Commun. Math. Phys. 76, 211–254 (1980)

Communications in Mathematical Physics

© by Springer-Verlag 1980

Universal Properties of Maps on an Interval

P. Collet[1], J.-P. Eckmann[2], and O. E. Lanford III[3]

[1] Harvard University, Cambridge, MA 02138, USA
[2] Departement de Physique Théorique, Université de Genève, CH-1211 Genève 4, Switzerland
[3] University of California, Berkeley, CA 94720, USA

Abstract. We consider iterates of maps of an interval to itself and their stable periodic orbits. When these maps depend on a parameter, one can observe period doubling bifurcations as the parameter is varied. We investigate rigorously those aspects of these bifurcations which are universal, i.e. independent of the choice of a particular one-parameter family. We point out that this universality extends to many other situations such as certain chaotic regimes. We describe the ergodic properties of the maps for which the parameter value equals the limit of the bifurcation points.

1. Introduction

Continuous mappings of intervals into themselves display some remarkable properties when regarded as discrete dynamical systems. (For a survey, see May [9] or Collet and Eckmann [14].) One much-studied example is the one-parameter family

$$x \to 1 - \mu x^2 \tag{1.1}$$

which maps $[-1, 1]$ into itself for $0 \leqq \mu \leqq 2$. In this and similar examples, what is interesting is not so much the behavior of any particular mapping; rather, it is the way this behavior changes with μ.

The example (1.1), and the more general one-parameter families $\mu \to \psi_\mu$ we will study, have a simplifying qualitative feature: Each ψ_μ has a unique (differentiable) maximum – at $x=0$ in the example – below which it is increasing and above which it is decreasing. We will consider mappings ψ which satisfy

P1) ψ is a continuously differentiable mapping of $[-1, 1]$ into itself.

P2) $\psi(0)=1$; ψ is strictly increasing on $[-1, 0]$ and strictly decreasing on $[0, 1]$.

P3) $\psi(-x)=\psi(x)$.

The space of all such mappings will be denoted by $\mathscr{P}$. (The condition that the maximum of ψ occurs at zero and that ψ sends zero to one can frequently be arranged, if necessary, by making an affine change of variables.) We have included

0010-3616/80/0076/0211/$08.80

condition P3) mostly for convenience; it simplifies matters and is satisfied by the ψ's we are able to analyze in detail.

One important property of such a transformation is having – or not having – an attracting periodic orbit. (Existence of periodic orbits which are not attracting is much less important, directly at least, in accounting for the behavior of typical orbits.) The fact that $[-1, 1]$ is ordered and connected gives rise to powerful and general methods for proving the existence of periodic orbits – see, for example, Stefan [13] – but these methods do not help very much with the existence of attracting periodic orbits. Note, however, that if 0 is periodic for $\psi \in \mathscr{P}$, then, since $\psi'(0)=0$, its orbit is necessarily attracting. We will say that ψ is *superstable of period p* if 0 is periodic of (minimal) period p for ψ. If ψ_0 is superstable of period p, then any $\psi \in \mathscr{P}$ which is near enough to ψ_0 in the C^1 topology will also have an attracting periodic orbit of period p. Thus for example if

$$\mu \to \psi_\mu$$

is a one-parameter family of elements of $\mathscr{P}$ with ψ_{μ_0} superstable of period p, there is an open interval about μ_0 in the parameter space such that each corresponding ψ_μ has an attracting periodic orbit of period p.

The existence of superstable ψ's can sometimes be proved by simple topological arguments. For example, with our normalization, ψ is superstable of period 2 if and only if $\psi(1)=0$. If we now consider a continuous one-parameter family ψ_μ defined on some interval of μ's, and if $\psi_\mu(1)$ is sometimes positive and sometimes negative, then there must be at least one μ for which ψ_μ is superstable of period 2. We give in Sect. 3 an elaboration of this simple argument which shows that, if $\psi_\mu(1)$ is near 1 for μ near the left end of the parameter interval and near -1 near the right end, there exists a sequence

$$\mu_1 < \mu_2 < \mu_3 < \ldots$$

such that ψ_{μ_j} is superstable of period 2^j. (See Guckenheimer [4] for an alternative approach to the existence of the μ_j's.) It is clear that, if we allow arbitrary (non-monotone) reparametrizations, we cannot hope to prove the existence of unique μ_j's. Moreover, similar topological considerations guarantee that such a parametrized family has, for each large j, many values of μ where ψ_μ is superstable of period 2^j. Nevertheless, in examples like

$$x \to 1 - \mu x^2$$

the first superstable values of μ appear to occur with periods

$$2, 4, 8, 16, \ldots$$

in that order. We will denote the corresponding values of μ by μ_j and $\lim_{j\to\infty} \mu_j$ by μ_∞.

By investigating numerically a number of one-parameter families, Feigenbaum [3] discovered a striking universality property: For large j, $\mu_\infty - \mu_j$ is asymptotic to

$$\text{const} \times \delta^{-j},$$

where $\delta = 4.66920\ldots$ *is apparently the same whatever one-parameter family is considered.* (Note that, encouragingly, this property of the μ_j's is not changed by making a differentiable change of parameter with derivative which does not vanish at μ_∞.)

Having discovered the universality of δ experimentally, Feigenbaum went on to propose an explanation for it which was inspired by the renormalization group approach to critical phenomena in statistical mechanics. The principal result of this paper is to show that Feigenbaum's explanation is correct, at least in a certain limiting regime to be explained below. We will next sketch our version of Feigenbaum's theory, ignoring numerous technical details which will need to be made precise later.

Consider a mapping $\psi \in \mathscr{P}$ and define

$$a = a(\psi) = -\psi(1)\,; \qquad b = b(\psi) = \psi(a)\,.$$

Assume

$$0 < a < b\,(< 1)$$

and assume also that

$$\psi(b) = \psi^2(a) < a\,.$$

ψ then maps

$$[-a, a] \quad \text{onto} \quad [b, 1] \quad \text{and} \quad [b, 1] \quad \text{onto} \quad [-a, \psi(b)] \subset [-a, a]\,,$$

i.e. it exchanges the two non-intersecting intervals $[-a, a]$ and $[b, 1]$. Hence $\psi \circ \psi$ maps $[-a, a]$ into itself, and $-\psi \circ \psi$ is again unimodal on $[-a, a]$ (see Fig. 1). If we reverse orientation and scale up by a factor of $\frac{1}{a}$, i.e. if we make the linear change of variables

$$x_{\text{old}} = -a x_{\text{new}}$$

then $\psi \circ \psi$ on $[-a, a]$ is transformed to

$$-\frac{1}{a}\psi \circ \psi(-ax) \equiv (\mathscr{T}\psi)(x)$$

on $[-1, 1]$. It is easy to verify that, with our hypotheses, $\mathscr{T}\psi$ again has properties P1)–P3) [but the condition $a(\mathscr{T}\psi) > 0$ or $a(\mathscr{T}\psi) < b(\mathscr{T}\psi)$ may fail]. We will refer to the transformation $\mathscr{T}$ as the *doubling transformation.* The doubling transformation is essentially just composition of ψ with itself, but combined with restriction of $\psi \circ \psi$ to a subdomain of the original domain and then a scaling (and reversal of orientation) chosen to preserve the "normalization" $\psi(0) = 1$. This combined operation, in contrast with composition alone, does not give rise to a more complicated-looking transformation. The utility of $\mathscr{T}$ in studying superstable ψ's lies largely in the remark that, provided ψ satisfies the conditions given above for $\mathscr{T}\psi$ to be defined, ψ is superstable of period p if and only if $\mathscr{T}\psi$ is superstable of period $p/2$ (and, in particular, p must be even).

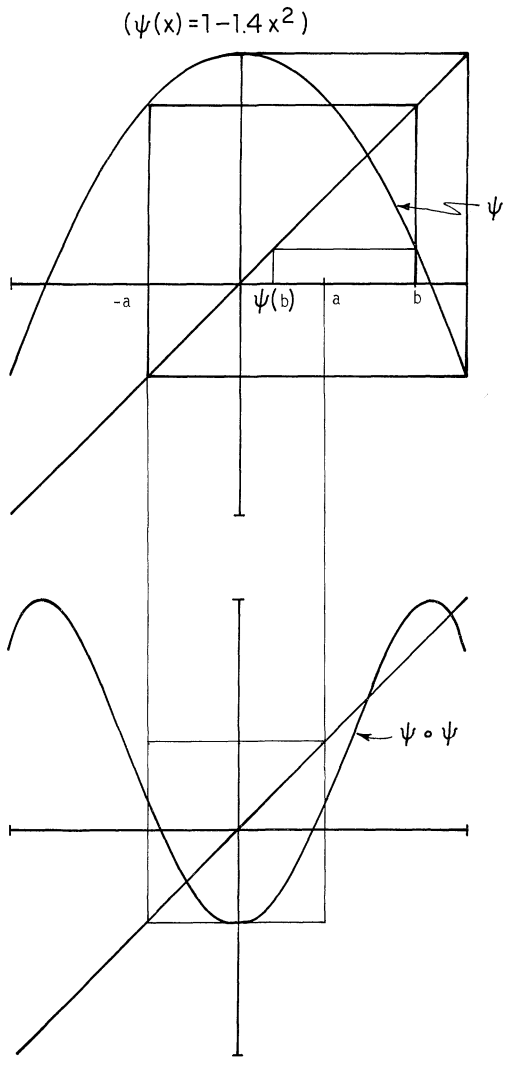

Fig. 1

We now, following Feigenbaum, propose some geometrical hypotheses about how $\mathscr{T}$ acts in the space $\mathscr{P}$ of transformations and show how these hypotheses account for the universality of δ. The picture is as follows:

a) $\mathscr{T}$ has a fixed point ϕ.

b) The derivative of $\mathscr{T}$ at the fixed point ϕ has a simple eigenvalue which is larger than one (and which will turn out to be δ); the remainder of its spectrum is contained in the open unit disk. $\mathscr{T}$ thus has a one-dimensional unstable manifold W_u and a codimension-one stable manifold W_s at ϕ.

c) The unstable manifold W_u intersects transversally the codimension-one surface Σ_1,

$$\Sigma_1 = \{\psi : \psi(1) = 0\}.$$

(Note that Σ_1 is exactly the set of ψ's which are superstable of period 2.) See Fig. 2.

Using this picture, we can account for the universality of δ as follows: Form successive inverse images $\Sigma_2, \Sigma_3, \ldots$ of Σ_1 under $\mathscr{T}$:

$$\Sigma_j = \mathscr{T}^{-(j-1)}\Sigma_1 .$$

Note that if $\psi \in \Sigma_j$ then $\mathscr{T}^{(j-1)}\psi \in \Sigma_1$, so $\mathscr{T}^{j-1}\psi$ is superstable of period 2, so ψ is superstable of period 2^j. The successive Σ_j's come closer and closer to W_s; in fact, a straightforward argument (which we will give in detail later) shows that the separation between Σ_j and W_s decreases exponentially like δ^{-j} for large j, where δ is the large eigenvalue of the derivative of $\mathscr{T}$ at ϕ.

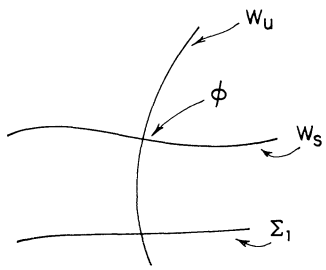

Fig. 2

Now consider a one-parameter family $\mu \to \psi_\mu$ of transformations and regard it as a curve in $\mathscr{P}$. Suppose this curve crosses the stable manifold W_s at $\mu = \mu_\infty$ with non-zero transverse velocity. It is then clear that, at least for large j, there will be a unique μ_j near μ_∞ such that $\psi_{\mu_j} \in \Sigma_j$ (which implies that ψ_{μ_j} is superstable of period 2^j) and that

$$\lim_{j \to \infty} \delta^j (\mu_\infty - \mu_j)$$

exists and is non-zero (see Fig. 3).

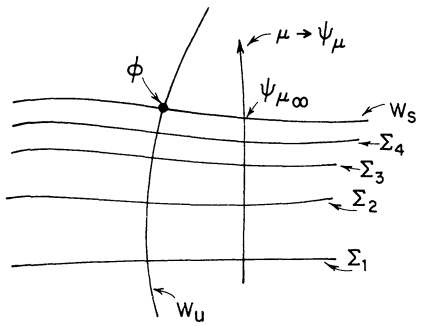

Fig. 3

Thus, Feigenbaum's hypotheses not only account for the universal rate at which μ_j approaches μ_∞; they also provide in principle an independent prescription for computing δ. They have other consequences as well; we will mention here just two of them:

1. For all j, $\mathscr{T}^{j-1}\psi_{\mu_j}$ is superstable of period 2. Because $\mathscr{T}$ contracts in the W_s direction, the $\mathscr{T}^{j-1}\psi_{\mu_j}$ converge as $j \to \infty$ to the point of intersection of Σ_1 with W_s,

which we will denote by ϕ_0^*. Thus, ϕ_0^* is also universal; for any one-parameter family as above, if we form

$$\psi_{\mu_j}^{2^{j-1}}$$

and scale properly, we get something near to ϕ_0^* for large j. Similarly, $\mathscr{T}^j \psi_{\mu_\infty}$ converges to ϕ.

2. Let $\tilde{\Sigma}_1$ denote the surface

$$\{\psi : \psi^3(1) = -\psi(1)\}$$

(i.e. the set of ψ's such that $-\psi(1)$ is a fixed point for ψ^2). Misiurewicz [11] has shown that there is an open set of ψ's on $\tilde{\Sigma}_1$ which admit an absolutely continuous invariant measure (and hence which have typical orbits which are not periodic). We will see that $\tilde{\Sigma}_1$ intersects W_u transversally, with point of intersection inside this open set. (The intersection point will lie *above* W_s in Figs. 2 and 3.) Again form successive inverse images of $\tilde{\Sigma}_1$ under $\mathscr{T}$

$$\tilde{\Sigma}_j = \mathscr{T}^{-(j-1)} \tilde{\Sigma}_1 ;$$

these surfaces converge to W_s, again exponentially with rate δ, from the side opposite to that of the Σ_j's. Again, for each large j, there will be a unique $\tilde{\mu}_j$ near μ_∞ with

$$\psi_{\tilde{\mu}_j} \in \tilde{\Sigma}_j ,$$

and the $\tilde{\mu}_j$'s converge to μ_∞ in the usual way:

$$\lim_{j \to \infty} (\tilde{\mu}_j - \mu_\infty) \cdot \delta^j$$

exists and is non-zero. For large enough j, $\mathscr{T}^{j-1} \psi_{\tilde{\mu}_j}$ will be near to the point of intersection of $\tilde{\Sigma}_1$ with W_u and hence will admit an absolutely continuous invariant measure. From this it is easy to show that $\psi_{\tilde{\mu}_j}$ itself admits an absolutely continuous invariant measure and hence also has orbits which are typically non-periodic.

Thus μ_∞ is the limit of values of μ for which ψ_μ is chaotic. We warn the reader, however, that not all ψ_μ's for μ just above μ_∞ are chaotic; for example, there is a sequence $\hat{\mu}_j$, again converging to μ_∞ from above, with the same exponential rate, such that

$$\hat{\psi}_{\hat{\mu}_j} \quad \text{is superstable with period} \quad 3 \cdot 2^j .$$

As indicated earlier, we are going to prove that Feigenbaum's hypotheses are correct in certain cases. As Feigenbaum has noted, the universality of δ is somewhat relative – its value depends on the function space in which the ψ's are assumed to lie. We will consider functions ψ of the form

$$\psi(x) = f(|x|^{1+\varepsilon}) .$$

where the function f is smooth. Except for a result on the uniqueness of the fixed point, we will in fact have to assume that f is analytic in a complex neighborhood of $[0, 1]$. We would of course like to deal with the case $\varepsilon = 1$, but the argument we are going to give is a perturbative analysis valid only for sufficiently small positive values of ε. Our results could be expressed in terms of convergent series expansions

in ε and various iterated logarithms of ε which are analogues of the ε-expansions occurring in the renormalization-group approach to the theory of critical phenomena. Work in progress, using quite different techniques, indicates that at least partial results can be obtained for $\varepsilon=1$ [8], see also note added in proof.

As an indication that there is some simple behavior at $\varepsilon=0$ about which we could hope to carry out a perturbative analysis, consider the family of functions

$$\psi_a(x)=1-(1+a)|x|,$$

for small positive a. A straightforward calculation shows that

$$(\mathscr{T}\psi_a)(x)=1-(1+a)^2|x|,$$

i.e.

$$\mathscr{T}\psi_a=\psi_{2a+a^2}.$$

Thus, the curve $a\to\psi_a$ is invariant under the action of $\mathscr{T}$, and the end point at $a=0$, although not in the domain of definition of $\mathscr{T}$, is a sort of virtual fixed point. If we consider instead

$$\psi(x)=1-(1+a)|x|^{1+\varepsilon}$$

we no longer get such a simple closed-form expression for $\mathscr{T}\psi$ but we do get

$$(\mathscr{T}\psi)(x)=1-(1+a)^2a^{\varepsilon}|x|^{1+\varepsilon}+O(\varepsilon).$$

This suggests that, for small ε, there might be a fixed point near ψ, where a is to be determined approximately by

$$1+a\simeq a^{\varepsilon}(1+a)^2,$$

i.e.

$$a^{\varepsilon}\simeq 1-a,$$

i.e.

$$\varepsilon\log a\simeq -a \quad \text{or} \quad \varepsilon\simeq\frac{a}{(-\log a)}.$$

Observe that this, if correct, implies that $\varepsilon\ll a$ and hence suggests that it should be possible to get a fixed point by adding to $1-(1+a)|x|^{1+\varepsilon}$ a correction which is small relative to a.

