

ON ITERATED MAPS OF THE INTERVAL

by

John Milnor
 Institute for Advanced Study
 Princeton, NJ 08540 USA

and

William Thurston
 Princeton University
 Princeton, NJ 08544 USA

Dedicated to the Memory of Rufus Bowen and Peter Štefan.

Introduction.

Mappings from an interval to itself provide the simplest possible examples of smooth dynamical systems. Such mappings have been widely studied in recent years since they occur in quite varied applications, and since their surprisingly complex behavior may provide a useful introduction to the study of higher dimensional situations.

The present paper sets up an effective calculus for describing the qualitative behavior of the successive iterates of a piecewise monotone mapping. Let I be a closed interval of real numbers. By definition, a continuous mapping f from I to itself is piecewise monotone if I can be subdivided into finitely many subintervals I_1, \dots, I_ℓ on which f is alternately strictly increasing or strictly decreasing. Each such maximal interval on which f is monotone is called a lap of f , and $\ell = \ell(f)$ is the lap number. The separating points $c_1, \dots, c_{\ell-1}$ at which f has a local minimum or maximum are called the turning points of f .

Section 1 studies the sequence $\ell(f), \ell(f^2), \ell(f^3), \dots$ of positive integers, where $f^k = f \circ \dots \circ f$ stands for the k -fold composition of the function f with itself. The limit

$$s = \lim_{k \rightarrow \infty} \ell(f^k)^{1/k} = \inf_k \ell(f^k)^{1/k}$$

is a real number in the interval $[1, \ell(f)]$ called the growth number

of f . (By a theorem of J. Rothschild and of Misiurewicz and Szlenk, its logarithm is precisely the "topological entropy" of f .) The special case of a quadratic map f is discussed in some detail.

Sections 2, 3 introduce an invariantly defined "formal coordinate function" $\theta(x)$ for $x \in I$. This is a formal power series of the form $\sum \theta_k(x)t^k$, which depends monotonely on x , under a suitable ordering. If $f^k(x)$ belongs to the interior of the lap I_j , then the coefficient $\theta_k(x)$ is a corresponding formal symbol I_j multiplied by a sign which is either ± 1 or 0 according as the function f^k is increasing, decreasing, or has a turning point at x .

Section 4 introduces a basic invariant called the kneading matrix of f . This is an $(\ell+1) \times \ell$ matrix with entries in the ring $\mathbb{Z}[[t]]$ of integer formal power series. If $k \geq 1$, the coefficient of t^k in its (i,j) -th entry is ± 2 if f^k maps a neighborhood of c_1 into the lap I_j , and is zero otherwise. Closely related is the kneading determinant, a formal power series $D(t) = 1 + D_1 t + D_2 t^2 + \dots$ with odd integer coefficients. In the simplest case, if f has just one turning point c_1 , which is say a minimum point, each coefficient D_k is either $+1$ or -1 according as f^{k+1} takes on a local minimum or a local maximum at c_1 .

Section 5 gives an explicit method for computing the sequence of lap numbers $\ell(f^k)$ in terms of the kneading matrix. As an example, if there is only one turning point, and if f maps both endpoints of the interval I to a single endpoint, then the formal power series $L(t) = \sum_{k=1}^{\infty} \ell(f^k)t^{k-1}$ is equal to $(1-t)^{-1} + D(t)^{-1}(1-t)^{-2}$.

Here is an explicit example. For the mapping $f(x) = (x^2-5)/2$ of Figure 1.4, it is easy to check that the iterates $f^k(0)$ belong alternately to the lap I_1 or I_2 . According to §4.5, this information determines the kneading determinant:

$$D(t) = 1 - t - t^2 + t^3 + t^4 - \dots = (1-t)/(1+t^2).$$

We can now compute the series $L(t) = 2(1-t+t^2)/(1-t)^3$, and it follows easily that $\ell(f^k) = k^2 - k + 2$ for $k \geq 1$, and that $s = 1$.

Section 6 studies convergence properties of these formal power series, showing for example that $D(t)$ is holomorphic in the open unit disk $|t| < 1$, and has smallest zero at $t = 1/s$ if $s > 1$, while $L(t)$ is meromorphic in the open unit disk with a pole at $1/s$.

Section 7, under the hypothesis $s > 1$, shows that f is "topologically semi-conjugate" to a piecewise linear map having slope $\pm s$ everywhere (see Figure 7.11). As a typical application, in the case $\ell = 2$ it is shown that f admits a periodic point of odd period $p \geq 3$ if and only if $s > \sqrt{2}$ (compare Štefan).

Section 8 surveys a number of known methods and theorems concerning periodic points, due to Julia and Fatou, Sarkovskii, Artin and Mazur, Bowen and Franks, Allwright, and D. Singer. In particular, the Artin-Mazur power series $\zeta(t)$ is a convenient way of enumerating the periodic points of f . Sections 9 and 10 state several versions of our main theorem which computes this Artin-Mazur zeta function in terms of the kneading determinant. As an application of this theorem, using Bowen and Franks, it is shown that $s > 1$ if and only if there exists a periodic point whose period is not a power of two. Furthermore, there are only finitely many distinct periods (all necessarily powers of two) if and only if the sequence of lap numbers $\ell(f^k)$ is bounded by a polynomial function of k .

Section 11 completes the proof of this main theorem. A basic tool is the statement that, under C^1 -smooth deformation of f , as the difference $f^p(c_i) - c_i$ changes sign the kneading determinant of f is changed by a factor of $(1-t^p)/(1+t^p)$.

Section 12 characterizes those power series $D(t)$ which can actually occur as kneading determinant for some map f with lap number $\ell(f) = 2$. It is shown that every such admissible kneading

determinant already occurs for some quadratic mapping, for example of the form $f(x) = bx(1-x)$. This section also discussed continuity properties of the growth number $s = s(f)$ under smooth deformation of f . (Compare Figure 13.3.) Section 13 proves a monotonicity theorem for quadratic maps. The last section constructs a number of illustrative examples.

An earlier version of this paper was widely circulated in 1977. The present version incorporates both Part I and Part II of the earlier one, as well as additional material. Most of it was written in 1981, but §13 was added in 1983 and a few minor corrections are more recent.

Contents

1. Laps and the growth number	469
2. The itinerary of a point	474
3. The invariant coordinate $\theta(x)$	477
4. The kneading matrix $[N_{ij}]$	482
5. Counting pre-turning points and laps	488
6. Convergent power series	493
7. Piecewise linear models	497
8. Periodic points: a review of the classics	507
9. Computation of zeta functions	515
10. Periodic points of negative type	521
11. Smooth deformations of f	529
12. An intermediate value theorem	539
13. More on quadratic maps	545
14. Constructing new functions out of old	553
References	560

1. Laps and the growth number

We are concerned with properties of a mapping which are invariant under topological conjugacy. By definition, two maps $f : I \rightarrow I$ and $g : J \rightarrow J$ are topologically conjugate if there exists a homeomorphism h from I onto J so that $g = h \circ f \circ h^{-1}$. (If h preserves orientation, then one speaks of an orientation preserving topological conjugacy.)

A mapping f from the closed interval I to itself is called piecewise-monotone if I can be subdivided into finitely many sub-intervals

$$I_1 = [c_0, c_1], I_2 = [c_1, c_2], \dots, I_\ell = [c_{\ell-1}, c_\ell]$$

in such a way that the restriction of f to each interval I_j is strictly monotone, either increasing or decreasing. Here it is to be understood that $c_0 < c_1 < \dots < c_\ell$ where I is the interval $[c_0, c_\ell]$. We will always assume that each I_j is a maximal interval on which f is strictly monotone. Such a maximal interval is called a lap of the function f , and the number $\ell = \ell(f)$ of distinct laps is called the lap number of f .

It follows that each of the points c_0, c_1, \dots, c_ℓ is either a local minimum or local maximum point of f . The interior local minimum and maximum points $c_1, \dots, c_{\ell-1}$ will be called the turning points of f , and will play an important role in what follows.

Together with f , we will also study the sequence

$$f^* = (f^0, f^1, f^2, \dots)$$

of successive iterates, where

$$f^0 = \text{identity map}, f^1 = f, f^2 = f \circ f,$$

and so on. Here we use the symbol $*$ to indicate a non-negative integer which indexes some sequence.

Caution. The reader must take care since our notation for iterates could be confusing: Superscripts on a function will always stand for iterates, but superscripts on a variable have the customary meaning as exponents.

A crude but powerful invariant of the topological conjugacy class of f is provided by the sequence

$$\ell(f^*) = (1, \ell(f), \ell(f^2), \dots)$$

of lap numbers associated with the successive iterates of f . Note that the number of laps of a composition of two functions satisfies the inequality

$$(1.1) \quad \ell(f \circ g) \leq \ell(f)\ell(g),$$

since $f \circ g$ can have at most $\ell(f)$ laps within each lap of g . Hence $\ell(f^n) \leq \ell(f)^n$. Using these inequalities, a standard argument proves the following.

Lemma 1.2. The n -th root $\ell(f^n)^{1/n}$ tends to a well-defined limit s as $n \rightarrow \infty$. Furthermore, this limit is precisely equal to the infimum of the numbers $\ell(f^n)^{1/n}$.

It follows that $\ell(f^n) \geq s^n$ for every n . In particular, s always belongs to the interval $[1, \ell(f)]$.

Definition. This limit s will be called the growth number of f .

It can be shown that the logarithm of s is equal to the invariant known as the topological entropy of f . Compare Rothschild, as well as Misiurewicz and Szlenk.

Proof of 1.2. Fixing some positive integer k , any n can be written as $n = kp + q$ with $0 \leq q < k$. Then

$$\ell(f^n) \leq \ell(f^k)^p \ell(f)^q$$

by repeated application of 1.1; hence

$$\ell(f^n)^{1/n} \leq \ell(f^k)^{p/n} \ell(f)^{q/n}.$$

As $n \rightarrow \infty$, the right hand side tends to the limit $\ell(f^k)^{1/k}$; hence

$$\limsup \{ \ell(f^n)^{1/n} \} \leq \ell(f^k)^{1/k}$$

for every k . Therefore

$$\limsup \{ \ell(f^n)^{1/n} \} \leq \inf \{ \ell(f^k)^{1/k} \} \leq \liminf \{ \ell(f^k)^{1/k} \},$$

which evidently completes the proof. ■

In order to illustrate these concepts, let us look at a family of explicit mappings.

Example 1.3. Given any constant $1 \leq b \leq 4$, consider the quadratic mapping

$$g(y) = by(1-y)$$

from the unit interval $[0,1]$ to itself, with a single turning point at $c_1 = \frac{1}{2}$. In spite of its innocuous appearance, this quadratic example can have a surprising rich structure, and has been studied by many authors, including biologists, physicists, meteorologists and mathematicians. See for example May, Collet and Eckmann, Lorenz, and Smale and Williams.

It is often convenient to make a canonical linear change of variable by setting x equal to the derivative $g'(y) = b(1-2y)$. In this way, we see that g is topologically conjugate to a quadratic mapping of the form

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

from the interval $-b \leq x \leq b$ to itself. Note that the derivative $f'(x)$ is identically equal to x . Here the constant

$$a = b^2 - 2b$$

is uniquely determined by b and ranges over the interval $-1 \leq a \leq 8$. Conversely, the constant $b = 1 + \sqrt{1+a}$ is uniquely and monotonely determined by a if we require that $1 \leq b \leq 4$. The mapping f is

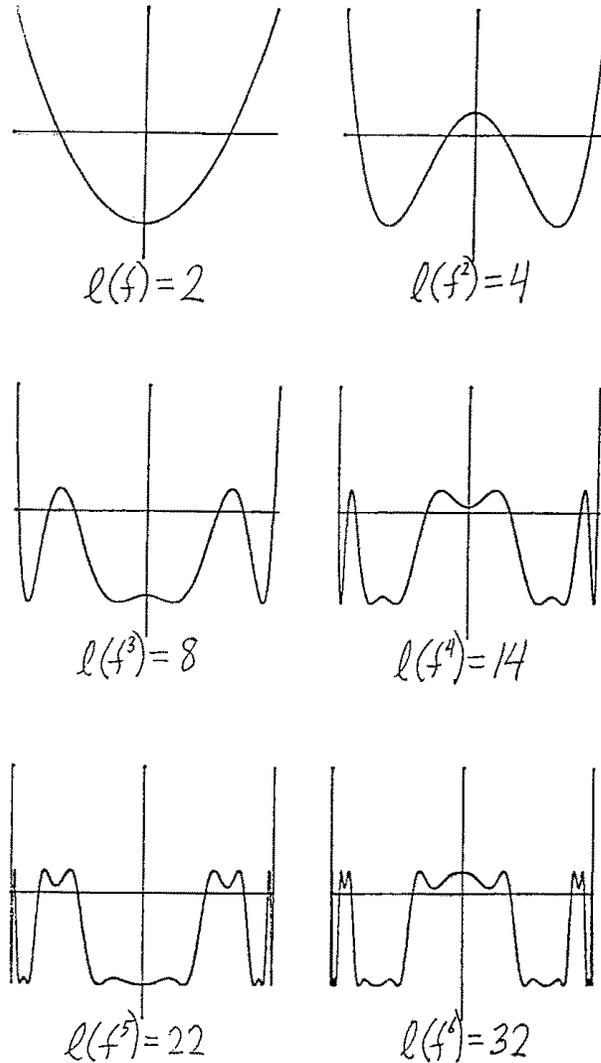


Figure 1.4.

Graphs of iterates of the function $f(y) = (y^2 - 5)/2$.

monotone decreasing on its first lap $[-b, 0]$ and monotone increasing on its second lap $[0, b]$. The special case $a = 5$ (or $b = 1 + \sqrt{6}$) is illustrated in Figure 1.4.

We will see in §13 that the behavior of this quadratic function f becomes more and more complicated as the parameter a increases

(or as $b = 1 + \sqrt{1+a}$ increases). Thus, for small values of a , the the lap number $\ell(f^n)$ has the order of magnitude of a polynomial function of n ; and it follows that the growth number s is equal to 1. The degree of this polynomial tends to infinity as a increases to the value 5.6046... . Here is a table giving more precise information. For later convenience, we have also listed the "kneading determinant" of f (compare §§4, 5, 12, 13).

