Thus, the multiplier λ at the fixed point is preserved by differentiable conjugacies. As we have seen, maps may behave dynamically the same despite having different multipliers at the fixed points. Thus the weakened notion of topological conjugacy is more appropriate for our purposes.

Exercises

1. Suppose $g(x)$ is as in Example 9.4. Prove that the intervals $g(10) < x \leq 10$ and $-10 \leq x < g(-10)$ form a fundamental domain for g.

2. Let $T_{-1}(x) = x^3 + x$. Prove that T_{-1} is not structurally stable.

3. Let $T_\lambda(x) = x^3 - \lambda(x)$. Prove that T_λ is structurally stable if $-1 < \lambda < 0$.

4. Prove that T_{λ_0} is topologically conjugate to T_{λ_1} if $-1 < \lambda_0, \lambda_1 < 0$.

5. Prove that $F_4(x) = 4x(1-x)$ is not structually stable.

6. Prove that $S(x) = \sin(x)$ is not structurally stable.

7. Prove that, if $f \sim g$ via h and f has a local maximum at x_0, then g has either a local maximum or minimum at $h(x_0)$.

8. Give an example to show that we may have $f \sim g$ via h and x_0 a local maximum for f and $h(x_0)$ a local minimum for g.

9. Let $S_\lambda(x) = \lambda \sin(x)$. If $0 < \lambda_1 < \lambda_2 < 1$, prove that $S_{\lambda_1} \sim S_{\lambda_2}$.

10. Show, however, that neither S_{λ_1} nor S_{λ_2} is structurally stable.

11. We may define a notion of linear structural stability for linear maps by replacing the notion of topological conjugacy by that of linear conjugacy. Two linear maps $T_1, T_2 \colon \mathbf{R} \to \mathbf{R}$ are linearly conjugate if there is a linear map L such that $T_1 \circ L = L \circ T_2$. $T_1(x) = ax$ is linearly stable if there is a neighborhood N about a such that if $b \in \mathbf{N}$, then $T_2(x) = bx$ is linearly conjugate to T_1. Find all linearly stable maps and identify all elements of a given conjugacy class.

12. (Sternberg's Theorem) Let p be a hyperbolic fixed point for f with $f'(p) = \lambda$ and $\lambda \neq 0$. Prove that f is locally topologically conjugate to its derivative map $x \to \lambda x$ as described in Theorem 9.8.

13. Combine Exercises 11 and 12 to prove that any small perturbation of a map near a hyperbolic fixed point is locally topologically conjugate to f.

14. Let $f\colon [0,1] \to [0,1]$ be a diffeomorphism. Prove that, if $f'(x) > 0$, then f has only fixed points and no periodic points. Prove that, if $f'(x) < 0$, then f has a unique fixed point and all other periodic points have period two.

15. A diffeomorphism $f\colon [0,1] \to [0,1]$ is called *Morse-Smale* if f has only hyperbolic periodic points. (Note that, since f is onto, the endpoints of $[0,1]$ are necessarily periodic.) Prove that a Morse-Smale diffeomorphism has only finitely many periodic points.

16. Prove that a Morse-Smale diffeomorphism of $[0, 1]$ is structurally stable.

17. Prove that the map $f(x) = x^3 + \frac{3}{4}x$ is a Morse-Smale diffeomorphism on the interval $[-\frac{1}{2}, \frac{1}{2}]$.

§1.10 SARKOVSKII'S THEOREM

In this section, we will prove a remarkable theorem due to Sarkovskii. The theorem only holds for maps of the real line, but nevertheless is amazing for its lack of hypotheses (f is only assumed continuous) and strong conclusion. We caution the reader that, as this is our first major theorem, the material in this section is a little "heavier" than in previous sections. As a warmup, and also as a means of highlighting the importance of period three points, we will prove a special case.