2. Statement of Results

We are going to consider functions $\psi\in\mathscr{P}$ of the form

$$\psi(x)=f(|x|^{1+\varepsilon}),$$

where f is the restriction to $[0,1]$ of a function analytic in some domain in the complex plane. Through most of our analysis, the domain Ω of analyticity will not affect our results very much – although it presumably affects how small we have to take ε in order to make our estimates work. There is one point in Sect. 7 where it is necessary to impose some conditions on Ω; these conditions are met if we take Ω to be an open disk with center 1/2 and radius not too much larger than 1/2. Aside

from this one argument, we can work with Ω any bounded connected open set in $\mathbb{C}$ containing $[0,1]$. We will mostly think of Ω as chosen once and for all and so will frequently suppress it from our notation. We will write $\mathfrak{H}(\Omega)$, or simply $\mathfrak{H}$, for the real Banach space of functions bounded and analytic on Ω, and real on $\Omega \cap \mathbb{R}$, equipped with the supremum norm. For $\varepsilon > 0$, we denote by $\mathscr{P}_\varepsilon (\subset \mathscr{P})$ the set of functions ψ on $[-1,1]$ of the form

$$\psi(x) = f(|x|^{1+\varepsilon}),$$

with $f \in \mathfrak{H}$ and satisfying

$$f(0) = 1; \quad \frac{df}{dt} < 0 \text{ on } [0,1]; \quad f(1) > -1.$$

We can identify $\mathscr{P}_\varepsilon$ in an obvious way with an open subset of the Banach space

$$\{g \in \mathfrak{H}(\Omega) : g(0) = 0\}.$$

Theorem 2.1. *For ε sufficiently small, $\mathscr{T}$ has a fixed point ϕ in $\mathscr{P}_\varepsilon$. If we write*

$$\phi_\varepsilon(x) = f_\varepsilon(|x|^{1+\varepsilon})$$

then $f_\varepsilon(t)$ extends to a function jointly analytic in (ε, t) for

$$\varepsilon \in \{z \in \mathbb{C} \setminus [-\infty, 0] : |z| < \varepsilon_0\}$$

and $t \in \Omega$. We denote $-\phi_\varepsilon(1)$ by λ_ε; then

$$\begin{aligned} \lambda_\varepsilon &= -\varepsilon \log \varepsilon + O(\varepsilon) \\ f_\varepsilon(t) &= 1 - (1 + \lambda_\varepsilon)t + O(\varepsilon^2 \log \varepsilon). \end{aligned} \tag{2.1}$$

ϕ_ε is an isolated fixed point for $\mathscr{T}$ in $\mathscr{P}_\varepsilon$; it has negative Schwarzian derivative (see Singer [12]), i.e.

$$\frac{\phi_\varepsilon'''}{\phi_\varepsilon'} - \frac{3}{2}\left(\frac{\phi_\varepsilon''}{\phi_\varepsilon'}\right)^2 < 0.$$

Theorem 2.2. *For ε sufficiently small, ϕ_ε is an isolated fixed point for $\mathscr{T}$ in the space of functions*

$$\psi(x) = f(|x|^{1+\varepsilon})$$

with f twice continuously differentiable on $[0,1]$.

Note that, if $\hat{\varepsilon} > \varepsilon$, then $\phi_{\hat{\varepsilon}}(x) = f_{\hat{\varepsilon}}(|x|^{1+\hat{\varepsilon}})$ can also be written as $g(|x|^{1+\varepsilon})$ with g *continuously differentiable on* $[0,1]$. Thus, $\mathscr{T}$ has at least a one-parameter family of fixed points of the form $g(|x|^{1+\varepsilon})$ with g only once continuously differentiable.

Notational Convention: From now on the symbols ϕ_ε and λ_ε are permanently reserved to denote the above objects. We will frequently suppress the subscript ε.

Theorem 2.3. *The transformation $\mathscr{T}$ is infinitely differentiable in a neighborhood of ϕ_ε in $\mathscr{P}_\varepsilon$. The derivative of $\mathscr{T}$ at ϕ_ε has one simple eigenvalue $\delta_\varepsilon > 1$ which approaches*

2 *as* ε *approaches zero. The diameter of the smallest disk centered at zero containing the rest of its spectrum goes to zero with* ε.

Theorem 2.4. *$\mathscr{T}$ has a smooth stable manifold, W_s, of codimension one and a smooth unstable manifold, W_u, of dimension one, at ϕ_ε. For each $a\in[-1,1]$ there is a unique point ϕ_a^* on W_u with*

$$\phi_a^*(1)=-a\,.$$

W_u crosses the surfaces Σ_1 and $\tilde{\Sigma}_1$ (defined in Sect. 1) transversally. Each ϕ_a^* has negative Schwarzian derivative.

Theorem 2.5. *Let $\mu\to\psi_\mu$ be a continuously differentiable parametrized curve in $\mathscr{P}_\varepsilon$ which crosses the stable manifold W_s with non-zero transverse velocity at $\mu=\mu_\infty$. There exist sequences μ_j and $\tilde{\mu}_j$ converging to μ_∞ from opposite sides such that*

$$\lim_{j\to\infty}\delta^j(\mu_\infty-\mu_j)\quad\text{and}\quad\lim_{j\to\infty}\delta^j(\mu_\infty-\tilde{\mu}_j)$$

are both finite and non-zero, and such that ψ_{μ_j} is superstable of period 2^j and $\psi_{\tilde{\mu}_j}$ admits an absolutely continuous invariant measure for each sufficiently large j. Moreover, the ratio of $\lim_{j\to\infty}\delta^j(\mu_\infty-\mu_j)$ to $\lim_{j\to\infty}\delta^j(\mu_\infty-\tilde{\mu}_j)$ is also universal, i.e. does not depend on the particular parametrized family under consideration.

Remark. One instance of such a parametrized family is

$$\psi_\mu(x)=\psi(\mu\cdot x)$$

for a fixed function ψ sufficiently near to ϕ_ε. We can then in particular take

$$\psi_\mu(x)=1-\mu|x|^{1+\varepsilon}.$$

[Actually, the first statement is not quite true. For if $\mu>1$, then $x\to\psi(\mu\cdot x)$ need not be in $\mathscr{P}_\varepsilon(\Omega)$, but it is in $\mathscr{P}_\varepsilon(\mu^{-(1+\varepsilon)}\Omega)$.]

Theorem 2.6. *If $\psi\in W_s$, then ψ has an invariant Cantor set J.*

1) *There is a decreasing chain of closed subsets of $[-1,1]$*

$$J^{(0)}\supset J^{(1)}\supset J^{(2)}\supset\ldots$$

each of which contains 0, and each of which is mapped onto itself by ψ.

2) *Each $J^{(i)}$ is a disjoint union of 2^i closed intervals. $J^{(i+1)}$ is constructed by deleting an open subinterval from the middle of each of the intervals making up $J^{(i)}$.*

3) *ψ maps each of the intervals making up $J^{(i)}$ onto another one; the induced action on the set of intervals is a cyclic permutation of order 2^i.*

We let J denote $\bigcap_i J^{(i)}$. ψ maps J onto itself in a one-one fashion. Every orbit in J is dense in J. If, besides being on W_s, ψ has negative Schwarzian derivative – for which it suffices that it be near ϕ_ε – then we have:

4) *For each $k=1,2,\ldots$ ψ has exactly one periodic orbit of period 2^{k-1}. This periodic orbit is repelling and does not belong to $J^{(k)}$; ψ has no periodic orbits other than these.*

5) *Every orbit of* ψ *either*

a) *lands after a finite number of steps exactly on one of the periodic orbits enumerated in* 4) *or*

or

b) *converges to the Cantor set* J *in the sense that, for each* k, *it is eventually contained in* $J^{(k)}$.

There are only countably many orbits of type a).

Theorem 2.7. *Again assume that* $\psi \in W_s$, *and let* $J^{(i)}$, J *be as in Theorem* 2.6. *Let* ν *denote the probability measure with support* J *which for each* i *assigns equal weight to each of the* 2^i *intervals making up* $J^{(i)}$.

1) ν *is invariant under the action of* ψ; *it is the only invariant probability measure on* J.

2) *The abstract dynamical system* (ν, ψ) *is ergodic but not weak mixing.*

3) *If* x *is any point of* $[-1,1]$ *whose orbit converges to* J, *and if* f *is any continuous function on* $[-1,1]$, *then*

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} f(\psi^n(x)) = \int f d\nu .$$

In particular, if ψ is close enough to ϕ_ε so that Theorem 2.6 holds, then this equality holds for all but countably many x's. Similar results were obtained by Misiurewicz [10]. The analysis leading to the Cantor set also gives an attractive picture of how the bifurcation at μ_∞ looks. This is described in detail at the end of Sect. 8.

The proofs will be organized as follows:

In Sect. 3 we develop the elementary theory of the doubling transformation $\mathscr{T}$, and prove for a fairly general class of one-parameter families $\{\psi_\mu\}$ in $\mathscr{P}$ the existence of an increasing sequence μ_j of parameter values such that ψ_{μ_j} is superstable of period 2^j.

Section 4 gives the proofs of Theorem 2.1 – except for the estimate (2.1) on the precise form of the fixed point, which is deferred to Sect. 7 – and Theorem 2.3. Theorem 2.2 is proved in Sect. 5.

Section 5 gives the precise definitions of global stable and unstable manifolds that we use and proves a general theorem, sketched in the introduction, permitting us to deduce Theorem 2.5 immediately from Theorem 2.4. Theorem 2.4 is proved in Sect. 7, and Theorems 2.6 and 2.7 in Sect. 8.

3. The Doubling Transformation

In this section we develop the elementary theory of the doubling transformation $\mathscr{T}$. Let $\psi \in \mathscr{P}$, and define

$$a = a(\psi) = -\psi(1) .$$

If $a > 0$ we also define

$$b = b(\psi) = \psi(a) .$$

The domain of $\mathscr{T}$, $\mathscr{D}(\mathscr{T})$, is the set of all $\psi \in \mathscr{P}$ such that:

$$1)\ a>0$$
$$2)\ b>a$$
$$3)\ \psi(\psi(a)) \leqq a,$$

and for $\psi \in \mathscr{D}(\mathscr{T})$ we define

$$(\mathscr{T}\psi)(x) = -\frac{1}{a}\psi \circ \psi(ax).$$

Remark. Although $\mathscr{D}(\mathscr{T})$ is defined by three conditions, the boundary of $\mathscr{D}(\mathscr{T})$ consists in fact of two surfaces:

$$a=0$$
$$\psi(\psi(a))=a.$$

This comes about because, in moving from $\mathscr{D}(\mathscr{T})$ to a region where 2) fails we must pass through a point where $b=a$, i.e. $\psi(a)=a$, and this implies $\psi(\psi(a))=a$. Normally, we would expect conditions 2) and 3) to fail simultaneously, but it is easy to find situations in which 3) fails and 2) does not.

Proposition 3.1. *Let $\psi \in \mathscr{P}$ satisfy $a(\psi)=0$, and let (ψ_n) be a sequence in $\mathscr{P}$ converging to ψ in the C^1 topology and with $a(\psi_n)>0$ for all n. Then*

1) *$\psi_n \in \mathscr{D}(\mathscr{T})$ for sufficiently large n.*
2) *$(\mathscr{T}\psi_n)(1) \to 1$ as $n \to \infty$.*

In other words: Every point on the surface $\{a(\psi)=0\}$ is part of the boundary of $\mathscr{D}(\mathscr{T})$, and $\mathscr{T}$ sends ψ's near this surface to functions near the constant function 1.

Proof. It is easy to see that, if $\psi^2(a)<0$, then $\psi \in \mathscr{D}(\mathscr{T})$. We will show:

$$\frac{\psi_n^2(a_n)-(-a_n)}{a_n} \to 0 \quad (a_n = a(\psi_n))$$

which implies both 1) [since it implies $\psi_n^2(a_n)<0$ eventually] and 2) [since $\mathscr{T}\psi_n(1) = -\psi_n^2(a_n)/a_n$].

In view of the facts that

$$a_n \to 0; \quad \psi_n \to \psi \text{ in } C^1; \quad \psi'(0)=0,$$

we get

$$\frac{\psi_n(a_n)-\psi_n(0)}{a_n} \to 0.$$

On the other hand,

$$|\psi_n'(x)| \leqq M \text{ uniformly in } x, n$$

so

$$\left|\frac{\psi_n^2(a_n)-(-a_n)}{a_n}\right| = \left|\frac{\psi_n(\psi_n(a_n))-\psi_n(\psi_n(0))}{a_n}\right| \leqq M\left|\frac{\psi_n(a_n)-\psi_n(0)}{a_n}\right| \to 0$$

as claimed.

It is on the other hand clear that if ψ satisfies $\psi\circ\psi(a)=a$ and if $\psi_n\in\mathscr{D}(\mathscr{T})$; $\psi_n\to\psi$, then

$$\mathscr{T}\psi_n(1)=-\frac{1}{a_n}\psi_n\circ\psi_n(a_n)\to-1\,.$$

Consider now a mapping $\mu\to\psi_\mu$ from an interval $(\mu_0,\tilde{\mu}_0)$ into $\mathscr{P}$, i.e., a one-parameter family of elements of $\mathscr{P}$. We assume the mapping to be continuous in the C^1 topology. We will say that such a one-parameter family is *full* if

$$\psi_\mu(1)\to 1 \quad \text{as} \quad \mu\to\mu_0$$

and

$$\psi_\mu(1)\to -1 \quad \text{as} \quad \mu\to\tilde{\mu}_0$$

(e.g. $\psi_\mu(x)=1-\mu x^2$; $\mu_0=0$; $\tilde{\mu}_0=2$).

We have already remarked that for any such one-parameter family there must be at least one μ such that $\psi_\mu(1)=0$, i.e. such that ψ_μ is superstable. There may be many such μ's; in any case, we denote by μ_1 the *largest* such. Proposition 1 shows that, for μ slightly larger than μ_1, $\psi_\mu\in\mathscr{D}(\mathscr{T})$. We will denote by $\tilde{\mu}_1$ the *smallest* $\mu>\mu_1$ such that $\psi_\mu\notin\mathscr{D}(\mathscr{T})$. [Since $b(\psi_\mu)=\psi_\mu^2(1)\to-1$ as $\mu\to\tilde{\mu}_0$, whereas $a(\psi_\mu)=-\psi_\mu(1)\to 1$, condition 2) in the definition of $\mathscr{D}(\mathscr{T})$ must fail before μ reaches $\tilde{\mu}_0$.] By our earlier remarks, $\psi_{\tilde{\mu}_1}^2(a_{\tilde{\mu}_1})=a_{\tilde{\mu}_1}$.

Proposition 3.2. *If $\mu\to\psi_\mu$ is a full one-parameter family, then*

$$\mu\to\mathscr{T}\psi_\mu, \qquad \mu_1<\mu<\tilde{\mu}_1$$

is also a full one-parameter family.

Proof. By Proposition 1, $\mathscr{T}\psi_\mu(1)\to 1$ as $\mu\downarrow\mu_1$, and by the remark following Proposition 1,

$$\mathscr{T}\psi_\mu(1)\to -1 \quad \text{as} \quad \mu\uparrow\tilde{\mu}_1\,.$$

By induction, then, there exist two sequences

$$\mu_0<\mu_1<\mu_2<\ldots<\tilde{\mu}_2<\tilde{\mu}_1<\tilde{\mu}_0$$

such that, for $\mu_j<\mu<\tilde{\mu}_j$, $\psi_\mu\in\mathscr{D}(\mathscr{T}^j)$ and

$$\mu\to\mathscr{T}^j\psi_\mu, \qquad \mu_j<\mu<\tilde{\mu}_j$$

is a full one-parameter family. In particular these sequences are constructed in such a way that

$$\mathscr{T}^{j-1}\psi_{\mu_j}(1)=0$$

i.e. $\mathscr{T}^{j-1}\psi_{\mu_j}$ is superstable of period 2, i.e. ψ_{μ_j} is superstable of period 2^j.