Table 1.5. Invariants of $f(x) = (x^2 - a)/2$ for $-1 \leq a \leq 5.6046\dots$

parameter a	lap number $\ell(f^n)$ for $n > 0$	kneading determinant $D(t)$
$-1 \leq a \leq 0$	2	$1/(1-t) = 1+t+t^2+\dots$
$0 < a \leq 4$	$2n$	$1/(1+t) = 1-t+t^2-\dots$
$4 < a \leq 5.2428\dots$	$n^2 - n + 2$	$(1-t)/(1+t^2) = 1-t-t^2+t^4-\dots$
$5.2428\dots < a \leq 5.5261\dots$	$(2n^3 - 3n^2 + 22n + (0 \text{ or } 3))/12$	$(1-t)(1-t^2)/(1+t^4)$
...	polynomial growth of degree $d + 1$	$(1-t)\dots(1-t^{2^{d-1}})/(1+t^{2^d})$
$a = 5.6046\dots$	faster than polynomial but less than exponential growth	$(1-t)(1-t^2)(1-t^4)(1-t^8)\dots$

Closely related is the periodicity behavior which is studied beginning in §8. For $-1 \leq a \leq 3$ the fixed point $x = 1 - \sqrt{1+a}$ is stable, and f has no periodic points with period greater than 1. But as a increases past 3 this stable fixed point "bifurcates"

into a stable orbit $(-1 \pm \sqrt{a-3})$ of period 2. At $a = 5$ there is a further bifurcation to a stable orbit of period 4, at $a = 5.4723\dots$ to a stable orbit of period 8, and so on. The stable periodic orbit bifurcates infinitely often as a increases towards the limit $a = 5.6046\dots$ which was mentioned above (compare Feigenbaum, Lanford).

For $a > 5.6046\dots$ it seems empirically that $s > 1$, so that the sequence $\ell(f^n)$ grows exponentially with n (compare §13). As a increases, the sequence $\ell(f^n)$ changes uncountably often, and the entire behavior of the iterates of f becomes much more complicated. For the largest parameter value $a = 8$, the map f carries each of its two laps onto the entire interval I , and it follows easily that $\ell(f^n) = 2^n$, hence $s = 2$.

More generally, consider any function f which maps each of its laps I_1, \dots, I_ℓ homeomorphically onto the entire interval I . Then an easy induction on n shows that $\ell(f^n) = \ell^n$, hence $s = \ell = \ell(f)$. The case $\ell = 3$ is illustrated in Figure 1.6.

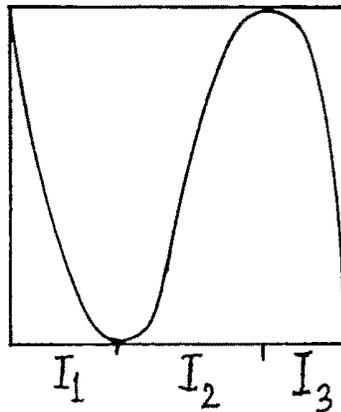


Figure 1.6. Graph of a function f with $\ell(f^n) = 3^n$.

2. The itinerary of a point

Let I_1, \dots, I_ℓ be the laps associated with the mapping f , numbered in their natural order, and let $c_1 < c_2 < \dots < c_{\ell-1}$ be the

turning points of f .

Definition. By the address $A(x)$ of a point $x \in I$ will be meant the formal symbol I_j if x belongs to the lap I_j and is not a turning point, or the formal symbol C_j if x is precisely equal to the turning point c_j . By the itinerary $A(f^*(x))$ will be meant the sequence of addresses

$$(A(x), A(f(x)), A(f^2(x)), \dots)$$

of the successive images of x . (As usual, the symbol $*$ stands for a variable integer $n \geq 0$.)

Here is a preliminary result which shows that the itinerary $A(f^*(x))$ contains significant qualitative information about the sequence of successive images $f^*(x)$. We will say that the itinerary $A(f^*(x))$ is eventually periodic if there exists an integer $p \geq 1$ so that the address $A(f^n(x))$ is equal to $A(f^{n+p}(x))$ for all large n . The smallest such p will be called the eventual period.

Lemma 2.1. The itinerary $A(f^*(x))$ is eventually periodic if and only if the sequence of points $f^*(x)$ converges towards a periodic orbit of f .

(See §8 for definitions.) If $A(f^*(x))$ has eventual period p , then the proof will show that the limit orbit has period either p or $2p$. Both cases can occur.

Here is an explicit numerical example to illustrate this lemma (compare 1.3).

Example 2.2. Suppose that $f(x) = (x^2 - 7)/2$. Then $c_1 = 0$ is the only turning point, and a somewhat delicate computation shows that the itinerary

$$A(f^*(0)) = (C_1, I_1, I_2, I_1, I_1, I_2, I_1, \dots)$$

is periodic of period 3, except for its initial entry. The successive

images $f^n(0)$ converge towards the limits

$$-.1099\dots \quad \text{if } n \equiv 0 \pmod{3}$$

$$-3.4939\dots \quad \text{if } n \equiv 1 \pmod{3}$$

$$2.6038\dots \quad \text{if } n \equiv 2 \pmod{3}$$

as $n \rightarrow \infty$. These three limit points, the roots of the equation $x^3 + x^2 - 9x - 1 = 0$, constitute a (one-sided stable) periodic orbit of period 3.

On the other hand, if $f(x) = (x^2 - 7.08)/2$, then the itinerary $A(f^*(0))$ again has period 3, except for its initial entry; but in this case the sequence $f^*(0)$ tends towards a periodic orbit of period 6.

Proof of 2.1. Clearly it suffices to consider the special case where the itinerary $A(f^*(x))$ is actually periodic of period p . Define intervals J_0, \dots, J_{p-1} as follows. Let J_r be the smallest closed interval containing all of the points $f^n(x)$ with $n \equiv r \pmod{p}$. Then J_r is contained in the single lap $A(f^r(x))$, so it follows that f maps J_r homeomorphically into J_{r+1} . It follows also that f^p maps the interval J_0 homeomorphically into itself. Excluding the trivial case where x itself is a periodic point, we may assume that J_0 is a nondegenerate interval. We now distinguish two cases.

Case 1. Suppose that the map f^p from J_0 into itself preserves orientation, i.e., is monotone increasing. If $x \leq f^p(x)$, then it follows inductively that

$$x \leq f^p(x) \leq f^{2p}(x) \leq \dots$$

Therefore, this subsequence converges to a limit in J_0 . Similarly, if $x > f^p(x)$, then the sequence

$$x > f^p(x) > f^{2p}(x) > \dots$$

converges to a limit in J_0 . In both cases, this limit point evident-

ly belongs to a periodic orbit whose period q divides p .

Case 2. Suppose that f^p reverses orientation. Then f^{2p} preserves orientation, and the argument proceeds as before. In this case, the limit orbit has period q dividing $2p$.

Conversely, if the sequence $f^*(x)$ converges to a periodic orbit of period q , then it is quite easy to check that the itinerary $A(f^*(x))$ is eventually periodic with eventual period p dividing q . The argument requires mild care only in the special case where the limit orbit contains a turning point of f .

Since p divides q and q divides $2p$, it follows that q is equal to either p or $2p$. ■

3. The invariant coordinate $\theta(x)$

In this section we will add just a little bit of information to the itinerary $A(f^*(x))$ in order to obtain a more useful sequence

$$\theta_* = (\theta_0, \theta_1, \theta_2, \dots)$$

which will be thought of as invariantly defined coordinate for the point x .

First some notations. Let V be the ℓ -dimensional vector space over the rational numbers \mathbb{Q} which has as basis the formal symbols I_1, \dots, I_ℓ . (Thus we are introducing a hopefully benign double meaning for I_j .) Our coordinate $\theta_* = \theta_*(x)$ will be a sequence of vectors in this vector space V .

For each symbol I_j , define the sign $\varepsilon(I_j)$ to be either $+1$ or -1 according as f is monotone increasing or monotone decreasing on the lap I_j . These signs alternate, so that we have either $(-1)^j \varepsilon(I_j) = +1$ for all j , or else $(-1)^j \varepsilon(I_j) = -1$ for all j . To each turning point c_j , lying between the laps I_j and I_{j+1} , we

assign the vector $C_j = (I_j + I_{j+1})/2$, and the sign $\varepsilon(C_j) = 0$.

Now to each itinerary $A(f^*(x)) = (A_0, A_1, A_2, \dots)$ we assign the sequence $(\theta_0, \theta_1, \theta_2, \dots)$ defined by

$$\theta_n = \varepsilon_0 \varepsilon_1 \dots \varepsilon_{n-1} A_n,$$

where ε_k is an abbreviation for $\varepsilon(A_k) = \varepsilon(A(f^k(x)))$. Thus

$$\theta_0 = A_0, \theta_1 = \varepsilon_0 A_1, \theta_2 = \varepsilon_0 \varepsilon_1 A_2, \dots$$

Remark. The sign $\varepsilon_0 \varepsilon_1 \dots \varepsilon_{n-1}$ of θ_n can be interpreted geometrically as follows. Note that $\varepsilon(A(x))$ is equal to the local degree of the mapping f at x , being ± 1 or 0 according as f is increasing, decreasing, or has a turning point at x . It follows inductively that the product $\varepsilon_0 \dots \varepsilon_{n-1} = \varepsilon(A(x)) \dots \varepsilon(A(f^{n-1}(x)))$ is equal to the local degree of the mapping f^n at x .

Now choose some translation invariant linear ordering of the vector space V (for example, by embedding V \mathbb{Q} -linearly into the real numbers) so that the basis vectors satisfy

$$I_1 < \dots < I_\ell.$$

Order the sequence θ_* lexicographically. Thus $\theta_* < \theta'_*$ if and only if

$$\theta_0 = \theta'_0, \dots, \theta_{n-1} = \theta'_{n-1}, \theta_n < \theta'_n$$

for some integer $n \geq 0$. These definitions have been concocted so that the following will be true.

Lemma 3.1. If $x < y$, then $\theta_*(x) \leq \theta_*(y)$.

In other words, the function $x \mapsto \theta_*(x)$, from the interval I to certain sequences of vectors, is monotone increasing.

Proof. Suppose, to fix our ideas, that $\theta_k(x) = \theta_k(y)$ for $k < n$. Then we must show that $\theta_n(x) \leq \theta_n(y)$.

Note first that the two points $f^k(x)$ and $f^k(y)$ have the same address, say A_k , for all $k < n$. Let $\varepsilon_k = \varepsilon(A_k)$ be the associated sign. Let J be the closed interval $[x, y]$. Since $J \subset A_0$, it follows that f maps J homeomorphically with degree $\varepsilon_0 \neq 0$ onto an interval $f(J) \subset A_1$. Similarly, f maps $f(J)$ homeomorphically with degree $\varepsilon_1 \neq 0$ onto an interval $f^2(J) \subset A_2$, and so on. It follows inductively that f^n maps J homeomorphically with degree $\varepsilon_0 \varepsilon_1 \dots \varepsilon_{n-1} \neq 0$ onto an interval $f^n(J)$. Evidently the two endpoints of this interval satisfy either $f^n(x) < f^n(y)$ or $f^n(x) > f^n(y)$ according as the sign $\varepsilon_0 \dots \varepsilon_{n-1}$ is $+1$ or -1 . In either case it follows that

$$\theta_n(x) = \varepsilon_0 \dots \varepsilon_{n-1} A(f^n(x)) \leq \varepsilon_0 \dots \varepsilon_{n-1} A(f^n(y)) = \theta_n(y),$$

as required. ■

It will often be convenient to introduce an indeterminate t , and to think of the vectors θ_n as the coefficients of a formal power series

$$\theta = \theta_0 + \theta_1 t + \theta_2 t^2 + \dots .$$

This series $\theta = \theta(x)$ will be called the invariant coordinate associated with the point x .

Of course, the series θ determines the sequence θ_* of coefficients, and conversely, so that it is just a matter of notational convenience whether we work with one or the other. For ordering purposes we should think of the indeterminate t as being positive, $t > 0$, but infinitesimally small so that a power series $v_0 + v_1 t + v_2 t^2 + \dots$ will be positive whenever its first nonzero coefficient is positive. With this convention, it follows from 3.1 that the function $x \mapsto \theta(x)$ is also monotone increasing.

Let $Q[[t]]$ be the ring of formal power series with rational coefficients, and let $V[[t]]$ be the module consisting of all formal

power series with coefficients in the vector space V . Thus each series $\theta = \theta(x)$ is an element of $V[[t]]$, which is a free module with basis I_1, \dots, I_ℓ over the ring $\mathbb{Q}[[t]]$. In other words, we can express θ uniquely as a sum

$$\theta = \theta_1 I_1 + \dots + \theta_\ell I_\ell$$

with coefficients θ_j which are formal power series with rational coefficients (in fact, usually with integer coefficients).

These coefficients θ_j satisfy one relation which is a formal consequence of the way in which the series θ was constructed.

Lemma 3.2. The sum

$$(1 - \varepsilon(I_1)t)\theta_1 + \dots + (1 - \varepsilon(I_\ell)t)\theta_\ell$$

in the formal power series ring $\mathbb{Q}[[t]]$ is identically equal to 1.

The coefficient of θ_j in this relation is either $1 - t$ or $1 + t$ according as the function f is increasing or decreasing on the lap I_j .

Before giving the proof, let us illustrate these constructions by considering the function $f(x) = (x^2 - 5)/2$ of Figure 1.4, which maps the interval $|x| \leq 1 + \sqrt{6}$ to itself. First consider the two fixed points $1 \pm \sqrt{6}$ of f . The interior fixed point $1 - \sqrt{6} = -1.44\dots$ has itinerary (I_1, I_1, I_1, \dots) . (It is often convenient to describe such an itinerary briefly by just listing the subscripts, in this case $1\ 1\ 1\ 1\ 1\ \dots$.) The corresponding invariant coordinate is

$$\theta(1 - \sqrt{6}) = I_1(1 - t + t^2 - \dots) = I_1/(1 + t).$$

Similarly, the rightmost point $1 + \sqrt{6} = 3.44\dots$ has itinerary corresponding to the sequence $2\ 2\ 2\ 2\ 2\ \dots$, with

$$\theta(1 + \sqrt{6}) = I_2(1 + t + t^2 + \dots) = I_2/(1 - t).$$

These fixed points are both unstable: nearby points always have

itineraries which are different from these.

The map f has just two other periodic points, constituting an orbit

$$\{-1-\sqrt{2}, -1+\sqrt{2}\} = \{-2.41\dots, 0.41\dots\}$$

of period 2. The itinerary of $-1-\sqrt{2}$ corresponds to the sequence $1\ 2\ 1\ 2\ 1\ 2\ \dots$, with

$$\theta(-1-\sqrt{2}) = I_1 - tI_2 - t^2I_1 + \dots = (I_1 - tI_2)/(1+t^2),$$

while the itinerary of $-1+\sqrt{2}$ corresponds to $2\ 1\ 2\ 1\ 2\ 1\ \dots$, with

$$\theta(-1+\sqrt{2}) = (tI_1 + I_2)/(1+t^2).$$

This orbit is stable, so that any point x which is sufficiently close to $-1-\sqrt{2}$ (or to $-1+\sqrt{2}$) will have the same itinerary and hence the same coordinate θ as $-1-\sqrt{2}$ (or as $-1+\sqrt{2}$).

Listing the four periodic points in their natural order, it is easy to check that

$$\theta(-1-\sqrt{2}) < \theta(1-\sqrt{6}) < \theta(-1+\sqrt{2}) < \theta(1+\sqrt{6}),$$

thus confirming 3.1 for these particular points. Also, it is easy to check that the relation $(1+t)\theta_1 + (1-t)\theta_2 = 1$ of 3.2 is satisfied for each of these points.