Theorem 10.1. *Let $f: \mathbf{R} \to \mathbf{R}$ be continuous. Suppose f has a periodic point of period three. Then f has periodic points of all other periods.*

Proof. The proof will depend on two elementary observations. First, if I and J are closed intervals with $I \subset J$ and $f(I) \supset J$, then f has a fixed point in I. This is, of course, a simple consequence of the Intermediate Value Theorem. See Fig. 10.1.

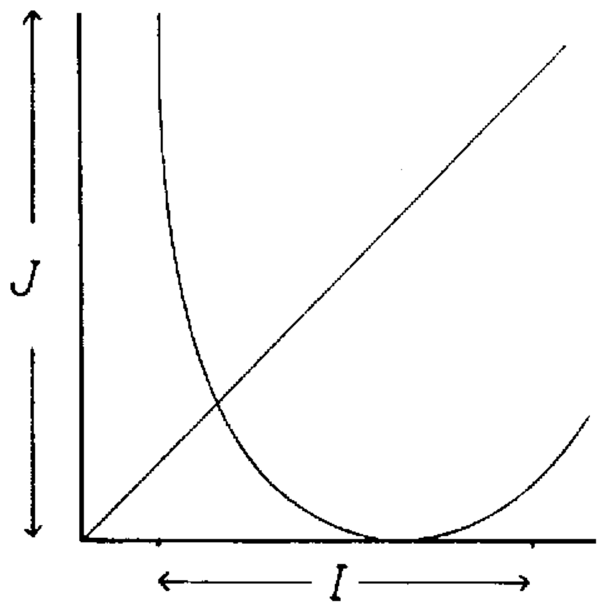

Fig. 10.1

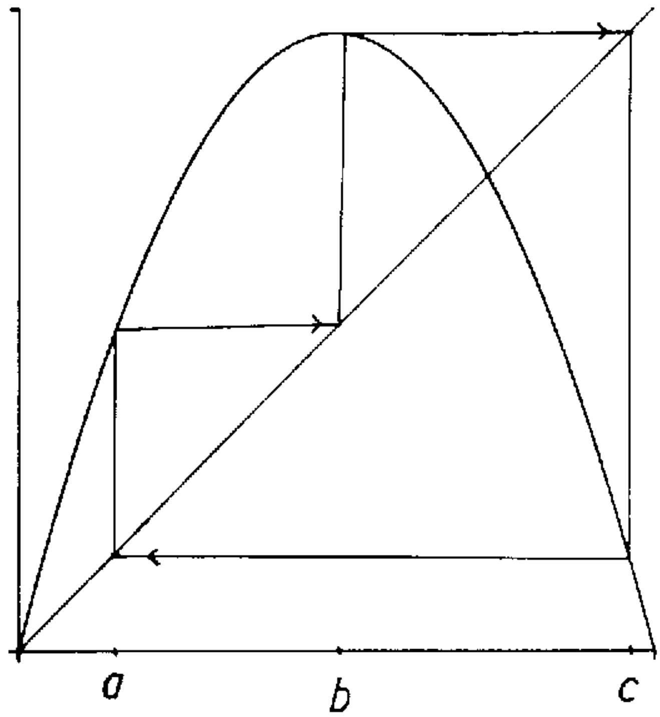

Fig. 10.2. The map $F_{3.839}(x) = 3.839x(1-x)$.

The second observation is the following: suppose $A_0, A_1, \ldots, A_n$ are closed intervals and $f(A_i) \supset A_{i+1}$ for $i = 0, \ldots, n-1$. Then there exists at least one subinterval J_0 of A_0 which is mapped onto A_1. There is a similar subinterval in A_1 which is mapped onto A_2, and thus there is a subinterval $J_1 \subset J_0$ having the property that $f(J_1) \subset A_1$ and $f^2(J_1) = A_2$. Continuing in this fashion, we find a nested sequence of intervals which map into the various A_i in order. Thus there exists a point $x \in A_0$ such that $f^i(x) \in A_i$ for each i. We say that $f(A_i)$ *covers* A_{i+1}. See Exercise 1.