4. Existence and Elementary Properties of the Fixed Point

If $\psi\in\mathscr{P}_\varepsilon$ is in $\mathscr{D}(\mathscr{T})$ as defined in Sect. 3, and if $\psi(x)=f(|x|^{1+\varepsilon})$, then

$$(\mathscr{T}\psi)(x)=F(|x|^{1+\varepsilon}) \qquad x\in[-1,1]\,,$$

where

$$F(t) = -\frac{1}{a} f(|(f(a^{1+\varepsilon}t))|^{1+\varepsilon}); \qquad a = -f(1) > 0.$$

The conditions for the definition of $\mathscr{T}\psi$ imply that

$$f(a^{1+\varepsilon}t) > 0 \quad \text{for} \quad 0 \leqq t \leqq 1$$

and we can therefore drop the absolute value sign. If we define $\mathscr{D}_\varepsilon$ to be the set of such functions ψ satisfying in addition

$$a^{1+\varepsilon}\bar{\Omega} \subset \Omega; \qquad f(a^{1+\varepsilon}\Omega) \cap (-\infty, 0) = \emptyset;$$
$$\overline{[f(a^{1+\varepsilon}\Omega)]^{1+\varepsilon}} \subset \Omega$$

then, if $\psi \in \mathscr{D}_\varepsilon$, $\mathscr{T}\psi$ is again in $\mathscr{P}_\varepsilon$. We are going to prove that $\mathscr{T}$ has a fixed point in $\mathscr{D}_\varepsilon$ for each sufficiently small positive ε.

It is convenient to introduce a new variable α related to ε by

$$\varepsilon = \frac{-\alpha}{1 + \log(\alpha)}.$$

Note that for each small positive α there corresponds exactly one small positive ε and vice versa. Any $\psi \in \mathscr{P}_\varepsilon$ can be written uniquely as

$$\psi(x) = f(|x|^{1+\varepsilon}); \qquad f(t) = 1 - t + \alpha t(g(t) - 1),$$

with $g \in \mathfrak{H}(\Omega)$.

Working with g rather than ψ is simply a (linear) change of variables in function space. If $\psi \in \mathscr{D}_\varepsilon$, we will write the g corresponding to $\mathscr{T}\psi$ as $T_\varepsilon g$. The domain of $\mathscr{T}$ is bounded on one side by the surface

$$\psi(1) = 0$$

which corresponds to

$$g(1) = 1.$$

We are going to show that, for small ε, T_ε is defined and well behaved on the open unit ball in $\mathfrak{H}$ and has a fixed point near zero.

To formulate our results concisely, we need some special terminology. If $\mathscr{X}$ is a normed space and ϱ a positive number, we write $\mathscr{X}_\varrho$ for the open ball in $\mathscr{X}$ with center 0 and radius ϱ. A mapping defined on $\mathscr{X}_1$ will be said to be *nearly bounded* if it is bounded on each $\mathscr{X}_\varrho$ with $\varrho < 1$. Similarly, functions will be said to converge *nearly uniformly* if they converge uniformly on each $\mathscr{X}_\varrho$ with $\varrho < 1$.

Proposition 4.1. *For $\varepsilon > 0$ sufficiently small, T_ε is defined on $\mathfrak{H}_1(\Omega)$. The mapping*

$$(\varepsilon, g) \to T_\varepsilon(g)$$

is jointly infinitely differentiable. For fixed ε, derivatives of all orders of T_ε with respect to g are nearly bounded on $\mathfrak{H}_1$. We can decompose T_ε as

$$T_\varepsilon(g) = T_0 g + r_\varepsilon(g),$$

where T_0 is a rank-one linear operator with range the constant functions:

$$(T_0 g)(t) = g(0) + g(1) + g'(1)$$

and r_ε and its g-derivatives of all orders converge almost uniformly to zero with ε.

We emphasize that: Although T_ε is highly non-linear, its zeroth order part T_0 is not only linear but very simple – dividing it by two gives a projection onto the constant functions. Its simplicity makes possible a detailed analysis of the behavior of T_ε for small ε.

Before proving the proposition we note its principal corollary.

Corollary 4.2. 1. *For each sufficiently small $\varepsilon > 0$, there is exactly one solution $g_\varepsilon^{(0)}$ for the fixed point problem*

$$g = T_\varepsilon(g)$$

in $\mathfrak{H}_{1/2}$. (Here, $\frac{1}{2}$ may be replaced by any number less than one.)

$$\varepsilon \to g_\varepsilon^{(0)}$$

is infinitely differentiable and $g_\varepsilon^{(0)}$ approaches zero with ε.

2. *$DT_\varepsilon(g_\varepsilon^{(0)})$ varies continuously with ε and approaches T_0 in operator norm as ε approaches zero.*

3. *Let $0 < \varrho < 1$. For sufficiently small ε, the only part of the spectrum of $DT_\varepsilon(g_\varepsilon^{(0)})$ at a distance greater than ϱ from 0 is a simple positive eigenvalue δ_ε which approaches two as ε approaches zero. The corresponding eigenspace converges to the space of constant functions.*

To prove 1., we write the fixed point problem as

$$g = T_0 g + r_\varepsilon(g)$$

or equivalently as

$$(I - T_0)g = r_\varepsilon(g).$$

Since $T_0^2 = 2T_0$, we have $(I - T_0)^2 = I$ and so the above equation is equivalent to

$$g = (I - T_0) r_\varepsilon(g).$$

Since r_ε and Dr_ε converge to zero nearly uniformly with ε,

$$g \to (I - T_0) r_\varepsilon(g)$$

is a contraction on $\mathfrak{H}_{1/2}$ for ε sufficiently small. The existence and uniqueness of $g_\varepsilon^{(0)}$ follows from the contraction mapping principle. The smoothness of the dependence of $g_\varepsilon^{(0)}$ on ε follows from the implicit function theorem in Banach space. (See, for example, Dieudonné [2].) That $g_\varepsilon^{(0)}$ approaches 0 with ε follows immediately from the nearly-uniform convergence of r_ε to zero.

Part 2 follows from the joint continuity of $DT_\varepsilon(g)$ in g, ε and the continuity of $g_\varepsilon^{(0)}$ in ε. Part 3 follows from 2 by standard perturbation theory (Kato [7]) and the fact that the spectrum of T_0 reduces to $\{0, 2\}$ with 2 a simple eigenvalue whose associated eigenspace is the constant functions.

The proof of the proposition is a relatively straightforward computation supported by some general theorems. We will give the computation first; then

sketch the justification that the remainder terms do indeed have the properties claimed.

We first do a computation whose result we will need to use again later. We have already seen that if

$$\psi(x)=f(|x|^{1+\varepsilon}); \qquad a=-\psi(1)$$

then the transformation $\psi\to\mathscr{T}\psi$ translates to

$$f\to-\frac{1}{a}f((f(a^{1+\varepsilon}t))^{1+\varepsilon}).$$

We will next write

$$f(t)=1-th(t)$$

and determine how h transforms. We will need the following notation: If $t_0\in\Omega$, we define a bounded linear operator Δ_{t_0} on $\mathfrak{H}(\Omega)$ by

$$\begin{aligned}(\Delta_{t_0}f)(t)&=\frac{f(t)-f(t_0)}{t-t_0} \qquad t\neq t_0\\ &=f'(t_0) \qquad\qquad t=t_0.\end{aligned}$$

Now define η_1, η_2 by

$$\begin{aligned}f(a^{1+\varepsilon}t)&=1-at\eta_1 \quad [\text{so } \eta_1=a^{\varepsilon}h(a^{1+\varepsilon}t)]\\ (1-at\eta_1)^{1+\varepsilon}&=1-at\eta_2.\end{aligned}$$

We now claim: Under the action of $\mathscr{T}$, h transforms as

$$h\to\eta_2\times\{h(1-at\eta_2)+(\Delta_1h)(1-at\eta_2)\}. \tag{4.1}$$

To verify this, write

$$\begin{aligned}f((f(a^{1+\varepsilon}t))^{1+\varepsilon})&=1-(1-at\eta_2)h(1-at\eta_2)\\ &=1-h(1-at\eta_2)+at\eta_2h(1-at\eta_2).\end{aligned}$$

Now use the following expression for the first $h(1-at\eta_2)$,

$$h(1-at\eta_2)=h(1)-at\eta_2(\Delta_1h)(1-at\eta_2)$$

and recall that $a=-f(1)=h(1)-1$, to get

$$-a+at\eta_2\{(\Delta_1h)(1-at\eta_2)+h(1-at\eta_2)\}.$$

Thus

$$-\frac{1}{a}f((f(a^{1+\varepsilon}t))^{1+\varepsilon})=1-t\eta_2\{(\Delta_1h)(1-at\eta_2)+h(1-at\eta_2)\}$$

from which the formula (4.1) for the action of $\mathscr{T}$ on h can be read off.

We must next insert the expression

$$h(t)=1-\alpha(g(t)-1)$$

and extract the principal terms for small ε. In so doing, we generate a large number of remainder terms, and it is convenient to have a systematic notation for the spaces in which the remainder terms lie. Let $\mathfrak{R}$ denote the space of all mappings

$$r:(\varepsilon, g)\to r(\varepsilon, g)$$

defined on a set of the form $(0,\varepsilon_0)\times\mathfrak{H}_1(\Omega)$ with values in $\mathfrak{H}(\Omega)$. Here ε_0 is a strictly positive number which may vary with r. These mappings are required to be jointly infinitely differentiable in ε, g, and derivatives of all orders with respect to g are required to be nearly bounded in $\mathfrak{H}_1(\Omega)$ for each ε and to converge to zero nearly uniformly with ε. We will use $\mathfrak{B}$ to denote the analogous space of functions which, together with their g derivatives, will merely be required to remain *bounded* as ε approaches zero and $\mathfrak{R}_0$, $\mathfrak{B}_0$ to denote the analogous spaces of functions taking values in $\mathbb{R}$ rather than $\mathfrak{H}$. Recall, also, that ε and α are related by

$$\varepsilon=\frac{-\alpha}{(1+\log\alpha)};$$

whenever α appears in one of our formulas it is to be regarded as a function of ε.

Since $a=h(1)-1$ and we are writing $h(t)=1-\alpha(g(t)-1)$ we have

$$a=\alpha(1-g(1))$$

and hence

$$a^\varepsilon=\alpha^\varepsilon(1-g(1))^\varepsilon.$$

Now $\alpha^\varepsilon=\exp[\varepsilon\log\alpha]=\exp\left(\frac{-\alpha\log\alpha}{1+\log\alpha}\right)=\exp[-\alpha-\varepsilon]$.

Thus

$$a^\varepsilon=e^{-\alpha}+\varepsilon b_1;\qquad b_1\in\mathfrak{B}_0.$$

Also,

$$g(a^{1+\varepsilon}t)=g(0)+a^{1+\varepsilon}t(\Delta_0 g)(a^{1+\varepsilon}t)=g(0)+\alpha b_2,\qquad b_2\in\mathfrak{B}.$$

Thus

$$\begin{aligned}\eta_1&=a^\varepsilon(1+\alpha(1-g(a^{1+\varepsilon}t)))\\&=(e^{-\alpha}+\varepsilon b_1)(1+\alpha-\alpha g(0)-\alpha^2 b_2)\\&=1-\alpha g(0)+\alpha r_1,\qquad r_1\in\mathfrak{R}.\end{aligned}$$

[We have used the fact that ε/α goes to zero with ε to replace εb_1 by $\alpha(\varepsilon/\alpha)b_1$ with $\varepsilon/\alpha b_1\in\mathfrak{R}_0$.] Now

$$1-at\eta_2=(1-at\eta_1)^{1+\varepsilon},$$

so

$$\eta_2=\eta_1+\varepsilon\alpha b_3,\qquad b_3\in\mathfrak{B},$$

so

$$\eta_2=1-\alpha g(0)+\alpha r_2,\qquad r_2\in\mathfrak{R}.$$

Also

$$h(1-at\eta_2)+(\Delta_1 h)(1-at\eta_2)=1+\alpha-\alpha g(1-at\eta_2)-\alpha(\Delta_1 g)(1-at\eta_2)$$
$$=1+\alpha(1-g(1)-g'(1)+r_3),\quad r_3\in\mathfrak{R}.$$

Thus

$$\eta_2\{h(1-at\eta_2)+(\Delta_1 h)(1-at\eta_2)\}=(1-\alpha g(0)+\alpha r_2)(1+\alpha(1-g(1)-g'(1)+r_3))$$
$$=1-\alpha(g(0)+g(1)+g'(1)-1+r_4),\quad r_4\in\mathfrak{R}$$

and hence by inspection

$$T_\varepsilon g=g(0)+g(1)+g'(1)+r_4,$$

as desired.

Remark. These computations show that $\mathscr{T}$ admits a fixed point $\phi_\varepsilon\in\mathscr{P}_\varepsilon$ with

$$\phi_\varepsilon(x)=1-(1+\alpha(\varepsilon))|x|^{1+\varepsilon}+\text{correction},$$

where the correction vanishes more rapidly than α as ε goes to zero. They also show that there is no other fixed point in a ball about ϕ_ε whose radius is bounded below by const$\alpha(\varepsilon)$. If we write

$$\lambda_\varepsilon=-\phi_\varepsilon(1),$$

then

$$\lambda_\varepsilon=\alpha(\varepsilon)+O(\alpha),\quad\text{or}\quad\lambda_\varepsilon=-\varepsilon\log\varepsilon+o(\varepsilon\log\varepsilon).$$

More detailed computations to be done in Sect. 7 show that in fact

$$\lambda_\varepsilon=-\varepsilon\log\varepsilon+O(\varepsilon)$$

and that

$$\phi_\varepsilon(x)=1-(1+\lambda_\varepsilon)|x|^{1+\varepsilon}+O(-\varepsilon^2\log\varepsilon).$$

We turn now to the problem of justifying the above computations, i.e. of showing that the remainder terms do indeed have the asserted properties. The verifications are tedious and we will not do all of them, but we will work through one in full detail. We wrote, in the course of the computation,

$$a^{1+\varepsilon}t(\Delta_0 g)(a^{1+\varepsilon}t)=\alpha b_2\quad\text{with}\quad b_2\in\mathfrak{B}. \tag{4.2}$$

We now want to prove this. The proof is based on a number of principles which we list here:

a) The mapping $(\varepsilon,g)\to g$ is in $\mathfrak{B}$.

b) For any $g_0\in\mathfrak{H}$, the constant mapping $(\varepsilon,g)\to g_0$ is in $\mathfrak{B}$.

c) If $b_0\in\mathfrak{B}_0$, the mapping $(\varepsilon,g)\to$(the constant function with value $b_0(\varepsilon,g)$) is in $\mathfrak{B}$.

d) Let $\mathfrak{V}_1,\ldots,\mathfrak{V}_n$ be open sets in $\mathfrak{H}$ and let

$$\Phi:\mathfrak{V}_1\times\ldots\times\mathfrak{V}_n\to\mathfrak{H}$$

be bounded, infinitely differentiable, and have bounded derivatives. Further, let $b_1, \ldots, b_n \in \mathfrak{B}$, with the range of b_i contained in $\mathfrak{B}_i$ for each ε. Then

$$(\varepsilon, g) \to \Phi(b_1(\varepsilon, g), \ldots, b_n(\varepsilon, g))$$

is in $\mathfrak{B}$.

e) (Corollary of d.) If $b_1, b_2 \in \mathfrak{B}$, then

$$(\varepsilon, g) \to b_1(\varepsilon, g) \cdot b_2(\varepsilon, g)$$

(pointwise product of analytic functions) is in $\mathfrak{B}$. Similarly, if b_1, b_2 are in $\mathfrak{B}_0$, so is their product.

f) If Ω_1, Ω_2 are bounded open sets in $\mathbb{C}$, we write $\mathfrak{H}(\Omega_1, \Omega_2)$ for the set of all $g \in \mathfrak{H}(\Omega_1)$ with $\overline{g(\Omega_1)} \subset \Omega_2$. [Hence, if Ω_2 is the open unit disk, $\mathfrak{H}(\Omega, \Omega_2) = \mathfrak{H}_1(\Omega)$.] $\mathfrak{H}(\Omega_1, \Omega_2)$ is an open subset of $\mathfrak{H}(\Omega_1)$. Now let $\Omega_2' \subset \bar{\Omega}_2' \subset \Omega_2$. Then composition $(g_1, g_2) \to g_2 \circ g_1$ is a C^∞ function with bounded derivatives from $\mathfrak{H}(\Omega_1, \Omega_2') \times \mathfrak{H}(\Omega_2, \Omega_3)$ into $\mathfrak{H}(\Omega_1, \Omega_3)$.

We omit the proofs of these statements. Note that a)–e) remain true if we replace $\mathfrak{H}$ by C^2, but that f) depends upon the properties of analytic functions, and fails in C^2.

We are now ready to verify 2).

1. The mappings

$$(\varepsilon, g) \to 1 - g(1), \quad (1 - g(1))^\varepsilon, \quad \alpha, \quad e^{-\alpha}, \quad e^{-\varepsilon}$$

are all in $\mathfrak{B}_0$. This is readily proved by direct verification. Note that the g-derivatives of $(1-g(1))^\varepsilon$ have singularities as $g(1) \to 1$; it is for this reason (only) that we have to work with functions with derivatives which are nearly bounded – rather than bounded – on the unit ball of $\mathfrak{H}$.