Proof of 3.2. Define a $\mathbb{Q}[[t]]$ -linear mapping h from the free module $V[[t]]$ to $\mathbb{Q}[[t]]$ by setting

$$h(I_j) = 1 - \varepsilon(I_j)t.$$

Then we must prove that $h(\theta) = 1$.

Note first that the vector $C_j = (I_j + I_{j+1})/2$ satisfies the corresponding identity

$$h(C_j) = 1 - \varepsilon(C_j)t = 1.$$

Hence for any possible address A_n , with sign $\varepsilon(A_n) = \varepsilon_n$, we have

$$h(A_n) = 1 - \varepsilon_n t.$$

Now setting $\theta = \sum \theta_n t^n = \sum \varepsilon_0 \dots \varepsilon_{n-1} A_n t^n$, to be summed over all $n \geq 0$, it follows that

$$h(\theta) = \sum \varepsilon_0 \dots \varepsilon_{n-1} (1 - \varepsilon_n t) t^n = (1 - \varepsilon_0 t) + \varepsilon_0 (1 - \varepsilon_1 t) t + \dots .$$

Evidently this adds up to $+1$. ■

Remark 3.3. For the reader who is only interested in the special case $\ell = 2$, much of the above discussion is unnecessary. It would be much easier just to work with the function

$$\bar{\theta}(x) = \theta_2(x) - \theta_1(x),$$

which is a formal power series of the form $\pm 1 \pm t \pm \dots$, with coefficients ± 1 or 0 . Evidently $\bar{\theta}(x)$ depends monotonely on x , and determines the series $\theta_1(x)$ and $\theta_2(x)$ uniquely and linearly. However, we will continue to develop the somewhat more complicated tools which are needed to deal with the case $\ell > 2$.

4. The kneading matrix $[N_{ij}]$

One basic invariant which describes the qualitative behavior of the map f and its iterates is given by the itineraries $A(f^*(c_1))$, $\dots, A(f^*(c_{\ell-1}))$ associated with the $\ell-1$ turning points of f . Out of these itineraries, we will construct $\ell-1$ formal power series

$$\nu_1, \dots, \nu_{\ell-1} \in V[[t]]$$

which will be called the kneading increments of f . These will play a fundamental role in all of our discussions.

Recall that $V[[t]]$ is the free module with basis I_1, \dots, I_ℓ over the ring $Q[[t]]$. We give $V[[t]]$ the formal power series topology, in which the submodules $t^n V[[t]]$ form a basis for the neigh-

borhoods of zero. As in §3, consider the invariant coordinate function $x \mapsto \theta(x)$ from I to $V[[t]]$. Since this function is monotone increasing, and since each coefficient $\theta_n(x)$ can take only finitely many distinct values, it follows easily that the right hand limit

$$\theta(x+) = \lim_{y \rightarrow x, y > x} \theta(y)$$

exists and is well defined for any x in the interior of the interval I . (To be more explicit, for each $n \geq 0$ there exists $\delta_n > 0$ so that $\theta_n(y)$ takes a constant value, denoted by $\theta_n(x+)$, for all y with $x < y < x + \delta_n$. Then $\theta(x+)$ is equal to the power series $\sum \theta_n(x+)t^n$.) Similarly, the left hand limit $\theta(x-)$ is well defined. The difference $\theta(x+) - \theta(x-)$ measures the discontinuity in the function $x \mapsto \theta(x)$ at x .

Definition. This measure of discontinuity $\theta(c_i+) - \theta(c_i-)$, evaluated at the turning point c_i , will be called the i -th kneading increment ν_i of f .

Closely related is the kneading matrix $[N_{ij}]$. This is the $(\ell-1) \times \ell$ matrix, with entries in the ring $\mathbb{Z}[[t]]$, which is obtained by setting

$$\nu_i = N_{i1}I_1 + \dots + N_{i\ell}I_\ell.$$

Remark 4.1. We can of course write this matrix $[N_{ij}]$ as a power series $\sum [N_{ij}^k]t^k$ where the coefficients $[N_{ij}^0], [N_{ij}^1], \dots$ are integer matrices. Each coefficient matrix $[N_{ij}^k]$ (after being rotated 90° counterclockwise to allow for the difference in orientation conventions for matrices and for graphs) can be thought of as a kind of approximate graph for the function f^k , evaluated only at the turning points $c_1, \dots, c_{\ell-1}$. Here is a precise statement. For $k \geq 1$, the entry N_{ij}^k is nonzero if and only if the address $A(f^k(c_i+))$ equals I_j , each nonzero entry being either $+2$ or -2 according as the

function f^k has a local minimum or a local maximum at c_i . On the other hand, for $k = 0$, the matrix

$$[N_{ij}^0] = \begin{bmatrix} -1 & 1 & 0 & 0 & & 0 & 0 & 0 \\ 0 & -1 & 1 & 0 & \dots & 0 & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot & & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & 0 & & 0 & -1 & 1 \end{bmatrix}$$

is completely independent of the mapping f . The proof is not difficult. ■

Using 3.2 we see easily that the linear relation

$$(4.2) \quad \sum_j N_{ij}(1-\varepsilon(I_j)t) = 0$$

must be satisfied for each i . In other words, the columns $\Gamma_1, \dots, \Gamma_\ell$ of the kneading matrix $[N_{ij}]$ are linearly dependent, satisfying the relation

$$\Gamma_1(1-\varepsilon t) + \Gamma_2(1+\varepsilon t) + \dots + \Gamma_\ell(1+(-1)^\ell \varepsilon t) = 0,$$

where $\varepsilon = \varepsilon(I_1)$.

Let $D_i = \det(\Gamma_1, \dots, \hat{\Gamma}_i, \dots, \Gamma_\ell)$ denote the determinant of the $(\ell-1) \times (\ell-1)$ matrix which is obtained from N_{ij} by deleting its i -th column. Using the relation 4.2 and elementary linear algebra we obtain the following.

Lemma 4.3. The ratio $D = (-1)^{i+1} D_i / (1-\varepsilon(I_i)t)$ is a fixed element of $\mathbb{Z}[[t]]$, independent of the choice of i . This power series D has leading coefficient $+1$, and hence is a unit of the ring $\mathbb{Z}[[t]]$.

Definition. We will call $D = D(t)$ the kneading determinant of f .

In the case $\ell = 2$ these constructions take a particularly simple form. In fact the equations

$$N_{12} = D_1 = (1-\varepsilon t)D$$

$$N_{11} = D_2 = (-1-\varepsilon t)D,$$

where $\varepsilon = \varepsilon(I_1)$, show that the determinant D completely determines

the kneading matrix $[N_{ij}]$. This power series D is sometimes called the kneading invariant of f when $\ell = 2$. Note that $N_{12} - N_{11} = 2D$. If we use the notation of §3.3, mapping $\theta(x) \in V[[t]]$ to a power series $\bar{\theta}(x)$ with integer coefficients by mapping the two basic elements I_1 and I_2 to -1 and $+1$ respectively, then $\bar{\theta}(c_1) = 0$, and it follows easily that

$$\begin{aligned}\bar{\theta}(c_{1+}) &= D \\ \bar{\theta}(c_{1-}) &= -D.\end{aligned}$$

Evidently D is always an infinite series of the form $1 + t + t^2 + \dots$.

Here is a more explicit description of how to compute the i -th kneading increment ν_i from the itinerary $A(f^*(c_{i+}))$. Let us use the notation A_n for the address $A(f^n(c_{i+})) = A(f^n(c_{i-}))$ for $n \geq 1$. Then evidently the right hand limit sequence $\theta_*(c_{i+})$ is given by

$$\theta_*(c_{i+}) = (I_{i+1}, \varepsilon_0 A_1, \varepsilon_0 \varepsilon_1 A_2, \dots)$$

where the signs

$$\varepsilon_0 = \varepsilon(I_{i+1}), \quad \varepsilon_n = \varepsilon(A_n) \quad \text{for } n \geq 1,$$

are all nonzero. Similarly

$$\theta_*(c_{i-}) = (I_i, -\varepsilon_0 A_1, -\varepsilon_0 \varepsilon_1 A_2, \dots).$$

Subtracting these two expressions, we find that ν_i is an infinite power series of the form

$$\nu_i = (I_{i+1} - I_i) + 2\varepsilon_0 A_1 t + 2\varepsilon_0 \varepsilon_1 A_2 t^2 + \dots$$

These formulas show that the right and left hand limits can be written as

$$(4.4) \quad \begin{cases} \theta(c_{i+}) = \theta(c_i) + \nu_i/2 \\ \theta(c_{i-}) = \theta(c_i) - \nu_i/2. \end{cases}$$

Thus the average of the right and left hand limits is always precisely

equal to the value $\theta(c_1) = (I_1 + I_{i+1})/2$ (compare §5.2). Any one of the three series $\theta(c_{i+})$, $\theta(c_{i-})$ and ν_i determines the other two uniquely.

If we are given the itinerary $A(f^*(c_i))$ together with the sign $\varepsilon(I_1)$, then the address $A_n = A(f^n(c_{i+}))$ can be computed as follows by induction on n . If the address $A(f^n(c_i))$ is an interval I_j , then $A(f^n(c_{i+}))$ equals I_j also, and there is no problem. However, if $f^n(c_i)$ is a critical point c_j , then $A(f^n(c_{i+}))$ can be either I_{j+1} or I_j . In fact, it will be I_{j+1} whenever the sign $\varepsilon_0 \varepsilon_1 \dots \varepsilon_{n-1}$ is $+1$ and I_j whenever $\varepsilon_0 \varepsilon_1 \dots \varepsilon_{n-1} = -1$. For this sign determines whether the function f^n is increasing or decreasing throughout some sufficiently small interval $[c_i, c_i + \delta)$ to the right of c_i .

In the special case $\ell = 2$, these arguments give rise to a quite simple recipe.

Lemma 4.5. If $\ell = 2$, then the kneading determinant can be computed by the formula

$$D = 1 + \varepsilon_1 t + \varepsilon_1 \varepsilon_2 t^2 + \varepsilon_1 \varepsilon_2 \varepsilon_3 t^3 + \dots$$

Here the sign $\varepsilon_n = \varepsilon(A(f^n(c_{i+})))$ is equal to $\varepsilon(A(f^n(c_i)))$ when $f^n(c_i) \neq c_i$, and satisfies $\varepsilon_1 \varepsilon_2 \dots \varepsilon_n = 1$ when $f^n(c_i) = c_i$.

The proof is easily supplied. ■

Thus, if $f^p(c_1) = c_1$ then the sequence of coefficients $1, \varepsilon_1, \varepsilon_1 \varepsilon_2, \dots$ will be periodic of period p . Briefly we will say that D has period p .

Example 4.6. Let us look at the quadratic map $f(x) = (x^2 - a)/2$ for a few representative values 4, 5, 6, 7, 8 of the parameter a . For

$$f(x) = (x^2 - 4)/2,$$

since $f^2(0) = 0$, the kneading determinant must be periodic of period 2. In fact, computation shows that

$$A(f^*(0)) = (C_1, I_1, C_1, I_1, \dots);$$

hence $\varepsilon_1 = \varepsilon_2 = \dots = -1$ and the determinant

$$D = 1 - t + t^2 - t^3 + \dots = 1/(1+t)$$

does indeed have period 2. (The kneading matrix $[N_{11}, N_{12}]$ is equal to

$$[-1+t, 1+t]D = [-1+2t-2t^2+\dots, 1].)$$

Next consider the map $f(x) = (x^2-5)/2$ of Figure 1.4. In this case

$$A(f^*(0)) = (C_1, I_1, I_2, I_1, I_2, \dots);$$

and the determinant

$$D = 1 - t - t^2 + t^3 + t^4 - \dots = (1-t)/(1+t^2)$$

has period 4. For the map $f(x) = (x^2-6)/2$, it is not known whether or not the kneading determinant is periodic. Certainly there is no small period. The map $f(x) = (x^2-7)/2$ of §2.2 has determinant

$$D = 1 - t - t^2 + t^3 - t^4 - t^5 + \dots = (1-t-t^2)/(1-t^3)$$

of period 3. Finally, the map $f(x) = (x^2-8)/2$ has a kneading determinant

$$D = 1 - t - t^2 - t^3 - \dots = (1-2t)/(1-t)$$

which is eventually periodic but not periodic.

In the case $\ell > 2$, the computations are of course more complicated. To give one example the function f graphed in Figure 1.5 has kneading matrix

$$[N_{ij}] = \begin{bmatrix} -1+2t+2t^3+\dots & 1 & -2t^2-2t^4-\dots \\ 2t^2+2t^4+\dots & -1 & 1-2t-2t^3 \end{bmatrix}$$

with kneading determinant

$$D = 1 - 3t + t^2 - 3t^3 + t^4 - 3t^5 + \dots = (1-3t)/(1-t^2).$$

5. Counting pre turning points and laps

Let us continue to study the continuity properties of the function $x \mapsto \theta(x)$ from I to $V[[t]]$, using the formal power series topology in $V[[t]]$. Let x_0 be some interior point of the interval I .

Lemma 5.1. If none of the successive images $x_0, f(x_0), f^2(x_0), \dots$ is a turning point of f , then the function θ is continuous at x_0 , that is, $\theta(x_0^-) = \theta(x_0) = \theta(x_0^+)$. In this case $\theta(x_0)$ is an infinite power series in t with all coefficients $\theta_n(x_0)$ different from zero.

The proof is easily supplied. ■

Now suppose that some $f^n(x_0)$ is a turning point of f . We will say that x_0 is a pre turning point of f . Let $n \geq 0$ be the smallest integer such that $f^n(x_0)$ is a turning point of f , and let $f^n(x_0) = c_i$.

Lemma 5.2. If x_0 is a pre turning point, where $f^n(x_0) = c_i$ with n minimal, then

$$\theta(x_0^+) = \theta(x_0) + t^n \nu_i / 2,$$

$$\theta(x_0^-) = \theta(x_0) - t^n \nu_i / 2.$$

Thus $\theta(x_0)$ is always equal to the average of the left and right hand limits.

Again, the proof is easily supplied. ■

Note that x_0 is a turning point of the iterated function f^k if and only if $k > n$. The integers $n \geq 0$ and i can be uniquely characterized by the fact that the difference $f^n(x) - c_i$ changes sign at x_0 . The equation $f^m(x_0) = c_j$ may have other solutions, with $m > n$, but if this happens, then x_0 will be a turning point of f^m , so that the function $f^m(x) - c_j$ will not change sign at x_0 .

Using this lemma, we will give an algorithm for counting the num-

bers of pre turning points of various kinds in any interval. Suppose that we are given two points $a < b$ of the interval I . Suppose first, to simplify the discussion, that the function θ is continuous at both a and b . Let J denote the interval $[a, b]$.

Lemma 5.3. The difference $\theta(b) - \theta(a)$ can be computed by summing the discontinuities $\theta(x+) - \theta(x-)$ over all points x in the interior of the interval J .