To prove the Theorem, let $a, b, c \in \mathbf{R}$ and suppose $f(a) = b$, $f(b) = c$, and $f(c) = a$. We assume that $a < b < c$. The only other possibility, $f(a) = c$, is handled similarly. This situation arises in the quadratic map F_μ for sufficiently large μ, and even for some $\mu < 4$. In fact, we will exploit this fact later when we discuss the case $\mu = 3.839$ in detail in §1.13. See Fig. 10.2.

Let $I_0 = [a, b]$ and $I_1 = [b, c]$ and note that our assumptions imply $f(I_0) \supset I_1$ and $f(I_1) \supset I_0 \cup I_1$. The graph of f shows that there must be a fixed point for f between b and c. Similarly, f^2 must have fixed points between a and b, and it is easy to see that at least one of these points must have period two. So we let $n \geq 2$; our goal then is to produce a periodic point of prime period $n > 3$.

Inductively, we define a nested sequence of intervals $A_0, A_1, \ldots, A_{n-2} \subset I_1$ as follows. Set $A_0 = I_1$. Since $f(I_1) \supset I_1$, there is a subinterval $A_1 \subset A_0$ such that $f(A_1) = A_0 = I_1$. Then there is a subinterval $A_2 \subset A_1$ such that $f(A_2) = A_1$, so that $f^2(A_2) = A_0 = I_1$. Continuing, we find a subinterval $A_{n-2} \subset A_{n-3}$ such that $f(A_{n-2}) = A_{n-3}$. According to our second observation above, if $x \in A_{n-2}$, then $f(x), f^2(x), \ldots, f^{n-2}(x) \subset A_0$ and, indeed, $f^{n-2}(A_{n-2}) = A_0 = I_1$.

Now since $f(I_1) \supset I_0$, there exists a subinterval $A_{n-1} \subset A_{n-2}$ such that $f^{n-1}(A_{n-1}) = I_0$. Finally, since $f(I_0) \supset I_1$ we have, $f^n(A_{n-1}) \supset I_1$ so that $f^n(A_{n-1})$ covers A_{n-1}. It follows from our first observations that f^n has a fixed point p in A_{n-1}.

We claim that p actually has prime period n. Indeed, the first $n-2$ iterations of p lie in I_1, the $(n-1)^{st}$ lies in I_0, and the n^{th} is p again. If $f^{n-1}(p)$ lies in the interior of I_0 then it follows easily that p has prime period n. If $f^{n-1}(p)$ happens to lie on the boundary, then $n = 2$ or 3, and again we are done.

q.e.d.

This theorem is just the beginning of the story. Sarkovskii's Theorem gives a complete accounting of which periods imply which other periods for continuous maps of **R**. Consider the following ordering of the natural numbers:

$$3 \rhd 5 \rhd 7 \rhd \cdots \rhd 2\cdot 3 \rhd 2\cdot 5 \rhd \cdots \rhd 2^2\cdot 3 \rhd 2^2\cdot 5 \rhd \cdots$$
$$\rhd 2^3\cdot 3 \rhd 2^3\cdot 5 \rhd \cdots\cdots \rhd 2^3 \rhd 2^2 \rhd 2 \rhd 1.$$

That is, first list all odd numbers except one, followed by 2 times the odds, 2^2 times the odds, 2^3 times the odds, etc. This exhausts all the natural numbers with the exception of the powers of two which we list last, in decreasing order. This is the Sarkovskii ordering of the natural numbers. Sarkovskii's Theorem is:

Theorem 10.2. *Suppose $f\colon \mathbf{R} \to \mathbf{R}$ is continuous. Suppose f has a periodic point of prime period k. If $k \rhd \ell$ in the above ordering, then f also has a periodic point of period ℓ.*

Before proving this Theorem, we note several consequences.

Remarks.

1. If f has a periodic point whose period is not a power of two, then f necessarily has infinitely many period points. Conversely, if f has only finitely many periodic points, then they all necessarily have periods which are powers of two. This fact will reappear when we discuss the period-doubling route to chaos in a later section.