We will from now on write, as a shorthand, statements like $(1-g(1))^\varepsilon \in \mathfrak{B}_0$ rather than the more logical

$$((\varepsilon, g) \to (1 - g(1))^\varepsilon) \in \mathfrak{B}_0 .$$

2. $a = \alpha(1-g(1)) \in \mathfrak{B}_0$; $a^\varepsilon = e^{-\alpha} e^{-\varepsilon}(1-g(1))^\varepsilon \in \mathfrak{B}_0$; $a^{1+\varepsilon} \in \mathfrak{B}_0$. [Use 1e).]

3. $a^{1+\varepsilon} t \in \mathfrak{B}$. [Use 2b), c), and e).] If ε_0 is small enough there exists a domain Ω_1 with $\bar{\Omega}_1 \subset \Omega$ such that $a^{1+\varepsilon} t \in \Omega_1$ for all $\varepsilon < \varepsilon_0$, $g \in \mathfrak{H}_1$, $t \in \Omega$.

4. $\Delta_0 g \in \mathfrak{B}$. [Use a), d).]

5. $\Delta_0 g(a^{1+\varepsilon} t) \in \mathfrak{B}$. [Use 3., 4., f), d).]

6. $\dfrac{1}{\alpha} a^{1+\varepsilon} \Delta_0 g(a^{1+\varepsilon} t) = (1-g(1)) a^\varepsilon \Delta_0 g(a^{1+\varepsilon} t) \in \mathfrak{B}$. [Use 1., 2., 5., c), e).]

This completes the proof of (4.1).

The arguments given so far show that the fixed point $g_\varepsilon^{(0)}$ varies smoothly with ε. We next show that $g_\varepsilon^{(0)}(t)$ is jointly analytic in ε, t. The logarithm appearing in the relation between $\lambda_\varepsilon = -\phi_\varepsilon(1)$ and ε shows that there must be a singularity at $\varepsilon = 0$, and we want to clarify the structure of that singularity. For this purpose it turns out to be useful to consider a somewhat contrived generalized fixed point problem in which the relation between ε and α is partly relaxed. Recall that $\mathscr{T}$ takes the form

$$f \to -\frac{1}{a} f((f(a^{1+\varepsilon} t))^{1+\varepsilon})$$

with $a=-f(1)=\alpha(1-g(1))$. We modify this transformation as follows: In the innermost argument we replace a^ε by $e^{-\varepsilon-\alpha}(1-g(1))^\varepsilon$. We then replace ε wherever it appears by $\mu\cdot\alpha$ and we regard μ,α as independent parameters. This gives a two-parameter family of transformations whose action we can again express in terms of g related to f as above. Expressed in terms of g, we will denote the transformations by

$$g\to T_{\alpha,\mu}(g).$$

Our original transformations T_ε are recovered through

$$T_{\varepsilon(\alpha)}=T_{\alpha,-\frac{1}{1+\log\alpha}}.$$

The advantage of considering α,μ as independent variables is that $T_{\alpha,\mu}$ turns out to be, heuristically, jointly analytic in α,μ at $(0,0)$; the non-analyticity appears only in the relation between μ and α.

We now, temporarily, drop the requirement of reality for real values of the argument in the definition of the space $\mathfrak{H}$, and we consider the transformations $T_{\alpha,\mu}$ for arbitrary small *complex* values of α,μ. Exactly the same computations as were done in the proof of Proposition 4.1 go through, and we obtain

$$T_{\alpha,\mu}(g)=T_0 g+r_{\alpha,\mu}(g),$$

where $r_{\alpha,\mu}$ and its derivatives of all orders with respect to g converge nearly uniformly to zero as α,μ both go to zero. Hence, as before, the fixed point problem

$$g=T_{\alpha,\mu}(g)$$

can be rewritten as

$$g=(I-T_0)r_{\alpha,\mu}(g)$$

and the right hand side is contractive on $\mathfrak{H}_{1/2}$ for sufficiently small $|\alpha|,|\mu|$. Thus, there exist $\alpha_0,\mu_0>0$ such that for all α,μ with $|\alpha|<\alpha_0$, $|\mu|<\mu_0$, $T_{\alpha,\mu}$ has a unique fixed point $g^{(0)}_{\alpha,\mu}$ in $\mathfrak{H}_{1/2}$. Moreover, the computations which show that $r_{\alpha,\mu}$ is small for small α,μ also show that if $\alpha,\mu\to g_{\alpha,\mu}$ is a mapping from $\{(\alpha,\mu):|\alpha|<\alpha_0,|\mu|<\mu_0\}$ into $\mathfrak{H}_{1/2}$ such that $g_{\alpha,\mu}(t)$ is jointly analytic in α,μ,t then $r_{\alpha,\mu}(g_{\alpha,\mu})(t)$ is again jointly analytic, and so the same is true for

$$(I-T_0)r_{\alpha,\mu}(g_{\alpha,\mu})(t).$$

From this it is easy to show – using the contraction mapping principle in an appropriate space of jointly analytic functions – that $g^{(0)}_{\alpha,\mu}(t)$ is jointly analytic in α,μ,t.

Since the parameter α was introduced artificially, it is desirable to express directly the dependence of $g^{(0)}$ on ε. To do this we need the following lemma.

Lemma 4.3. *There exists a function $r(z_1,z_2)$, analytic on a neighborhood of zero in $\mathbb{C}^2$ and with $r(0,0)=1$ such that for small positive ε*

$$\alpha=-\varepsilon\log\varepsilon\, r\left(\frac{1}{\log\varepsilon},\frac{\log(-\log\varepsilon)}{\log\varepsilon}\right)$$

is the unique small positive solution of the equation

$$\varepsilon = \frac{-\alpha}{1+\log\alpha}. \tag{4.3}$$

Proof. Write $\alpha = -\varepsilon\cdot\log\varepsilon\cdot r$ and insert in (4.3). The result is

$$1+\frac{1}{\log\varepsilon}+\frac{\log(-\log\varepsilon)}{\log\varepsilon}+\frac{\log r}{\log\varepsilon}=r.$$

For fixed ε in $(0,1)$ this equation has a unique solution r; for ε small, the solution is near 1. On the other hand, by the implicit function theorem, there is a uniquely determined function $r(z_1,z_2)$ defined and analytic on a neighborhood of zero in $\mathbb{C}^2$ and satisfying

$$1+z_1+z_2+z_1\log r=r;$$

and $r(0,0)=1$. This is thus our desired function r.

We can express the conclusion of the lemma more concisely by saying simply that $\alpha/\varepsilon\log\varepsilon$ is an analytic function of $1/\log\varepsilon$ and $\log(-\log\varepsilon)/\log\varepsilon$. It follows immediately that ε/α is also an analytic function of $1/\log\varepsilon$ and $\log(-\log\varepsilon)/\log\varepsilon$. Thus we get:

Proposition 4.4. *$g_\varepsilon^{(0)}(t)$ is an analytic function of $\varepsilon\log\varepsilon, \dfrac{1}{\log\varepsilon}, \dfrac{\log(-\log\varepsilon)}{\log\varepsilon}$, and t. In particular, $g_\varepsilon^{(0)}(t)$ is jointly analytic and bounded in ε, t for ε in a small disk about zero in $\mathbb{C}$ and off the negative real axis and t in Ω.*

5. Existence and Uniqueness in C^2

Write

$$f(t)=1-(1+\lambda)t+\lambda^2\ell(t). \tag{5.1}$$

We are looking for (ε,ℓ) such that

$$\lambda f(t)+f(f(\lambda^{1+\varepsilon}t)^{1+\varepsilon})=0. \tag{5.2}$$

We study the equation to be satisfied by the second derivative of f. If (5.2) holds, then

$$f'(t)=-\lambda^\varepsilon f'(f(\lambda^{1+\varepsilon}t)^{1+\varepsilon})f(\lambda^{1+\varepsilon}t)^\varepsilon(1+\varepsilon)f'(\lambda^{1+\varepsilon}t), \tag{5.3}$$

and

$$\begin{aligned} f''(t)=-\lambda^{1+2\varepsilon}\{&f''(f(\lambda^{1+\varepsilon}t)^{1+\varepsilon})f(\lambda^{1+\varepsilon}t)^{2\varepsilon}(1+\varepsilon)^2f'(\lambda^{1+\varepsilon}t)^2\\ &+f'(f(\lambda^{1+\varepsilon}t)^{1+\varepsilon})f(\lambda^{1+\varepsilon}t)^\varepsilon(1+\varepsilon)f''(\lambda^{1+\varepsilon}t)\\ &+f'(f(\lambda^{1+\varepsilon}t)^{1+\varepsilon})f(\lambda^{1+\varepsilon}t)^{\varepsilon-1}(1+\varepsilon)\varepsilon f'(\lambda^{1+\varepsilon}t)^2\}. \end{aligned} \tag{5.4}$$

Setting $t=1$ in (5.3) we get

$$\begin{aligned} 1+\lambda-\lambda^2\ell'(1)=&\lambda^\varepsilon\{(1+\lambda)-\lambda^2\ell'(f(\lambda^{1+\varepsilon})^{1+\varepsilon})\}\\ &\cdot\{1-(1+\lambda)\lambda^{1+\varepsilon}+\lambda^2\ell(\lambda^{1+\varepsilon})\}^\varepsilon(1+\varepsilon)\\ &\cdot\{1+\lambda-\lambda^2\ell(\lambda^{1+\varepsilon})\}, \end{aligned} \tag{5.5}$$

or

$$0=\lambda+\varepsilon+\varepsilon\log\lambda+\lambda^2\hat{N}_0(\ell,\varepsilon,\lambda), \tag{5.6}$$

i.e.

$$\varepsilon=-\frac{\lambda}{1+\log\lambda}+\lambda^2\hat{N}_1(\ell,\varepsilon,\lambda)\equiv N_\lambda(\ell,\varepsilon). \tag{5.7}$$

We also rewrite (5.4):

$$\ell''=\lambda K_\lambda(\ell,\varepsilon)\ell''+M_\lambda(\ell,\varepsilon), \tag{5.8}$$

where for $h\in C^0[0,1]$, and f defined as in (5.1),

$$\begin{aligned}(K_\lambda(\ell,\varepsilon)h)(t)=-\lambda^{2\varepsilon}\{&h(f(\lambda^{1+\varepsilon}t)^{1+\varepsilon})f(\lambda^{1+\varepsilon}t)^{2\varepsilon}(1+\varepsilon)^2f'(\lambda^{1+\varepsilon}t)\\&+h(\lambda^{1+\varepsilon}t)f'(f(\lambda^{1+\varepsilon}t)^{1+\varepsilon})f(\lambda^{1+\varepsilon}t)^{\varepsilon}(1+\varepsilon)\},\end{aligned} \tag{5.9}$$

and

$$M_\lambda(\ell,\varepsilon)(t)=-\lambda^{2\varepsilon}\frac{\varepsilon}{\lambda}f'(f(\lambda^{1+\varepsilon}t)^{1+\varepsilon})f(\lambda^{1+\varepsilon}t)^{\varepsilon-1}(1+\varepsilon)f'(\lambda^{1+\varepsilon}t)^2. \tag{5.10}$$

Finally, we define $\mathscr{L}$ by

$$(\mathscr{L}h)(t)=\int_0^t(t-\tau)h(\tau)d\tau-t\int_0^1(1-\tau)h(\tau)d\tau. \tag{5.11}$$

Instead of solving (5.2) directly, we rather discuss first the set of equations (for fixed, small $\lambda>0$)

$$\varepsilon=N_\lambda(\ell,\varepsilon), \tag{5.12a}$$

and

$$\ell=\mathscr{L}(I-\lambda K_\lambda(\ell,\varepsilon))^{-1}M_\lambda(\ell,\varepsilon). \tag{5.12b}$$

We claim that the solutions (ε,ℓ) of (5.12a, b) solve actually (5.2).

[*Proof*. (5.12a) implies (5.6) and (5.5), and hence (5.3) at $t=1$. Equation (5.12b) implies (5.4). Integrating, we find that (5.3) must hold up to an additive constant, but this constant is zero since we have already seen that (5.3) holds at $t=1$. Integrating again, we find that (5.12) holds up to an additive constant, which however must be zero, since, by (5.12b) and the definition of $\mathscr{L}$, $\ell(0)=\ell(1)=1$.]

We now discuss the existence and uniqueness of solutions of (5.12). [This also implies uniqueness of the solution of (5.2) since every solution of (5.2) solves (5.12).] The appropriate function space is

$$\mathfrak{E}=\{(\ell,\varepsilon),\ \ell\in C^2[0,1],\ \ell(0)=\ell(1)=0,\ \varepsilon\in\mathbb{C}\}$$

equipped with the norm

$$\|(\ell,\varepsilon)\|_{\mathfrak{E}}=\sup_{x\in[0,1]}|\ell''(x)|-|\varepsilon|\frac{\log\lambda+1}{2\lambda};\quad(\lambda\text{ is small}).$$

Let $\mathfrak{E}_1$ be the unit ball in $\mathfrak{E}$. $\mathfrak{E}$ is a Banach space. In fact $\ell = \mathscr{L}(\ell'')$, and $|(\mathscr{L}h)'(t)| + |(\mathscr{L}h)(t)| \leqq \text{const} \sup_{t\in[0,1]} |h(t)|$. (All spaces C^0, C^1 are on $[0,1]$.)

We claim that for sufficiently small $\lambda > 0$,

a) N_λ is a contraction[1] from $\mathfrak{E}_1$ to $\mathbb{C}$.

b) M_λ is a contraction from $\mathfrak{E}_1$ to C^1.

c) For $(\ell, \varepsilon) \in \mathfrak{E}_1$, $K_\lambda(\ell, \varepsilon)$ is a bounded linear map from C^0 to C^0 and from C^1 to C^1.

d) For $X_1, X_2 \in \mathfrak{E}_1$,

$$|K_\lambda(X_1)h - K_\lambda(X_2)h|_{C^0} \leqq \text{const}\, \|X_1 - X_2\|_{\mathfrak{E}} |h|_{C^1}.$$

e) $\mathscr{L}$ is bounded linear from C^0 to C^2, and $(\mathscr{L}h)(0) = (\mathscr{L}h)(1) = 0$.

Note that from c), d) above it follows that

$$\begin{aligned} &|(1-\lambda K_\lambda(X_1))^{-1}h - (1-\lambda K_\lambda(X_2))^{-1}h|_{C^0} \\ &\quad = \lambda |(1-\lambda K_\lambda(X_1))^{-1}(K_\lambda(X_1) - K_\lambda(X_2))(1-\lambda K_\lambda(X_2))^{-1}h|_{C^0} \\ &\quad = \lambda\, \text{const}\, \|X_1 - X_2\|_{\mathfrak{E}} |h|_{C^1}, \end{aligned}$$

and hence we see that (5.12) has a unique solution. It remains to verify the claims a)–e). This is tedious but straightforward. Each time we have to estimate

$$h(f_1(\lambda^{1+\varepsilon_1} t)^{1+\varepsilon_1}) - h(f_2(\lambda^{1+\varepsilon_2} t)^{1+\varepsilon_2}),$$

we use the formula

$$|h(u) - h(v)| \leqq |h|_{C^1} |u - v|. \tag{5.13}$$

This is responsible for the fact that the contractions in b) and d) lose a derivative.

We only comment on the otherwise trivial verifications:

a): Obvious from (5.5), (5.6), (5.7).

b): Obvious from (5.10). [Note that $f(\lambda^{1+\varepsilon} t) > \frac{1}{2}$ for small $\lambda \geqq 0$.]

c): Obvious from (5.9) since no derivatives of h occur.

d): Obvious from (5.9) using (5.13).

e): By construction, $(\mathscr{L}h)(0) = (\mathscr{L}h)(1) = 0$, cf. (5.11). On the other hand, a direct computation shows

$$|\mathscr{L}h|_{C^0} \leqq |h|_{C^0}; \quad |(\mathscr{L}h)'|_{C^0} \leqq |h|_{C^0}; \quad |(\mathscr{L}h)''|_{C^0} \leqq |h|_{C^0}.$$

This proves existence and uniqueness in C^2 and hence Theorem 2.2. The same proof works on any interval $[0, A]$ provided $\lambda > 0$ is sufficiently small, and $A > 1$.

6. Stable and Unstable Manifolds

In this section, we present a general argument deriving some analytic consequences from a geometric situation. The geometric situation is as follows:

We consider a twice continuously differentiable mapping $\mathscr{T}$ defined on an open set $\mathfrak{D}$ in a Banach space $\mathfrak{H}$ and taking values in $\mathfrak{H}$. We do not assume that $\mathscr{T}$

1 By contraction we mean a bounded, contractive map

maps $\mathfrak{D}$ into itself, but we do assume that it has a fixed point ϕ. We further assume that $D\mathscr{T}(\phi)$, the derivative of $\mathscr{T}$ at ϕ (which is a bounded linear operator on $\mathfrak{H}$) has spectrum which, except for a single simple eigenvalue $\delta > 1$, is entirely contained in the open unit circle. It then follows from invariant manifold theory that $\mathscr{T}$ admits a stable manifold W_s of codimension one and an unstable manifold W_u of dimension one. We will define these submanifolds precisely later; for present purposes, it is good enough to think of W_s as an invariant surface and W_u an invariant curve (the two of them crossing at the fixed point ϕ) with $\mathscr{T}$ acting in a purely contractive way on W_s and in a purely expansive way on W_u.