In particular, it is asserted that this infinite sum converges in the formal power series topology. Of course, only countably many of the summands can be nonzero.

Proof. First consider the function $\theta(x)$ modulo t^n . Evidently, $\theta(x) \bmod t^n$ is a step function, with only finitely many discontinuities in the interval J , so the formula

$$\theta(b) - \theta(a) \equiv \sum (\theta(x+) - \theta(x-)) \bmod t^n \text{ in } \mathbb{Z}[[t]]$$

follows immediately. Taking the limit as $n \rightarrow \infty$, we are done. ■

With $J = [a, b]$ as above, we will next define formal power series

$$\gamma_1(J), \dots, \gamma_{\ell-1}(J) \in \mathbb{Z}[[t]],$$

as follows. For each $n \geq 0$, let $\gamma_{in}(J)$ be the number of times that the function $f^n(x) - c_i$ changes sign in the interior of the interval J , and let $\gamma_i(J) = \sum \gamma_{in}(J)t^n$. Then we will prove the following.

Theorem 5.4. The difference $\theta(b) - \theta(a)$ is equal to $\gamma_1(J)\nu_1 + \dots + \gamma_{\ell-1}(J)\nu_{\ell-1}$.

Proof. If $f^n(x) - c_i$ changes sign at x_0 , then evidently the point x_0 makes a contribution of t^n to the power series $\gamma_i(J)$, and hence a contribution of $t^n \nu_i$ to the sum $\gamma_1(J)\nu_1 + \dots + \gamma_{\ell-1}(J)\nu_{\ell-1}$. But, by 5.2 and 5.3, the contribution $\theta(x_0+) - \theta(x_0-)$ of x_0 to the

difference $\theta(b) - \theta(a)$ is also precisely $t^n \nu_i$. Summing over all x_0 in the interior of J , the conclusion follows. ■

Remark 5.5. If we drop the hypothesis that θ is continuous at a and b , then the formula becomes

$$\theta(b-) - \theta(a+) = \gamma_1(J)\nu_1 + \dots + \gamma_{\ell-1}(J)\nu_{\ell-1},$$

as can easily be established.

Corollary 5.6. Given the kneading matrix $[N_{ij}]$ and given the difference $\theta(b-) - \theta(a+)$, we can compute the number $\gamma_{in}(J)$ of times that the function $f^n(x) - c_i$ changes sign in the interior of J , and also the number $\ell(f^n|J)$ of laps of the function f^n restricted to J .

Proof. If we delete say the last column of the matrix $[N_{ij}]$, recall from §4.2 that the determinant D_ℓ of the resulting square matrix is a unit of the ring $\mathbb{Z}[[t]]$. Hence there is a unique two-sided inverse matrix, say $[M_{jk}]$, with entries in $\mathbb{Z}[[t]]$. Writing 5.5 as

$$\theta_j(b-) - \theta_j(a+) = \sum_i \gamma_i(J) N_{ij},$$

we can multiply on the right by M_{jk} and sum over j to obtain

$$(5.7) \quad \sum_{j=1}^{\ell-1} (\theta_j(b-) - \theta_j(a+)) M_{jk} = \gamma_k(J).$$

This computes the power series $\gamma_k(J)$, and hence its coefficients $\gamma_{kn}(J)$.

But x_0 is a turning point of f^n if and only if $f^m(x) - c_i$ changes sign at x_0 for some $m < n$ and some c_i . The number of such points in the interior of J is $\gamma_{im}(J)$. Hence the number of turning points of f^n in the interior of J is equal to the sum of $\gamma_{im}(J)$ over all turning points c_i and over all $m < n$. Adding +1 to this sum, we obtain the required number $\ell(f^n|J)$ of laps of f^n

on J . ■

For later use, we also record a formal power series version of this last computation.

Corollary 5.8. The formal power series $\sum_{n=1}^{\infty} \ell(f^n|J)t^{n-1}$ is equal to
 $(1+\gamma_1(J)+\dots+\gamma_{\ell-1}(J))/(1-t)$.

The proof is straightforward. ■

In particular, these results apply to the case where J is the entire interval I . We use the notations c_0 and c_ℓ for the two endpoints of I and the notation $\dot{I} = \{c_0, c_\ell\}$ for the boundary of I . (Caution: c_0 and c_ℓ are not turning points.) Suppose that the following is satisfied.

Hypothesis. The mapping f carries the boundary \dot{I} into itself.

Then evidently both $\theta(c_0)$ and $\theta(c_\ell)$ can be expressed as linear combinations of the form $\theta_1 I_1 + \theta_\ell I_\ell$. Hence 5.7 takes the somewhat simpler form

$$(\theta_1(c_\ell) - \theta_1(c_0))M_{1k} = \gamma_k(I).$$

In order to compute $\theta_1(c_\ell) - \theta_1(c_0)$, we need only know the precise way in which f maps \dot{I} to itself. There are just four possibilities, which can be listed as follows.

Case Λ . Both end points map to the initial point c_0 .

Case N . Both end points are fixed by f .

Case V . Both end points map to c_ℓ .

Case h . The endpoints are interchanged by f .

(The symbols Λ , N , V , and h have been chosen to resemble schematic graphs for the respective functions f .)

In each of these four cases, the computation can easily be carried out. For example, in Case Λ we have

$$\theta(c_0) = I_1/(1-t), \theta(c_\ell) = -tI_1/(1-t)+I_\ell;$$

hence

$$\theta_1(c_\ell) - \theta_1(c_0) = -(1+t)/(1-t).$$

Similarly, in Case V this difference is -1 and in Cases N and h it is $-1/(1-t)$. (The apparent asymmetry between Cases A and V is an artifact of our method of computation, which is based on canceling the last column of the kneading matrix.)

If $\ell = 2$, then M_{11} is just the reciprocal of $N_{11} = -(1+\varepsilon(I_1)t)D$, so these computations take a particularly simple form.

Corollary 5.9. If $\ell = 2$, and if $f(i) < i$, then the power series $\gamma_1(I)$ is equal to $(1-t)^{-1}D^{-1}$ where D is the kneading determinant. Hence $L(t) = \sum \ell(f^n)t^{n-1}$ is equal to $(1-t)^{-1} + (1-t)^{-2}D(t)^{-1}$.

The proof is easily supplied. ■

Consider for example the function $f(x) = (x^2-4)/2$ mapping the interval $|x| \leq b = 1+\sqrt{5}$ to itself, with kneading determinant $D = (1+t)^{-1}$ (compare 1.3, 4.6). Since $f(\pm b) = b$, we are in Case V. By 5.9 we have

$$\gamma_1(I) = (1-t)^{-1}D^{-1} = (1+t)/(1-t) = 1 + 2t + 2t^2 + \dots$$

and

$$L(t) = 2/(1-t)^2 = 2 + 4t + 6t^2 + \dots$$

In other words $\ell(f^n) = 2n$ for $n \geq 1$. The growth number $s = \lim(2n)^{1/n}$ equals 1.

On the other hand, for the function $f(x) = (x^2-7)/2$ with determinant $D = (1-t-t^2)/(1-t^3)$, we obtain

$$1 + \gamma_1(I) = 2/(1-t-t^2) = 2 + 2t + 4t^2 + 6t^3 + 10t^4 + 16t^5 + \dots$$

The coefficients in this sequence are just the doubles of the Fibonacci numbers

$$1, 1, 2, 3, 5, 8, \dots$$

The associated lap numbers $\ell(f^n)$ can now easily be computed. For example, $\ell(f^6) = 2(1+1+2+3+5+8) = 40$. Using the formula

$$\left(\left(\frac{1+\sqrt{5}}{2} \right)^{n+1} - \left(\frac{1-\sqrt{5}}{2} \right)^{n+1} \right) / \sqrt{5}$$

for the n -th Fibonacci number, or using a radius of convergence argument as in §6, one finds that $s = (1+\sqrt{5})/2 = 1.61\dots$

6. Convergent power series

In this section we will re-examine the various formal power series which have been introduced, showing that they can be profitably studied as convergent series. Thus we think of the symbol t no longer as a formal indeterminate, but now rather as a real or complex variable.

As a first example, consider the series $L(t) = \sum \ell(f^n) t^{n-1}$ studied in §§5.8, 5.9. Using the definition $s = \lim \ell(f^n)^{1/n}$, we see that the radius of convergence of this series is precisely $1/s$. In other words, the function $L(t)$ is defined and complex analytic for $|t| < 1/s$, but cannot be extended as an analytic function over any larger disk.

As a second example, consider an entry $N_{ij} = \sum_k N_{ij}^k t^k$ of the kneading matrix. Since the coefficients N_{ij}^k are bounded integers, this series converges for $|t| < 1$. Thus we will now think of the kneading matrix as a matrix $[N_{ij}(t)]$ of functions, all defined and analytic throughout the unit disk $|t| < 1$. It follows that the determinant $D(t)$ is also defined and analytic for $|t| < 1$.

As another example, fixing some point x_0 of I , the coordinate $\theta(x_0) = \sum \theta_j(x_0) I_j$ will now be thought of as a function

$$\theta(x_0, t) = \sum \theta_j(x_0, t) I_j$$

which is defined and analytic for $|t| < 1$, and takes values in the real or complex vector space with basis I_1, \dots, I_ℓ .

As a final example, the various power series $\gamma_i(J) = \sum \gamma_{in}(J)t^n$ considered in §5.4 will now be thought of as functions $\gamma_i(J, t)$.

Using the easily verified inequality

$$0 \leq \gamma_{in}(J) \leq \ell(f^n|J) \leq \ell(f^n),$$

we see that these functions are all defined and analytic for $|t| < 1/s$. However, we can do better than this.

Lemma 6.1. For each i and J this complex function $\gamma_i(J, t)$, which a priori is only defined for $|t| < 1/s$, actually extends to a function which is defined and meromorphic throughout the disk $|t| < 1$. This meromorphic function has poles at most at the zeros of the analytic function $D(t)$.

Proof. This follows easily from the identity

$$\gamma_k(J, t) = \sum (\theta_j(b^-, t) - \theta_j(a^+, t)) M_{jk}(t)$$

of 5.7., where the θ_j are analytic for $|t| < 1$. Multiplying both sides by $D_\ell(t)$, note that each $D_\ell(t)M_{jk}(t)$ is equal to an appropriate minor determinant of the matrix $[N_{ij}(t)]$, and hence is analytic for $|t| < 1$. Therefore, the product $D(t)\gamma_k(J, t)$ is analytic for $|t| < 1$. ■

Using 5.8., it follows that the function $L(t) = \sum \ell(f^n)t^{n-1}$ also extends to a function which is meromorphic throughout the disk $|t| < 1$, with $D(t)L(t)$ analytic for $|t| < 1$.

Corollary 6.2. If $s > 1$, then the analytic function $D(t)$ has a zero at $t = 1/s$.

Proof. Since the power series $\sum \ell(f^n)t^{n-1}$ has radius of convergence precisely $1/s$, the meromorphic function $L(t)$ must have a pole somewhere on the circle $|t| = 1/s$. But the coefficients $\ell(f^n)$ are

nonnegative; hence

$$|L(t)| \leq L(|t|)$$

for $|t| \leq 1/s$, and it follows that there must be a pole precisely at $t = 1/s$. ■

In fact, this is the smallest zero of $D(t)$.

Theorem 6.3. The function $D(t)$ is nonzero for $t < 1/s$. Therefore, the first zero of $D(t)$ as t varies along the interval $[0,1)$ occurs precisely at $t = 1/s$ if $s > 1$, while $D(t)$ has no zeros at all in the open unit disk if $s = 1$.

As examples, consider the five kneading determinants computed in §4.6:

$$D(t) = 1/(1+t), \quad (1-t)/(1+t^2), \quad (1-t-t^2)/(1-t^3), \\ (1-2t)/(1-t), \quad \text{and} \quad (1-3t)/(1-t^2).$$

Looking for the first zero of $D(t)$ along the interval $[0,1)$, we find that the growth numbers s for the corresponding functions f are respectively 1, 1, $(1+\sqrt{5})/2$, 2, and 3.

The proof of 6.3 begins as follows. For the moment we will need the supplementary hypothesis that f maps the boundary \dot{I} into itself. This extra hypothesis will be removed in §7.7.

In the special case $\ell = 2$, the conclusion follows immediately from 5.9 and the definition of s . However, the general case requires more work. We first construct a rather trivial nonsingular matrix.

Lemma 6.4. Suppose that $f(\dot{I}) \subset \dot{I}$. Then the square matrix

$$[A_{jk}(t)] = [\theta_j(c_k, t) - \theta_j(c_0, t)]$$

of analytic functions, where j and k range from 1 to $\ell-1$, is nonsingular for every t in the disk $|t| < 1$.

Proof. Introducing the abbreviation $a = -2\theta_1(c_0, t)$, inspection shows that twice this matrix is equal to

$$\begin{bmatrix} 1+a & a & a & & a & a & a \\ 1 & 1 & 0 & \dots & 0 & 0 & 0 \\ 0 & 1 & 1 & & 0 & 0 & 0 \\ & \dots & & & & \dots & \\ 0 & 0 & 0 & & 1 & 1 & 0 \\ 0 & 0 & 0 & & 0 & 1 & 1 \end{bmatrix}$$

Using elementary row operations to remove an even number of a 's, it follows that the determinant of this doubled matrix is equal to $+1$ when ℓ is odd, and is equal to

$$1+a = 1 - 2\theta_j(c_0, t)$$

when ℓ is even. Computation shows that this last expression equals -1 in Case V, and equals $-(1+t)/(1-t)$ in Case A (compare §5). Thus in all cases, the determinant is nonzero for $|t| < 1$. ■

Proof of 6.3, assuming that $f(\dot{i}) \subset \dot{i}$. In this proof, i, j, k will vary from 1 to $\ell-1$. Consider the square matrix with entries

$$\Gamma_{ik}(t) = \gamma_i([c_0, c_k], t) + \delta_{ik}/2,$$

where $[\delta_{ik}]$ is the identity matrix. Using 5.5 and 5.2 we easily obtain the formula

$$\theta(c_k, t) - \theta(c_0, t) = \sum_i \nu_i(t) \Gamma_{ik}(t),$$

or in other words,

$$(6.5) \quad \theta_j(c_k, t) - \theta_j(c_0, t) = \sum_i N_{ij}(t) \Gamma_{ik}(t),$$

where all of these functions are defined and analytic for $|t| < 1/s$. For every such t , the left hand matrix is nonsingular; hence the matrix $[N_{ij}(t)]$ with determinant $D_\ell(t)$ must also be nonsingular. This proves that $D_\ell(t) \neq 0$; hence $D(t) \neq 0$, for $|t| < 1/s$. ■

7. Piecewise-linear models

Throughout this section we assume that the growth number s is strictly greater than 1. We set $r = 1/s$. Thus $0 < r < 1$.

