DYNAMICAL SYSTEMS
Lent Term, 2001

1. General notions and the logistic map

This course, like Dynamics of Differential Equations, is concerned with the following situation. We have a system, defined by the values of finitely many parameters, which varies with time in a well-defined manner. If there are n parameters, then we can think of the system as described by a point P in n-dimensional space $\mathbf{R}^n$. If we regard time as continuous, then the motion is described by the system of differential equations

$$\frac{dP}{dt} = f(P) \tag{1}$$

for some function f; the course Dynamics of Differential Equations provides an introduction to such systems. But if we only observe the system at discrete time intervals, we can index successive observations by integers; and if the rth observation of the system is described by the point P_r, then we shall be interested in its behaviour under a transformation law of the type

$$P_{r+1} = F(P_r) \tag{2}$$

for some function F which does not depend on r. It is with the behaviour of a system under such laws that this course is concerned.

These two topics are in fact closely linked. For example, suppose that (1) admits a periodic solution $\mathcal{C}$. Let A be a point of $\mathcal{C}$, Π a hyperplane through A transversal to $\mathcal{C}$, and P a point of Π close to A. Because P is close to A, the trajectory of (1) through P remains close to $\mathcal{C}$ for a long time, and in particular it crosses Π again at a point close to A. Denote this point by $F(P)$; this defines a function F from a neighbourhood of A in Π to Π, which is continuous at and near A and maps A to itself. Studying the behaviour of trajectories of (1) near $\mathcal{C}$ is clearly equivalent to studying the behaviour of the map F near A.

Again, consider the system

$$\frac{dP}{dt} = f(P, t) \tag{3}$$

where f is periodic in t with period τ. For any point P, define F by the condition that the trajectory starting at $P \times 0$ passes through $F(P) \times \tau$ —

in other words, following the trajectory along one period maps P to $F(P)$. Then studying the long-term behaviour of trajectories of (3) is the same as studying the behaviour of large powers of F. Note that in both these examples F must be one-one into, because of the uniqueness theorem for the solutions of differential equations.

It is intuitively plausible that under suitable conditions we should be able to reverse the argument; in other words, given F we should be able to find a (non-unique) f which gives rise to it. This is in fact true, though the proof is messy. But the condition that F is one-one is crucial. One therefore expects, what turns out to be true, that the study of systems (2) in $n-1$ dimensions with F **invertible** has much in common with the study of systems (1) in n dimensions, or of systems (3) in $n-1$ dimensions; and the study of one of these turns out to be about as difficult as the study of the other. However, if we drop the condition of invertibility the situation changes. We first get qualitatively interesting behaviour of (1) in $\mathbf{R}^3$, and of invertible systems (2) in $\mathbf{R}^2$; but the behaviour of non-invertible systems (2) is already interesting in one-dimensional spaces, and is much easier to study. This is the motivation for this course, that it provides the simplest models for behaviour which is interesting and appears to be very common.

The definition (2) makes sense for a map F from any space $\mathcal{S}$ to itself; and there is no natural family of spaces to which the theory should be restricted, though it is desirable to have a topology on $\mathcal{S}$. Some definitions and results do not even require $\mathcal{S}$ to be metric; others involve very strong hypotheses on $\mathcal{S}$. The right attitude seems to be that $\mathcal{S}$ should be a sensible space in the context being considered, but what "sensible" means will vary from one part of the subject to another. One loses very little by always thinking of $\mathcal{S}$ as metric. In this course $\mathcal{S}$ will usually (but not always) be n-dimensional for some n, in the sense that each point of $\mathcal{S}$ has a neighbourhood homeomorphic to an open set in $\mathbf{R}^n$ — that is, $\mathcal{S}$ looks like $\mathbf{R}^n$ locally. An important special case is when $\mathcal{S}$ is **one-dimensional**, which is to say homeomorphic to $\mathbf{R}$, $[0,\infty)$, $[0,1]$ or the circle $\mathbf{R}/\mathbf{Z}$. If $\mathcal{S}$ looks like $\mathbf{R}^n$ locally, we shall require F to be piecewise smooth though it may have points of discontinuity. If P_r is a point of $\mathcal{S}$ we usually write $F(P_r) = P_{r+1}$. The map $P_0 \mapsto P_r$ is written F^r; and if F is invertible we can allow r here to be negative. The main questions to address are:

- What is the behaviour of P_r as r grows?

- How is this affected by a small change in P_0?
- How do the answers change as we vary F?

A point P is called a **fixed point** of F if $F(P) = P$; it is called a **periodic point** if $F^r(P) = P$ for some $r > 0$, and in this case we say that it has period r. If P_0 has least period r then $\{P_0, \ldots, P_{r-1}\}$ is called a **cycle**. The set $\{P_0, P_1, \ldots\}$ is called the **orbit** of P_0; but if F is invertible the orbit sometimes instead means the set $\{\ldots, P_{-1}, P_0, P_1, \ldots\}$. An **omega-point** of the orbit is a point of accumulation of the sequence $\{P_r\}$ — that is to say, a point Q such that for any neighbourhood $\mathcal{N}$ of Q and any integer $R > 0$ there are points P_r with $r > R$ in $\mathcal{N}$. If F is invertible there is a corresponding definition of an **alpha-point**; but these will not appear in this course. The set of omega-points of a given orbit is closed, and is mapped onto itself by F. But even if $\mathcal{S}$ and F are respectable, the set of omega-points of an orbit can be highly pathological; it is indeed this fact that gives the subject its glamour.

A related idea is that of an attractor. A fixed point Q is called an **attractor** of F if

(i) given any neighbourhood $\mathcal{N}$ of Q there is a neighbourhood $\mathcal{N}^*$ of Q such that P in $\mathcal{N}^*$ implies that $F^r(P)$ is in $\mathcal{N}$ for all $r \geq 0$;

(ii) there is a neighbourhood $\mathcal{N}^\flat$ of Q such that P in $\mathcal{N}^\flat$ implies $F^r(P) \to Q$ as $r \to \infty$.

In other words, for any P near Q the $F^r(P)$ stay near Q and tend to Q. Of course we can take $\mathcal{N}^* = \mathcal{N}^\flat$ by replacing each of them by their intersection. The condition in (ii) is in metric form; but it is equivalent to the non-metric requirement that for any open neighbourhood $\mathcal{N}^\sharp$ of Q there exists r_0 depending only on P and $\mathcal{N}^\sharp$ such that $F^r(P)$ is in $\mathcal{N}^\sharp$ for all $r > r_0$.

In many books, what is here called an attractor is instead called **asymptotically stable**. If Q satisfies (i) it is called **quasi-asymptotically stable**, and if Q satisfies (ii) it is called **Liapounov stable**; but it turns out that neither of these properties is very interesting.

The opposite of an attractor is a repellor; a fixed point Q is a **repellor** of F if there is a neighbourhood $\mathcal{N}$ of Q such that for any P in $\mathcal{N}$ other than Q not all $F^r(P)$ lie in $\mathcal{N}$.

Proposition 1 *If Q is a fixed point of F which is an attractor for F, then F is continuous at Q.*

Proof It follows from (i) that, given any neighbourhood $\mathcal{N}$ of Q, there is a neighbourhood $\mathcal{N}^*$ of Q such that $F(\mathcal{N}^*) \subset \mathcal{N}$. □

More generally, let $\mathcal{A}$ be a closed set mapped into itself by F; then it is natural to call $\mathcal{A}$ an **attractor** if

(i) given any neighbourhood $\mathcal{N}$ of $\mathcal{A}$ there is a neighbourhood $\mathcal{N}^*$ of $\mathcal{A}$ such that P in $\mathcal{N}^*$ implies that $F^r(P)$ is in $\mathcal{N}$ for all $r \geq 0$;

(ii) there is a neighbourhood $\mathcal{N}^\flat$ of $\mathcal{A}$ such that for any P in $\mathcal{N}^\flat$ and any $\mathcal{N}^\sharp$ containing $\mathcal{A}$ there exists r_0 such that $F^r(P)$ is in $\mathcal{N}^\sharp$ for all $r > r_0$.

Note that making $\mathcal{N}$ smaller in (i), or $\mathcal{N}^\sharp$ smaller in (ii), only makes these conditions harder to satisfy.

If an attractor $\mathcal{A}$ is both minimal and pathological, early writers called it a **strange attractor** without giving an exact definition. This was a very sensible decision; no one has yet given a generally accepted formal definition and all definitions known to me allow some very un-strange sets to be called strange attractors. For example, we shall see in Theorem 8 that for $F(x) = 4x(1-x)$ on $[0,1]$ the only attractor is the entire space.

In particular, let $\mathcal{C} = \{P_0, \ldots, P_{M-1}\}$ be a cycle of least period M; we would expect the cycle $\mathcal{C}$ to be an attractor for F if and only if each point of $\mathcal{C}$ is an attractor for F^M. This is almost true, but we need an additional hypothesis.

Proposition 2 *Let $\mathcal{C}$ be a cycle of period M, and let F be continuous at each point of $\mathcal{C}$. Then $\mathcal{C}$ is an attractor for F if and only if each point of $\mathcal{C}$ is an attractor for F^M.*

Proof Suppose first that $\mathcal{C} = \{P_0, \ldots, P_{M-1}\}$ is an attractor for F; and let $M = M'M''$ where M' is the least period of $\mathcal{C}$. Using the continuity of F, we can choose disjoint open neighbourhoods $\mathcal{N}_i$ of P_i such that (taking all subscripts mod M') $F(\mathcal{N}_i)$ meets no $\mathcal{N}_j$ other than $\mathcal{N}_{i+1}$. Given any neighbourhood $\mathcal{N}$ of $\mathcal{C}$ we can first replace the $\mathcal{N}_i$ by $\mathcal{N}_i \cap \mathcal{N}$ and then replace $\mathcal{N}$ by $\cup \mathcal{N}_i$, which is smaller. The $\mathcal{N}^*$ in the definition is contained in $\mathcal{N}$, so it is the disjoint union of the open sets $\mathcal{N}_i^* = \mathcal{N}_i \cap \mathcal{N}^*$; and $F(\mathcal{N}_i^*)$ is contained in $\mathcal{N}$ and does not meet any $\mathcal{N}_j$ other than $\mathcal{N}_{i+1}$, so that $F(\mathcal{N}_i^*) \subset \mathcal{N}_{i+1}$. By induction on r, $F^r(\mathcal{N}_i^*) \subset F(\mathcal{N}_{i+r-1}) \cap \mathcal{N} = \mathcal{N}_{i+r}$. We can replace the $\mathcal{N}_i^*$ by their intersections with $\mathcal{N}^\flat$ without changing any of these properties, and it then follows that for all P in any $\mathcal{N}_i^*$ we have $F^{Mr+j}(P) \to P_{i+j}$ as $r \to \infty$. This implies that each P_i is an attractor for F^M.

Conversely, suppose that each P_i is an attractor for F^M. With the $\mathcal{N}_i$ as above, we can choose $\mathcal{N}_i^*$ so that $F^{Mr}(\mathcal{N}_i^*) \subset \mathcal{N}_i$ for all r. By continuity we can choose neighbourhoods $\mathcal{N}_i^{**}$ of P_i such that $F^j(\mathcal{N}_i^{**}) \subset \mathcal{N}_{i+j}^*$ for $0 \leq j < M$; and now

$$F^{Mr+j}(\mathcal{N}_i^{**}) = F^{Mr}(F^j(\mathcal{N}_i^{**})) \subset F^{Mr}(\mathcal{N}_{i+j}^*) \subset \mathcal{N}_{i+j}.$$

This proves the first condition for $\mathcal{C}$ to be an attractor. For the second, let $\mathcal{N}_i^\flat$ be such that P in $\mathcal{N}_i^\flat$ implies $F^{Mr}(P) \to P_i$. If we apply F^j with $0 \leq j < M$ and use the continuity of F at each P_k, we deduce $F^{Mr+j}(P) \to P_{i+j}$. □

Proposition 3 *Suppose that $\mathcal{S}$ is one-dimensional and F has a continuous first derivative at and near the fixed point y. Then*

$$|F'(y)| < 1 \implies y \textit{ is an attractor } \implies |F'(y)| \leq 1.$$

If $|F'(y)| > 1$ then y is a repellor.