2. Period 3 is the greatest period in the Sarkovskii ordering and therefore implies the existence of all other periods, as we saw above.

3. The converse of Sarkovskii's Theorem is also true! There are maps which have periodic points of period p and no "higher" period points according to the Sarkovskii ordering. We give several examples of this at the end of this section.

We will give an elementary proof of Sarkovskii's Theorem due to Block, Guckenheimer, Misiurewicz and Young. The proof rests mainly on the two observations which we used above. For two closed intervals, I_1 and I_2, we will introduce the notation $I_1 \to I_2$ if $f(I_1)$ covers I_2. If we find a sequence of intervals $I_1 \to I_2 \to \ldots \to I_n \to I_1$, then our previous observations show that there is a fixed point of f^n in I_1.

We first assume that f has a periodic point x of period n with n odd and $n > 1$. Suppose that f has no periodic points of odd period less than n. Let $x_1, \ldots, x_n$ be the points on the orbit of x, enumerated from left to right. Note that f permutes the x_i. Clearly, $f(x_n) < x_n$. Let us choose the largest i for which $f(x_i) > x_i$. Let I_1 be the interval $[x_i, x_{i+1}]$. Since $f(x_{i+1}) < x_{i+1}$, it follows that $f(x_{i+1}) \leq x_i$ and so we have that $f(I_1) \supset I_1$. Therefore, $I_1 \to I_1$.

Since x does not have period 2, it cannot be that $f(x_{i+1}) = x_i$ and $f(x_i) = x_{i+1}$ so that $f(I_1)$ contains at least one other interval of the form $[x_j, x_{j+1}]$. A priori, there may be several such intervals, but we will see below that in fact there is only one. Let $\mathcal{O}_2$ denote the union of intervals of the form $[x_j, x_{j+1}]$ that are covered by $f(I_1)$. Hence we have $\mathcal{O}_2 \supset I_1$ but $\mathcal{O}_2 \neq I_1$, and if I_2 is any interval in $\mathcal{O}_2$ of the form $[x_j, x_{j+1}]$, then $I_1 \to I_2$.

Now let $\mathcal{O}_3$ denote the union of intervals of the form $[x_j, x_{j+1}]$ that have the property that they are covered by the image of some interval in $\mathcal{O}_2$. Continuing inductively, we let $\mathcal{O}_{\ell+1}$ be the union of intervals that are covered by the image of some interval in $\mathcal{O}_\ell$. Note that, if $I_{\ell+1}$ is any interval in $\mathcal{O}_{\ell+1}$, there is a collection of intervals $I_2, \ldots, I_\ell$ with $I_j \subset \mathcal{O}_j$ which satisfy $I_1 \to I_2 \to \ldots \to I_\ell \to I_{\ell+1}$.

Now the $\mathcal{O}_\ell$ form an increasing union of intervals. Since there are only finitely many x_j, it follows that there is an ℓ for which $\mathcal{O}_{\ell+1} = \mathcal{O}_\ell$. For this ℓ we must have that $\mathcal{O}_\ell$ contains all intervals of the form $[x_j, x_{j+1}]$, for otherwise x would have period less than n.

We claim that there is at least one interval $[x_j, x_{j+1}]$ different from I_1 in some $\mathcal{O}_k$ whose image covers I_1. This follows since there are more x_i's on one side of I_1 than on the other (n is odd.) Hence some x_i's must change sides under the action of f, and some must not. Consequently, there is at least one interval whose image covers I_1.

Now let us consider chains of intervals $I_1 \to I_2 \to \ldots \to I_k \to I_1$ where each I_ℓ is of the form $[x_j, x_{j+1}]$ for some j and where $I_2 \neq I_1$. We do not assume that $I_\ell \subset \mathcal{O}_\ell$. By the above observations, there is at least one such chain. Let us choose the smallest k for which this happens, i.e., $I_1 \to I_2 \to \ldots \to I_k \to I_1$ is the shortest path from I_1 to I_1 except, of course, $I_1 \to I_1$. We therefore find a diagram as in Fig. 10.3.