We also give ourselves two further objects:

A submanifold Π_1 of $\mathfrak{H}$ of codimension one which intersects W_u transversally at some point $\phi^* \neq \phi$.

A continuously differentiable parameterized curve $\mu \to \psi_\mu$ in $\mathfrak{H}$ which crosses the stable manifold W_s at $\mu = \mu_\infty$ with non-zero transverse velocity.

Although the notation in this section will be chosen to suggest the application of the results obtained here to the main topic of the paper, the reader should note that these results depend only on a few explicit assumptions about the objects under consideration. Symbols like $\mathscr{T}$, ϕ, W_u, etc. are used in a more general sense here than in the remainder of the paper. Moreover, Π_1 will be later on identified with Σ_1 (see Fig. 3).

From this set-up we want to conclude:

a) There exists a sequence μ_n (perhaps defined only for large n), converging to μ_∞, with $\mathscr{T}^{n-1}\psi_{\mu_n} \in \Pi_1$, and such that $\lim_{n\to\infty} \delta^n(\mu_n - \mu_\infty)$ exists and is non-zero.

b) The sequence $\mathscr{T}^{n-1}\psi_{\mu_n}$ converges to ϕ^*.

The significance of these conclusions has been discussed in the introduction. (We would like to be able to make the more precise assertion that μ_n is the unique value of μ near μ_∞ such that $\mathscr{T}^{n-1}\psi_\mu \in \Pi_1$. Whether this is true or not depends on relatively inaccessible global properties of $\mathscr{T}$; but we can say, informally, that μ_n is the unique such μ for which $\mathscr{T}\psi_\mu, \mathscr{T}^2\psi_\mu, \ldots, \mathscr{T}^{n-2}\psi_\mu$ all lie between W_s and Π_1.)

The first step in our analysis will be to define precisely what we mean by stable and unstable manifolds. This is not entirely routine since, in the application we have in mind, the transformation $\mathscr{T}$ is not invertible; in fact, it is not even locally one-one near ϕ.

If $\mathfrak{V}$ is a sufficiently small open ball in $\mathfrak{H}$ with center ϕ then

$$\{\psi \in \mathfrak{V} : \mathscr{T}^j\psi \in \mathfrak{V} \quad \text{for} \quad j = 1, 2, 3, \ldots\}$$

may be shown to be a smooth connected submanifold of $\mathfrak{V}$ of codimension one. We will call this set a *local stable manifold* for $\mathfrak{V}$ at ϕ and denote it by $W_s^{(0)}$. It passes through ϕ and is tangent there to the stable eigenspace for $D\mathscr{T}(\phi)$, i.e. the spectral subspace for the part of the spectrum which is inside the unit disk. The set $W_s^{(0)}$ is mapped into itself by $\mathscr{T}$, and the sequence of sets $\mathscr{T}^j W_s^{(0)}$ shrinks to $\{\phi\}$, i.e. is eventually contained in any neighborhood of ϕ. The proofs of these facts, as well as those to be cited in the next paragraph on local unstable manifolds, can be found in the monograph of Hirsch et al. [6].

If we define (again for $\mathfrak{V}$ a small open ball with center ϕ)

$$\mathfrak{V}_0 = \mathfrak{V}; \qquad \mathfrak{V}_{j+1} = (\mathscr{T}\mathfrak{V}_j) \cap \mathfrak{V},$$

then $\bigcap_j \mathfrak{V}_j$ is a smooth connected one-dimensional submanifold of $\mathfrak{V}$, passing through ϕ and tangent there to the eigenspace of $D\mathscr{T}(\phi)$ corresponding to the large eigenvalue δ. We call this set a *local unstable manifold* for $\mathscr{T}$ at ϕ and denote it by $W_u^{(0)}$.

We have:

$$\mathscr{T} W_u^{(0)} \supset W_u^{(0)}$$

and, for any $\psi \in W_u^{(0)}$ and any $j=1,2,3,\ldots$, there is a unique $\psi_j \in W_u^{(0)}$ such that

$$\mathscr{T}^j \psi_j = \psi;$$

moreover, the sequence (ψ_j) converges to ϕ.

The globalization of the stable and unstable manifolds is complicated by the non-invertibility of $\mathscr{T}$. We will simply define what we mean by *a* stable or unstable manifold without investigating the existence of a unique largest one. Thus we define:

A *stable manifold* for $\mathscr{T}$ is a smooth codimension-one submanifold W_s of the domain of $\mathscr{T}$ such that:

a) $\mathscr{T} W_s \subset W_s$.

b) If $\psi \in W_s$, then $\lim_{j\to\infty} \mathscr{T}^j \psi = \phi$. (Note that this implies that $\mathscr{T}^j \psi \in W_s^{(0)}$ for sufficiently large j.)

c) (Transversality.) For any ψ in W_s, the range of $D\mathscr{T}(\psi)$ is not contained in the tangent space to W_s at $\mathscr{T}\psi$.

An *unstable manifold* for $\mathscr{T}$ is a smooth one-dimensional submanifold W_u of $\mathfrak{H}$ (not necessarily contained in the domain of $\mathscr{T}$) such that

a) $\mathscr{T}(W_u \cap \mathfrak{D}(\mathscr{T})) \supset W_u$.

b) If $\psi \in W_u$, there is a sequence ψ_j converging to ϕ such that $\psi = \mathscr{T}^j \psi_j$. (This implies that $W_u \subset \bigcup_{j=1}^{\infty} \mathscr{T}^j W_u^{(0)}$.)

c) For any $\psi \in W_u \cap \mathfrak{D}(\mathscr{T})$, the tangential derivative of $\mathscr{T}$ along W_u at ψ does not vanish.

Since $W_s^{(0)}$ and $W_u^{(0)}$ are, respectively, stable and unstable manifolds, stable and unstable manifolds do exist.

We now need some special terminology. Let Π_j, $j=1,2,3,\ldots$ and W be submanifolds of $\mathfrak{H}$ of codimension one. We will say that the sequence Π_j *converges to* W *exponentially with rate* δ (δ a real number larger than one) if, for each $\psi \in W$ there is a diffeomorphism from $\mathscr{X}_1 \times (-1,1)$, $\mathscr{X}_1$ the open unit ball in some Banach space $\mathscr{X}$, onto a neighborhood $\mathfrak{V}$ of ψ (i.e. a set of local coordinates at ψ) such that

1. ψ is the image of $(0,0)$.
2. $W \cap \mathfrak{V}$ is the image of $\mathscr{X}_1 \times \{0\}$.
3. For each sufficiently large j, $\Pi_j \cap \mathfrak{V}$ is the image of the graph of a mapping $\hat{\Pi}_j : \mathscr{X}_1 \to (-1,1)$,

where

4. $\delta^j \hat{\Pi}_j$ converges in the C^1 topology on $\mathscr{X}_1$ to a nowhere vanishing limit.

Intuitively, this means that the separation between Π_j and W varies asymptotically (for large j) like δ^{-j} multiplied by a differentiable function of position on W.

The following proposition is nearly obvious:

Proposition 6.1. *Let Π_j converge exponentially to W with rate δ, and let $\mu \to \psi_\mu$ be a continuously differentiable parametrized curve in $\mathfrak{H}$ crossing W with non-zero transverse velocity at $\mu = \mu_\infty$. There is then a sequence $\mu_j \to \mu_\infty$ (defined for sufficiently large j) such that $\psi_{\mu_j} \in \Pi_j$; the quantity $\delta^j(\mu_\infty - \mu_j)$ converges as $j \to \infty$ to a finite non-zero limit.*

Returning to the principal objective of this section, we see that the proof of a) p. 233 now reduced to constructing appropriately localized preimages Π_j of Π_1 under $\mathscr{T}^{j-1}$ and showing that they converge exponentially to W_s with rate δ. The following theorem asserts that this is possible; it also asserts that b) holds.

Theorem 6.2. *Let $\mathscr{T}$, ϕ, W_s, W_u, δ, Π_1, and ϕ^* be as above. Then there exists a sequence (Π_j) of codimension-one submanifolds of $\mathfrak{H}$, converging exponentially to W_s with rate δ, such that*

$$\mathscr{T}^{j-1}\Pi_j \subset \Pi_1 .$$

Moreover, if $\psi \in W_s$ and if $\mathfrak{W}$ is a sufficiently small neighborhood of ψ in $\mathfrak{H}$, then

$$\mathscr{T}^{j-1}(\Pi_j \cap \mathfrak{W}) \to \{\phi^*\} \quad \text{as} \quad j \to \infty .$$

The first step in proving this theorem is to reduce it to a statement which is local at ϕ. More precisely, we claim that the theorem as stated is true if we can find an open neighborhood $\mathfrak{V}$ of ϕ such that it is true for W_s and W_u replaced by $W_s \cap \mathfrak{V}$ and $W_u \cap \mathfrak{V}$ respectively, with the added assumption that $\Pi_1 \subset \mathfrak{V}$. Proof of this claim is straight-forward, using (notably) the transversality conditions in the definition of stable and unstable manifolds. We will sketch one part of the argument, showing that there is no loss of generality in assuming that $\Pi_1 \subset \mathfrak{V}$.

As before, we let ϕ^* denote the (first) point where W_u intersects Π_1. Since $\phi^* \in W_u$, there is an integer k and a point $\phi_k^* \in W_u \cap \mathfrak{V}$ such that $\mathscr{T}^{k-1}\phi_k^* = \phi^*$. By our definition of W_u, the tangential derivative of $\mathscr{T}^{k-1}$ along W_u at ϕ_k^* does not vanish. From this (and the implicit function theorem) it follows that, for $\mathfrak{U}'$ a sufficiently small open ball about ϕ_k^*,

$$\Pi_1' = \{\psi \in \mathfrak{U}' : \mathscr{T}^{k-1}\psi \in \Pi_1\}$$

is a smooth codimension-one submanifold of $\mathfrak{V}$ intersecting W_u transversally at ϕ_k^*. The localized version of the theorem implies the existence of a sequence of surfaces Π_j' converging exponentially to $W_s \cap \mathfrak{V}$ with rate δ and with

$$\mathscr{T}^{j-1}\Pi_j' \subset \Pi_1' .$$

We can thus take

$$\Pi_{j+1} = \Pi_{j+1}' .$$

This establishes localizability in the expanding direction. A similarly straightforward argument, which we omit, establishes localizability in the contracting direction.

Thus, we have only to prove the localized version of the theorem. We do this by choosing special coordinates in which $\mathscr{T}$ and Π_1 take particularly simple forms. The result we need is the following:

If ϕ^* is close enough to ϕ, then there exists a C^1 diffoemorphism from a set of the form $\mathscr{X}_1 \times [-1,1]$, $\mathscr{X}_1$ the unit ball in some Banach space, onto a neighborhood $\mathfrak{V}$ of ϕ such that

ϕ is the image of $(0,0)$,

$W_s \cap \mathfrak{V}$ is the image of $\mathscr{X}_1 \times \{0\}$,

$W_u \cap \mathfrak{V}$ is the image of $\{0\} \times (-1,1)$,

$\Pi_1 \cap \mathfrak{V}$ is the image of $\mathscr{X}_1 \times \{1\}$.

If we regard $x \in \mathscr{X}_1$, $y \in [-1,1]$ as coordinates for their image in $\mathfrak{V}$, then in these coordinates $\mathscr{T}$ takes the form

$$\mathscr{T}: (x,y) \to (M(x,y), \delta y),$$

where

$$\|M(x_1,y) - M(x_2,y)\| \leqq \alpha \|x_1 - x_2\|,$$

with $\alpha < 1$, and $M(0,y) = 0$.

In terms of these coordinates we can take simply

$$\Pi_j = \text{image of } \mathscr{X}_1 \times \{\delta^{-(j-1)}\}$$

and this sequence of surfaces converges exponentially to $W_s \cap \mathfrak{V}$ with rate δ. Moreover, in view of the contractivity of $M(x,y)$ in x, the diameter of $\mathscr{T}^{j-1}\Pi_j$ goes to zero as $j \to \infty$. Since this set always contains ϕ^*, we have

$$\mathscr{T}^{j-1}\Pi_j \to \{\phi^*\} \quad \text{as} \quad j \to \infty.$$

The proof of the theorem is thus reduced to proving the existence of the indicated "normal coordinates" for $\mathscr{T}$ and Π_1. We will concentrate on showing the existence of coordinates in which $\mathscr{T}$ has the desired form, since this result may be of interest in other contexts; we will then at the end sketch how the argument can be modified to bring Π_1 into normal form as well. The proof of the following theorem is based on an analogous but more complicated result proved in Collet and Eckmann [1].

Theorem 6.3. *Let $\mathscr{T}$ be a twice continuously differentiable mapping from an open set $\mathfrak{D}$ in a Banach space into the Banach space. Let ϕ be a fixed point for $\mathscr{T}$. Assume that $D\mathscr{T}(\phi)$ has a single simple eigenvalue $\delta > 1$ and that the rest of its spectrum is in the interior of the unit disk. Then there exists a C^1 diffeomorphism of $\mathscr{X}_1 \times (-1,1)$ ($\mathscr{X}_1$ denoting the open unit ball in some Banach space) onto a neighborhood $\mathfrak{V}$ of ϕ – i.e. a set of local coordinates at ϕ – such that*

$(0,0)$ represents ϕ,

$\mathscr{X}_1 \times \{0\}$ represents $W_s \cap \mathfrak{V}$,

$\{0\} \times (-1,1)$ represents $W_u \cap \mathfrak{V}$,

$\mathscr{T}$ takes the form $(x,y) \to (M(x,y), \delta y)$,

where

$$M(0,y) = 0; \quad \|D_x M(x,y)\| \leqq \alpha < 1 \quad \text{for} \quad (x,y) \in \mathscr{X}_1 \times (-1,1).$$

We emphasize

1. In these coordinates the action of $\mathscr{T}$ in the y direction has been made *exactly* linear.

2. We do not require – and it is generally not possible – that $\mathscr{T}$ be linearized smoothly in the x direction as well.

The first step in the proof is to introduce coordinates which are approximately right. We do this in a sequence of steps:

– Make a translation to put ϕ at the origin.

– Write the Banach space as the direct sum of the one-dimensional eigenspace corresponding to the eigenvalue δ (y direction) and the complementary spectral subspace (x direction).

– Carry out an x-dependent translation in the y direction to bring the stable manifold to the surface $\{y=0\}$.

– Carry out a y-dependent translation in the x direction to bring the unstable manifold to the line segment $x=0$.

– Reparametrize the y coordinate in such a way that the action of $\mathscr{T}$ *on the unstable manifold* becomes exactly multiplication by δ. The possibility of doing this follows from a trivial case of the Sternberg Linearization Theorem. (See, for example, Hartmann [5].)

Thus we can assume that:

1. The domain of $\mathscr{T}$ is $\mathscr{X}_1 \times(-1,1)$ and $\mathscr{T}(x,y)=(M(x,y), \delta y+N(x,y))$ where M, N are continuously differentiable and $M(0,y)=0$; $N(x,0)=0$; $N(0,y)=0$.

In the course of the argument, we will need to assume that the nonlinear terms in $\mathscr{T}$ are small. This can be accomplished by *magnification*, i.e. by replacing (x,y) with new coordinates $x'=\lambda x$, $y'=\lambda y$ with λ large (and restricting the domain to the set $\|x'\| \leqq 1$, $|y'| \leqq 1$). This transformation leaves the linear terms unchanged but shrinks the nonlinear terms by a factor of at least $1/\lambda$. In this way (and possibly also renorming the space $\mathscr{X}$) we can arrange that, for some $\beta>0$,

2. $$\left.\begin{aligned} &\|D_x M(x,y)\| \leqq \alpha<1 \\ &1<\beta \leqq \frac{\delta y+N(x,y)}{y} \end{aligned}\right\} \quad \text{for} \quad \|x\|<1, \quad |y|<1.$$

It will also be convenient to assume that $\mathscr{T}(x,y)$ is defined and well behaved for *all* y. To extend it, we choose a smooth cut-off function $\varrho(y)$,

$$0 \leqq \varrho(y) \leqq 1,$$

with

$$\varrho(y)=1 \quad \text{for} \quad |y| \leqq \frac{1}{\beta}$$

$$\varrho(y)=0 \quad \text{for} \quad |y| \geqq 1$$

and modify $\mathscr{T}$ to the mapping

$$\begin{aligned} (x,y) &\to (\varrho(y)M(x,y), \delta y+\varrho(y)N(x,y)), && |y|<1 \\ &\to (0, \delta y), && |y| \geqq 1. \end{aligned}$$

The modified $\mathscr{T}$ agrees with the original $\mathscr{T}$ on all (x,y) with $\mathscr{T}(x,y) \in \mathscr{X}_1 \times[-1,1]$; it maps $\mathscr{X}_1 \times \mathbb{R}$ into itself and satisfies the inequalities (2.) for all y. We will work, from now on, with this modified $\mathscr{T}$.