Proof If x is near y then $F(x) = F(y) + (x-y)F'(\xi)$ for some ξ between x and y. If $|F'(y)| < 1-\epsilon$ and x is near y, then so is ξ; so $|F'(\xi)| < 1 - \frac{1}{2}\epsilon$ and $|F(x) - y| < (1-\frac{1}{2}\epsilon)|x-y|$. The first condition in the definition of an attractor follows at once. Repeating this argument, we obtain

$$|F^r(x) - y| < (1-\tfrac{1}{2}\epsilon)^r|x-y|, \quad \text{whence} \quad F^r(x) \to y \text{ as } r \to \infty.$$

Conversely, if $|F'(y)| > 1+\epsilon$ let $\mathcal{N}$ be a bounded neighbourhood of y at all points of which $|F'(x)| > 1 + \frac{1}{2}\epsilon$; then a similar argument shows that if x, $F(x), \ldots,$ $F^r(x)$ are all in $\mathcal{N}$ then

$$|F^r(x) - y| > (1+\tfrac{1}{2}\epsilon)^r|x-y|.$$

Hence some $F^s(x)$ is outside $\mathcal{N}$, so that y fails the first condition for an attractor but satisfies the condition for a repellor. □

It is useful to have a similar test for whether a cycle $\mathcal{C}$ is an attractor. Suppose that $\mathcal{S}$ is one-dimensional, that y_0 is periodic with period M and that F has a continuous first derivative at and near each point $y_m = F^m(y_0)$. Then

$$(F^M)'(y_0) = \prod_{m=0}^{M-1} F'(F^m(y_0)) = \prod_{m=0}^{M-1} F'(y_m) = (F^M)'(y_i)$$

for any i. This proves the analogue of Proposition 3 with $F'(y)$ replaced by $\prod_{m=0}^{M-1} F'(y_m)$.

The following result overlaps considerably with Proposition 3; one reason why it is of interest is that the method of proof is so different. Note that it treats 'sign' as a three-valued function, the values being $+, -$ and 0.

Proposition 4 *Suppose that $\mathcal{S}$ is $(0,1)$ or $[0,1]$ and F is continuous; and let y be a fixed point of F such that for all x near enough to y*

$$F(x) \leq y \text{ if } x < y \quad \text{and} \quad F(x) \geq y \text{ if } x > y.$$

Then y is an attractor if and only if, for all x near enough to y, $F(x) - x$ has the same sign as $y - x$; and y is a repellor if and only if, for all x near enough to y, $F(x) - x$ has the opposite sign to $y - x$.

Proof If there are fixed points arbitrarily close to y then y cannot be either an attractor or a repellor; and at such points $F(x) - x$ is zero and hence cannot have the same sign as $y - x$ or the opposite one. Hence we may assume that y is an isolated fixed point. Denote by $\mathcal{N}$ a neighbourhood of y which contains no other fixed point. Because F is continuous, for x in $\mathcal{N}$ the sign of $F(x)-x$ depends only on the sign of $y-x$. Suppose for example that x is in $\mathcal{N}$ and $F(x)-x > 0$ and $y-x > 0$; then $x < F(x) \leq y$ and it follows by induction on r that the sequence of $F^r(x)$ is monotone non-decreasing and bounded above by y. Hence it tends to a limit ξ; and since

$$F(\xi) = F(\lim F^r(x)) = \lim F(F^r(x)) = \xi,$$

we must have $\xi = y$. Similarly if we suppose that $F(x)-x < 0$ and $y-x > 0$ the sequence of $F^r(x)$ is decreasing as long as its elements remain in $\mathcal{N}$. If the sequence never left $\mathcal{N}$, it would have a limit ξ in $\mathcal{N}$ with $\xi < x$ and $F(\xi) = \xi$; and no such ξ exists. Hence the sequence eventually leaves $\mathcal{N}$. Combining these results with the corresponding ones when $y < x$ we obtain the Proposition. □

For any space $\mathcal{S}$ and map $F : \mathcal{S} \to \mathcal{S}$ with a fixed point Q, we define the **domain of attraction** of Q to be the set of points P in $\mathcal{S}$ such that $F^r(P) \to Q$ as $r \to \infty$; we need not require Q to be an attractor, but this seems to be the only case in which the domain of attraction is bound to have sensible properties. We can generalize this definition by replacing Q by any set which is mapped into itself by F; but it appears that we get nothing novel by doing so. Note that if $\mathcal{D}$ is the domain of attraction of Q then F maps $\mathcal{D}$ into $\mathcal{D}$, and $F^{-1}(\mathcal{D}) = \mathcal{D}$.

Lemma 5 *If Q is an attractor for a continuous map F, the domain of attraction of Q is an open set.*

Proof By (ii) in the definition of an attractor, there is a neighbourhood $\mathcal{N}$ of Q which lies in the domain of attraction of Q. Let P be any point in the domain of attraction of Q; then there exists r depending on P such that $F^r(P)$ is in $\mathcal{N}$. By continuity, if P^* is close enough to P then $F^r(P^*)$ is in $\mathcal{N}$; so $F^{r+s}(P^*) \to Q$ as $s \to \infty$, and P^* is in the domain of attraction of Q. Note that this argument still works if we replace Q by any attractor $\mathcal{A}$. $\square$

Now suppose also that $\mathcal{S}$ is connected and has more than one attracting fixed point. Then the set of points of $\mathcal{S}$ which do not tend to any attracting fixed point is closed, being the complement of the union of the domains of attraction of these attracting fixed points. It is not empty, for otherwise $\mathcal{S}$ would be the union of these domains of attraction, which are disjoint open sets; since $\mathcal{S}$ is connected, this is impossible. Intuitively, this closed set separates the various domains of attraction; but it can be highly pathological. If $\mathcal{C}$ is an M-cycle with $M > 1$, the domain of attraction of $\mathcal{C}$ usually turns out to be just the union of the domains of attraction under F^M of the various points of $\mathcal{C}$. However, this is a simple way of obtaining the situation which we have just been considering.

The simplest pair $\mathcal{S}, F$ of interest is given by the following Lemma.

Lemma 6 *Suppose that $\mathcal{S}$ is the closed unit interval $[0,1]$, on which F is continuous and invertible. Then F is strictly monotone.*

Proof Since F is invertible, it takes no value more than once. Suppose that $0 \le x < y < z \le 1$; I claim that $F(y)$ lies between $F(x)$ and $F(z)$. For if not, it is enough to consider the typical case when $F(x) < F(z) < F(y)$. By continuity, there exists ξ with $x < \xi < y$ and $F(\xi) = F(z)$, which is impossible. Now suppose for example that $F(0) < F(1)$, and let $0 < x < y < 1$. Then $F(x)$ lies between $F(0)$ and $F(1)$, so that $F(x) < F(1)$; and $F(y)$ lies between $F(x)$ and $F(1)$, so that $F(x) < F(y)$. $\square$

Suppose first that F is monotone increasing. The fixed points of F are the points where the graph of F meets the diagonal; and since these are both closed sets, so is their intersection. Denote the open set of non-fixed points in [0,1] by $\mathcal{N}$ and let y be a point of $\mathcal{N}$. There is a maximal open interval $I = (x, z)$ of $\mathcal{N}$ containing y; here z is a fixed point of F or $z = 1$, or possibly both, and similarly for x. By continuity, $F(y') - y'$ has the same sign for all y' in I. We are now in the position considered in the proof of Proposition 4.

Let y' be any point of I; then $F^r(y')$ tends to z or x according as the graph of F on I lies above or below the diagonal. If 0 or 1 is not a fixed point, we can apply a similar argument to it. It follows that every point of [0,1] tends to a fixed point under iteration of F; in particular F admits no periodic points other than fixed points.

If F is monotone decreasing the situation is slightly more complicated. Now $F(x) - x$ is continuous, strictly monotone decreasing, non-negative at $x = 0$ and non-positive at $x = 1$; so it has just one zero. In other words, F has just one fixed point. On the other hand, F^2 is monotone increasing, and we can apply to it all the results in the previous paragraph. In particular, F can have any number of cycles of period 2, and every point tends either to the unique fixed point or to a cycle of period 2.

Now go back to the case when F is monotone increasing, and let y be a fixed point of F. In general, we expect that the graph of F will cross the diagonal at y and that a small (but non-pathological) change in F will merely move y slightly. But suppose that the graph touches the diagonal at y; then a small perturbation of F will in general have one of two effects:

- The graph of F may no longer meet the diagonal near y; so there is no longer a fixed point of F near y.
- The graph of F may meet the diagonal in two points near y; and using Proposition 4 one can see that one of them will be an attractor and the other a repellor.

(Of course, much worse things can happen, but this is the general non-pathological case.) If F depends on a parameter, one expects to move from one of these cases to the other as the parameter passes through the critical value at which the graph touches the diagonal at y. A change of this kind in the qualitative picture is called a **bifurcation**; the study of bifurcations is an important tool in analysing changes in the behaviour of a system.

The bifurcation described here, in which a pair of fixed points appear out of nothing (or alternatively, in which they annihilate each other) is called a **saddle-node bifurcation**, for reasons related to differential equations. (If maps had been studied first, we would probably be calling it an attractor-repellor bifurcation.) It can happen in virtually any space $\mathcal{S}$. It is one of the two simplest sorts of bifurcation; the other is the period-doubling bifurcation, which will appear shortly.

We obtain the next simplest example if we take $\mathcal{S}$ to be the closed unit interval $0 \leq x \leq 1$ and F to be given by

$$F(x) = cx(1-x) \quad \text{where} \quad 0 \leq c \leq 4,$$

This is often called the **logistic** map. If $0 \leq c \leq 1$ this map has just one fixed point, at $x = 0$. Moreover $F(x) < x$ for $x \neq 0$, so x_r is monotone decreasing as $r \to \infty$ and must tend to a limit ξ such that $F(\xi) = \xi$. Thus $\xi = 0$, and the entire space $\mathcal{S}$ is the domain of attraction of the origin.

Now suppose that $1 < c \leq 4$. Since $|F'(0)| = c > 1$, it follows from Proposition 3 that 0 is a repellor. However there is a second fixed point, at $x = (c-1)/c$. Notice that as c increase through 1 this fixed point emerges from the fixed point 0; this too is a bifurcation, but a slightly spurious one, for if we regard F as a map $\mathbf{R} \to \mathbf{R}$ then it will have two fixed points for all $c \neq 1$, and $c = 1$ is merely the moment when they coincide. Since $F'((c-1)/c) = 2 - c$, Proposition 3 shows that $(c-1)/c$ is an attractor if $c < 3$ and may be one also if $c = 3$. We actually have a much stronger result — that it is almost a global attractor.

Theorem 7 *If* $1 < c \leq 3$ *then* $(c-1)/c$ *is an attractor; and if* $0 < x < 1$ *then* $x_r \to (c-1)/c$ *as* $r \to \infty$.