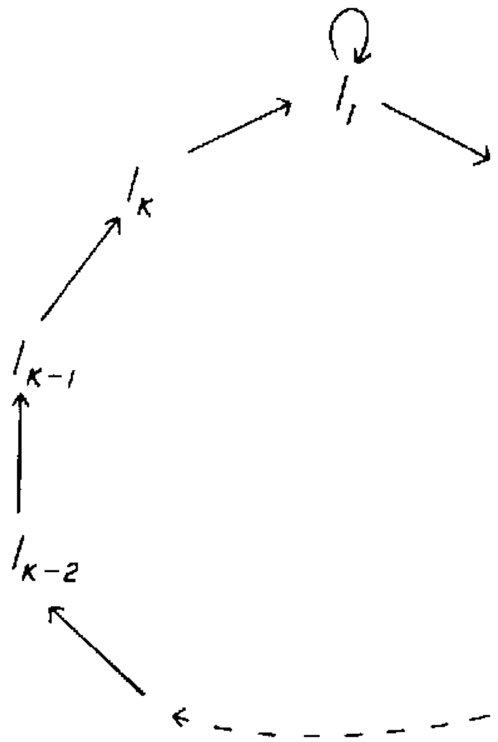

Fig. 10.3.

Now, if $k < n-1$, then one of the loops $I_1 \to I_2 \to \ldots \to I_k \to I_1$ or $I_1 \to \ldots \to I_k \to I_1 \to I_1$ gives a fixed point of f^m with m odd and $m < n$. This point must have prime period $< k$ since $I_1 \cap I_2$ consists of only one point, and that point has period $> m$. Therefore $k = n-1$.

Since k is the smallest integer that works, we cannot have $I_\ell \to I_j$ for any $j > \ell + 1$. It follows that the orbit of x must be ordered in $\mathbf{R}$ in one of two possible ways, as depicted in Fig. 10.4.

It follows that we can extend the diagram depicted in Fig. 10.3 to that shown in Fig. 10.5. Sarkovskii's Theorem for the special case of n odd is now immediate. Periods larger than n are given by cycles of the form $I_1 \to \ldots \to I_{n-1} \to I_1 \to \ldots I_1$. The smaller even periods are given by cycles of the form

$$I_{n-1} \to I_{n-2} \to I_{n-1},$$

$$I_{n-1} \to I_{n-4} \to I_{n-3} \to I_{n-2} \to I_{n-1}$$

and so forth. For the case of n even, we first note that f must have a periodic point of period 2. This follows from the above arguments provided we can

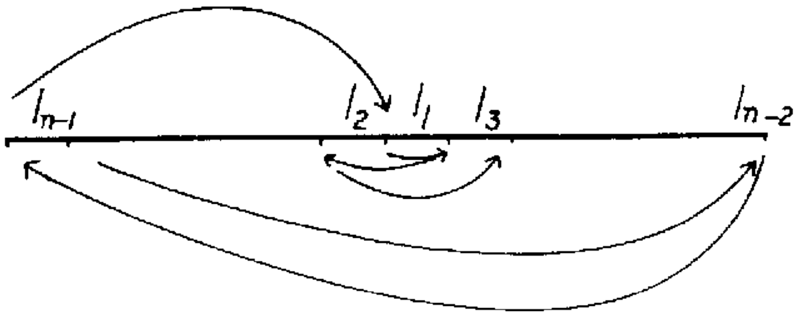

Fig. 10.4. One possible ordering of the I_j.
The other is the mirror image.

guarantee that some x_i's change sides under f and some do not (use the facts that $I_{n-1} \leftarrow I_{n-2}$ and $I_{n-1} \rightarrow I_{n-2}$). If this is not the case, then all of the x_i's must change sides and so $f[x_1, x_i] \supset [x_{i+1}, x_n]$ and $f[x_{i+1}, x_N] \supset [x_1, x_i]$. But then, our observation above produces a period 2 point in $[x_1, x_i]$.