To prove the theorem, it suffices to find a continuously differentiable function z such that

$$z \circ \mathscr{T} = \delta \cdot z, \qquad z(0, y) = y.$$

The inverse function theorem then assures us that (x, z) is a set of local coordinates at $(0, 0)$ and $\mathscr{T}$ evidently has the desired form in these coordinates. To construct z, we let $y_n(x, y)$ denote the y coordinate of $\mathscr{T}^n(x, y)$, and we put

$$z_n(x, y) = \delta^{-n} y_n(x, y).$$

It is immediate that $z_n(x, y)$ is continuously differentiable and that $z_n(0, y) = 0$. Also,

$$z_n \circ \mathscr{T} = \delta \cdot z_{n+1}$$

so if

$$\lim_{n \to \infty} z_n \equiv z$$

exists, we have immediately

$$z \circ \mathscr{T} = \delta \cdot z$$

as desired. Evidently, $z_n(x, 0) = 0$, so the limit exists in this case. On the other hand, if $y \neq 0$, then, since

$$|y_n| \geqq \beta^n |y|,$$

$|y_n|$ is eventually larger than one. But because of the way $\mathscr{T}$ is cut off, $y_{n+1} = \delta y_n$ if $|y_n| \geqq 1$, and so $z_{n+1}(x, y) = z_n(x, y)$. Hence, for all x, y

$$z(x, y) = \lim_{n \to \infty} z_n(x, y)$$

exists, and the only problem is to prove that it is continuously differentiable. Note, incidentally, that if $y \neq 0$ and n is sufficiently large, $z = z_n$ on a neighborhood of (x, y) and so z is continuously differentiable on that neighborhood; the only place where differentiability could fail is for $y = 0$.

We will write

$$z_n(x, y) = y + r_n(x, y);$$

then a simple computation using

$$z_{n+1} = \frac{1}{\delta} z_n \circ \mathscr{T}$$

and the expression (1.) for $\mathscr{T}$ gives:

$$r_{n+1}(x, y) = \frac{1}{\delta} N(x, y) + \frac{1}{\delta} r_n \circ \mathscr{T}(x, y).$$

Defining a linear operator L by

$$Ls(x, y) = \frac{1}{\delta} s \circ \mathscr{T}(x, y)$$

we see that

$$r_{n+1} = \frac{1}{\delta} \sum_{j=0}^{n} L^j N ,$$

so, if we can find a normed space containing N on which L is a contraction, we can write

$$z - y = \frac{1}{\delta} \sum_{j=0}^{\infty} L^j N .$$

We will use the norm

$$|||s||| = \max \left\{ \sup_{x,y} \frac{|s(x,y)|}{\|x\|}, \sup_{x,y} \frac{|D_y s(x,y)|}{\|x\|}, \sup_{x,y} \|D_x s(x,y)\| \right\}$$

on the space of those continuously differentiable functions vanishing for $x = 0$ for which the norm is finite. (It is to guarantee that N belongs to this space that we have to assume that $\mathscr{T}$ is twice continuously differentiable. We could make do with less, but simple continuous differentiability does not seem to be enough.)

The proof that L is a contraction in this norm is straightforward; we will describe explicitly only the most sensitive of the estimates,

$$D_y(Ls)(x,y) = \frac{1}{\delta} (D_x s)(M(x,y), \delta y + N(x,y)) \cdot D_y M(x,y)$$

$$+ \frac{1}{\delta} (D_y s)(M(x,y), \delta y + N(x,y))(\delta + D_y N(x,y)),$$

$$\frac{1}{\delta} \frac{|D_x s(M(x,y), \delta y + N(x,y)) D_y M(x,y)|}{\|x\|} \leqq \frac{1}{\delta} |||s||| \sup_{x,y} \|D_x D_y M(x,y)\| .$$

Now $D_x D_y M(x,y)$ comes from the nonlinear terms in $\mathscr{T}$ which can be made small by magnification, so we can estimate this expression by an arbitrarily small multiple of $|||s|||$. From

$$M(0,y) = 0 ; \qquad \|D_x M(x,y)\| \leqq \alpha$$

it follows that

$$\|M(x,y)\| \leqq \alpha \cdot \|x\|$$

and thus

$$\left| \frac{1}{\delta} \frac{D_y s(M(x,y), \delta y + N(x,y))(\delta + D_y N(x,y))}{\|x\|} \right| \leqq \alpha |||s||| \left(1 + \frac{1}{\delta} \sup_{x,y} \|D_y N(x,y)\| \right).$$

Again, the last term can be made arbitrarily small by magnification. The final result is, then, that we have an estimate

$$\sup_{x,y} \left\{ \frac{|D_y Ls(x,y)|}{\|x\|} \right\} \leqq |||s||| \times (\alpha + \text{something small}).$$

The other two terms are estimated similarly.

Hence, L is a contraction so the series

$$\frac{1}{\delta}\sum_{j=0}^{\infty} L^j N$$

converges in this normed space; the sum is $z-y$ so z is continuously differentiable.

Remark. Similar estimates show that if $\alpha\cdot\delta^{r-1}<1$ then z is C^r (assuming that $\mathscr{T}$ is C^{r+1}).

On the other hand, no matter how smooth $\mathscr{T}$ is, it is usually not possible to find a C^r solution to $z\circ\mathscr{T}=\delta\cdot z$, $z(0,y)=y$ if one of $\frac{1}{\delta}, \frac{1}{\delta^2}, \ldots, \frac{1}{\delta^{r-2}}$ can be written as a product of points of the spectrum of $D_x M(0,0)$.

Finally, we have to show how to modify the above argument to bring a codimension-one surface crossing the unstable manifold transversally away from the fixed point simultaneously into the desired normal form, a flat horizontal surface. We start as before, choosing coordinates in a neighborhood of ϕ in which $\mathscr{T}$ takes the form (1.) and (2.) holds. Let $\hat{\Pi}_1$ intersect the unstable manifold inside this coordinate neighborhood; then a part of $\hat{\Pi}_1$ near the intersection point can be represented in our coordinate system as the graph of a function $x\to\hat{\Pi}_1(x)\in(-1,1)$ with $\hat{\Pi}_1$ defined in a neighborhood of 0. By magnifying in the x dimension we can assume that $\hat{\Pi}_1$ is defined and non-zero on all of $\mathscr{X}_1$.

Now define a new y coordinate by

$$y_{\text{new}}=\frac{y_{\text{old}}}{\hat{\Pi}_1(x)}.$$

In terms of the new coordinates, (1.) still holds; if (2.) doesn't then it can be made to hold by a further magnification in the x direction. We have thus arranged so that (1.) and (2.) hold, and, in addition, Π_1 corresponds to the surface $\{y=1\}$. We then proceed to cut off and extend N, M as indicated. Note that, because of the way we have done the cutting off, the inverse image under $\mathscr{T}$ of any set in $\mathscr{X}_1\times[-1,1]$ is exactly the same as it was before the modification of $\mathscr{T}$ but, on the other hand,

$$\mathscr{T}^n(x,1)=(0,\delta^n).$$

From this last equation, it follows that

$$z(x,1)=1$$

i.e. that Π_1 corresponds to the surface $\{z=1\}$.

It still has to be shown that z is continuously differentiable. To prove that, one needs to know that

$$\sup_{x,y}\|D_x D_y M(x,y)\|, \quad \sup_{x,y}\|D_y N(x,y)\|, \quad \sup_{x,y}\|D_x N(x,y)\|$$

are all small. This would normally be arranged by magnifying sufficiently *before* cutting off in the y direction. We note, however, that the magnification can just as well be done *after* cutting off, and that this does not spoil the flatness of Π_1 in the z coordinate.

7. The Global Unstable Manifold

The preceding section shows that the universality of the rate of period doubling can be understood if the unstable manifold for the fixed point ϕ_ε crosses the surface $\Sigma_1 = \{\psi : \psi(1) = 0\}$ transversally. In this section, we estimate the global structure of the unstable manifold. In essence, we show that it remains close to the line segment $\{\psi(x) = 1 - (1+a)|x|^{1+\varepsilon} : -1 < a < 1\}$ throughout the full length of this segment, provided that ε is small enough.

We need yet another realization of the action of $\mathscr{T}$ in convenient coordinates. Recall that we showed that if we write

$$\psi(x) = f(|x|^{1+\varepsilon}); \quad f(t) = 1 - th(t);$$

then in terms of h, $\mathscr{T}$ takes the form

$$h \to \eta_2 \{h(1 - at\eta_2) + \Delta_1 h(1 - at\eta_2)\}, \tag{7.1}$$

where the notation is as in Sect. 4. Deviating from the notation of that section, we will write

$$h(t) = 1 + a + (t-1)g(t) \tag{7.2}$$

and determine the form of $\mathscr{T}$ expressed in terms of the coordinates $a (\in (-1, 1))$ and $g (\in \mathfrak{H})$. Note that

$$1 + a = h(1)$$
$$g(t) = (\Delta_1 h)(t)$$

so it is easy, in principle, to read off this form from (7.1). We will need, however, to look with some care at the expression for η_2. Recall

$$\eta_1 = a^\varepsilon h(a^{1+\varepsilon} t)$$
$$(1 - at\eta_1)^{1+\varepsilon} = 1 - at\eta_2 .$$

A straightforward computation shows that

$$(1-z)^{1+\varepsilon} = 1 - z(1 + \varepsilon s(z, \varepsilon)),$$

where $s(z, \varepsilon)$ (which can easily be written explicitly) is analytic for all ε and all $z \notin [1, \infty)$. Hence

$$\eta_2 = \eta_1 (1 + \varepsilon s(-at\eta_1, \varepsilon)).$$

Also,

$$\begin{aligned} h(1 - at\eta_2) + \Delta_1 h(1 - at\eta_2) &= 1 + a - at\eta_2 g(1 - at\eta_2) + g(1 - at\eta_2) \\ &= 1 + a + (1 - at\eta_2) g(1 - at\eta_2). \end{aligned}$$

Thus the transformed h becomes

$$a^\varepsilon (1 + a + (a^{1+\varepsilon} t - 1) g(a^{1+\varepsilon} t))(1 + \varepsilon s(-at\eta_1, \varepsilon))(1 + a + (1 - at\eta_2) g(1 - at\eta_2)). \tag{7.3}$$

We will write $A(\varepsilon, a, g)$ and $G(\varepsilon, a, g)$ for the a, g components of the representation of (7.3) in the form (7.2). Thus, to get $A(\varepsilon, a, g)$ we have simply to insert $t = 1$ in (7.3)

and then subtract 1. Doing this, and grouping terms in a straightforward way, we find

$$A = a^{\varepsilon}(1+a)^2 - 1 + A_1 + \varepsilon A_2,$$

where $A_1 = 0$ for $g = 0$ and where A_1, A_2 are both "regular". The meaning of "regular" has to be specified with a bit of care. The formulas are full of factors of a^{ε} whose a-derivatives become infinite as a approaches zero. The idea is that these are the only singularities for small ε, a, g. One way to formulate this is indicated in the following proposition.

Proposition 7.1. *We can write*

$$\begin{aligned} A &= a^{\varepsilon}(1+a)^2 - 1 + A_1(\varepsilon, a, a^{\varepsilon}, g) + \varepsilon A_2(\varepsilon, a, a^{\varepsilon}, g) \\ G &= aG_1(\varepsilon, a, a^{\varepsilon}, g) + a\varepsilon G_2(\varepsilon, a, a^{\varepsilon}, g) \end{aligned} \tag{7.4}$$

(A_1 and A_2 take values in $\mathbb{R}$; G_1 and G_2 in $\mathfrak{H}(\Omega)$) where A_1, A_2, G_1, G_2 are all infinitely differentiable with bounded derivatives on

$$(0, \varepsilon_0) \times (0, a_0) \times (0, 1) \times \{g \in \mathfrak{H} : \|g\| < g_0\}$$

(for sufficiently small ε_0, a_0, g_0) and where A_1 and G_1 vanish identically for $g = 0$.

We have already indicated the proof for A. To get the formula for G, apply Δ_1 to (7.3) using the "product rule"

$$\Delta_1 g_1 g_2 = g_1 \Delta_1 g_2 + g_2(1) \Delta_1 g_1;$$

group the contributions from differencing in the first and third factors to form $a \cdot G_1$ and the contribution from the second factor to form $a \cdot \varepsilon \cdot G_2$. To extract the indicated explicit factors of a, make repeated use of the "chain rule"

$$(\Delta_{z_0}(g_1 \circ g_2))(z) = (\Delta_{g_2(z_0)} g_1)(g_2(z)) \Delta_{z_0} g_2(z)$$

and observe that every t in (7.3) is accompanied by a multiplicative factor of a.

Corollary 7.2. *There exist constants B_1, B_2 such that*

$$\|G\| \leqq a \cdot B_1 \|g\| + a\varepsilon B_2 \qquad (0 < a < a_0,\ 0 < \varepsilon < \varepsilon_0,\ \|g\| < g_0). \tag{7.5}$$

In particular, at the fixed point, $a = \lambda_{\varepsilon}$ and we write $g_{\varepsilon}^{(0)}$ for the corresponding g. Then

$$\|g_{\varepsilon}^{(0)}\| \leqq \lambda_{\varepsilon} B_1 \|g_{\varepsilon}^{(0)}\| + \lambda_{\varepsilon} \varepsilon B_2$$

from which it follows at once that

$$\|g_{\varepsilon}^{(0)}\| \leqq \frac{\varepsilon \lambda_{\varepsilon} B_2}{(1 - \lambda_{\varepsilon} B_1)} = O(\varepsilon \lambda_{\varepsilon})$$

for small ε. Since $\lambda_{\varepsilon} = O(-\varepsilon \log \varepsilon)$ for small ε, this estimate gives exactly the estimate

$$\phi_{\varepsilon}(x) = 1 - (1 + \lambda_{\varepsilon})|x|^{1+\varepsilon} + O(-\varepsilon^2 \log \varepsilon)$$

announced in Sect. 2.

Corollary 7.3. *If*

$$a(B_1 + B_2) \leqq 1,$$

and if

$$\|g\| \leqq \varepsilon,$$

then

$$\|G\| \leqq \varepsilon.$$

In other words, $\mathscr{T}$ can push a point out of the cylinder $\{(a, g) : 0 < a < a_1, \|g\| \leqq \varepsilon\}$, [with a_1 the smaller of a_0, $1/(B_1 + B_2)$], only through the ends.

We have noted already that singularities appear when we differentiate a^ε for a near zero. However, since

$$\frac{d}{da} a^\varepsilon = \frac{\varepsilon}{a} a^\varepsilon$$

the singularities are not much in evidence in the first derivative for $a \geqq \varepsilon$. Since the fixed point occurs at $\lambda_\varepsilon \gg \varepsilon$, we don't have to worry about the singularities when looking at first derivatives near and beyond the fixed point. Thus:

Corollary 7.4. *There exist constants* B_3, B_4, B_5 *such that*

$$\left\| \frac{d}{da} G(a, g_a) \right\| \leqq B_3 \|g_a\| + B_4 \varepsilon + B_5 a \left\| \frac{dg}{da} \right\|$$

for any differentiable mapping $a \to g_a$ *provided*

$$\varepsilon < a < a_0; \qquad \|g_a\| \leqq g_0.$$

Proof. The B_3 term comes from the explicit aa^ε dependence in aG_1; the B_4 term from the explicit aa^ε dependence in aG_2; the B_5 term from the g dependence of the sum.