For any closed interval $J \subset I$, consider the power series

$$L(J, t) = \sum \ell(f^n|_J) t^{n-1}$$

of 5.8. Recall from 6.1 that this series converges for $|t| < r$, and extends to a meromorphic function which is defined throughout the disk $|t| < 1$. Taking $J = I$, the meromorphic function $L(I, t)$ definitely does have a pole at $t = r$.

Evidently $0 \leq L(J, t) \leq L(I, t)$ for $0 \leq t < r$. Hence the ratio $L(J, t)/L(I, t)$ remains bounded as t tends to r from the left. This implies that the apparent singularity of this meromorphic function at $t = r$ is removable. The limit

$$\Lambda(J) = \lim_{t \rightarrow r} L(J, t)/L(I, t)$$

exists, and satisfies $0 \leq \Lambda(J) \leq 1$. (Perhaps one should think of this limit intuitively as the probability that a randomly chosen lap of the function f^n , with n very large, will be contained in the interval J .)

Here are some basic properties of this construction.

Lemma 7.1. If the intervals J_1 and J_2 intersect only at a common end point, then $\Lambda(J_1 \cup J_2) = \Lambda(J_1) + \Lambda(J_2)$.

Lemma 7.2. The number $\Lambda(J)$ depends continuously on the end points of the interval J .

Lemma 7.3. If J is contained in a single lap of f , or in other words if $f|_J$ is a homeomorphism, then $\Lambda(f(J)) = s\Lambda(J)$.

The first two lemmas imply that the set function Λ gives rise to a measure on the interval I , each point having measure zero and the whole interval having measure $\Lambda(I) = 1$. The third lemma says

that f multiplies the measure of any subset of a lap by the constant factor $s > 1$. Thus the growth number s appears again in a somewhat different guise.

Proof of 7.1. Clearly the sum $\ell(f^n|_{J_1}) + \ell(f^n|_{J_2})$ differs from $\ell(f^n|_{J_1 \cup J_2})$ by at most 1. Hence the difference

$$L(J_1, t) + L(J_2, t) - L(J_1 \cup J_2, t)$$

remains bounded as $t \rightarrow r$. Dividing by $L(I, t)$ and passing to the limit as $t \rightarrow r$, we obtain the required equality $\Lambda(J_1) + \Lambda(J_2) - \Lambda(J_1 \cup J_2) = 0$. ■

Proof of 7.3. If $f|_J$ is a homeomorphism, then $\ell(f^{n+1}|_J) = \ell(f^n|_{f(J)})$, which implies that

$$L(J, t) = 1 + tL(f(J), t).$$

The required equation follows easily. ■

Proof of 7.2. If the interval J is small enough to be contained in just one lap of the function f^n , then it follows inductively from 7.3 that

$$\Lambda(J) = s^{-n} \Lambda(f^n(J)) \leq s^{-n}.$$

This upper bound tends to zero as $n \rightarrow \infty$. Making use of 7.1, the conclusion follows easily. ■

Now define a map $\lambda : I \rightarrow [0, 1]$ by setting $\lambda(x) = \Lambda([c_0, x])$. Evidently λ is continuous, monotone, and maps I onto the unit interval.

Theorem 7.4. There exists one and only one map F from the unit interval $[0, 1]$ into itself so that

$$F(\lambda(x)) = \lambda(f(x))$$

for every x in I . This map F is piecewise-linear, with slope equal to $\pm s$ everywhere. The growth number of F is precisely equal

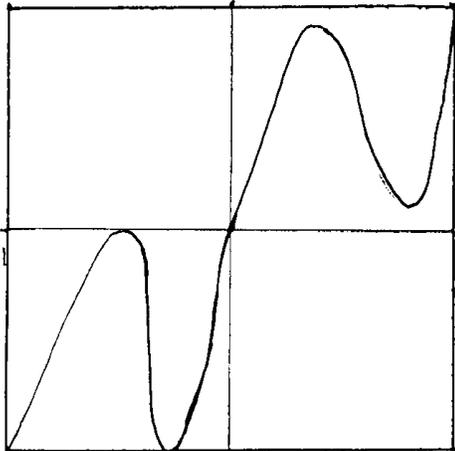
to s .

(Compare Parry.) Thus the following diagram is commutative:

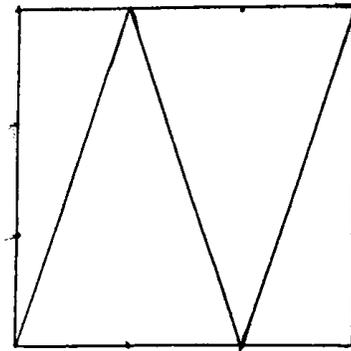
$$\begin{array}{ccc}
 I & \xrightarrow{f} & I \\
 \lambda \downarrow & & \downarrow \lambda \\
 [0,1] & \xrightarrow{f} & [0,1],
 \end{array}$$

where the function λ is continuous, monotone, and onto, but is not necessarily a homeomorphism (see Figure 7.11). Briefly, we will say that f is topological semiconjugate to a piecewise-linear function F having the same growth number s , and having slope $\pm s$.

Remark. The map λ may collapse some entire lap I_j to a point. If this happens, then F will have fewer laps than f . An example, in which λ maps the entire right hand half of the interval I to $+1$, is illustrated in Figure 7.5. In this case the growth number s equals 3, so the slope of F is ± 3 everywhere. The growth number of f restricted to the right hand half of I would be strictly less than 3, so this half gets lost in the piecewise-linear model.



Graph of f



Graph of F

Figure 7.5.

Proof of 7.4. We must first show that F exists. That is, if $\lambda(x) = \lambda(y)$, we must show that $\lambda(f(x)) = \lambda(f(y))$ so that $F(\lambda(x)) = \lambda(f(x))$ will be well defined. But if $\lambda(x) = \lambda(y)$ with say $x < y$, then evidently $\Lambda([x, y]) = 0$. Using 7.3, it follows that $\Lambda(f([x, y])) = 0$; hence $\lambda(f(x)) = \lambda(f(y))$, as required.

Next, let us compute $F(\lambda(x))$ as x varies through some lap $I_k = [c_{k-1}, c_k]$ of f . Suppose, to fix our ideas, that $\varepsilon(I_j) = +1$, so that $f|_{I_k}$ preserves orientation. If J denotes the subinterval $[c_{k-1}, x]$, then

$$F(\lambda(x)) = \lambda(f(x)) = \lambda(f(c_{k-1})) + \Lambda(f(J)).$$

But

$$\Lambda(f(J)) = s\Lambda(J) = s\lambda(x) - s\lambda(c_{k-1})$$

by 7.3. Setting A_k equal to the constant $\lambda(f(c_{k-1})) - s\lambda(c_{k-1})$, this proves that

$$F(\lambda(x)) = A_k + s\lambda(x)$$

for all x in I_k . In other words, the function F is linear with slope s throughout the interval $\lambda(I_k)$. If $\varepsilon(I_k) = -1$, then a completely analogous argument shows that F is linear with slope $-s$ throughout $\lambda(I_k)$.

To compute the growth number $s(F)$, we note that F^n has slope $\pm s^n$ everywhere. Hence every lap of F^n has length $\leq s^{-n}$, and the number of laps must satisfy $\ell(F^n) \geq s^n$. Therefore, $s(F) \geq s = s(f)$.

On the other hand, λ maps each lap of f^n into a lap of F^n , so it follows that $\ell(F^n) \leq \ell(f^n)$, and therefore, $s(F) \leq s(f)$. ■

Remark. More generally, it is true that any piecewise-linear map F with slope $\pm s$ everywhere has growth number s , providing that $s \geq 1$. Compare Misiurewicz and Szlenk.

As an easy corollary of 7.4, we can complete the proof of 6.3.

Consider a map $f_0 : J \rightarrow J$, with growth number $s_0 \geq 1$, which does not map the boundary of J into itself. Embedding J as the middle third of a larger interval I , we can extend f_0 as illustrated in Figure 7.6 to a map f with the same kneading matrix which does satisfy the hypothesis $f(I) \subset I$. (Case Λ is illustrated, but the construction in Case N , V , or h is completely analogous.)

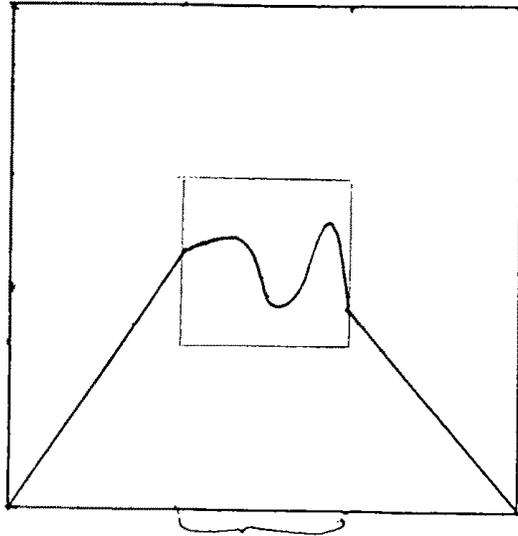


Figure 7.6.

Lemma 7.7. The growth number s of this extended function f is precisely equal to the growth number s_0 of f_0 .

Proof. Clearly $\ell(f^n) \geq \ell(f_0^n)$; hence $s \geq s_0$. Suppose that s were strictly greater than s_0 . Then $s > 1$, so Theorem 7.4 applies. Setting $r = 1/s$, since $r < 1/s_0$, the power series $L(J, r)$ converges. Hence the limit

$$\Lambda(J) = \lim_{t \rightarrow r} L(J, t) / L(I, t)$$

equals zero. Therefore, the function $\lambda(x) = \Lambda([c_0, x])$ maps the entire interval J into a single point λ_0 .

The associated piecewise-linear function $F : \lambda(x) \mapsto \lambda(f(x))$ must actually be linear, with $|\text{slope}| = s > 1$, throughout the inter-

val $[0, \lambda_0]$, and also throughout the interval $[\lambda_0, 1]$. Furthermore, since f maps the interval J into itself, we must have $F(\lambda_0) = \lambda_0$. Inspection shows that these conditions are impossible. Therefore, s must equal s_0 . ■

Proof of Theorem 6.3. The functions f and f_0 have the same kneading matrix and the same growth invariant. Since we have already proved 6.3 for the function f , it follows that it is true for f_0 also. ■

7.8. Discussion of Uniqueness. It should be noted that the function λ constructed above may not be the only semiconjugacy from f to a piecewise-linear mapping with $|\text{slope}| = \text{constant}$. As an example, let us single out some index k , and consider the power series $\gamma_k(I, t)$ of §6.1. Evidently its radius of convergence r_k satisfies $r \leq r_k \leq 1$ (where $r_k = r$ for at least one value of k). We will need the following.

Hypothesis. This radius of convergence r_k should be strictly less than 1.

Defining $\Lambda_k(J) = \lim_{t \rightarrow r_k} \gamma_k(J, t) / \gamma_k(I, t)$, the analogues of 7.1,

7.2, 7.3, and 7.4 are then easily verified. Thus the function

$$\lambda_k(x) = \Lambda_k([c_0, x])$$

has properties completely analogous to those of λ . In particular, $\lambda_k(f(x)) = F_k(\lambda_k(x))$, where F_k has slope $\pm r_k^{-1}$ everywhere.

As an example, for the function f illustrated in Figure 7.5, the function λ_3 will map the left half of the interval I to a point, and will provide a semiconjugacy from f to a piecewise-linear map F_3 which has $|\text{slope}|$ equal to a constant r_3^{-1} which is strictly less than 3.

This construction is, of course, closely related to the original

construction of λ . Using 5.8 one sees that λ can be expressed as a sum

$$\lambda(x) = a_1 \lambda_1(x) + \dots + a_{\ell-1} \lambda_{\ell-1}(x),$$

where the coefficient

$$a_k = (r/(1-r)) \lim_{t \rightarrow r} \gamma_k(I, t) / L(I, t)$$

is zero unless the function $\gamma_k(I, t)$ has a pole of precisely the right order, precisely at r .

In order to actually compute $\lambda(x)$, we proceed as follows.

Lemma 7.9. This real valued function $\lambda(x)$, which semiconjugates f to a piecewise linear map, can be expressed as a linear combination

$$\lambda(x) = a_1 \theta_1(x, r) + \dots + a_{\ell-1} \theta_{\ell-1}(x, r)$$

with constant coefficients.

The proof will provide an explicit procedure for determining the coefficients a_i . This formula is well adapted to numerical computations, since the power series $\theta_1(x, t)$ converges nicely at $t = r = 1/s$.

Proof of 7.9. Recall that the $(\ell-1) \times (\ell-1)$ kneading matrix $[N_{ij}(t)]$ has an inverse matrix $M_{jk}(t)$, whenever $D(t) \neq 0$. Using 6.4 and 6.5, it is not difficult to show that the limit

$$\bar{M}_{jk} = \lim_{t \rightarrow r} M_{jk}(t) / L(I, t)$$

exists, and is finite. Note that the matrix product

$$\sum N_{ij}(r) \bar{M}_{jk} = \lim_{t \rightarrow r} \delta_{ij} / L(I, t)$$

is identically zero. Using 5.7 and 5.2, we can now compute the limit of the ratio $\gamma_k(J, t) / L(I, t)$, as $t \rightarrow r$, for any interval $J = [a, b]$. We find that $\lim_{t \rightarrow r} \gamma_k(J, t) / L(I, t) = \sum_j (\theta_j(b, r) - \theta_j(a, r)) \bar{M}_{jk}$.

Using 5.8, this implies that

$$\Lambda(J) = (r/(1-r)) \sum_{j,k} (\theta_j(b,r) - \theta_j(a,r)) \bar{M}_{j,k}.$$

Hence

$$(7.10) \quad \lambda(x) = (r/(1-r)) \sum_{j,k} \theta_j(x,r) \bar{M}_{j,k} + (\text{constant}).$$

This is the required formula.

(It is conjectured that the functions $\lambda_k(x)$ of §7.8 can be expressed analogously as linear combinations of $\theta_1(x, r_k), \dots, \theta_{\ell-1}(x, r_k)$.)

In the case $\ell = 2$, formula 7.10 takes the simple form

$$\lambda(x) = \theta_2(x, r)$$

in Case A. Similarly, $\lambda(x) = 1 - \theta_1(x, r)$ in Case V.

As an example to illustrate this computational technique, the function $\lambda(x)$ associated with the mapping $f(x) = (x^2-7)/2$ is plotted in Figure 7.11.