Proof We deal first with the case $1 < c < 3$, so that we know that $(c-1)/c$ is an attractor. There is a largest open interval $\xi^- < (c-1)/c < \xi^+$ which lies in the domain of attraction of $(c-1)/c$. By direct calculation, neither 0 nor 1 lies in the domain of attraction of $(c-1)/c$. If say ξ^+ lay in that domain we would therefore have $\xi^+ \neq 1$; so by Lemma 5 there would be an open interval containing ξ^+ and lying in this domain, and this contradicts the maximality of ξ^+. Hence neither ξ^- nor ξ^+ lies in the domain of attraction. The image under F of the open interval (ξ^-, ξ^+) contains $(c-1)/c$, because that point is its own image. It is connected (because F is continuous) and lies in the domain of attraction; so it must lie in the open interval (ξ^-, ξ^+). But the images of ξ^- and ξ^+ do not lie in this interval; so by continuity each of these images must be either ξ^- or ξ^+. But the only fixed point other than $(c-1)/c$ is 0; so either $\xi^- = 0$, $\xi^+ = 1$ or $\{\xi^-, \xi^+\}$ is a cycle. In the former case we have proved the theorem. In the latter case

$$F^2(x) = c^2x(1 - (c+1)x + 2cx^2 - cx^3),$$

and since we already know the two roots 0 and $(c-1)/c$ of the quartic equation $F^2(x) = x$, its other two roots ξ^-, ξ^+ must be the roots of

$$c^2x^2 - c(c+1)x + (c+1) = 0. \tag{4}$$

The discriminant of this is $c^2(c+1)^2 - 4c^2(c+1) = c^2(c+1)(c-3) < 0$, so its roots are not real.

When $c = 3$ the same calculation gives

$$F^2(y + \tfrac{2}{3}) = \tfrac{2}{3} + y - 18y^3 - 27y^4,$$

so if y is small of either sign this lies between $\frac{2}{3}$ and $\frac{2}{3} + y$. Hence we can apply Proposition 4 to F^2, which shows that $\frac{2}{3}$ is an attractor for F^2; and it follows from the continuity of F at $\frac{2}{3}$ that it is also an attractor for F. Now both roots of (4) are $\frac{2}{3}$, so the rest of the proof works as before. We used F^2 rather than F because if y is small $F^2(\frac{2}{3} + y)$ is on the same side of $\frac{2}{3}$ as $\frac{2}{3} + y$ is, whereas $F(\frac{2}{3} + y)$ is on the opposite side. □

There is an alternative and simpler proof by drawing pictures.

How do things change as c increases through 3? We have just seen that a 2-cycle comes into existence at $c = 3$, given by the roots of (4) and initially coincident with the fixed point $(c-1)/c$; for obvious reasons this is called a **period-doubling** bifurcation. We can use Proposition 3 to show that this 2-cycle is stable for $3 < c < 1 + \sqrt{6}$; and an argument like that above (but more laborious) shows that it is stable when $c = 1 + \sqrt{6}$ also. It can be shown that for this range of values of c the domain of attraction of this 2-cycle consists of all points of $[0, 1]$ with countably many exceptions. (See the Examples Sheet.)

When c increases through $1 + \sqrt{6}$ the 2-cycle ceases to be an attractor. At $c = 1 + \sqrt{6}$ there is another period-doubling bifurcation; this creates a 4-cycle which is initially an attractor. As c increases, we have a sequence of period-doubling bifurcations; the limit of the corresponding values of c is approximately $3{\cdot}569942$. What happens after that will become clearer later in the course; but we can get some idea by considering the case $c = 4$, which by pure luck happens to be straightforward. But in order to study it, we need some more definitions.

Suppose we have two maps, $F : \mathcal{X} \to \mathcal{X}$ and $G : \mathcal{Y} \to \mathcal{Y}$. We shall say that F and G are **topologically conjugate** if there is a homeomorphism $\phi : \mathcal{X} \to \mathcal{Y}$ such that $\phi(F(x)) = G(\phi(x))$ for all x in $\mathcal{X}$ — a relation which can also be written $G = \phi \circ F \circ \phi^{-1}$. This implies that, so far as properties which can be described topologically are concerned, F and G behave in exactly the same way. We can define C^1 conjugacy by imposing the additional condition that ϕ should be a diffeomorphism — that is to say, that both ϕ and ϕ^{-1} should be once continuously differentiable. But it turns out that this requirement is more restrictive and therefore less useful. (See the Examples Sheet.) Both kinds of conjugacy are clearly equivalence relations.

In applications, we only expect to know F approximately; so we would like to say that F is **structurally stable** if small changes in F do not induce significant changes in the behaviour of the system. This is a vague requirement, and there are two separate phrases in it which need to be made more precise. If G is near F, we can regard the behaviour of the two systems as being essentially the same if G is topologically conjugate to F by a conjugacy which is close to the identity. But even on $\mathbf{R}$ we cannot simply consider all G for which $F(x) - G(x)$ is uniformly small, because that would allow G to have arbitrarily many fixed points near an isolated fixed point of F and so G could not be topologically conjugate to F. The straightforward remedy for this is to require F and G to be C^1 and use the C^1 metric, given on $\mathbf{R}^1$ by

$$d(F, G) = \max |F(x) - G(x)| + \max |F'(x) - G'(x)|.$$

If we want to consider more general F, we require $F + H$ to be topologically conjugate to F whenever H is C^1 with $(\max |H(x)| + \max |H'(x)|)$ small enough. The generalization to other one-dimensional spaces and indeed to any n-dimensional space is obvious; and one can give an ad hoc extension to any other space one is likely to encounter in this context.

Theorem 8 *If $c = 4$ the logistic map $F(x) = 4x(1 - x)$ on $\mathcal{S} = [0, 1]$ has periodic solutions of every period. None of them is an attractor. Moreover there are points x whose omega-set is the entire space $\mathcal{S}$.*

Proof The topological conjugacy given by $x = \sin^2(\frac{1}{2}\pi y)$ from $\mathcal{S}$ to itself implies $F(x) = \sin^2(\pi y)$; so F is conjugate to the map G given by

$$G(y) = \begin{cases} 2y \text{ if } 0 \le y \le \frac{1}{2}, \\ 2 - 2y \text{ if } \frac{1}{2} \le y \le 1. \end{cases}$$

This is most easily handled if y is written in binary, and it will be helpful to introduce the following notation. We use greek letters to denote finite or semi-infinite sequences of 0's and 1's, and put a bar over a greek letter to denote the transformation which replaces 0's by 1's and 1's by 0's. Then I claim that $G^m(y)$ for any m can be described as follows:

- Delete the first m digits of y, and put a bar over the rest if the original mth digit of y was a 1.

The proof of this is by induction on m, using $F^m(y) = F(F^{m-1}(y))$.

It is easy to see that the following recipe generates all periodic points of period m: take any sequence α of length m, and set $y = \alpha\alpha\ldots$ if the last digit of α is 0, or $y = \alpha\bar{\alpha}\alpha\bar{\alpha}\ldots$ if the last digit of α is 1. There are periodic points of least period m; take for example $\alpha = 0$ for $m = 1$, or $\alpha = 0\ldots01$ for $m > 1$ where there are $m-1$ 0's. If you have met Euler's function in number theory, it is not hard to write down a formula for the exact number of points of least period m; but the argument is complicated by the rule about the use of the bar.

Now suppose that y' and y'' have their first m symbols the same; then with the standard metric on $[0,1]$ we have

$$|G^m(y') - G^m(y'')| = 2^m|y' - y''|.$$

It follows easily that every periodic point of period m is a repellor for G^m; so no cycle can be an attractor.

To construct a point y_0 whose omega-set is the entire space, make a list $\alpha_1, \alpha_2, \ldots$ of all finite sequences of 0's and 1's and write

$$\alpha = \alpha_1 0 \alpha_2 0 \ldots .$$

(The 0's here are simply to ensure that the bar operator does not complicate the argument.) Now each α_m is the leading part of some $G^r(\alpha)$, so we can approximate as closely as we like to any given y by suitable $G^r(\alpha)$ with arbitrarily large r. □

Without straining the normal meaning of the word, this kind of behaviour can reasonably be described as **chaotic**. I shall not give a formal definition of **chaos** in this course, for two reasons: there are already a large number of not-quite-equivalent definitions in the literature; and it seems in general very difficult to prove that behaviour is chaotic even when there can be no reasonable doubt that it is so. There are however several useful definitions which belong to the same family of ideas.

Let $F : \mathcal{S} \to \mathcal{S}$ be a map. We say that F is **transitive** if there is a point x in $\mathcal{S}$ whose orbit is dense in $\mathcal{S}$. We say that F is **topologically transitive** if $\bigcup_{r \geq 0} F^r V$ is dense in $\mathcal{S}$ for every non-empty open subset V of $\mathcal{S}$.

Proposition 9 *If a map F is transitive then it is topologically transitive.*

Proof If the orbit of P is dense in $\mathcal{S}$ and V is a non-empty open subset of $\mathcal{S}$, then $F^m P$ lies in V for some m. But now $\bigcup_{r \geq m} F^r P \subset \bigcup_{r \geq 0} F^r V$ and the left hand side is dense in $\mathcal{S}$. □

It is also true, though more difficult to prove, that under moderate conditions on $\mathcal{S}$ topological transitivity implies transitivity.

Now assume that $\mathcal{S}$ is a metric space. We shall say that $F : \mathcal{S} \to \mathcal{S}$ has **sensitive dependence on initial conditions** if there exists $d > 0$ such that for any x in $\mathcal{S}$ and any $\epsilon > 0$ there exists y in $\mathcal{S}$ and $n > 0$ such that

$$|x - y| < \epsilon \quad \text{and} \quad |F^n x - F^n y| > d.$$

In other words, points which start close together do not necessarily stay close together. It is usually said that both transitivity and sensitive dependence on initial conditions ought to be part of any sensible definition of chaos. But note that sensitive dependence on initial conditions is not preserved under topological conjugacy. For let $\mathcal{X}$ be $(0, \infty)$ with $F(x) = 2x$ and let $\mathcal{Y}$ be $\mathbf{R}$ with $G(y) = y + \log 2$; the map $\mathcal{X} \to \mathcal{Y}$ given by $y = \log x$ is a topological conjugacy (and even a C^1 conjugacy) $\mathcal{X} \to \mathcal{Y}$ which takes F to G, but F has sensitive dependence on initial conditions whereas G does not.

2. Maps of the circle

We define a **circle map** to be a continuous map of the circle $\mathbf{R}/\mathbf{Z}$ onto itself; the theory is almost entirely concerned with invertible maps, but we do not need to require invertibility immediately. Some authors prefer to define the circle as the point set in the complex plane given by $|z| = 1$; this makes no real difference but causes a factor $2\pi i$ to appear in many places. There is a canonical covering map $\psi : \mathbf{R} \to \mathbf{R}/\mathbf{Z}$. If f is a circle map, we shall say that $F : \mathbf{R} \to \mathbf{R}$ is a **lift** of f if F is continuous and $f\psi = \psi F$.

Proposition 10 *Every circle map has a lift, which is unique up to translation by an integer.*

Proof Choose A in $\mathbf{R}$ which is a lift of $f(\psi 0)$; in this paragraph we confine ourselves to lifts F such that $F(0) = A$. If x is near enough to 0 then by continuity $f(\psi x)$ has a unique lift which lies in the interval $(A - \frac{1}{3}, A + \frac{1}{3})$; this interval is homeomorphic to its image in $\mathbf{R}/\mathbf{Z}$, so in a neighbourhood of 0 we can lift f to a function F which is clearly continuous. Moreover if F_1 and F_2 are defined in a common interval $I \subset \mathbf{R}$ containing 0 and satisfy $\psi F_1(x) = \psi F_2(x) = f(\psi x)$ in that interval, then $F_1(x) - F_2(x)$ is continuous, integer-valued and zero at 0; so it is zero throughout I. Hence if we have lifts F_1 in I_1 and F_2 in I_2, where I_1, I_2 are intervals both containing 0, then they agree on $I_1 \cap I_2$ and hence we can patch them together to obtain a lift to $I_1 \cup I_2$. Now consider the union I of all intervals containing 0 to which we can lift f, and let F be the lift to I. If I is not the whole of $\mathbf{R}$ it has an end-point x_0. But in the same way we can find an interval I^* containing x_0 and a lift F^* of f to I^*. Since $F^*(x) - F(x)$ is continuous and integer-valued in $I \cap I^*$, it is constant; and we can subtract this constant (which is an integer) from F^* and then patch F and F^* together. This extends F beyond x_0, so we have a contradiction; in other words our lift was to the whole of $\mathbf{R}$.