Corollary 7.5. *Consider a curve given as the graph of a function* $a \to g_a$, *defined in a subinterval of* (ε, a_0). *Assume*

$$\frac{dA(a, g_a)}{da} \geqq 1. \tag{7.6}$$

Then the image of this curve under the action of $\mathscr{T}$ *is again representable as the graph of a function*

$$\hat{a} \to \hat{g}_{\hat{a}}$$

and

$$\left| \frac{d\hat{g}}{d\hat{a}} A(\varepsilon, a, g) \right| \leqq B_3 \|g_a\| + B_4 \varepsilon + B_5 a \left\| \frac{dg_a}{da} \right\|. \tag{7.7}$$

We can apply this corollary in particular to a (possibly very small) local unstable manifold. Such a manifold can be represented as the graph of a mapping defined

on a small interval I about λ_ε; condition (7.6) is satisfied and we can assume

$$N = \sup_{a \in I} \|g_a\| \ll \varepsilon .$$

Put

$$M = \sup_{a \in I} \left\| \frac{dg}{da} \right\| .$$

Since the local unstable manifold is mapped onto itself by $\mathscr{T}$, we get

$$M \leqq B_3 N + B_4 \varepsilon + 2 B_5 \lambda_\varepsilon M$$

or, for small ε,

$$M \leqq 2 B_4 \varepsilon .$$

Consider now a manifold specified by

$$b \to g_b$$

defined on a subinterval of (ε, a), and satisfying

$$\|g_a\| \leqq \varepsilon ; \qquad \left\| \frac{dg_a}{da} \right\| \leqq 2(B_3 + B_4)\varepsilon . \tag{7.8}$$

From (7.4) it is easy to check that there exists a constant B_6 such that, under these hypotheses,

$$\frac{dA(a, g_a)}{da} \geqq 2 - B_6 \varepsilon$$

(and so $\geqq 1$ for small ε); then from (7.7) the image curve satisfies

$$\left\| \frac{d\hat{g}}{d\hat{a}} \right\| \leqq (B_3 + B_4)\varepsilon + B_5 a 2(B_3 + B_4)\varepsilon$$

so if

$$a < a_0 ; \qquad a(B_1 + B_2) < 1 ; \qquad 2aB_5 < 1$$

we get again

$$\left\| \frac{dg}{da} \right\| \leqq 2(B_3 + B_4)\varepsilon .$$

Summarizing and simplifying the notation, we get:

Proposition 7.6. *There exist constants $\varepsilon_1 > 0$; $a_1 > 0$, B_7 such that, if $0 < \varepsilon < \varepsilon_1$, and if we have a curve in $\mathscr{P}_\varepsilon$ specified by*

$$a \to g_a$$

defined on a subinterval of (ε, a_1) and satisfying

$$\|g_a\| \leqq \varepsilon ; \qquad \left\| \frac{dg_a}{da} \right\| \leqq B_7 \varepsilon \tag{7.9}$$

then

$$\frac{d}{da} A(\varepsilon, a, g_a) \geqq 1.5, \tag{7.10}$$

so the image curve also admits a representation as

$$a \to \hat{g}_a$$

and the bounds (7.9) *hold with g replaced by* $\hat{g}$. *A sufficiently small local unstable manifold satisfies these hypotheses.*

Now let $\varepsilon < \varepsilon_1$, a small local unstable manifold, and apply $\mathscr{T}$ to it repeatedly, throwing away at each step the part of the image curve with a outside of (ε, a_1). In view of (7.10), we obtain after a finite number of iterations a curve $a \to g_a^*$ defined on all of (ε, a_1) contained in the unstable manifold and satisfying the inequalities (7.9).

It remains to show that the unstable manifold in the form $a \to g_a^*$ can be extended to all $a \in (-1, 1)$. We will consider only the problem of extending it to values of $a \geqq a_1$; the extension to $a \in (-1, \varepsilon)$ is similar but easier.

If we write

$$f_a^*(t) = 1 - t(1 + a + (t-1)g_a^*), \qquad \phi_a^*(x) = f_a^*(|x|^{1+\varepsilon}),$$

it will suffice, by the general theory of Sect. 3, to show that g_a^* can be extended to a value of a such that

$$\phi_a^*(a) = a$$

(i.e. such that $(\phi_a^*)^4(0)$ is the fixed point of ϕ_a^* in $[0, 1]$); then one more iteration of $\mathscr{T}$ gives values of a running all the way to 1. By continuity, it will suffice to extend it to a value of a such that

$$\phi_a^*(a) < a. \tag{7.11}$$

As above, we will write ϕ^* in the form

$$\phi_a^*(x) = f_a^*(|x|^{1+\varepsilon}); \qquad f_a^*(t) = 1 - t[1 + a + (t-1)g(t)].$$

Ignoring g and terms of order ε, we get

$$\phi_a^*(a) \approx 1 - a(1 + a);$$

from this it is easy to see that (7.11) will hold provided $a(2 + a) > 1$, i.e. $a > \sqrt{2} - 1$ and provided $\|g\|$ and ε are small enough.

Next we will argue that we can take $a_0 > \sqrt{2} - 1$ in Proposition 7.1. To do this we need, for the first time, to be careful about our choice of the domain Ω. Examination of the proof of Proposition 7.1 shows that the only limitations on ε_0, a_0, and g_0 are that

1) $a^{1+\varepsilon}\bar{\Omega} \subset \Omega$

2) $f(a^{1+\varepsilon}\bar{\Omega}) \cap (-\infty, 0] = \emptyset$

3) $[f(a^{1+\varepsilon}\bar{\Omega})]^{1+\varepsilon} \subset \Omega$

for $0<a<a_0$; $0<\varepsilon<\varepsilon_0$; $\|g\|<g_0$. If these three conditions are satisfied with $g=0$ and $\varepsilon=0$, it follows by continuity (and compactness) that they will be satisfied for all sufficiently small ε and g. Thus, a_0 is limited by:

$$1') \quad a\bar{\Omega}\subset\Omega$$

$$2') \quad (1-a(1+a)\bar{\Omega})\subset\Omega\backslash(-\infty,0]$$

for $0<a<a_0$. It is easy to check that these conditions hold, for example, for $a_0=1/2>\sqrt{2}-1$ and for Ω the open disk with radius 3/4 and center 1/2.

We can thus assume that estimates (7.5), (7.7), and (7.10) hold for $\varepsilon<a<1/2$, and that the unstable manifold has been extended to a curve of the form $a\to g_a^*$ satisfying (7.9) and defined on (ε, a_1) where a_1 may be taken to be independent of ε. Because of (7.10), a number of iterations of $\mathscr{T}$ which is bounded uniformly in ε for small ε suffices to extend the unstable manifold to a curve of the above form defined on $(\varepsilon, 1/2)$. The bounds (7.9) are no longer necessarily propagated by application of $\mathscr{T}$, but each such application worsens them by only a finite amount [because of (7.5), (7.7)]. In this way (and also extending similarly in the direction $a\to 0$) we get:

Proposition 7.7. *Let Ω be the open disk of radius 3/4 and center 1/2. There exist constants B_8, B_9, and $\varepsilon_0>0$, such that for $0<\varepsilon<\varepsilon_0$, the unstable manifold for $\mathscr{T}$ in $\mathscr{P}_\varepsilon$ contains a curve*

$$a\to\phi_a^*: \phi_a^*(x)=1-(1+a)|x|^{1+\varepsilon}-|x|^{1+\varepsilon}(1-|x|^{1+\varepsilon})g_a^*(|x|^{1+\varepsilon})$$

defined on $a\in[0,\tilde{a}]$ and satisfying:

$$\|g_a^*\|\leqq B_8\varepsilon; \qquad \left\|\frac{dg_a^*}{da}\right\|\leqq B_9\varepsilon; \qquad \phi_{\tilde{a}}^*(\tilde{a})=\tilde{a}.$$

Moreover if $A(a)=-(\mathscr{T}\phi_a^*)(1)$, then $\dfrac{dA}{da}>1.5$ on $[0,\tilde{a}]$.

Remarks. The estimates developed in this section show that any ψ near ϕ_ε and not on the stable manifold will be driven out of $\mathfrak{D}(\mathscr{T})$ by a finite number of iterations of $\mathscr{T}$. This justifies the restriction in Sect. 6 to a non-recurrent fixed point. It also permits us to clarify the uniqueness of the μ_j's. Consider the cylinder in $\mathscr{P}_\varepsilon$ corresponding to

$$0<a<1; \qquad \|g\|\leqq g_1.$$

For fixed small ε, and sufficiently small g_1, the stable manifold cuts across this cylinder and thus divides it into two parts. We will refer to the part on the side of $a=0$ as "above" the unstable manifold and the other part as "below" it (see Fig. 4). The surfaces Σ_j further divide the part of the cylinder above the stable manifold into slabs. If ψ lies between Σ_j and Σ_{j+1}, then $\psi_j\equiv\mathscr{T}^j\psi$ is defined and $(\psi_j)(1)>0$. Thus, ψ_j maps all of $[-1,1]$ into the interval $[\psi_j(1),1]$, which does not contain 0, and hence $\psi_j^p(0)\neq 0$ for $p=1,2,\ldots$. But ψ_j differs from ψ^{2^j} only by a scale factor, so $\psi^{p2^j}(0)\neq 0$ for all p. This implies that $\psi^p(0)\neq 0$, so ψ is *not* superstable. Thus: The

only superstable ψ's lying in the part of the cylinder above W_s are those on the surfaces $\Sigma_j, j=1,2,\ldots$. If ψ_μ is a parametrized curve crossing W_s from above when $\mu=\mu_\infty$, at a point inside the cylinder, with non-zero vertical velocity, then for sufficiently large j, the μ_j are uniquely determined by the conditions

$$\psi_{\mu_j} \text{ is superstable of period } 2^j; \quad \mu_j<\mu_\infty; \quad \mu_\infty-\mu_j \text{ is small.}$$

On the other hand, it is not hard to see that, for large j, there are very many values of μ, larger than but near to μ_∞, where ψ_μ is superstable with period 2^j.

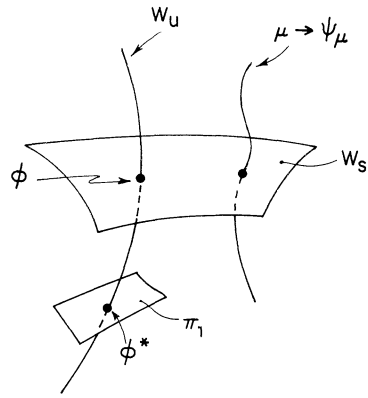

Fig. 4

In another direction: It is easy to verify (using the implicit function theorem) that, for small ε, g there is a uniquely determined $a=\hat{a}(g)$ such that ψ corresponding to (a,g) is superstable of period 3. Moreover $g\to\hat{a}(g)$ is smooth and defines a codimension-one surface crossing W_u transversally. Call this surface $\hat{\Sigma}_1$, and apply the theory of the preceding section to show that ots successive inverse images

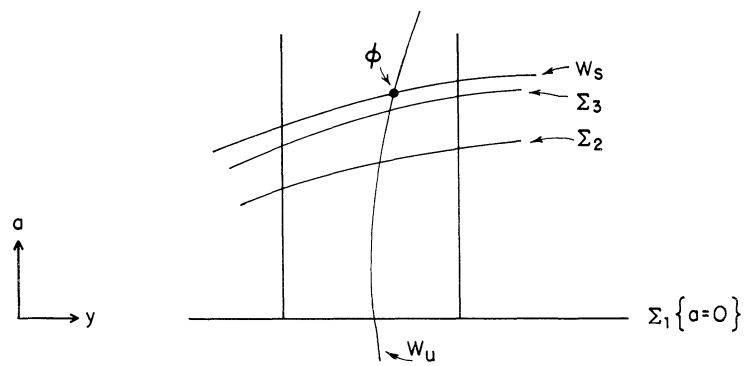

Fig. 5

$\hat{\Sigma}_2, \hat{\Sigma}_3, \ldots$ (under $\mathscr{T}$) converge to W_s exponentially with rate δ. If $\mu\to\psi_\mu$ is a parametrized curve as before, there exists a sequence $\hat{\mu}_j$, converging down to μ_∞, such that $\psi_{\hat{\mu}_j}\in\hat{\Sigma}_j$, i.e. $\psi_{\hat{\mu}_j}$ is superstable of period $3\cdot 2^{j-1}$. Moreover, just as before

$$(\hat{\mu}_j-\mu_\infty)\delta^j$$

converges to a finite non-zero limit. A warning, however: There is another sequence, say $\hat{\hat{\mu}}_j$, with $\psi_{\hat{\hat{\mu}}_j}$ superstable of period $3\cdot 2^{j-1}$ and $\hat{\mu}_{j-2}>\hat{\hat{\mu}}_j>\hat{\mu}_{j-1}$. In

fact, there are infinitely many more distinct, interleaved, sequences of periods $3 \cdot 2^{j-1}$.

Similarly, the equation

$$\psi(a) = a \quad [\text{i.e. } \psi^3(0) = \psi^4(0)]$$

defines a surface, say $\tilde{\Sigma}_1$, of codimension one crossing W_u transversally. A general result of Misiurewicz implies that any $\psi \in \tilde{\Sigma}_1$ sufficiently near to the unstable manifold admits an absolutely continuous invariant measure. It is easy to see, however, that if $\mathcal{T}\psi$ admits an absolutely continuous invariant measure, then so does ψ, so, applying the machinery described above, we see that, for each $\mu \to \psi_\mu$ as above, there is a sequence $\tilde{\mu}_j$ converging to μ_∞ such that each $\psi_{\tilde{\mu}_j}$ admits an absolutely continuous invariant measure and such that

$$\delta^j(\tilde{\mu}_j - \mu_\infty)$$

approaches a finite non-zero limit. Many other such examples could be considered, e.g., take for the initial surface the set of ψ's such that $\psi'(x_0) = -1$, where x_0 denotes the unique fixed point of ψ in $[0, 1]$. The μ_j's in this case will correspond to bifurcation points, where the orbit of period 2^{j-1} becomes unstable and the orbit of period 2^j appears.

8. Attracting Cantor Sets

Let $\psi \in \mathfrak{D}(\mathcal{T})$. As in Sect. 1, we write

$$a = -\psi(1); \qquad b = \psi(a);$$

and we will also write

$$c = \psi(b),$$

and we will assume $c \geqq 0$. [This would follow automatically if $\psi \in \mathfrak{D}(\mathcal{T}^2)$.] Since ψ maps $[-1, 1]$ into $[-a, 1]$, we may as well restrict its definition to the interval $[-a, 1]$. We have seen that ψ maps $[-a, a]$ onto $[b, 1]$ and it evidently maps $[b, 1]$ in a one-to-one fashion onto $[-a, c]$. Thus, the set

$$[-a, c] \cup [b, 1]$$

is mapped onto itself by ψ. Note that this invariant set is constructed out of the original interval $[-a, 1]$ by deleting an open subinterval (c, b) in the middle, i.e. as in the first step of constructing a Cantor set. Note also that $[-a, c]$ is mapped onto itself by $\psi \circ \psi$ and that $\mathcal{T}\psi$ is obtained from the restriction of $\psi \circ \psi$ to $[-a, c]$ by a linear change of variable $x \to -ax$. Observe, finally, also that if $K \subset [-a, c]$ is mapped onto itself by $\psi \circ \psi$, then

$$J = K \cup (\psi^{-1} K \cap [b, 1])$$

is mapped onto itself by $\mathcal{T}\psi$.

If $\mathcal{T}\psi$ is also in $\mathfrak{D}(\mathcal{T}^2)$, we can apply the same operation to $\mathcal{T}\psi$ and thus obtain an invariant set for ψ by deleting an open subinterval from the middle of each of $[-a, c]$ and $[b, 1]$. Continuing, if $\psi \in W_s$ and hence $\psi \in \mathfrak{D}(\mathcal{T}^j)$ for all j, we

can repeat this operation infinitely often and so obtain an invariant Cantor set for ψ. In this section, we will analyze the construction of this Cantor set in more detail. In particular, we determine the action of ψ on the Cantor set, show that orbits of ψ which converge to the Cantor set have simple statistical properties, and show that if ψ is near enough to the fixed point all but countably many orbits do indeed converge to the Cantor set.

We deal first with some combinatorial aspects of the construction of the Cantor set. For ψ as above we write

$$J_1^{(1)}=[b,1]\,; \qquad J_2^{(1)}=[-a,c]\,; \qquad J^{(1)}=J_1^{(1)}\cup J_2^{(1)}\,.$$

Observe that ψ maps $J_1^{(1)}$ onto $J_2^{(1)}$ and $J_2^{(1)}$ back onto $J_1^{(1)}$; also that the end-points of $J_1^{(1)}$ are $\psi(0)$ and $\psi^3(0)$ while those of $J_2^{(1)}$ are $\psi^2(0)$ and $\psi^4(0)$. The following proposition is proved in a straightforward way by induction, using the remarks already made.