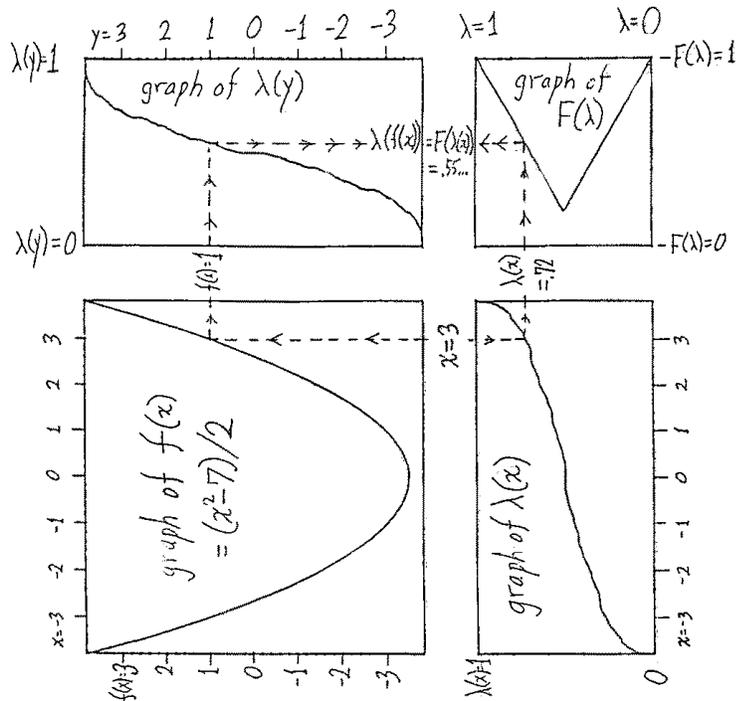
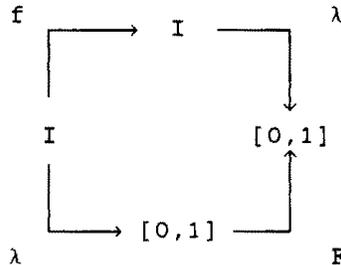


Figure 7.11. Piecewise-linearization of the function $f(x) = (x^2-7)/2$.

Note that λ has intervals of constancy, corresponding to points x whose forward images $f^*(x)$ converge towards the one-sided stable periodic orbit which was described in §2. The commutative diagram



is illustrated for the particular value $x = 3$.

We conclude this section with three applications of 7.4. Let f be a function with just one turning point, having growth number s .

Corollary 7.12. If s is not an algebraic unit satisfying a polynomial equation of the form $s^n + s^{n-1} + \dots + 1 = 0$, then the kneading determinant of f is uniquely determined by s .

We will see in §13.8 that some such hypothesis on s is definitely needed.

Proof. Evidently $s > 1$, so f is topologically semiconjugate to a map which can be put in the form

$$F(y) = s|y| - 1$$

by a suitable choice of variable. For any given s , the successive iterates $F^k(0) = \pm s F^{k-1}(0) - 1$ can easily be computed. Our hypothesis implies that these numbers are all nonzero. Hence, using 4.5, the successive coefficients of the kneading determinant can also be computed. ■

Now assume, to fix our ideas, that f takes its minimum at $c_1 = 0$. We continue to assume that there are no other turning points.

Corollary 7.13. If the growth number satisfies $s > 1$, then $s \leq \sqrt{2}$ if and only if $f^k(0) < 0$ for all odd k , or if and only if the

kneading determinant of f takes the form $(1-t)(1+t^2+t^4+t^6+\dots)$.

Proof. With F as above, let y_0 denote the fixed point $-1/(1+s)$.

If $s \leq \sqrt{2}$, then it is not difficult to check that

$$F(0) < F^3(0) \leq y_0 < 0 < F^2(0).$$

Hence F maps each of the two intervals $[f(0), y_0]$ and $[y_0, F^2(0)]$ into the other. Therefore $F^k(0) \leq y_0 < 0$, and hence $f^k(0) < 0$, for all odd k . Using 4.5, this implies that the kneading determinant has the required form. Conversely, if $D(t)$ has this form, then the estimate $s \leq \sqrt{2}$ follows easily from 6.3. ■

To conclude this section, we provide a prelude to the study of periodic points. According to Stefan, any map of the interval which has a periodic point of odd period $p \geq 3$ must satisfy $s > \sqrt{2}$. We will prove a converse statement for maps with just one turning point, which we may assume is a minimum.

Corollary 7.14. If f has just one turning point, then $s > \sqrt{2}$ if and only if f admits a periodic point of odd period $p \geq 3$.

Proof. If $s > \sqrt{2}$, then the preceding corollary implies that $f^p(0) \geq 0$ for some odd p . If x_0 is the unique negative fixed point of f , note that $f^p(x) < x_0 < x$ for x just to the right of x_0 . Hence, by the intermediate value theorem, the difference $f^p(x) - x$ must vanish at some point \bar{x} of the interval $(x_0, 0]$. Evidently the period of \bar{x} is odd, and this period must be at least 3 since $f(\bar{x}) \leq x_0 < \bar{x}$.

Conversely, if $s \leq \sqrt{2}$, then according to Stefan every periodic point must have a period which is either even or equal to 1. (In the case $1 < s \leq \sqrt{2}$, this could easily be proved by an argument similar to the proof of 7.13. However, for the case $s = 1$, this method would break down.) ■

8. Periodic points: a review of the classics

This section will begin the study of periodic points by giving an outline of a number of important results from the literature.

First, some general definitions and notations. Given a function $f : X \rightarrow X$ from an arbitrary set to itself, a point x is periodic if $f^p(x) = x$ for some integer $p \geq 1$. The smallest such p is the period of x , and the finite set $o = \{x, f(x), \dots, f^{p-1}(x)\}$ is a periodic orbit of period $p = p(o)$.

Suppose now that each iterate f^k has only finitely many fixed points. Following Artin and Mazur, the zeta function of f is defined to be the formal power series

$$\zeta(t) = \exp \sum_{k \geq 1} \tilde{n}(f^k) t^k / k,$$

where $\tilde{n}(f^k)$ is the number of fixed points of f^k and where $\exp a = 1 + a + a^2/2! + \dots$. (The idea of this formalism goes back to Weil. For a survey of applications, see Williams.)

In order to understand this definition, first consider the special case where X consists of a single periodic orbit of period p . Then evidently

$$\zeta(t) = \exp(t^p + t^{2p}/2 + \dots) = \exp(-\log(1-t^p))$$

is equal to

$$(1-t^p)^{-1} = 1 + t^p + t^{2p} + \dots$$

In the general case, expressing the set of periodic points of f as a disjoint union of finite orbits o with periods $p(o)$, it follows that

$$\zeta(t) = \prod_o (1 + t^{p(o)} + t^{2p(o)} + \dots).$$

Thus the zeta function of f , if defined at all, is always a formal power series with nonnegative integer coefficients.

In practice, it is often more convenient to work with the recip-

rocal series

$$\zeta(t)^{-1} = \prod_0 (1 - t^{p(o)}).$$

If the set of all periodic points is finite with d elements, note that $\zeta(t)^{-1}$ is a polynomial of degree d . Another alternate expression, which may be more convenient, is the logarithmic derivative multiplied by t . The identities

$$t\zeta'(t)/\zeta(t) = \sum_{k \geq 1} n(f^k) t^k = \sum_0 p(o) (t^{p(o)} + t^{2p(o)} + \dots)$$

are easily verified.

If X is a smooth compact manifold, then Artin and Mazur show that the space of smooth maps from X to itself admits a dense subset consisting of maps whose associated zeta function has a positive radius of convergence. In other words, for maps f in this subset, the numbers $n(f^k)$ grow at most exponentially with k . This statement applies also to maps of the unit interval

A most remarkable theorem concerning maps from the unit interval to itself was proved by A.N. Šarkovskii, in a paper which was published in 1964 but was little known in the west until much later. Define the Šarkovskii ordering of the positive integers by

$$3 < 5 < 7 < \dots < 6 < 10 < 14 < \dots < 12 < 20 < \dots \quad \dots < 4 < 2 < 1,$$

taking first the odd integers other than one in their natural order, then their doubles, then their quadruples, ..., and finally the powers of two in descending order.

Šarkovskii's Theorem. If a continuous map f from an open or closed interval to itself has a periodic point of period p , and if $p < q$ in the ordering described above, then f must have a periodic point of period q also.

For example, the function $f(x) = (x^2 - 7)/2$, since it possesses a

periodic point of period 3, must have periodic points of all possible periods (compare Li and Yorke).

For proofs of Šarkovskii's Theorem, see for example Targonski or Block, Guckenheimer, Misiurewicz and Young, as well as Štefan, Straffin, or Collet and Eckmann.

This theorem is related to the growth parameter s of §1, or to the topological entropy $\log s$, as follows. If f has a periodic point of some period p which is not a power of 2, then according to Bowen and Franks the topological entropy is strictly positive, hence $s > 1$. (More precisely, they show that $s > 2^{1/p}$.) For a converse result, and further discussion, see 9.6 below.

In studying periodic points of a map f , it is often important to single out these periodic orbits $o = \{x_0, f(x_0), \dots, f^{p-1}(x_0)\}$ which are stable, in the sense that for any point x sufficiently close to o the successive images $f^k(x)$ converge to o , uniformly in x . If f is differentiable, then clearly f is stable whenever the inequality

$$-1 < df^p(x)/dx < 1$$

holds at x_0 , and hence at every point of the orbit o . On the other hand, if the absolute value of the derivative of f^p at x_0 is strictly greater than one, then o is unstable; that is, there exists a neighborhood U of o and points x_i arbitrarily close to x_0 so that some forward image $f^k(x_i)$ does not belong to U .

If the derivative of f^p at x_0 is precisely equal to ± 1 , then the orbit through x_0 may be stable or unstable or may exhibit some intermediate behavior. For example, there may be stability as x approaches x_0 from one side but instability for the other side. As illustrative examples, the four functions $x - x^3$, $x + x^3$, $x + x^2$, and $-x/(2x+1)$ all exhibit different types of behavior at the origin.

In practice, it seems that "most" periodic orbits are unstable.

However, if f possesses stable periodic orbits, then these orbits tend to dominate the qualitative behavior of the successive iterates of f .

In the case of a polynomial function, a basic result was proved some 65 years ago. Adapting the hypotheses to our situation, it can be stated as follows.

Theorem of Julia and Fatou. If $f : I \rightarrow I$ is a polynomial function of degree $n \geq 2$, then f has at most $n - 1$ periodic orbits which are stable or nearly stable in the sense that the associated derivative satisfies

$$-1 \leq df^P(x)/dx \leq 1.$$

In fact, any such orbit can be obtained as the limit of the forward images $f^k(z_0)$ starting at some critical point of f in the complex plane, that is, at some complex number z_0 at which the derivative $f'(z_0)$ vanishes.

As an immediate corollary, a quadratic function has at most one stable periodic orbit, which can always be located by starting at the critical point and iterating the function.

A sketch of the proof is given at the end of this section. This theorem is proved with complex variable methods. If the derivative f' has non-real roots, then we definitely need to look at nonreal values of the variable, even though we are presumably interested only in real stable periodic orbits.

The Julia-Fatou proof applies also to rational functions of degree $n \geq 2$. In this case there are at most $2n - 2$ critical points in the extended complex plane $\mathbb{C} \cup \infty$, hence at most $2n - 2$ stable periodic orbits. The condition of rationality is essential. As an example, the transcendental function $f(x) = (e^x - 1)/2$ has a stable fixed point, but has no critical points at all.

An extremely simple and useful alternative condition, which

eliminates the need for complex variable computations has recently been introduced by D.J. Allwright and by D. Singer. Suppose first that f is smooth of class C^3 and satisfies the inequality

$$f'(x)f'''(x) < \frac{3}{2}(f''(x))^2$$

everywhere in the interval I . For example, this condition will always be satisfied if f is a polynomial whose derivative f' has distinct real roots. (This can be proved, making use of the fact that the second derivative of $\log|f'|$ is strictly negative wherever $f' \neq 0$.) If this inequality is satisfied, then clearly the following is also satisfied.

Hypothesis. The map f is piecewise monotone, C^3 -smooth, and the Schwarzian derivative

$$f'''/f' - \frac{3}{2}(f''/f')^2$$

is defined and strictly negative throughout the interior of each lap of f .

Theorem of Singer. Let f be such a piecewise monotone map with negative Schwarzian derivative inside each lap, and let c_0, \dots, c_ℓ be its local extrema. Then f has at most $\ell+1$ periodic orbits o which are stable or nearly stable in the sense that

$$-1 \leq df^{p(o)}(x)/dx \leq 1$$

for $x \in o$. Any such orbit can be obtained as the limit of the successive images $f^k(c_i)$ as $k \rightarrow \infty$, starting at one of the $\ell+1$ local extreme points c_i .

The proof can be outlined in three steps as follows.

Step 1. If f satisfies this Hypothesis, then each iterate $g = f^p$ does also. This is proved inductively by a straightforward computation.

Step 2. The derivative g' cannot have any local minimum within

an open interval where $g' > 0$. For the inequalities $g' > 0$, $g'' = 0$, $g''' \geq 0$ would contradict the hypothesis of negative Schwarzian derivative.

Step 3. If x_0 is a fixed point of g with $0 < g'(x_0) \leq 1$, lying in the interior of a lap $[c', c'']$ of g , then either the successive images $g^k[c', x_0]$ or the successive images $g^k[x_0, c'']$ converge uniformly to x_0 .

We must show either that $x < g(x) < x_0$ for all $x \in [c', x_0)$ or that $x_0 < g(x) < x$ for all $x \in (x_0, c'']$. But otherwise there would exist points x_1 between c' and x_0 and x_2 between x_0 and c'' with $g'(x_1) \geq 1$ and $g'(x_2) \geq 1$. Hence g' would attain a local minimum somewhere between x_1 and x_2 , contradicting Step 2. Further details will be left to the reader (compare Singer, as well as Allwright and Collet-Eckmann). ■

In the case $\ell = 2$, a similar argument proves the following statement. Suppose that $f(\dot{I}) \subset \dot{I}$, where the end fixed point is unstable, and suppose that the Hypothesis above is satisfied. Then there exists a stable or one-sided stable periodic orbit if and only if the kneading determinant of f is periodic.

To conclude this section, here is an outline proof of the Julia-Fatou Theorem. Let f be a rational function of degree ≥ 2 mapping the sphere $\mathbb{C} \cup \infty$ to itself, and let z_0 be a periodic point which we may assume finite. Let $g = f^p$, so that $g(z_0) = z_0$. First assume that the derivative $g'(z_0)$ has absolute value strictly less than 1. Then clearly there exists a small disk D_0 centered at z_0 so that the successive images

$$D_0 \supset g(D_0) \supset g^2(D_0) \supset \dots$$

converge to z_0 .

Let $g^{-n}(D_0)$ denote the full inverse image of D_0 under the iterate g^n . We assert that some $g^{-n}(D_0)$ contains a critical point

of g . For if this statement were false, then we could derive a contradiction as follows.