Now drop the hypothesis $F(0) = A$; since $F_1(x) - F_2(x)$ is continuous and integer-valued for any two lifts F_1, F_2 of f, it is a constant integer. □

Proposition 11 *If F, G are lifts of the circle maps f, g then FG is a lift of fg. In particular F^m is a lift of f^m for all $m > 0$, and for all m if f is invertible.*

Proof The first sentence is obvious since FG is continuous and $F(G(x))$ is a lift of $f(g(x))$ by considering the commutative diagram

$$\begin{array}{ccccc}
\mathbf{R} & \xrightarrow{G} & \mathbf{R} & \xrightarrow{F} & \mathbf{R} \\
\downarrow & & \downarrow & & \downarrow \\
\mathbf{R}/\mathbf{Z} & \xrightarrow{g} & \mathbf{R}/\mathbf{Z} & \xrightarrow{f} & \mathbf{R}/\mathbf{Z}.
\end{array}$$

If f is invertible, F is strictly monotone in $(0,1)$; for otherwise there would be x_1, x_2 with $0 < x_1 < x_2 < 1$ such that $F(x_1) = F(x_2)$, and f would not be invertible. By continuity F is monotone in the closed interval $[0,1]$, and $F(1) - F(0) > 0$. For any integer n we have

$$F(x+n) - F(x) = F(n) - F(0) = n(F(1) - F(0)) \tag{5}$$

because the left hand side is continuous and integer-valued, and in

$$F(n)-F(0) = (F(n)-F(n-1)) + (F(n-1)-F(n-2)) + \ldots + (F(1)-F(0))$$

each difference on the right is $F(1) - F(0)$. It follows easily that F is strictly monotone in $(-\infty, \infty)$ and maps $(-\infty, \infty)$ onto itself; hence F is invertible. Now $F^{-1}F$ is the identity map $x \mapsto x$ and is therefore a lift of $f^{-1}f$. For any z in $\mathbf{R}$ write $F^{-1}(z) = y$; then

$$\psi F^{-1}(z) = \psi F^{-1}(F(y)) = f^{-1}(f(\psi y)) = f^{-1}(\psi z),$$

so that F^{-1} is a lift of f^{-1}. $\square$

In the course of these proofs we saw that $d = F(x+1) - F(x)$ is an integer independent of x and the choice of F; it is called the **degree** of f. We can write (5) as

$$F(x+n) = F(x) + n\deg(f).$$

Moreover if F, G are lifts of f, g then

$$F(G(x+1)) = F(G(x) + \deg(g)) = F(G(x)) + \deg(f)\deg(g)$$

so that

$$\deg(fg) = \deg(f)\deg(g) = \deg(gf).$$

If f is invertible, this implies $\deg(f)\deg(f^{-1}) = 1$ whence $\deg(f) = \pm 1$.

The interesting circle maps are those which are invertible and have degree $+1$; they are generally called **orientation-preserving** and we shall confine ourselves to them for the rest of the section.

Theorem 12 *A continuous map $F : \mathbf{R} \to \mathbf{R}$ is a homeomorphism if and only if F is strictly monotone and onto. Moreover, the following three statements are equivalent:*

(*i*) *F is a lift of an orientation-preserving homeomorphism of $\mathbf{R}/\mathbf{Z}$.*

(*ii*) *$F : \mathbf{R} \to \mathbf{R}$ is a homeomorphism and $F(x+1) = F(x) + 1$.*

(*iii*) *$F : \mathbf{R} \to \mathbf{R}$ is a homeomorphism and $F(x) = x + \phi(x)$ where ϕ is periodic with period 1. If F is any lift of the orientation-preserving map f then f determines ϕ up to addition of an arbitrary integer.*

Proof Suppose that F is strictly monotone increasing (for example) and onto; then $F(x_1) < F(x) < F(x_2)$ for any $x_1 < x < x_2$, and because F is continuous, every point of $(F(x_1), F(x_2))$ is an $F(y)$ for some y; because F is strictly monotone y cannot lie outside (x_1, x_2). Similarly the image of (x_1, x_2) is in $(F(x_1), F(x_2))$. It follows that F maps open intervals to open intervals and therefore open sets to open sets; this completes the proof that F is a homeomorphism. Conversely suppose that F is not strictly monotone; then we have some situation like $x_1 < x_2 < x_3$ and $F(x_1) \le F(x_3) \le F(x_2)$. By continuity there exists x with $x_1 \le x \le x_2$ and $F(x) = F(x_3)$; so F is not a homeomorphism.

Now assume (i). That F is strictly monotone increasing was shown in the proof of Proposition 11, and we have just shown that $\deg(f) = \pm 1$; here we must take the upper sign since F is increasing. Because F is continuous it takes every value in $(F(0), F(1) = F(0) + 1)$. Let y be in $\mathbf{R}$ and choose n so that $F(0) \le y - n \le F(0) + 1$; then there is an x in $[0, 1]$ such that $F(x) = y - n$, whence $y = F(x + n)$. This completes the proof of (ii). Conversely if (ii) holds then F induces a map $f : \mathbf{R}/\mathbf{Z} \to \mathbf{R}/\mathbf{Z}$ which is continuous and has degree 1. It is invertible because $f(\psi x_1) = f(\psi x_2)$ implies $F(x_1) = F(x_2) + n = F(x_2 + n)$ for some integer n; so $x_1 = x_2 + n$, and this last equality is equivalent to $\psi x_1 = \psi x_2$. Finally (ii) is obviously equivalent to (iii). $\square$

It follows from (ii) that if f, g are orientation-preserving then so are fg and f^{-1}. The importance of (iii) is that it makes it easy to construct circle maps. For choose a differentiable function $\phi : \mathbf{R} \to \mathbf{R}$ which is periodic with period 1 and satisfies $\phi'(x) > -1$ for all x; then F is strictly increasing and continuous, so that F is a lift of an invertible f.

We can associate with any orientation-preserving circle map a **rotation number** which, roughly speaking, measures the rotational effect of f averaged over its iterates. To define it we proceed as follows. Given f, choose a lift F and define ϕ as in Theorem 12(iii). I claim that

$$F^r(x) = x + \phi(x) + \phi(F(x)) + \ldots + \phi(F^{r-1}(x)) \tag{6}$$

for $r > 0$. The proof is by induction, the inductive step being

$$F^{r+1}(x) = F^r(F(x)) = F(x) + \phi(F(x)) + \ldots + \phi(F^r(x));$$

the case $r = 1$ is trivial.

Theorem 13 *With the notation above, there is a real number $\rho(F)$ such that*

$$\left| \frac{F^q(x) - x}{q} - \rho(F) \right| < \frac{1}{q}$$

for all x in $\mathbf{R}$ and all $q > 0$.

Proof Take a positive integer q. Since F^q is a lift of the orientation-preserving circle map f^q, $x \mapsto F^q(x) - x$ is continuous and periodic with period 1. Hence it attains its minimum, say A_q, at some point u. Now $u \le x \le u + 1$ implies

$$A_q \le F^q(x) - x \le F^q(u+1) - u = F^q(u) + 1 - u = A_q + 1;$$

and since $F^q(x) - x$ is periodic with period 1, this implies

$$A_q \le F^q(x) - x \le A_q + 1 \tag{7}$$

for all x. Choose any real y and any integer $n > q$, and write $n = mq + r$ where $0 \le r < q$; taking $x = y, F^q(y), \ldots, F^{(m-1)q}(y)$ in (7) and adding, we obtain

$$mA_q \le F^{mq}(y) - y \le m(A_q + 1).$$

In exactly the same way we can obtain

$$rA_1 \le F^{mq+r}(y) - F^{mq}(y) \le r(A_1 + 1);$$

adding these last two equations and using $m = (n - r)/q$ gives

$$\left(\frac{n-r}{q}\right) A_q + rA_1 \le F^n(y) - y \le \left(\frac{n-r}{q}\right)(A_q + 1) + r(A_1 + 1).$$

Dividing by n and letting n tend to infinity, we obtain

$$\frac{A_q}{q} \leq \liminf_{n\to\infty} \frac{F^n(y)-y}{n} \leq \limsup_{n\to\infty} \frac{F^n(y)-y}{n} \leq \frac{A_q+1}{q}. \tag{8}$$

Here the two middle terms are independent of q and their difference is bounded by q^{-1}; so they must be equal. In other words

$$\lim \frac{F^n(y)-y}{n} = \rho(F,y)$$

exists for each y. Now (8) reduces to

$$\frac{A_q}{q} \leq \rho(F,y) \leq \frac{(A_q+1)}{q}; \tag{9}$$

and it follows that

$$|\rho(F,x)-\rho(F,y)| \leq 2/q$$

for all x, y. Since q is arbitrary and the left hand side is independent of q, the left hand side must vanish. Hence $\rho(F,x)$ is independent of x and can be written simply as $\rho(F)$. Finally (9) and (7) give the Theorem. $\square$

However, $\rho(F)$ does depend on the choice of the lift F of f. Suppose that F_1 is another lift of f, so that $F_1 = F + N$ for some integer N; then $F_1^n = F^n + nN$ by induction, and therefore $\rho(F_1) = \rho(F) + N$. To get a rotation number which depends only on f and not on the choice of lift, we have two closely related ways to proceed. One, which I shall use, is to define $\rho(f)$ to be the image of $\rho(F)$ in $\mathbf{R}/\mathbf{Z}$; the other, which some authors use, is to define $\rho(f)$ to be $\exp(2\pi i\rho(F))$. The justification for calling $\rho(f)$ a rotation number is that for the simple rotation map $x \mapsto x + \alpha$, we have

$$F(x) = x+\alpha, \quad F^n(x) = x+n\alpha, \quad \rho(F) = \alpha, \quad \rho(f) = \alpha.$$

It is trivial to check that

$$\rho(F^m) = \lim \frac{F^{mn}(x)-x}{n} = \lim m\frac{F^{mn}(x)-x}{mn} = m\rho(F)$$

and therefore $\rho(f^m) = m\rho(f)$. A more interesting result is

Proposition 14 *If f, g are orientation-preserving then $\rho(gfg^{-1}) = \rho(f)$.*

Proof Write $h = gfg^{-1}$ and let F, G be lifts of f, g respectively; then G^{-1} is a lift of g^{-1} and $H = GFG^{-1}$ is a lift of h. Since g is orientation-preserving, $G(x) = x + \phi(x)$ for some ϕ which is periodic and therefore bounded. Now

$$\rho(H) = \lim \frac{H^n(x) - x}{n} = \lim \frac{GF^nG^{-1}(x) - x}{n}.$$

Putting $x = G(0)$ we obtain

$$\begin{aligned}\rho(H) &= \lim \frac{GF^n(0) - G(0)}{n} = \lim \frac{F^n(0) + \phi(F^n(0)) - G(0)}{n} \\ &= \lim \frac{F^n(0)}{n} = \rho(F)\end{aligned}$$

because $\phi(F^n(0)) - G(0)$ is bounded as $n \to \infty$. □

Theorem 15 *The map $F \mapsto \rho(F)$ is continuous. In other words, given any lift F and any $\epsilon > 0$ there exists $\delta > 0$ such that for any lift G satisfying $|G(x) - F(x)| < \delta$ for all x we have $|\rho(G) - \rho(F)| < 3\epsilon$.*

Proof Choose n such that $n\epsilon > 1$; then

$$|\rho(G) - \rho(F)| \leq \left|\rho(G) - \frac{G^n(0)}{n}\right| + \left|\frac{G^n(0) - F^n(0)}{n}\right| + \left|\frac{F^n(0)}{n} - \rho(F)\right|.$$

By Theorem 13 the first and third terms on the right are bounded by $n^{-1} < \epsilon$; so we have only to deal with the second. Each map $x \mapsto F^k(x) - x$ for $0 < k \leq n$ is periodic and continuous on $\mathbf{R}$ and hence uniformly continuous there; so there exists $\delta > 0$ such that

$$|y - x| < \delta \quad \text{implies} \quad |F^k(y) - F^k(x)| < \epsilon \text{ for } 0 < k \leq n.$$

But $|G^n(0) - F^n(0)| \leq \sum_{m=0}^{n-1} |F^mGG^{n-m-1}(0) - F^mFG^{n-m-1}(0)|$ and each term on the right has the form $|F^mG(x) - F^mF(x)|$ for some x. Hence if $|G(x) - F(x)| < \delta$ for all x each term in the sum is bounded by ϵ, and $|G^n(0) - F^n(0)| < n\epsilon$. □

There is a corresponding result for the map $f \mapsto \rho(f)$, which is more complicated only because of the need to define what it means for two circle maps to be near; the result then follows from Theorem 15 and the obvious fact that two circle maps which are near have lifts which are near.