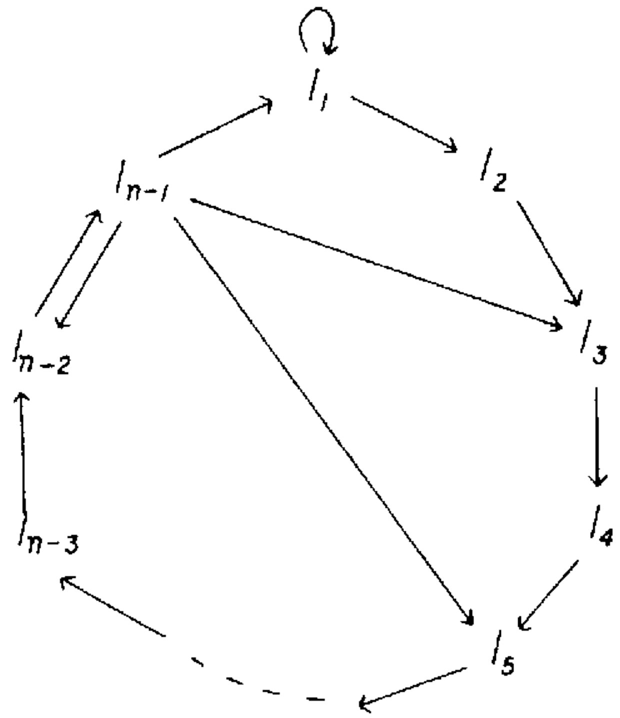

Fig. 10.5.

The Theorem now will be proved for $n = 2^m$ as follows. Let $k = 2^\ell$ with $\ell < m$. Consider $g = f^{k/2}$. By assumption, g has a periodic point of period $2^{m-\ell+1}$. Therefore, g has a point which has period 2. This point has period 2^ℓ for f.The final case is now $n = p \cdot 2^m$ where p is odd. This case can be reduced to the previous two. We leave these reductions as Exercises.

q.e.d.

We now turn to the converse of Sarkovskii's Theorem. To produce a map with period 5 and no period 3, consider a map $f\colon [1,5] \to [1,5]$ which satisfies

$$\begin{aligned} f(1) &= 3 \\ f(3) &= 4 \\ f(4) &= 2 \\ f(2) &= 5 \\ f(5) &= 1 \end{aligned}$$

so that 1 is periodic of period 5. Suppose that f is linear between these integers, i.e., the graph is as shown in Fig. 10.6.

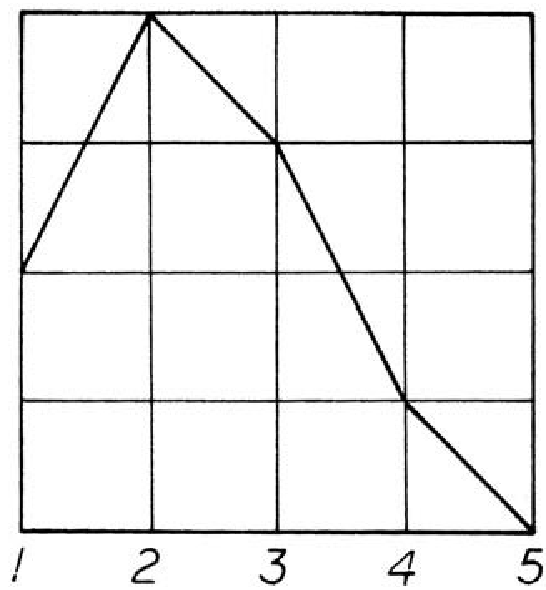

Fig. 10.6.