The function g^n maps the compact set $g^{-n}(D_0)$ onto D_0 by a mapping which is assumed to be locally biholomorphic. It follows that each component of $g^{-n}(D_0)$ is a covering space of D_0 , and hence maps biholomorphically onto D_0 . In particular, this is true for that component D_n of $g^{-n}(D_0)$ which contains D_0 . In this way we construct topological disks $D_0 \subset D_1 \subset D_2 \subset \dots$, together with a single-valued branch $D_0 \xrightarrow{\cong} D_n$ of the many-valued function g^{-n} . These disks D_n cannot exhaust the sphere since g has degree ≥ 2 , and hence is many-to-one. Therefore there must exist an open set disjoint from all of the D_n . After a change of coordinate we may assume that this open set contains the infinite point. In other words, we may assume that these single-valued holomorphic functions $g^{-n} : D_0 \xrightarrow{\cong} D_n$ are uniformly bounded. By a classical theorem, it follows that there exists a subsequence $\{g^{-n_i}\}$ which converges on the interior of D_0 , uniformly on each compact interior subset, to a holomorphic limit (see for example Ahlfors, p. 216). It follows that the derivative $dg^{-n_i}(z)/dz$ must also converge on the interior of D_0 as $i \rightarrow \infty$. But this is impossible since the derivative of g^{-n} at z_0 tends to infinity with n .

Now consider the more difficult case where $g'(z_0) = 1$. After a careful change of coordinate, we may assume that $z_0 = 0$ and that

$$g(z) = z - z^{k+1} + (\text{higher terms}),$$

with $k \geq 1$. The local behavior of the iterates of g can be indicated roughly by the sketch below.

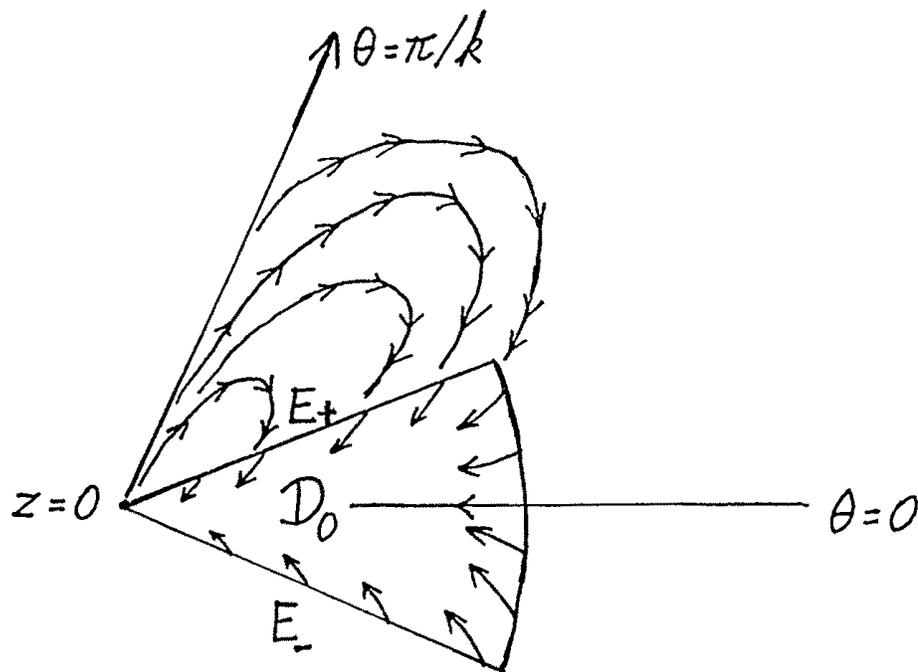


Figure 8.1.

Each arrow leads from some point $z = re^{i\theta}$ to $g(z) \approx z - z^{k+1}$.

In particular, if D_0 is a small "pie slice" consisting of all $z = re^{i\theta}$ with $r \leq \varepsilon$ and $|\theta| \leq \pi/3k$, then it can be shown that the successive images

$$D_0 \supset g(D_0) \supset g^2(D_0) \supset \dots$$

converge uniformly to zero. As in the argument above, if the full inverse images $g^{-n}(D_0)$ do not contain any critical point of g , then, after a change of coordinate, we can construct a single-valued branch

$$g^{-n} : D_0 \xrightarrow{\cong} D_n \supset D_0,$$

and a subsequence $\{g^{-n_i}\}$ which converges uniformly on every compact interior subset of D_0 to a finite holomorphic limit.

Let E_+ and E_- be the two straight edges of the region D_0 . A

novel feature of the new situation is that the successive images $g^{-n}(E_+)$ converge uniformly to zero (compare the figure). Therefore g^{-n_i} converges to zero on a set $g(E_+)$ containing an entire curve of interior points, so it follows that the holomorphic limit of the sequence $\{g^{-n_i}\}$ is identically zero throughout the interior of D_0 .

Combining these two statements, we see that the image under g^{-n_i} of the entire boundary of the region $g(D_0)$ converges uniformly to zero. Hence $g^{-n_i}(g(D_0))$ converges to zero. But this is impossible, since the sets $g^{-n}(g(D_0))$ increase with n . ■

9. Computation of zeta functions.

This section will state two different forms of our main theorem, relating the Artin-Mazur zeta function of §8 to the kneading determinant of §4. It will also derive several applications.

First consider a piecewise monotone map $f : I \rightarrow I$ satisfying the following.

Instability Hypothesis. All but finitely many of the periodic points of f are unstable.

This is needed to avoid badly behaved examples such as $f(x) = 1 - x$. Recall from §8 that this hypothesis is satisfied whenever f is a polynomial or rational function of degree ≥ 2 , or whenever f has negative Schwarzian derivative in the interior of each lap.

It will be convenient to use the phrase cyclotomic polynomial for any product of finitely many factors of the form $1 - t^n$.

Theorem 9.1. If f satisfies this Instability Hypothesis, then the reciprocal $\zeta(t)^{-1}$ of its zeta function is equal to its kneading determinant $D(t)$ multiplied by a cyclotomic polynomial.

The proof in §10.4 will provide an explicit recipe for computing this cyclotomic polynomial in terms of stable periodic orbits.

Here is an alternative version which does not require any special hypothesis, other than piecewise monotonicity. Let us say that two fixed points $x < y$ of some iterate f^k are monotonely equivalent if f^k maps the entire closed interval $[x, y]$ homeomorphically onto itself. It is easy to check that this is an equivalence relation. Let $\hat{n}(f^k)$ be the number of monotone equivalence classes of fixed points of f^k . These numbers are finite with $\hat{n}(f^k) \leq \ell(f^k)$, since each lap of f^k clearly contains at most one monotone equivalence class of fixed points. The associated formal power series

$$\hat{\zeta}(t) = \exp \sum \hat{n}(f^k) t^k / k$$

will be called the reduced zeta function of f .

Theorem 9.2. The reciprocal $\hat{\zeta}(t)^{-1}$ of this reduced zeta function is equal to $D(t)$ multiplied by a cyclotomic polynomial.

The proof in §10.6 will provide an explicit way of computing this cyclotomic polynomial in terms of the kneading matrix.

The reduced zeta function satisfies a product formula, which can be written as

$$(9.3) \quad \hat{\zeta}(t) = \prod_{\hat{o}} (1 - t^{p(\hat{o})})^{-1},$$

where \hat{o} varies over all monotone equivalence classes of periodic orbits. Here two periodic orbits o' and o'' are said to be monotonely equivalent if they contain representative points x' and x'' which are monotonely equivalent under some iterate of f .

Caution: It may happen that the period of one of these two monotonely equivalent orbits is twice the period of the other. (This happens, for example, for the function $f(x) = 1 - x$.) We define the period $p(\hat{o})$ to be the smaller of the periods $p(o)$ for orbits o in the monotone equivalence class \hat{o} . With this understanding, the proof of the product formula 9.3 proceeds just as in §8.

The rest of this section will be devoted to applications of Theorem 9.1 or 9.2. To fix our ideas, we will concentrate on the reduced zeta function $\hat{\zeta}(t)$. However, similar arguments clearly apply to the original zeta function $\zeta(t)$, whenever the Instability Hypothesis is satisfied.

It follows immediately from 9.2 that the reciprocal $\hat{\zeta}(t)^{-1}$ extends as an analytic function of the complex variable t throughout the disk $|t| < 1$, and has zeros precisely at the zeros of $D(t)$.

Corollary 9.4. The radius of convergence of the series $\hat{\zeta}(t)$ is precisely $1/s$. In fact, if $s > 1$ then $\hat{\zeta}(t)$ has a pole at $t = 1/s$.

Proof. This follows from 6.3. ■

Using the standard formula for the radius of convergence of a power series, this implies that

$$\limsup_{k \rightarrow \infty} \hat{n}(f^k)^{1/k} = s.$$

In other words, at least some of the numbers

$$\hat{n}(f^k) \leq \ell(f^k)$$

grow almost as fast as the lap numbers $\ell(f^k)$ as $k \rightarrow \infty$.

Here is a much more precise estimate for $\hat{n}(f^k)$. Note that the analytic function $D(t)$ can have only finitely many zeros within any disk of radius less than 1. (For examples illustrating the behavior of $D(t)$ near the boundary of the unit disk, see §14.) Choose some number $\varepsilon > 0$.

Corollary 9.5. Suppose that $D(t)$ has zeros at the points a_1, \dots, a_q (listing an m -fold zero m times) within the disk $|t| < (1+\varepsilon)^{-1}$.

Then $\hat{n}(f^k) = a_1^{-k} + \dots + a_q^{-k} + o((1+\varepsilon)^k)$.

Here the symbol $o((1+\varepsilon)^k)$ stands for an error term which has absolute value less than $(1+\varepsilon)^k$ times some constant independent of k .

Note: If the Instability Hypothesis of 9.1 is satisfied, then we can substitute $\underline{n}(f^k)$ in place of $\hat{n}(f^k)$ in this estimate.

As an example, for the function $f(x) = (x^2 - 7)/2$, with $s = (1 + \sqrt{5})/2$, the determinant $D(t) = (1 - t - t^2)/(1 - t^3)$ has just one simple zero in the unit disk, at the point $1/s$ (compare 6.3 and the discussion following 5.9). Hence in this case

$$\underline{n}(f^k) = ((1 + \sqrt{5})/2)^k + o((1 + \varepsilon)^k)$$

for any $\varepsilon > 0$. It follows easily from this estimate that f has points of period p for all sufficiently large p . (Of course, Šarkovskii's theorem implies the sharper statement that f has points of all periods.)

On the other hand, for the function $f(x) = (x^2 - 5.9)/2$ computation shows that

$$D(t) = (1 - t)(1 - t^2 - t^4)/(1 - t^6),$$

with simple zeros at two points $1/s$ and $-1/s$ in the unit disk, where now $s = ((1 + \sqrt{5})/2)^{1/2}$ (compare 14.4). Applying 9.5, it follows that

$$\underline{n}(f^k) = 2s^k + o((1 + \varepsilon)^k)$$

for k even and for any $\varepsilon > 0$, but

$$\underline{n}(f^k) = o((1 + \varepsilon)^k)$$

for k odd. In fact, it follows from 7.13 and 7.14 that f has no points at all of odd period, other than the two fixed points.

Proof of 9.5. If $D(t)$ has a zero of order m at a , then $\hat{\zeta}(t)$ has a pole of order m at a , and it follows easily that the logarithmic derivative of $\hat{\zeta}(t)$ has a simple pole of the form

$$\hat{\zeta}'(t)/\hat{\zeta}(t) = m/(a - t) + (\text{holomorphic function})$$

near a . Hence the function

$$t\hat{\zeta}'(t)/\hat{\zeta}(t) = \sum \hat{n}(f^k)t^k$$

is equal to

$$tm/(a-t) = m(t/a + t^2/a^2 + \dots)$$

plus a function which is analytic near t . Subtracting off these correction terms for all zeros within a slightly larger circle, we obtain a power series which must converge throughout the circle $|t| \leq (1+\varepsilon)^{-1}$. The conclusion follows easily. ■

Next let us describe a relationship between the set of periods p of a mapping and its growth number s . According to Bowen and Franks, if there exists an orbit whose period p is not a power of 2, then

$$s > \sqrt[p]{2} > 1.$$

A somewhat sharper version of this inequality has been given by Štefan. See also Jonker and Rand, and Targonski. Conversely, using 9.5, we can prove the following.

Corollary 9.6. If $s > 1$, then f possesses a periodic orbit whose period is not a power of 2.

Proof. Choose $1+\varepsilon$ just a little smaller than s , so that all the zeros of $D(t)$ in the disk $|t| < (1+\varepsilon)^{-1}$ lie on the circle of radius s^{-1} . Let a_1, \dots, a_q be these zeros. (As usual, a zero of multiplicity m is to be listed m times.) Then

$$\hat{n}(f^k) = a_1^{-k} + \dots + a_q^{-k} + o((1+\varepsilon)^k)$$

by 9.5. Since the q -tuple (sa_1, \dots, sa_q) lies on the q -dimensional torus, which is a compact topological group, there exist arbitrarily large integers r satisfying the inequality

$$(*) \quad |(sa_j)^{-r} - 1| < 1/10$$

for each j . It follows that the unit complex number $(sa_j)^{-3r}$ has real part greater than $1/2$, hence $\operatorname{Re}(a_j^{-3r}) > \frac{1}{2}s^{3r}$ for each j ,

and

$$\hat{n}(f^{3r}) > \frac{1}{2}qs^{3r} + o((1+\varepsilon)^{3r});$$

although

$$\hat{n}(f^r) \leq qs^r + o((1+\varepsilon)^r).$$

Subtracting these two inequalities, it follows that

$$\hat{n}(f^{3r}) > \hat{n}(f^r)$$

whenever r is large enough and satisfies (*). Hence f has a monotone-equivalence class of periodic points whose period divides $3r$ but not r . Evidently such a period cannot be a power of 2. ■

Note: This proof is quite nonconstructive. The following example shows that one cannot be too explicit as to which periods will occur. Start with a mapping f with growth number $s > 1$ such that all periods are divisible by some high power 2^N (compare 14.4 below). Then the iterate f^{2m+1} has growth number s^{2m+1} which (for suitable choice of m) can be arbitrarily large. Yet every period of f^{2m+1} is divisible by 2^N .

Here is a different but closely related consequence of 9.2.

Corollary 9.7. The mapping f has only finitely many distinct points, or in other words only finitely many periodic points up to monotone-equivalence, if and only if the sequence of lap numbers $\ell(f^k)$ is bounded by a polynomial function of k .

If these conditions are satisfied, then every period must be a power of 2 by Šarkovskii's theorem. Evidently the growth number s must be 1 (compare Block).

Proof. It will be convenient to say that a formal power series $A(t) = \sum a_k t^k$ satisfies a polynomial growth condition of degree d if its coefficients satisfy

$$|a_k| \leq \varphi(k)$$

for all k , where φ is some polynomial of degree d . If $A(t)$ and $B(t)$ satisfy such conditions of degrees d and d' , then evidently the product $A(t)B(t)$ satisfies a growth condition of degree $d+d'+1$. Given integers $k_0, \dots, k_d \geq 1$, the product

$$(1-t^{k_0})^{-1} \dots (1-t^{k_d})^{-1}$$

provides an example of a power series which satisfies a growth condition of degree d , but not of degree $d-1$.