Proposition 8.1. *Let $\psi\in\mathfrak{D}(\mathscr{T}^n)$ and assume $(\mathscr{T}^n\psi)(1)\leqq 0$. There then exists a decreasing sequence of closed sets*

$$J^{(1)}\supset J^{(2)}\supset J^{(3)}\supset \ldots \supset J^{(n)}$$

with the following properties:

1. *Each $J^{(i)}$ is mapped onto itself by ψ.*
2. *Each $J^{(i)}$ is a union of 2^i disjoint closed intervals which we can label $J_1^{(i)},\ldots,J_{2^i}^{(i)}$ in such a way that ψ maps $J_j^{(i)}$ onto $J_{j+1}^{(i)}$, where addition is understood modulo 2^i. The interval $J_{2^i}^{(i)}$ contains 0.*
3. *$J^{(i+1)}$ is constructed by removing an open subinterval from the middle of each of the intervals $J_j^{(i)}$. The resulting two intervals are labelled $J_j^{(i+1)}$ and $J_{j+2^i}^{(i+1)}$. The interval to be removed and the labelling are determined as follows: ψ^{2^i} restricted to $J_{2^i}^{(i)}$ differs from $\psi_i=\mathscr{T}^i\psi$ only by a scale factor. Remove from $J_{2^i}^{(i)}$ the interval corresponding, under this scaling, to $(c(\psi_i),b(\psi_i))$. Call the remaining subinterval which contains zero $J_{2^{i+1}}^{(i+1)}$ and the other $J_{2^i}^{(i+1)}$. (Note that ψ^{2^i} interchanges these intervals.) Then put $J_j^{(i+1)}=\psi^j[J_{2^{i+1}}^{(i+1)}]$, $j=1,2,\ldots,2^{i+1}-1$.*
4. *The end points of $J_j^{(i)}$ are $\psi^j(0)$ and $\psi^{j+2^i}(0)$.*

Corollary 8.2. *If $\psi\in W_s$ (so $\psi\in\mathfrak{D}(\mathscr{T}^n)$ for all n) then ψ admits an invariant Cantor set $J=\bigcap_i J^{(i)}$. This Cantor set is homeomorphic to $\{0,1\}^N$, with a correspondence such that $x\leftrightarrow(i_1,i_2,\ldots)\in J_j^{(i)}$ if and only if $j=i_1+2i_2+2^2i_3+\ldots+2^{j-1}i_j$.*

In this representation $\psi|_J$ takes the following form

$$\psi:(1,1,1,\ldots)\to(0,0,0,\ldots)$$

$$\psi:(1,1,\ldots,1,0,i_{n+1},\ldots)\to(0,0,0,\ldots,0,1,i_{n+1},\ldots)\,.$$

The orbit of each element of J is dense in J; ψ is invertible on J. Heuristically, we can think of the sequence $(i_1,i_2,\ldots)$ as the binary representation of a (usually infinite) integer

$$k=\sum_{\ell=1}^{\infty} i_\ell 2^{\ell-1}\,.$$

In terms of k, the action of ψ is simply $k\to k+1$.

Next we continue to assume that $\psi \in W_s$ and we look at the ergodic theory of the action of ψ on J. Let ν be the unique probability measure on J defined by

$$\nu(J_j^{(i)}) = 2^{-i} \quad \text{for all} \quad i, j$$

(i.e. ν assigns equal weight to each of the intervals making up $J^{(i)}$.) Since

$$\psi^{-1}(J_j^{(i)}) \cap J^{(i)} = J_{j-1}^{(i)},$$

the uniqueness of ν implies that it is invariant under the action of ψ; on the other hand, this same equation shows that any ψ-invariant probability measure assigning measure zero to the complement of J must assign equal measure to each $J_j^{(i)}$ (i fixed but arbitrary, $j = 0, 1, 2, \ldots, 2^i - 1$) and hence must coincide with ν.

Proposition 8.3. *ν is the only ψ-invariant probability measure on J. If $x \in [-1, 1]$ has the property that $\psi^n(x)$ approaches J as $n \to \infty$ and if f is any continuous function on $[-1, 1]$ then*

$$\lim_{N \to \infty} \frac{1}{N} \sum_{n=0}^{N-1} f \circ \psi^n(x) = \int f d\nu.$$

The abstract dynamical system (ν, ψ) is ergodic but not weak mixing.

Proof. The first statement has already been proved. To prove the second, suppose it is not true. By standard compactness arguments, there then exists a sequence N_j going to ∞ with j such that

$$\bar{f} = \lim_{j \to \infty} \frac{1}{N_j} \sum_{n=0}^{N_j - 1} f \circ \psi^n(x)$$

exists for all continuous functions f on $[-1, 1]$ but such that $\bar{f} \neq \int f d\nu$ for some f. But $f \to \bar{f}$ is a positive linear functional on the space of continuous functions, taking the value 1 on the constant function 1, and vanishing if $f = 0$ on J [since $\psi^n(x) \to J$ by assumption]. Thus, there is a probability measure $\bar{\nu}$ on J such that

$$\bar{f} = \int f d\bar{\nu}.$$

A standard argument shows that $\overline{f \circ \psi} = \bar{f}$, so $\bar{\nu}$ is ψ-invariant, so $\bar{\nu} = \nu$, contradicting the fact that $\bar{f} \neq \int f d\nu$ for some f.

The ergodicity of (ν, ψ) follows at once from the fact that ν is the unique ψ-invariant probability measure on J. On the other hand, the set

$$J \cap J_0^{(1)}$$

is invariant under ψ^2 and has measure 1/2, so ψ^2 is not ergodic, so ψ is not weak mixing. One can also show that the spectrum is discrete.

We next show that, if ψ is sufficiently near to the fixed point (and on W_s), then all but countably many orbits of ψ converge to J. In general, if $\psi \in \mathscr{P}_\varepsilon$, it has a unique fixed point in $[0, 1]$, which we will denote by x_0.

Lemma 8.4. *For ε sufficiently small and ψ sufficiently close to the fixed point ϕ (but ψ not necessarily on W_s), the orbit of every $x \neq \pm x_0$ is eventually in $J^{(1)}$.*

Proof. Note first that, since $\psi(x_0) = x_0$ and $\psi(-x_0) = \psi(x_0) = x_0$, the orbits of x_0 and $-x_0$ are extremely simple. Note also that, for any $x \in [-1, 1]$, $\psi^n(x) \in [-a, 1]$

for all $n \geqq 1$ and that

$$[-a, 1] = J^{(1)} \cup \{x_0\} \cup (c, x_0) \cup (x_0, b).$$

Now ψ maps (x_0, b) onto (c, x_0) and (c, x_0) onto an interval containing (x_0, b). Any orbit must therefore either

land on x_0 after no more than one step

or

land in $J^{(1)}$ after no more than two steps (and remain there, by invariance)

or

land in (c, x_0) after no more than two steps.

(These are not mutually exclusive.)

We have to show that, in the third case, $\psi^n(x)$ is eventually in $J^{(1)}$. Now $\psi \circ \psi$ leaves x_0 fixed and maps (c, x_0) onto a larger interval contained in $(-a_0, x_0)$. The idea is that $\psi \circ \psi$ pushes points further and further from x_0 and we want to show that every orbit for $\psi \circ \psi$ starting in (c, x_0) eventually reaches $[-a_0, c]$. If it reaches $[c, a]$, then one more iteration of $\psi \circ \psi$ will put it in $[-a, c]$, so what have to show is that it is impossible that

$$(\psi \circ \psi)^n(x) \in (a, x_0) \quad \text{for all } n.$$

Since $\psi \circ \psi(x_0) = x_0$, it will suffice to prove

$$(\psi \circ \psi)'(x) > 1 \quad \text{on} \quad [a, x_0].$$

The proof of this last statement is straightforward, using the fact that since

$$\phi(x) = 1 - (1 + \lambda_\varepsilon)|x|^{1+\varepsilon} + O(\lambda_\varepsilon)$$

we can, by making ψ sufficiently close to ϕ, arrange that (for example)

$$\psi'(x) \leqq -(1 + \tfrac{2}{3}\lambda_\varepsilon)(1 + \varepsilon)|x|^\varepsilon \quad \text{on} \quad [0, 1]$$

and also that

$$a \geqq \lambda_\varepsilon / 2; \qquad x_0 \geqq 1/3.$$

We will now iterate this arrangement to prove:

Proposition 8.4. *If $\psi \in W_s$ is near enough to ϕ, then*

1) *ψ has exactly one periodic orbit of each period $1, 2, 4, 8, \ldots$, and no periodic orbits of other periods. All these periodic orbits are repelling.*

2) *Every orbit of ψ which does not eventually fall exactly on one of the repelling periodic orbits converges to the invariant Cantor set J.*

Proof. The preceding lemma tells us that x_0 is a repelling fixed point for ψ and that every orbit which does not eventually fall exactly on x_0 is eventually in $J^{(1)}$. $J^{(1)}$ consists of the two pieces $J_0^{(1)}$ and $J_1^{(1)}$ which are exchanged by ψ, so to analyze orbits which are eventually in $J^{(1)}$ it suffices to analyze orbits of $\psi \circ \psi$ in $J_0^{(1)}$, i.e., of ψ in $[-1, 1]$. Since ψ is again in W_s and near ϕ, we can apply Lemma 3 to it. Thus, $\psi \circ \psi$ has a repelling fixed point in $J_0^{(1)}$, which corresponds to a repelling orbit of period 2 for ψ, and every orbit for $\psi \circ \psi$ in $J_0^{(1)}$ which does not eventually land on

the fixed point is eventually in $J_0^{(2)} \cup J_2^{(2)}$. Expressed in terms of ψ, every orbit which does not land eventually on either the fixed point x_0 or the orbit of period 2 just described is eventually in $J^{(2)}$.

Continuing in this way we find, for each n, a repelling periodic orbit of period 2^{n-1} and show that every orbit which does not fall exactly on one of the constructed orbits of periods $1, 2, 4, \ldots, 2^{n-1}$ is eventually in $J^{(n)}$.

This proves everything in the proposition except for the non-existence of periodic orbits other than the ones enumerated. From 2) any such orbit, if one exists, must be in J. But for any n, J can be broken into 2^n disjoint pieces

$$J \cap J_j^{(n)}, \quad j = 0, 1, 2, \ldots, 2^n - 1,$$

which are permuted cyclically by ψ. It follows that periodic orbits in J would have to have a period which is divisible by 2^n for all n. This is impossible, so there are no further periodic orbits.

Remarks. 1. We note that λ_ε appears as an asymptotic scaling parameter for the Cantor set J. Specifically, $J_0^{(n)} \cap J$ and $J_0^{(n+1)} \cap J$ asymptotically differ by a scale factor of λ_ε. This means the following: If we write

$$A_n = a(\psi) \ldots a(\mathscr{T}\psi) \ldots a(\mathscr{T}^{n-1}\psi)$$

then

$$J_0^{(n)} \cap J = A_n J(\mathscr{T}^n \psi).$$

As $n \to \infty$, $\mathscr{T}^n \psi \to \phi$, so, in a sense which is easy to make precise, $J(\mathscr{T}^n \psi) \to J(\phi)$. Thus,

$$A_n^{-1}(J_0^{(n)} \cap J) \quad \text{and} \quad A_{n+1}^{-1}(J_0^{(n+1)} \cap J)$$

look essentially the same for large n. In other words, $J_0^{(n+1)} \cap J$ looks almost the same as $J_0^{(n)} \cap J$ multiplied by a scale factor of $A_{n+1}/A_n = a(\mathscr{T}^n \psi)$. Again, since $\mathscr{T}^n \psi \to \phi$, this scale factor converges to $a(\phi) = \lambda_\varepsilon$. Observe, however, that this scaling is different for other pieces of the Cantor set. For example, successive terms in the decreasing sequence $J \cap J_1^{(1)} \supset J \cap J_1^{(2)} \supset J \cap J_1^{(3)} \supset \ldots$ differ asymptotically by a numerical factor of $\lambda_\varepsilon^{1+\varepsilon}$ rather than λ_ε, and the same is true for any of the sequences $J \cap J_j^{(n)}$ for fixed, non-zero j.

2. It is easy to see that, for ψ near ϕ, $J_0^{(2)}$ is longer than $J_1^{(2)}$. More generally for any n, $J_0^{(n)}$ is the longest of the 2^n intervals making up $J^{(n)}$ and $J_1^{(n)}$ is the shortest. The length of $J_0^{(n)}$ is $A_n(1 + a(\mathscr{T}^n \psi))$ which behaves asymptotically like const $\times \lambda_\varepsilon^n$. The conclusion that the longest interval in $J_0^{(n)}$ has length bounded by const $\times \lambda_\varepsilon^n$ holds even for all ψ on W_s since $\mathscr{T}^n \psi$ still converges to ϕ.

3. Proposition 4 remains true even for ψ not near ϕ provided that ψ is near the unstable manifold W_u, that $\psi \in \mathfrak{D}(\mathscr{T})$, and that $\mathscr{T}\psi(1) \leqq 0$. This leads to the following picture of the "bifurcation" which occurs on the stable manifold:

If $\psi \in \mathfrak{D}(\mathscr{T}^n)$, and if $(\mathscr{T}^n \psi)(1) \leqq 0$, then ψ admits a finite decreasing chain

$$J^{(1)} \supset J^{(2)} \supset \ldots \supset J^{(n)}$$

of invariant sets; $J^{(n)}$ is a sort of approximate Cantor set; it is a union of 2^n disjoint closed intervals permuted cyclically by ψ. If in addition ψ is not too far from ϕ, then the space between successive pairs of these intervals contains exactly one

periodic point of ψ. These periodic points have periods $1, 2, 4, \ldots, 2^{n-1}$; there is exactly one cycle of each of these periods, and they are all repelling. There are countably many orbits which fall onto one of these repelling orbits after finitely many steps; all others converge to $J^{(n)}$. If we collapse each of the intervals making up $J^{(n)}$ to a point, all such ψ's look the same – they have an attracting periodic orbit of period 2^n together with the simplest set of repelling periodic orbits between them required by simple considerations of connectedness. Each such ψ can thus be thought of as a sort of semi-direct product of the simplest possible ψ which is superstable of period 2^n with the transformation $\mathscr{T}^n\psi$ scaled down and made to act on $J_0^{(n)}$. These $\mathscr{T}^n\psi$'s can of course be very different – e.g., may on the one hand be superstable of period 2 or on the other hand admit an absolutely continuous invariant measure – but the differences act on a small spatial scale and will therefore not be very noticeable for large n. In the limit $n \to \infty$ the approximate Cantor set becomes a true Cantor set which remains attracting and which can crudely be thought of as a single attracting periodic orbit of period 2^∞; at the same time, the spatial scale of the difference between ψ's goes to zero and so the difference disappears entirely.

Even if ψ is not near enough to W_u for Lemma 4 to apply, it will still be true that $\mathscr{T}^n\psi$ is near enough for n larger than some n_0. Thus, although we cannot be sure that the gaps between the intervals in $J^{(n_0)}$ are free of extraneous recurrent behavior for ψ, each gap produced in passing from $J^{(n_0)}$ to $J^{(n)}$, $n > n_0$ will indeed contain exactly one repelling periodic point. Furthermore, any extraneous recurrent behavior has to be fairly tame. It is not hard to see that any orbit which never enters $J^{(n_0)}$ must be asymptotically periodic with period $1, 2, 4, \ldots$, or 2^{n_0}.

Note. a) In Lemma 8.4 and Proposition 8.5 the condition that ψ be near to ϕ can be replaced by the condition that ψ have negative Schwarzian derivative.

b) The computational proof of Lemma 8.4 can be replaced by a simple conceptual proof using the negativity of the Schwarzian derivative.

Acknowledgements. This work was made possible through the hospitality of the following institutions. EPF – Lausanne, IHES – Bures sur Yvette, Rockefeller University, Harvard University, ETH – Zurich, University of Geneva, and further financial support from 3e Cycle de Physique de la Suisse Romande (O.L.), Grant NSF-MCS78-06718 (O.L.), Grant NSF-PHY-77-18762 (P.C.), and Fonds National Suisse (P.C.).

References

1. Collet, P., Eckmann, J.-P.: A renormalization group analysis of the hierarchical model in statistical physics. Lecture Notes in Physics, Vol. 74. Berlin, Heidelberg, New York: Springer 1978
2. Dieudonné, J.: Foundations of modern analysis. New York, London: Academic Press 1969
3. Feigenbaum, M.: Quantitative universality for a class of nonlinear transformations. J. Stat. Phys. **19**, 25 (1978); **21**, 669 (1979)
4. Guckenheimer, J.: Bifurcations of dynamical systems. C.I.M.E. Lectures 1978
5. Hartmann, P.: Ordinary differential equations. New York, London: Wiley 1964
6. Hirsch, M.W., Pugh, C.C., Shub, M.: Invariant manifolds. Lecture Notes in Mathematics, Vol. 583. Berlin, Heidelberg, New York: Springer 1977
7. Kato, T.: Perturbation theory for linear operators. Berlin, Heidelberg, New York: Springer 1966
8. Lanford III, O.E.: To appear, and Lecture Notes in Physics, Vol. 116, Berlin, Heidelberg, New York: Springer 1980

9. May, R.M.: Simple mathematical models with very complicated dynamics. Nature **261**, 259–467 (1976)
10. Misiurewicz, M.: Structure of mappings of the interval with zero entropy. Preprint I.H.E.S. (1978)
11. Misiurewicz, M.: Absolutely continuous measures for certain maps on an interval. Preprint I.H.E.S./M/79/293 (1979)
12. Singer, D.: Stable orbits and bifurcations of maps of the interval. S.I.A.M. J. Appl. Math. **35**, 260–267 (1978)
13. Stefan, P.: A theorem of Sharkovskii on the existence of periodic orbits of continuous endomorphisms of the real line. Commun. Math. Phys. **54**, 237–248 (1977)
14. Collet, P., Eckmann, J.-P.: Iterated maps on the interval as dynamical systems. Progress in Physics. Birkhäuser Boston 1980 (to appear)

Communicated by D. Ruelle

Received February 28, 1980

Note added in proof. An alternate proof of existence of a fixed point for $\varepsilon = 1$ has been provided by M. Campanino, H. Epstein, D. Ruelle, (to appear). For the extension of the results to multidimensional dissipative maps, see P. Collet, J.-P. Eckmann, H. Koch, J. Stat. Phys. (to appear)