It is easy to check that

$$\begin{aligned} f^3[1,2] &= [2,5] \\ f^3[2,3] &= [3,5] \\ f^3[4,5] &= [1,4] \end{aligned}$$

so f^3 has no fixed points in any of these intervals. It is true that $f^3[3,4] = [1,5]$ so that f^3 has at least one fixed point in $[3,4]$. But we claim that this point is unique, and therefore must be the fixed point for f, not the period 3 point. Indeed, $f\colon [3,4] \to [2,4]$ is monotonically decreasing, as is $f\colon [2,4] \to [2,5]$ and $f\colon [2,5] \to [1,5]$. Therefore f^3 is monotonically decreasing on $[3,4]$ and the fixed point is unique.

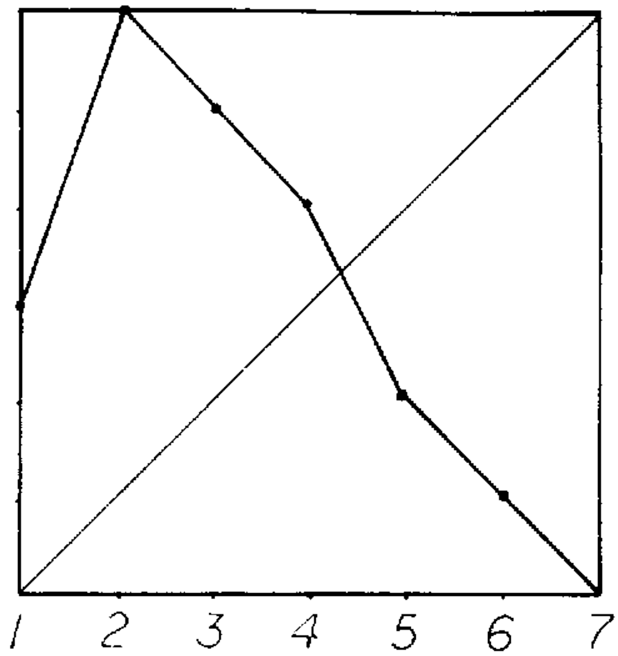

Fig. 10.7.

The graph, shown in Fig. 10.7, produces period 7 but not period 5.

This process is easily generalized to give the first portion of the Sarkovskii ordering. For the even periods, we will introduce a trick. Let $f: I \to I$ be continuous. We will construct a new function F, the double of f, whose periodic points will have exactly twice the period of those of f, plus one additional fixed point. The procedure for producing F is as follows. Divide the interval I into thirds. Compress the graph of f into the upper left corner of $I \times I$ as shown on Fig. 10.8.a. The rest of the graph is filled in as in Fig. 10.8.b.