There is a fundamental dichotomy between the cases when $\rho(f)$ is rational and when it is irrational. Since the definition of $\rho(f)$ involves taking limits, we do not expect to be able to compute $\rho(f)$ from its definition except in very favourable cases; but the next theorem shows that we can expect to compute it whenever it is rational, and the discussion after the proof of Theorem 17 shows that it is rational more often than one might expect.

Theorem 16 *An orientation-preserving circle map f has periodic points of least period q if and only if $\rho(F) = p/q$ for some integer p prime to q.*

Proof Suppose first that f has periodic points of period m and let ψy be one of them. Then $f^m(\psi y) = \psi y$ and therefore $F^m(y) = y + k$ for some integer k. It follows that $F^{mn}(y) = y + nk$, whence $\rho(F) = k/m$ and $\rho(f)$ is rational with least numerator a factor of m.

Conversely, suppose that $\rho(F) = p/q$ where p, q are coprime integers. Since F^q is a lift of f^q, $F^q(x) - x$ is periodic and continuous, and hence bounded; let its minimum and maximum in $[0, 1]$, and therefore also in $\mathbf{R}$, be α, β respectively. Thus $\alpha \leq F^q(x) - x \leq \beta$ and there exist values of x for which we have equality on whichever side we choose. It follows by iteration that $n\alpha \leq F^{nq}(x) - x \leq n\beta$; dividing by nq and taking limits, we obtain

$$\frac{\alpha}{q} \leq \rho(F) = \frac{p}{q} \leq \frac{\beta}{q}$$

and therefore $\alpha \leq p \leq \beta$. By continuity $F^q(y) - y = p$ for some y, so that the image of y in $\mathbf{R}/\mathbf{Z}$ is periodic with period q. Moreover, q is the least period of y, by the first part of the proof.

Now suppose that f also had a periodic point with least period $q_1 \neq q$, and let its lift be y_1. By the first part of the proof $F^{q_1}(y_1) = y_1 + \rho(F)q_1$, so $\rho(F)q_1$ is an integer and q_1 is a multiple of q. Since y_1 does not have period q, by hypothesis, $F^q(y_1) \neq y_1 + p$; without loss of generality we can assume that $F^q(y_1) > y_1 + p$. Since all powers of F are monotone increasing, we obtain by induction

$$F^{qr}(y_1) > F^{q(r-1)}(y_1 + p) = F^{q(r-1)}(y_1) + p \geq y_1 + rp$$

for all $r > 0$. Setting $r = q_1/q$ we obtain a contradiction. □

If $\rho(F) = p/q$ is rational, we have a result which is usually very much stronger than Theorem 15. Recall that in this case we know that $F^q(x) - x$ takes the value p for some x.

Theorem 17 *Suppose that F is a lift of an orientation-preserving circle map f, that $\rho(F) = p/q$ and that $F^q(x) - x - p$ takes both positive and negative signs. Then $\rho(G) = p/q$ for any G which is a lift and is close enough to F.*

Proof Suppose that G is near enough to F; then G^q is near to F^q as in the proof of Theorem 15. Let x_1, x_2 be such that

$$F^q(x_1) - x_1 - p > 0 > F^q(x_2) - x_2 - p;$$

then by the result in the previous sentence

$$G^q(x_1) - x_1 - p > 0 > G^q(x_2) - x_2 - p.$$

Applying $G^{q(n-1)}$, which is strictly monotone, to $G^q(x_1) > x_1 + p$ gives

$$G^{qn}(x_1) > G^{q(n-1)}(x_1 + p) = G^{q(n-1)}(x_1) + p,$$

whence by induction $G^{qn}(x_1) - x_1 > np$. This implies that $\rho(G) \geq p/q$; and the opposite inequality comes in the same way by considering x_2. □

Suppose that f and therefore F depends continuously on a parameter λ. Then $\rho(F)$ is a continuous function of λ, by Theorem 15. When $\rho(F)$ takes a rational value we can normally expect F to satisfy the conditions of Theorem 17, for otherwise the graphs of $y = F^q(x)$ and $y = x + p$ would have to touch. So we normally expect $\rho(F)$ to have an interval of constancy whenever it takes a rational value. For obvious reasons, the graph of $\rho(F)$ against λ is called the Devil's Staircase.

We now turn to the case when $\rho(F)$ is irrational and therefore f has no periodic points. For each z in $\mathbf{R}/\mathbf{Z}$, no two terms of the sequence $z, f(z), f^2(z), \ldots$ are equal; so the set $\Omega(z)$ of limit points of this sequence is closed and non-empty. If x is in $\Omega(z)$ there is a subsequence of the $f^n(z)$ which tends to x. Applying f to this statement, there is a subsequence which tends to $f(x)$; and applying f^{-1} there is a subsequence which tends to $f^{-1}(x)$. Thus f maps $\Omega(z)$ onto itself; we shall call any such set **f-invariant**. It will also be convenient to say that z is a **wandering point** under f if there is an open neighbourhood $\mathcal{N}$ of z such that $\mathcal{N} \cap f^n(\mathcal{N})$ is empty for all $n > 0$. All other points are called **non-wandering**.

Theorem 18 *Let f be an orientation-preserving circle map with $\rho(f)$ irrational. Then* (*i*) *the set* $\Omega(z) = \Omega$ *is independent of* z *and is the smallest closed, non-empty,* f*-invariant subset of* $\mathbf{R}/\mathbf{Z}$*;*

(*ii*) Ω *is the set of non-wandering points of* $\mathbf{R}/\mathbf{Z}$*;*

(*iii*) Ω *is either the whole of* $\mathbf{R}/\mathbf{Z}$ *or a Cantor set in* $\mathbf{R}/\mathbf{Z}$.

Proof We have already shown that every $\Omega(z)$ is closed, non-empty and f-invariant. Let E be any closed, non-empty, f-invariant subset of $\mathbf{R}/\mathbf{Z}$, and let I be a component of its complement; thus I is a maximal interval disjoint from E. If $I \cap f^n(I)$ is non-empty for some $n > 0$ then f^n is an orientation-preserving map of I into itself which does not map the end-points of I into I because they lie in E and therefore map into E. By continuity (and preservation of orientation) f^n must therefore map each end-point of I to itself, contrary to the hypothesis that f has no periodic points. Thus I is disjoint from $f^n(I)$, so that I is contained in the wandering set W.

Now let x be in $\Omega(z)$; then there is a subsequence of the $f^n(z)$ which tends to x. Let $\mathcal{N}$ be any neighbourhood of x and let r, s with $r > s$ be such that $f^r(z), f^s(z)$ both lie in $\mathcal{N}$. Then $f^{r-s}(f^s(z)) = f^r(z)$ lies in $f^{r-s}(\mathcal{N})$ and therefore $\mathcal{N} \cap f^{r-s}(\mathcal{N})$ is not empty; in other words x lies in the non-wandering set. In view of the result in the previous paragraph, $\Omega(z)$ is precisely the non-wandering set, which proves (i) and (ii).

Since Ω is f-invariant, so is its interior and therefore its boundary $\partial\Omega$. Since Ω is a closed f-invariant set, either it is empty or it contains Ω. In the first case Ω is the entire circle. In the latter case Ω contains no arcs. It also contains no isolated points; for if x_0 was an isolated point of Ω and I an interval containg x_0 but having no other point of Ω in its closure then there would exist $m > 0$ such that $f^m(I)$ meets I because x_0 is non-wandering. Now $f^m(I)$ contains exactly one point of Ω and its end-points are not in Ω, because I has the same properties. The one point of $f^m(I) \cap \Omega$ cannot be x_0 because then x_0 would be periodic; so it must be the next point of Ω to x_0 in one direction or the other. Also x_0 cannot be an end-point of $f^m(I)$ because its end-points do not lie in Ω. So we can choose an interval $I_1 \subset I$ and an integer m_1 with the same properties as I, m and such that I_1 is disjoint from $f^m(I)$. Necessarily $m_1 \neq m$. Repeating this again if necessary, we obtain integers $m' \neq m''$ such that both $f^{m'}(x_0)$ and $f^{m''}(x_0)$ are the next point of Ω to x_0 on the same side of x_0, and are therefore equal. This again contradicts the fact that f has no periodic points.

We have therefore shown that Ω is compact, non-empty and totally disconnected. This is just the definition of a Cantor set. $\square$

It can be shown by examples that both of these cases do happen. The obvious example of the first is the rotation map $x \mapsto x + \rho$ where ρ is irrational; examples of the second are always pathological, for the reason which follows. Clearly topological conjugacy acts in the natural way on Ω, because each of the defining properties (i) and (ii) in Theorem 18 is purely topological. This proves the easy part of the following theorem; the rest is too difficult for this course.

Theorem 19 (Denjoy) *An orientation-preserving circle map f with $\rho(f)$ irrational is topologically conjugate to a rotation if and only if $\Omega = \mathbf{R}/\mathbf{Z}$. This certainly holds whenever f is C^2.*

A rotation can never be structurally stable, because a rotation f has periodic points of period q if and only if $\rho(f) = p/q$ for some integer p; hence it is C^1-close to rotations which are not topologically conjugate to it. Since structural stability is invariant under topological conjugacy, it follows that if f is C^2 and $\rho(f)$ is irrational, then f cannot be topologically stable. (Not surprisingly, the f with $\rho(f)$ irrational and Ω a Cantor set are also not structurally stable.) Also if Ω is to be a Cantor set then f cannot be C^2; so one expects it to be rather ugly.

3. Sequence spaces and the shift map

We consider an abstract dynamical system which is conjugate to (and hence serves as a model for) many concrete examples. Let $\mathcal{A}$ be a finite alphabet of symbols, and let Σ be the space of all semi-infinite sequences $\mathbf{a} = (a_1, a_2, \ldots)$ where all the a_j are in $\mathcal{A}$. We shall frequently call the elements of $\mathcal{A}$ **letters**, and finite sequences of them **words**. We choose any topology on this space such that two sequences $\mathbf{a}, \mathbf{b}$ are close if and only if they agree on a long initial segment; for example we can use the metric

$$d(\mathbf{a}, \mathbf{b}) = 2^{-n}$$

where the sequences $\mathbf{a}, \mathbf{b}$ differ for the first time in the $(n+1)$th position. The natural map to apply to Σ is the **shift map**

$$\sigma : \ (a_1, a_2, \ldots) \mapsto (a_2, a_3, \ldots)$$

which consists of deleting the first element of a sequence; clearly it is continuous.