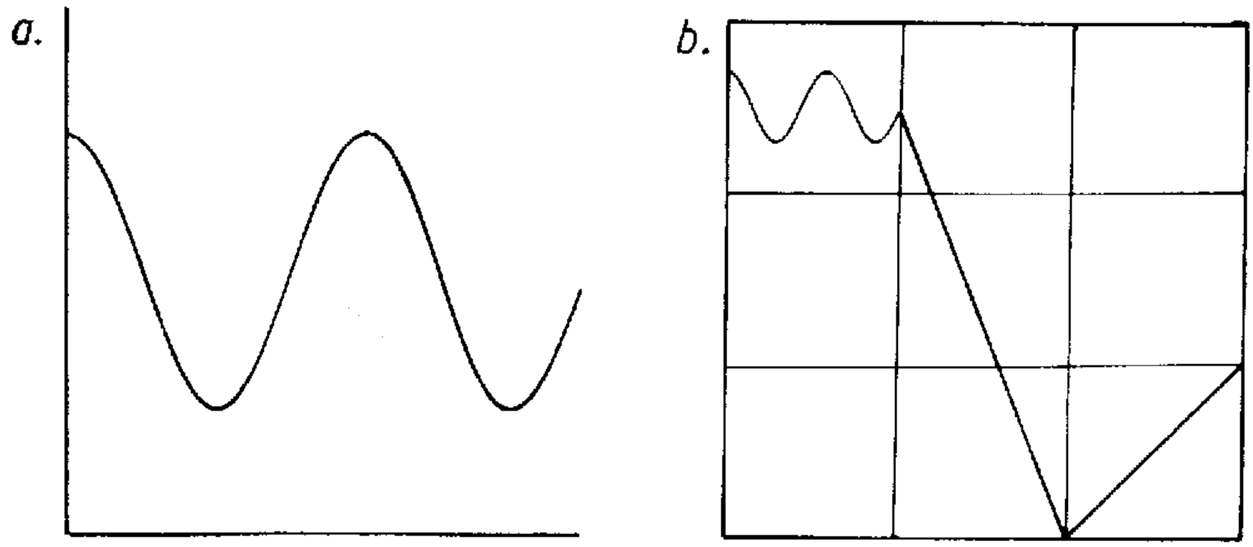

Fig. 10.8. Fig. 10.8.a. gives the graph of $f(x)$ while Fig. 10.8.b. gives the graph of its double, $F(x)$.

The map F is piecewise linear on $[1/3, 2/3]$ and $[2/3, 1]$. Moreover, $F(\frac{2}{3}) = 0$, $F(1) = \frac{1}{3}$, and F is continuous.

Note that F maps $[0, \frac{1}{3}]$ into $[\frac{2}{3}, 1]$ and vice versa. Also note that if $x \in [\frac{1}{3}, \frac{2}{3}]$ and x is not the fixed point, then there exists n so that $F^n(x) \in [0, \frac{1}{3}] \cup [\frac{2}{3}, 1]$. This implies that there are no other F-periodic points in $(\frac{1}{3}, \frac{2}{3})$. Exercise 7 shows that if x is a periodic point of period n for f, then $x/3$ is periodic of period $2n$ for F. On the other hand, if y is F-periodic then either y or $F(y)$ lies in $[0, \frac{1}{3}]$ and Exercise 9 shows that $3y$ or $3F(y)$ is f-periodic. Thus to produce a map with period 10 but not period 6, we need only double the graph of a function with period 5 but not period 3.

As a final remark, we must emphasize that Sarkovskii's Theorem is very definitely only a one-dimensional result. There is no higher dimensional analogue of this result. In fact, the Theorem does not even hold on the circle. For example, the map which rotates all points on the circle by 120^o makes all points periodic with period three. There are no other periods whatsoever.

Exercises

1. Suppose $A_0, A_1, \ldots, A_n$ are closed intervals and $f(A_i) \supset A_{i+1}$ for $i = 0, \ldots, n-1$. Prove that there exists a point $x \in A_0$ such that $f^i(x) \in A_i$ for each i.

2. Prove that if f has period $p \cdot 2^m$ with p odd, then f has period $q \cdot 2^m$ with q odd, $q > p$.

3. Prove that if f has period $p \cdot 2^m$ with p odd, then f has period 2^ℓ, $\ell \leq m$.

4. Prove that if f has period $p \cdot 2^m$ with p odd, then f has period $q \cdot 2^m$ with q even.

5. Construct a piecewise linear map with period $2n + 1$.

6. Give a formula for $F(x)$ in terms of $f(x)$, where $F(x)$ is the double of $f(x)$.

7. Prove that $F(x)$, the double of $f(x)$, has a periodic point of period $2n$ at $x/3$ iff x has f-period n.

8. Construct a map that has periodic points of period 2^j for $j < \ell$ but not period 2^ℓ.

9. Prove that if $F(x)$, the double of $f(x)$, has a periodic point p that is not fixed, then either p or $F(p)$ lies in $[0, \frac{1}{3}]$. Prove that, in this case, either $3p$ or $3F(p)$ is a periodic point for f.

Comment: An Introduction to Dynamic System.pdf

No.1 Theme:Comments

note：I0和I1是区间[0，1]除去A0的左半区间和右半区间