Proposition 20 *If $\mathcal{A}$ contains 2 elements then Σ can be identified with the classical Cantor set.*

Proof Let the alphabet be 0,2 and map the sequence $(a_1, a_2, \ldots)$ onto the ternary number $\cdot a_1 a_2 \ldots$. All the numbers which we obtain lie in the interval [0,1], and the map is one-one into; for although numbers of the form $a/3^n$ have two ternary representations, at least one of them uses the digit 1 and is therefore forbidden. Since $a_1 \neq 1$, the open interval $(\frac{1}{3}, \frac{2}{3})$ is not in the image; since $a_2 \neq 1$, the open intervals $(\frac{1}{9}, \frac{2}{9})$ and $(\frac{7}{9}, \frac{8}{9})$ are not in the image, and so on. Under this identification the map σ corresponds to multiplying by 3 and then reducing mod 1. □

For an arbitrary finite alphabet $\mathcal{A}$ the space Σ is a Cantor set in a more general sense; if $\mathcal{A}$ contains n elements a similar argument works in the scale of $2n-1$. The spaces we shall consider will all be subspaces of Σ which are mapped into themselves by σ, and with trivial exceptions each of them will be a Cantor set in the sense of Theorem 18.

Proposition 21 *The map $\sigma : \Sigma \to \Sigma$ is continuous and onto. It has periodic points of every period, and the set of all periodic points is dense. There are points whose omega-set is the entire space Σ.*

Proof We have $d(\sigma\mathbf{a}, \sigma\mathbf{b}) \leq 2d(\mathbf{a}, \mathbf{b})$, so that σ is continuous. In an obvious notation, any $\mathbf{a}$ is the image of $(a, \mathbf{a})$ for a in $\mathcal{A}$, so the map is onto. If α is a sequence of n letters in which some letter occurs just once (so that α is not obtained by repeating a shorter sequence) then $(\alpha, \alpha, \ldots)$ is periodic with period exactly n. To find a periodic point near a given point $\mathbf{a}$, let α be a long initial segment of $\mathbf{a}$; then $(\alpha, \alpha, \ldots)$ does what we want. Finally, the construction of a point whose omega-set is the entire space is the same as the construction for the corresponding result in Theorem 8. □

We can also consider the **doubly infinite sequence space** Σ' cosisting of all $(\ldots, a_{-1}, a_0, a_1, \ldots)$ where all the a_j are in $\mathcal{A}$. We define the topology on Σ' by means of the metric

$$d(\mathbf{a},\, \mathbf{b}) = 2^{-n}$$

where now n is the least non-negative integer such that $a_n \neq b_n$ or $a_{-n} \neq b_{-n}$. The shift map σ, which is now a homeomorphism, is given by $\sigma\mathbf{a} \mapsto \sigma\mathbf{b}$ where $b_j = a_{j+1}$. Almost all our results for infinite sequences hold (with essentially the same proof) for doubly infinite sequences.

The theory becomes more interesting if we restrict ourselves to subspaces $\mathcal{S}$ of Σ which are mapped onto themselves by σ. There are two natural ways in which this can come about.

First, consider a finite network $\mathcal{G}$ consisting of M vertices, numbered from 1 to M and joined by a certain number of directed edges. (Note that we allow edges which run from a vertex to itself.) If $\mathcal{A} = \{1, \ldots, M\}$ we can represent this network by an $M \times M$ matrix A with elements a_{ij}, where $a_{ij} = 1$ if there is an edge from vertex i to vertex j and $a_{ij} = 0$ otherwise. Conversely, any $M \times M$ matrix whose elements are 0 or 1 represents such a network. An infinite or doubly infinite path in such a network can be specified by the list of vertices visited, and therefore corresponds to an element of Σ or Σ' respectively; conversely such an element corresponds to a path in the network if and only if $a_{ij} = 1$ for any two consecutive elements i, j of the sequence. The action of σ on Σ corresponds to the same path with its first step removed, so if $\mathcal{S}$ is the space of sequences corresponding to the paths on the network, σ maps $\mathcal{S}$ to itself.

This representation enables us to count the number of periodic paths of any given period — or more generally the number of paths of length ℓ from i to j. For let A be the matrix which represents the network; in forming A^ℓ each path of length ℓ from i to j contributes 1 to the element in the (i, j)-th position. Thus the number of periodic paths of period (but not necessarily least period) ℓ is just the trace of A^ℓ; here a periodic path of least period $\ell'|\ell$ is counted ℓ' times, once for each of its possible starting points. But A satisfies its characteristic equation $\phi(A) = 0$; if we multiply this by A^m and take traces, we obtain a linear recurrence relation of degree M with constant coefficients for the $\operatorname{tr}(A^\ell)$

Let $c_\ell = c_\ell(A) = \operatorname{tr}(A^\ell)$ be the number of periodic paths of period ℓ, counted according to the convention above. We treat the c_ℓ as coefficients, defining the **counting function** or **zeta-function** of the network to be

$$\zeta(x) = \sum_{\ell=0}^{\infty} c_\ell x^\ell.$$

(There is some arbitrariness in the choice of the constant term here; but for what follows it hardly matters.)

Theorem 22 *The zeta-function associated with a network is rational.*

Proof Let the characteristic polynomial of A be

$$\phi(X) = \sum_{m=0}^{M} b_m X^m,$$

so that $\phi(A) = 0$. Then

$$x^M \phi(x^{-1})\zeta(x) = \operatorname{tr} \sum_{\ell=0}^{\infty} x^{\ell+M} \phi(x^{-1}) A^\ell = \operatorname{tr} \sum_{\ell=0}^{\infty} \sum_{m=0}^{M} x^{\ell+M-m} b_m A^\ell.$$

In this we interchange the two summations, split off the terms with $\ell < m$ (which give a polynomial $g(x)$ of degree $M-1$) and write $\ell + m$ for ℓ in the remaining terms. In this way we obtain

$$x^M \phi(x^{-1})\zeta(x) - g(x) = \operatorname{tr} \sum_{m=0}^{M} \sum_{\ell=0}^{\infty} x^{\ell+M} b_m A^{\ell+m} = \operatorname{tr} \sum_{\ell=0}^{\infty} x^{\ell+M} A^\ell \phi(A).$$

Since $\phi(A) = 0$, this proves the Theorem. The denominator of the expression for $\zeta(x)$ is essentially the characteristic polynomial; but there is nothing obvious that one can say about the numerator. □

There is an obvious generalization of the subshift space associated with a network; choose a set of words $\mathcal{W}$, and consider the space $\Sigma_{\mathcal{W}}$ consisting of those sequences none of whose subsequences is in $\mathcal{W}$. Clearly σ maps $\Sigma_{\mathcal{W}}$ to itself, and our previous example is just the special case when all the forbidden words have length 2. We call $\Sigma_{\mathcal{W}}$ a **subshift space**, and when $\mathcal{W}$ is finite (which is by far the most interesting case) we call it a **subshift space of finite type**.

Proposition 23 *$\Sigma_{\mathcal{W}}$ is a closed subset of Σ for any $\mathcal{W}$. Conversely, let Π be a closed subset of Σ which is mapped into itself by σ; then $\Pi = \Sigma_{\mathcal{W}}$ for some set $\mathcal{W}$.*

Proof Suppose that $\mathbf{a}$ is in the closure of $\Sigma_{\mathcal{W}}$ but not in $\Sigma_{\mathcal{W}}$. Thus $\mathbf{a}$ contains a forbidden word α, and for some n this word lies entirely within the first n letters of $\mathbf{a}$. But by hypothesis there is an element of $\Sigma_{\mathcal{W}}$ so close to $\mathbf{a}$ that the two agree in their first n letters — so that this element of $\Sigma_{\mathcal{W}}$ would also contain the forbidden word α.

To prove the converse we construct $\mathcal{W}$. For each $\mathbf{x}=(x_1, x_2, \ldots)$ not in Π, there is a neighbourhood of $\mathbf{x}$ which does not meet Π, and hence an r such that no element of Π begins $x_1, x_2, \ldots, x_r$. If an element of Π contained the word $x_1, x_2, \ldots, x_r$ then by repeated application of σ we could bring this word to the left hand end, contrary to hypothesis; so $x_1, x_2, \ldots, x_n$ is a forbidden word for Π. Take $\mathcal{W}$ to be the set of all words constructed in this way; then every element of Σ not in Π starts with a forbidden word. So $\Pi = \Sigma_{\mathcal{W}}$. □

Theorem 24 *Any subshift space of finite type is topologically conjugate to a subshift space coming from a network (though with a larger alphabet).*

Proof Let $\Sigma_{\mathcal{W}}$ be a subshift space on an alphabet $\mathcal{A}$ with $\mathcal{W}$ finite. If M is such that each word in $\mathcal{W}$ has length at most M, let $\mathcal{A}^*$ be the alphabet whose letters correspond to those words of length M in $\mathcal{A}$ which do not have any element of $\mathcal{W}$ as a sub-word. Let $\mathcal{W}^*$ consist of all those words of length 2 in $\mathcal{A}^*$ whose corresponding word in $\mathcal{A}$ contains some element of $\mathcal{W}$ as a sub-word. Consider the subshift space on the alphabet $\mathcal{A}^*$ with $\mathcal{W}^*$ as its set of forbidden words; then the corresponding sequences in the original alphabet $\mathcal{A}$ are just those sequences which do not contain any word in $\mathcal{W}$; for a word of length at most M in the alphabet $\mathcal{A}$ will always be contained within a pair of consecutive letters in the alphabet $\mathcal{A}^*$. It is also obvious that the one-one correspondence obtained in this way presrves the topology, and is therefore a homeomorphism. □

This is a highly wasteful process, in that the alphabet $\mathcal{A}^*$ will be much bigger than $\mathcal{A}$ and the same will be true of the set of forbidden words. The point of it is that we can now apply Theorem 22:

Corollary *The zeta-function of any subshift space of finite type is rational.*

4. An application of the Schwarzian derivative

The object of this section is to prove that for any c with $0 < c \leq 4$ the logistic map $x \mapsto cx(1-x)$ has at most one attractive periodic cycle, and that if there is one it is easy to compute. (As we have seen in the case $c = 4$, it need not have one at all.) The path to this is rather roundabout, and starts with something apparently totally irrelevent.

We can make the group $\mathrm{GL}_2(\mathbf{R})$ act on any $\mathbf{R}$-valued function $f(x)$ by

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \mapsto \frac{af(x)+b}{cf(x)+d}.$$

Can we find an expression which depends only on the orbit generated by $f(x)$, and not on $f(x)$ itself? If we assume that $f(x)$ is C^3, then the **Schwarzian derivative**, defined as

$$S(f,x) = \frac{f'''(x)}{f'(x)} - \frac{3}{2}\left(\frac{f''(x)}{f'(x)}\right)^2 = F'(x) - \tfrac{1}{2}(F(x))^2 \quad \text{where } F = \frac{f''}{f'},$$

has this property.

Lemma 25 *If $f_1(x) = \frac{af(x)+b}{cf(x)+d}$ then $S(f_1,x) = S(f,x)$. Moreover*

$$S(fg,x) = S(f,g(x))g'(x)^2 + S(g,x). \tag{10}$$

Proof If $F_1 = f_1''/f_1'$ then

$$f_1' = \frac{(ad-bc)f'}{(cf+d)^2}, \quad f_1'' = \frac{(ad-bc)f''}{(cf+d)^2} - 2\frac{c(ad-bc)f'^2}{(cf+d)^3}$$

and therefore

$$F_1 = \frac{f_1''}{f_1'} = \frac{f''}{f'} - \frac{2cf'}{cf+d} = F - \frac{2cf'}{cf+d}.$$

Hence

$$F_1' = F' - \frac{2cf''}{cf+d} + \frac{2c^2f'^2}{(cf+d)^2};$$

and finally $F_1' - \frac{1}{2}F_1^2 = F' - \frac{1}{2}F^2$. This proves the first assertion. For the second, write $h = f \circ g$, $H = h''/h'$. Then

$$h' = (f' \circ g)g', \quad h'' = (f'' \circ g)g'^2 + (f' \circ g)g'',$$

whence

$$H(x) = F(g(x))g'(x) + G(x),$$
$$H'(x) = F'(g(x))g'(x)^2 + F(g(x))g''(x) + G'(x).$$

Thus finally we obtain

$$H' - \tfrac{1}{2}H^2 = (F' - \tfrac{1}{2}F^2)g'^2 + G' - \tfrac{1}{2}G^2$$

as asserted. □

Corollary *If f is such that $S(f,x) < 0$ for all x, then $S(f^r,x) < 0$ for all x and all $r > 0$.*

Proof Writing $y = g(x)$ and $g = f^{n-1}$ in (10) shows that if $S(f^{n-1},x) < 0$ for all x then $S(f^n,x) < 0$ for all x. The Corollary now follows by induction. □

Lemma 26 *Let $f(x)$ be a real polynomial of degree d such that $f'(x) = 0$ has $d-1$ real roots; then $S(f,x) < 0$ for all x.*

Proof Let $f'(x) = (x - c_1)\dots(x - c_{n-1})$ where the c_i are real; then

$$F(x) = (\log f'(x))' = \sum (x - c_i)^{-1}, \quad F'(x) = -\sum (x - c_i)^{-2} < 0$$

and therefore $S(f,x) = F'(x) - \frac{1}{2}F(x)^2 < 0$. □

Theorem 27 *Let $f : \mathbf{R} \to \mathbf{R}$ be a C^3 map such that $S(f,x) < 0$ for all x and $|f(x)| > |x|$ for all large enough $|x|$. If $\mathcal{D}$ is the basin of attraction of an attractoring cycle $\mathcal{C}$ of f, then $\mathcal{D}$ contains a root of $f'(x) = 0$.*

Proof We consider first the case when $\mathcal{C}$ is an attracting fixed point x_0. Since $\mathcal{D}$ is open by Lemma 5, there is a maximal open subinterval I of $\mathcal{D}$. By the second condition on f, for all large enough $|x|$ we have $|f^n(x)| \to \infty$ as $n \to \infty$; hence $\mathcal{D}$ is bounded and $I = (x^-, x^+)$ for some x^-, x^+. Just as at the beginning of the proof of Theorem 7, since I contains the fixed point x_0 and maps into $\mathcal{D}$, it must map into I; and by continuity each of x^-, x^+ must map to a point which is in the closure of I but is not in I — that is to say, each of them must map to x^- or x^+. If $f(x^-) = f(x^+)$ then I contains a zero of f', by the Mean Value Theorem. If $f(x^-) = x^-$ and $f(x^+) = x^+$ then by the Mean Value Theorem there are point ξ^- in (x^-, x_0) and ξ^+ in

(x_0, x^+) such that $f'(\xi^-) = f'(\xi^+) = 1$. But by Proposition 3, $|f'(x_0)| \leq 1$; hence either f' vanishes somewhere in (ξ^-, ξ^+) or it has a positive minimum (not necessarily strict) there. But the latter is impossible because at such a point we would have $f' > 0$, $f'' = 0$, $f''' \geq 0$ and this contradicts $S(f) < 0$. Finally suppose that $f(x^-) = x^+$, $f(x^+) = x^-$; then x_0 is a fixed point under f^2 with domain of attraction $\mathcal{D}$ again. So by what we have already proved, $(f^2)'$ has a zero x^* in $\mathcal{D}$ and indeed in I. But $(f^2)'(x^*) = (f'f(x^*))f'(x^*)$, so one of the factors on the right must vanish; and if x^* is in I so is $f(x^*)$. This completes the proof in the case when $\mathcal{C}$ is a single attractive fixed point.

Now let $\mathcal{C}$ be an attractive cycle of order M, and let x_0 be one of its points; then x_0 is an attractive fixed point for f^M. By what we have already proved, the domain of attraction $\mathcal{D}_0$ of x_0 for f^M contains a zero of

$$(f^M)' = \prod_{m=0}^{M-1} f'f^m.$$

Let x^* be this zero, so that $f'(f^m(x^*)) = 0$ for some m; then $f^m(x^*)$ is in $f^m\mathcal{D}_0 \subset \mathcal{D}$. $\square$

Corollary *For any c with $0 < c \leq 4$ the logistic map $f_c : x \mapsto cx(1-x)$ has at most one attractive cycle, and the domain of attraction of that cycle contains the point $x = \frac{1}{2}$.*

Proof Though in §1 we defined the logistic map f_c as a map of $[0, 1]$ to itself, we can obviously extend it to a map of $\mathbf{R}$ to itself. As such, it satisfies both the conditions in the Theorem. $\square$

If we try to apply the same argument to a map of $[0, 1]$ or $(0, 1)$ to itself, there is an additional complication: I may be coterminous at one or both ends with the space we are considering. In the case of $\mathbf{R}$ we ruled this out by means of the second hypothesis in Theorem 27, but now the analogue of this would be unnatural. For $[0, 1]$ it is simplest to allow explicitly for the possibility that 0 or 1 is in I, giving the following result.

Theorem 28 *Let f be a C^3 map of $[0, 1]$ to itself such that $S(f, x) < 0$ for all x. If $\mathcal{D}$ is the basin of attraction of an attractoring cycle $\mathcal{C}$ of f, then $\mathcal{D}$ contains 0, 1 or a root of $f'(x) = 0$.*

For maps f of $(0, 1)$ to itself, we can confine ourselves to the case when f' has only finitely many zeroes; for otherwise a result like Theorem 27 would be of no interest even if we could prove it. It now follows that f is monotone in $(0, \epsilon)$ for some $\epsilon > 0$, so that f can be extended by continuity to 0. (It may not be differentiable at 0, but it can be checked that this does not affect the argument.) A similar statement holds for 1, and we can now obtain an analogue of Theorem 28.

5. Sharkovsky's Theorem

Consider the map $f_c : x \mapsto cx(1-x)$ of $[0,1]$ to itself, where $0 < c \leq 4$. If $c < 3$ we know that there is just one attracting fixed point, and that all points except perhaps 0 and 1 tend to it. On the other hand, we know that when $c = 4$ there are periodic cycles of every order; and Theorem 31 below, together with a tedious but straightforward calculation of the two cycles of period 3, shows that this continues to be true throughout a certain interval $c_0 \leq c \leq 4$. We have seen in §4 that for any c in this interval at most one of the periodic cycles is an attractor; and the numerical evidence does not suggest any simple pattern in the way in which the period of that attractor changes with c.

Can we say anything about the transition from the well-behaved situation when $c \leq 3$ to the somewhat chaotic situation when $c = 4$? Most of what is known is described in two famous theorems, the second one being a special case of the first but one which is much simpler to prove. The first one is Sharkovsky's Theorem. The second one (which is also later in time) is due to Li and Yorke, but is always known by the title of their paper: "Period three implies chaos". Both of them are concerned with continuous maps of a closed interval to itself; nothing analogous is known for any other space.

Both of these theorems can be regarded as a collection of statements of the form: if there is a cycle of least period M' then there is a cycle of least period M''. Applied say to the logistic map f_c, they tell us about the order of the first appearances of cycles of any particular periods as c increases. Suppose for simplicity that p is prime; then the points of period p come from the roots of $f_c^p(x) = x$, which is an equation of degree 2^p, but two of these roots are the fixed points of f_c. All the real roots of this equation lie in $[0,1]$, but the roots need not all be real; so we have at most $(2^p - 2)/p$ cycles of exact period p. When $c = 4$ it is easy to see that we have exactly this many, by using conjugacy with the tent map. But if $p > 3$ we cannot expect them all to come into existence at the same time.

Lemma 29 *Suppose that f is continuous on a closed subinterval I of* $\mathbf{R}$.
(i) *If $I \subset f(I)$ then f has a fixed point in I.*
(ii) *If J is a closed subinterval of $f(I)$ then there is a closed subinterval K of I with $f(K) = J$.*

Proof. (i) Let $I = [a,b]$ and let the maximum and minimum of f on I be attained at u, v respectively. Then $f(u) \geq b \geq u$ and $f(v) \leq a \leq v$. Thus

$g(x) = f(x) - x$ satisfies $g(u) \geq 0 \geq g(v)$; so g takes the value 0 at some point w in $[u, v]$, and w is a fixed point of f.

(ii) I is compact, so $f(I)$ is closed. Let $\alpha_1 = f(a_1)$ and $\beta_1 = f(b_1)$ be the least and greatest values which f takes in I; then f takes every value between α_1 and β_1, by the Mean Value Theorem, and so $f(I)$ is precisely the closed interval $[\alpha_1, \beta_1]$. Now let J be $[\alpha, \beta]$; the case $\alpha = \beta$ is trivial (and will be irrelevent), so we can assume $\alpha < \beta$. There exist a, b in I such that $f(a) = \alpha$, $f(b) = \beta$; and without loss of generality we can assume $a < b$. The set $\mathcal{S}_\alpha$ of solutions in I of $f(x) = \alpha$ is closed and contains $a < b$; so let a' be the greatest element of $\mathcal{S}_\alpha$ in $x \leq b$. Similarly let b' be the least solution of $f(x) = \beta$ not less than a'. The set $f([a', b'])$ contains α and β — and hence the whole of $[\alpha, \beta]$ by the Mean Value Theorem. Suppose it contained a point γ outside $[\alpha, \beta]$ — say with $\gamma = f(c) > \beta$. Then there would exist x in $[a', c]$ with $f(x) = \beta$ by the Mean Value Theorem, contrary to the definition of b'. □

In what follows, we shall say that I' **covers** I'' if $f(I') \supset I''$. Note that if $I = [x', x'']$ then I covers $[f(x'), f(x'')]$ and hence also every subinterval of it. We shall call a sequence of closed intervals $B_1, \ldots, B_r$ a **covering sequence** if B_i covers B_{i+1} for $1 \leq i < r$, and a **cyclic covering sequence** if also B_r covers B_1. We use cyclic covering sequences to construct periodic cycles by means of Lemma 30; the second alternative in the Lemma has to be in the statement, but in applications we always ensure that it does not happen.

Lemma 30 *Suppose that f is continuous on a closed subinterval Y of $\mathbf{R}$ and there are subintervals $B_1, \ldots, B_n$ of Y which form a cyclic covering sequence. Then either Y contains a periodic cycle of exact order n or it contains a cycle whose order m is a proper factor of n. In the latter case, there is a point y of the cycle such that the intersection $B_i \cap B_{i+m}$ contains the point $f^{i-1}(y)$ of the cycle for each i, where the subscripts are to be interpreted $\bmod m$.*

Proof. Write $A_n = B_n$ and use Lemma 29(ii) successively for $i = n-1, \ldots, 1$ to choose $A_i \subset B_i$ so that $f(A_i) = A_{i+1}$. Since $f^n(A_1) = \ldots = f(A_n) \supset A_1$, it follows from Lemma 29(i) that f^n has a fixed point y in A_1 and $f^i(y)$ is in A_{i+1} for $i = 0, 1, \ldots, n-1$. If y had period $m < n$ under f then $f^{i-1}(y) = f^{i+m-1}(y)$ would lie in $A_i \cap A_{i+m}$. □

Theorem 31 *Let f be a continuous map of the closed interval X to itself. Suppose that f has a cycle of exact period 3; then it has a cycle of exact period n for every $n \geq 1$.*

Proof. By reversing the ordering on X if necessary, we can assume that the cycle of order 3 is $\{x_0, x_1, x_2\}$ with $x_0 < x_1 < x_2$, where f induces $x_0 \mapsto x_1 \mapsto x_2 \mapsto x_0$. The case $n = 1$ needs special treatment; but it is just Lemma 29(i) applied to $[x_1, x_2]$ since $f([x_1, x_2])$ is an interval which contains $f(x_1) = x_2$ and $f(x_2) = x_0$. Henceforth we can assume that $n > 1$. We construct intervals $B_1, \ldots, B_n$ having the properties stated in Lemma 30. For this we take B_1 to be $[x_0, x_1]$ and each subsequent B_i to be $[x_1, x_2]$. To complete the proof of Theorem 31, we need only apply Lemma 30. For if the second alternative in the Lemma happened, there would be a point of the cycle in $A_1 \cap A_{m+1} = \{x_1\}$; so n would be a multiple of 3 and $f^2(x_1) = x_0$ would be in $A_3 \subset B_3$, which is false. □

Theorem 31 tells us that for the map f_c in the first paragraph, as c increases from 3 to 4 the first cycle of period 3 is created after the first cycle of any other period. To specify the order in which the others are created, we need a stronger result. Define the **Sharkovsky ordering** on the positive integers by

$$3 \succ 5 \succ 7 \succ \ldots \succ 3.2 \succ 5.2 \succ \ldots \succ 3.2^2 \succ 5.2^2 \succ \ldots \succ 2^3 \succ 2^2 \succ 2 \succ 1.$$

Theorem 32 *Let f be a continuous map of the closed interval X to itself. Suppose that f has a cycle of exact period n, and that $n_1 \prec n$ in the Sharkovsky ordering; then f has a cycle of exact period n_1.*

Note that this Theorem is not examinable; but I give the proof because it illustrates most of the ideas which have already appeared in the course. The major step in the proof is Lemma 33 below, which was trivial in the case $n = 3$. Note that the forbidden values of m in Lemma 33 are just those for which $m \succ n$ in the Sharkovsky ordering. None of the individual steps in the proof of Lemma 33 is difficult in itself; the difficulty is that at each stage there appear to be several equally plausible ways to proceed, and only one of them provides useful information. Once we have Lemma 33, the proof of the theorem consists of a subdivision into a number of cases, none of which is really difficult.

Lemma 33 *Suppose that $n > 1$ is an odd integer and that f has a periodic cycle $\mathcal{C}$ of order precisely n but no periodic cycle of any precise order m where m is an odd integer with $1 < m < n$. After reversing the ordering on X if necessary, we can choose x_0 in $\mathcal{C}$ so that*

$$x_{n-1} < x_{n-3} < \ldots < x_2 < x_0 < x_1 < x_3 < \ldots < x_{n-2} \qquad (11)$$

where as usual $x_{i+1} = f(x_i)$.

Proof. Let α, β be the least and greatest of the x_i, and recall that

$$f([x_i, x_j]) \supset [f(x_i), f(x_j)] = [x_{i+1}, x_{j+1}]$$

for any i, j. Now $x_{i+1} = \alpha$, $x_{j+1} = \beta$ gives $f([\alpha, \beta]) \supset [\alpha, \beta]$. The points of $\mathcal{C}$ break $[\alpha, \beta]$ up into $n-1$ closed non-overlapping intervals; for the rest of this section, a letter I will always denote such an interval. Let a be the largest element y of $\mathcal{C}$ such that $f(y) > y$, and let b be the least element of $\mathcal{C}$ strictly greater than a. Then $[a, b]$ is an I; we shall denote it by I_1. We break the proof of the Lemma up into a number of steps.

(i) $f(I_1) \supset I_1$.

For $f(a) > a$, so that $f(a) \geq b$; and in $f(b) \leq b$ equality is impossible, so that $f(b) \leq a$. The assertion now follows from the continuity of f.

(ii) Let $\mathcal{S}$ be a set of consecutive intervals I'; then the intervals I'' each of which is covered by some I' in $\mathcal{S}$ also form a consecutive set.

Since $f(I')$ is connected, the intervals covered by I' form a consecutive set; and if y' is an end-point of I' this set includes an I'' which has $f(y')$ as an end-point. The assertion follows easily.

(iii) Let $\mathcal{S}_r$ be the set of intervals I'_r such that there is a covering sequence $I'_1 = I_1, I'_2, \ldots, I'_r$ of length r. Then $\mathcal{S}_{r+1} \supset \mathcal{S}_r$ and $\mathcal{S}_{n-1}$ contains every interval I.

The first assertion follows from (i), because we can put an extra I_1 in front of any covering sequence of length r. As in the proof of (ii), if y' is an end-point of an interval in $\mathcal{S}_r$ then $f(y')$ is an end-point of an interval in $\mathcal{S}_{r+1}$. But each of α and β is equal to $f^r(a)$ or $f^r(b)$ for some r with $0 \leq r \leq n-2$; so the first and last intervals I lie in $\mathcal{S}_{n-1}$. By (ii), so do the others.

(iv) There is an $I \neq I_1$ which covers I_1.

Since n is odd, there are more elements of $\mathcal{C}$ on one side of the interior of I_1 than on the other; without loss of generality we can assume that there are more below than above. Hence there exists y in $\mathcal{C}$ such that both y and $f(y)$ are below the interior of I_1. Let y' be the greatest such y and let y'' be the element of $\mathcal{C}$ next above y'; since $y' \neq a$, y'' is also below the interior of I_1 and hence $f(y'')$ must be above the interior of I_1. So we can take $I = [y', y'']$.

(v) There are cyclic covering sequences $I_1, I_2, \dots, I_r$ with $I_2 \neq I_1$; and the shortest of these has $r = n - 1$.

By (iii), there is such a sequence with $I_r = I$ where I is as in (iv); for if a sequence such as in (iii) started with several consecutive copies of I_1 we could delete all but the last of them. Now consider the shortest one. It can have no repeats; for if $I_i = I_j$ with $i > j$ we could delete $I_j, \dots, I_{i-1}$. Thus $r \leq n-1$. Suppose that $r < n-1$; if r is odd set $s = r$, but if r is even set $s = r+1$ and $I_s = I_1$. In either case $I_1, \dots, I_s$ is a cyclic covering sequence of odd order s with $1 < s < n$. Applying Lemma 30 we obtain a periodic point whose order is odd and less than n; and it cannot be a fixed point because otherwise it would lie in $I_1 \cap I_2$ which is either empty or a single point of $\mathcal{C}$. This contradicts the hypothesis of Lemma 33. We now fix one shortest such sequence; it contains all the I, so we can use it to give them individual names I_i.

(vi) In the notation just defined, if I_i covers I_1 then $i = 1$ or $n-1$. If $j > i+1$ then I_i does not cover I_j. In particular I_1 only covers I_1 and I_2.

All these assertions follow from the minimality of the sequence of I_i.

We can now complete the proof of the Lemma. By (vi) and (ii), I_1 and I_2 are consecutive; without loss of generality we can assume that they abut at a. (If instead they abut at b, we first reverse the ordering on X.) Thus $f(a) = b$ and $f(b)$ is the lower end-point of I_2; in other words, $f^2(a) < a < f(a)$ and these three points are consecutive points of $\mathcal{C}$.

We now take $x_0 = a$ and proceed by induction, assuming that for some r with $2 \leq r < n-1$ we have proved the parts of the assertion (11) which relate to $x_0, \dots, x_r$ including the implicit statement that no x_i with $i > r$ lies between x_r and x_{r-1}. Thus I_j is $[x_{j-2}, x_j]$ for $2 \leq j \leq r$ and the I_i with $i > r$ lie outside $[x_{r-1}, x_r]$. It follows from the second and third assertions in (vi) that x_{r+1} must be the other neighbour of x_{r-1}; and this completes the proof of the Lemma. □

We can now prove Theorem 32 when n is odd. We need to split cases. If $n_1 = 1$ then I_1 covers itself; so there is a fixed point of f by Lemma 29(i). If $n_1 < n$ is even then $I_{n-1} = [x_{n-1}, x_{n-3}]$ covers $[x_0, x_{n-2}]$; so $I_{n-n_1}, \dots, I_{n-1}$ is a cyclic covering sequence of length n_1, so by Lemma 30 there is a periodic cycle of exact period n_1 — for the second alternative in Lemma 30 merely gives $\mathcal{C}$ which cannot have period n_1. If $n_1 > n$ then $I_1, I_2, \dots, I_{n-1}, I_1, \dots, I_1$

(where the last $n_1 - n + 1$ terms are all I_1) is a cyclic covering sequence of order n_1. Again this gives rise to a periodic point of exact period n_1; for the second alternative in Lemma 30 would imply that the corresponding periodic cycle was $\mathcal{C}$ and this contradicts the fact that three consecutive terms of our cyclic covering sequence are equal to I_1.

To deal with the case when n is even, we need a further Lemma.

Lemma 34 *Suppose that f has a cycle of precise order m where m is even; then f has a periodic cycle of precise order* 2*.*

Proof. Let n be the least integer greater than 1 such that f has a periodic cycle of order precisely n. If n is odd, the Lemma is contained in the part of Theorem 32 which we have already proved; so we can take n even with $n > 2$. We follow the proof of Lemma 33 as far as the proof of (iii), and then split cases.

Suppose there is an $I \neq I_1$ such that $f(I) \supset I_1$. Then (iv) of Lemma 33 is still true, the proof of (v) becomes simpler, and the proofs of (vi) and of the Lemma itself still hold. Thus $[x_{n-1}, x_{n-3}]$, $[x_{n-4}, x_{n-2}]$ is a cyclic covering sequence which yields a periodic point of exact period 2.

We may therefore suppose that there is no such I. We cannot have $a = \alpha$ because then each interval I' which had $f^{n-1}(a)$ as an end-point would cover I_1; either there are two such or $f^{n-1}(a) = \beta$ and the one I' whose end-point is $f^{n-1}(a)$ would not be I_1. So $a > \alpha$ and similarly $b < \beta$. The interval next below I_1 covers an interval abutting $f(a) = b$; so this must be above I_1. It now follows from (ii) of Lemma 33, applied to the set of I' below I_1, that if I' is below I_1 then every I'' covered by I' is above I_1; and a similar argument using the fact that $f(b) \leq a$ shows that if I' is above I_1 then every I'' covered by I' is below I_1. But every I'' is covered by some I'; so

$$f([\alpha, a]) \supset [b, \beta], \quad f([b, \beta]) \supset [\alpha, a],$$

and this covering sequence yields a periodic cycle of exact period 2. □

For the remainder of the proof of Theorem 32 we write $n = 2^s p$, $n_1 = 2^t q$ where s, t are positive and p, q are odd. We have three cases to consider:

- $q = 1$ and $t \leq s$;
- $t > s$ and $p > 1$;
- $t = s$ and $q > p > 1$.

In the first case, n_1 is a power of 2. If $n_1 = 1$ the conclusion follows from $f(I_1) \supset I_1$ by Lemma 29(i), and if $n_1 = 2$ the conclusion is Lemma 34; so we can take $n_1 \geq 4$. Write $g = f^{n_1/2}$, so that any point of period n for f is a point of period $2n/n_1$ for g; then g has a point y of exact period 2, by Lemma 34. But $f^{n_1}(y) = y$ and $f^{n_1/2}(y) = g(y) \neq y$; so y has exact period n_1.

In the second and third cases we write $g = f^{2^s}$, so that g has a periodic cycle of exact period p. By the part of the Theorem first proved, g has a point y with exact period $2^{t-s}q$. If $t > s$ this is even, so y must have exact period $2^t q$ under f. If however $t = s$ all we can deduce is that y has exact period $2^u q$ under f for some u with $0 \leq u \leq s$. If $u = s$ the proof is complete; if not, by starting from y we are in the second of our three cases instead of the third — and the second case has already been proved.