Now consider the following possible conditions on f .

(1) The associated power series $\hat{\zeta}(t) = \prod_{\hat{o}} (1-t^{p(\hat{o})})^{-1}$ satisfies

a polynomial growth condition;

(2) the power series $D(t)^{-1}$ does;

(3) the coefficients $M_{jk}(t)$ of the inverse kneading matrix (§5.6) all satisfy such conditions;

(4) the series $\gamma_i(J, t)$ of §5.4 all do;

(5) the series $L(t) = \sum \ell(f^k) t^{k-1}$ satisfies such a condition.

These conditions are all equivalent to each other. In fact (1) \Rightarrow (2) by 9.2; (2) \Rightarrow (3) since the entries of the kneading matrix $N_{ij}(t)$ all satisfy a polynomial growth condition; (3) \Rightarrow (4) by 5.7; (4) \Rightarrow (5) by 5.8; and (5) \Rightarrow (1) by the inequality $\hat{n}(f^k) \leq \ell(f^k)$. But, using the remarks above, we see that (1) is satisfied if and only if the product formula for $\hat{\zeta}(t)$ has only finitely many factors, or in other words if and only if there are only finitely many monotone equivalence classes \hat{o} of periodic orbits. This completes the proof. ■

10. Periodic points of negative type

This section will state a third and more precise form of our main theorem relating zeta functions and kneading determinant, and then derive Theorems 9.1 and 9.2 from it.

Definition. Let x be a fixed point lying in the interior of some lap of the piecewise monotone function f . We will say that s is of positive type or of negative type according as f is increasing or decreasing throughout a neighborhood of x . If x is an endpoint of some lap of f , then it will be called a fixed point of critical type.

Similarly, a periodic orbit $o = \{x, f(x), \dots, f^{p-1}(x)\}$ will be called of positive, negative, or critical type according as x is of positive, negative, or critical type as a fixed point of f^p .

Let $N(f) = 2\hat{n}_-(f) - 1$ where \hat{n}_- is the number of fixed points of negative type. Evidently $N(f) \leq \hat{n}(f)$, since between any two fixed points of negative type there must be at least one other fixed point. Hence the numbers $N(f^k) \leq \hat{n}(f^k) \leq \ell(f^k)$ are all finite.

Definition. The formal power series

$$Z(t) = \exp \sum N(f^k) t^k / k$$

will be called the modified zeta function of f .

Our main theorem can now be stated as follows.

Theorem 10.1. For any piecewise monotone function f we have $Z(t) = D(t)^{-1}$.

The proof, in §11, will be quite indirect. (A more conceptual proof has recently been given in an unpublished manuscript by J.B. Rogers in Melbourne.) For the remainder of this section, we simply assume 10.1 without proof.

As is usual with zeta functions, we can also express $Z(t)$ as a product. If there are no periodic orbits at all of negative type, then $N(f) = N(f^2) = \dots = -1$, so that

$$Z(t) = \exp(-t - t^2/2 - \dots) = 1 - t.$$

Evidently each periodic orbit o of negative type with period $p = p(o)$ contributes a factor of

$$\exp 2(t^p + t^{3p}/3 + t^{5p}/5 + \dots) = (1+t^p)/(1-t^p)$$

to the series $Z(t)$. Thus we obtain the following.

Lemma 10.2. The modified zeta function $Z(t)$ is equal to

$$(1-t)\prod(1+t^{p(o)})/(1-t^{p(o)}) = (1-t)\prod(1+2t^{p(o)}+2t^{2p(o)}+\dots),$$

taking the product over all periodic orbits o of negative type.

One immediate consequence is that all of the coefficients of the formal power series $D(t)^{-1}(1-t)^{-1} = Z(t)/(1-t)$ are nonnegative.

This would surely be difficult to check directly from the definition of the kneading determinant.

Next let us show how theorem 9.1 follows from 10.1. For this argument only, we make a special definition. Let us say that an isolated fixed point x_0 of f , not of negative type, is formally left stable if $f(x) > x$ throughout some interval $(x_0 - \epsilon, x_0)$, or also if x_0 is the leftmost point of the entire interval I . Similarly x_0 is formally right stable if $f(x) < x$ throughout an interval to the right of x_0 , or if x_0 is the rightmost point of I . Fixed points of negative type must definitely be excluded in this definition.

Remark. For fixed points of positive type, this corresponds precisely to the intuitive notion of left or right stability. If f is smooth of class C^1 , then the same is true for fixed points of critical type. (In fact, every turning point c_i with $f(c_i) = c_i$ is automatically both left and right stable in the C^1 case.) However, the non-differentiable function $f(x) = 2|x|$ provides an example of a critical fixed point which is formally left stable, but is not left stable in any ordinary sense.

Lemma 10.3. If f is any piecewise monotone function with isolated fixed points, then the difference $n(f) - N(f)$ is equal to the number

of formally left stable fixed points plus the number of formally right stable fixed points.

The proof is completely elementary. Let $x_1 < \dots < x_n$ be the fixed points of f . Then x_1 is either formally left stable or of negative type and x_n is either formally right stable or of negative type. For each intermediate interval (x_{i-1}, x_i) on which $f(x) > x$, we know that x_i is either formally left stable or of negative type. Similarly, if $f(x) < x$ in this interval then x_{i-1} is either formally right stable or of negative type. Thus the sum

number of formally left stable fixed points
+ number of formally right stable fixed points
+ twice number of fixed points of negative type

is equal to $n+1$, which proves the lemma. ■

Now suppose that each iterate f^k has isolated fixed points, so that the zeta function $\zeta(t)$ of f is defined.

Corollary 10.4. The ratio $Z(t)/\zeta(t) = D(t)^{-1}\zeta(t)^{-1}$ is equal to a product of cyclotomic polynomials as follows. Each periodic orbit of period p contributes a factor of

$(1-t^p)^2$ if it is stable of positive or critical type
 $1-t^{2p}$ if it is stable of negative type, and
 $1-t^p$ if it is unstable of critical type, or if it is stable on one side but unstable on the other.

There is no contribution at all from an unstable orbit of positive or negative type.

Note that a one-sided stable periodic orbit is necessarily of positive type. Thus the appropriate factors can be tabulated as follows.

	stable	one-sided stable	unstable
positive type	$(1-t^p)^2$	$1-t^p$	1
critical type	$(1-t^p)^2$		$1-t^p$
negative type	$1-t^{2p}$		1

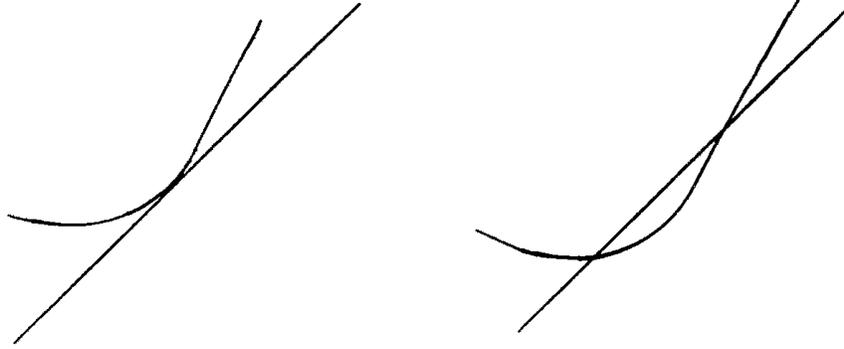
If f is differentiable, note that an unstable orbit of critical type can only consist of one or both endpoints of the interval I , with p equal to 1 or 2. However, for nondifferentiable functions, like $2|x|$ near the origin, there can definitely be other unstable orbits of critical type.

Evidently theorem 9.1 is an immediate consequence of this corollary. For if all but finitely many periodic orbits are unstable, there can be only finitely many correction factors.

Here is an explicit example. As described in 2.2 and 4.6, the map $f(x) = (x^2 - 7)/2$ from the interval $|x| \leq 1 + \sqrt{8}$ to itself (or equivalently the map $g(y) = (1 + \sqrt{8})y(1 - y)$ from the unit interval to itself) has a one-sided stable orbit of period 3, and also an unstable fixed point $1 + \sqrt{8}$ of critical type. By the Julia-Fatou Theorem or by Singer's Theorem, there are no other stable, one-sided stable, or critical periodic orbits. Since $D(t) = (1 - t - t^2)/(1 - t^3)$, it follows that

$$\zeta(t)^{-1} = (1 - t)(1 - t^3)D(t) = (1 - t)(1 - t - t^2)$$

is a polynomial of degree 3. On the other hand, for the map $f(x) = (x^2 - (7 + \epsilon))/2$ on a corresponding maximal interval, where ϵ is small and positive, the kneading determinant is unchanged but the one-sided stable orbit of period 3 splits into one stable and one unstable orbit of period 3. Hence $\zeta(t)^{-1} = (1 - t)(1 - t^3)(1 - t - t^2)$. This change in f corresponds to a perturbation in which the graph of f^3 passes through the diagonal as sketched schematically in Figure 10.5 (compare Jonker and Rand).



one-sided stable

stable + unstable

Figure 10.5.

Proof of 10.4 (assuming 10.1). Let $\sigma(f^k)$ be the number of fixed points of f^k which are formally left stable plus the number which are formally right stable. Then

$$Z(t)/\zeta(t) = \exp \sum (-\sigma(f^k)t^k/k)$$

by 10.3. Separating out the contributions from the various periodic orbits of f , we can obtain a corresponding product formula for $Z(t)/\zeta(t)$.

Consider for example a one-sided stable orbit of period p . Evidently the points of this orbit contribute precisely p to the number $\sigma(f^k)$ of formal left or right stabilities of k is a multiple of p , and contribute nothing at all otherwise. Therefore such an orbit contributes a factor of

$$\exp(-t^p - t^{2p}/2 - \dots) = 1 - t^p$$

to the ratio $Z(t)/\zeta(t)$. The contributions from other types of periodic orbits can be computed similarly. Details will be left to the reader. ■

The proof of 9.2 (assuming 10.1) is completely analogous. We will say that a monotone equivalence class of fixed points of f , with leftmost point x_- , is formally left stable if x_- is not of negative type yet $f(x) > x$ for all x just to the left of x_- , or also if x_- is the leftmost point of the entire interval I . Formal right stability is defined similarly. In analogy with 10.3, the difference $\hat{p}(f) - N(f)$ is equal to the number of formally left stable monotone equivalence classes of fixed points plus the number of formally right stable ones. Expressing $Z(t)/\hat{\zeta}(t)$ as a product, in analogy with 10.4, the proof of 9.2 can easily be computed. ■

In order to obtain a more explicit computation of this product, in terms of kneading invariants and the behavior of f on the boundary of the interval I , we proceed as follows.

Let $I_j = [c_{j-1}, c_j]$ be a lap of f . Recall that the invariant coordinate $\theta_*(c_{j-1}^+)$ of §4 is a certain infinite sequence of formal symbols $\pm I_r$.

Assertion. This infinite sequence $\theta_*(c_{j-1}^+)$ is strictly periodic of period 1, being equal to (I_j, I_j, \dots) , if and only if f is increasing on I_j and this lap contains a (necessarily unique) monotone equivalence class of fixed points of f which is formally left stable. Similarly the sequence $\theta_*(c_{j-1}^+)$ has period dividing p if and only if the lap of f^p which is bounded on the left by c_{j-1} has the analogous property. Similar statements hold for $\theta_*(c_j^-)$ and formal right stability.

The proof is straightforward. However, in order to apply this assertion, we must take account of the fact that two different periodic invariant coordinates $\theta_*(c_j^\pm)$ may correspond to the same monotone equivalence class of periodic orbits.

Definition. We will say that two different periodic infinite sequences θ_* and θ'_* are cyclically equivalent if $\theta'_i = \theta_{i+k}$ for

some constant $k \geq 0$ and for all $i \geq 0$.

The precise formula can now be stated as follows.

Theorem 10.6. The ratio $Z(t)/\hat{\zeta}(t) = D(t)^{-1}\hat{\zeta}(t)^{-1}$ can be expressed as a product of at most $\ell+1$ factors of the form $1-t^p$ as follows. There is one factor of $1-t^p$ corresponding to each cyclic equivalence class of sequences of period p among the 2ℓ infinite sequences $\theta_*(c_{j-1}^+)$ and $-\theta_*(c_j^-)$ for $1 \leq j \leq \ell$.

The proof is not difficult, and will be omitted. ■

The following special case is somewhat easier to state.

Corollary 10.7. Suppose that f has just one turning point (so that $\ell = 2$), and suppose that $f(\dot{I}) \subset \dot{I}$. Then

$$\hat{\zeta}(t)^{-1} = (1-t)D(t),$$

if the sequence of coefficients of the power series $D(t)$ is not strictly periodic, while

$$\hat{\zeta}(t)^{-1} = (1-t)(1-t^p)D(t)$$

if the sequence of coefficients is periodic of period p .

This follows immediately, since the fixed endpoint contributes a factor of $1-t$, and it is easy to check that one of the two invariant coordinates $\pm\theta_*(c_{1\pm})$ is periodic of period p if and only if the coefficients of D are periodic of period p . ■

In the periodic case, it is interesting to note that $\hat{\zeta}(t)^{-1}$ is a polynomial of degree p .

In both 10.6 and 10.7 it is important to emphasize that the sequences must be precisely periodic. Eventual periodicity does not count. As an example, for the function $f(x) = (x^2-8)/2$ (or the equivalent function $g(y) = 4y(1-y)$) the coefficients of the kneading determinant

$$D(t) = (1-2t)(1-t) = 1-t-t^2-t^3-\dots$$

are not strictly periodic, hence there is no extra factor, and

$$\hat{\zeta}(t)^{-1} = (1-t)D(t) = 1 - 2t.$$

In fact

$$\hat{\zeta}(t) = \zeta(t) = (1-2t)^{-1} = 1 + 2t + 4t^2 + \dots,$$

and

$$t\zeta'(t)/\zeta(t) = \sum_{k \geq 1} n(f^k)t^k = 2t + 4t^2 + 8t^3 + \dots$$

for this example.

11. Smooth deformations of f

This section will complete the arguments of the previous two sections by proving the equality $D(t) = Z(t)^{-1}$ of Theorem 10.1. The idea of the proof is first to check this equation for some particularly simple mapping f with given lap number ℓ and then to check that both sides change in precisely the same way as we deform f . We will first concentrate on changes in the kneading determinant as we change f .

Let $C^0(I, \mathbb{R})$ denote the Banach space consisting of all continuous functions f from the interval I to the real numbers \mathbb{R} , with norm

$$\|f\|_0 = \text{Max}_{x \in I} |f(x)|.$$

We will also make use of the Banach space $C^r(I, \mathbb{R})$ consisting of functions f which are smooth of class C^r , with norm

$$\|f\|_r = \text{Max}(\|f\|_0, \|df/dx\|_0, \dots, \|d^r f/dx^r\|_0).$$

We will be particularly interested in the closed convex subset $C^r(I, I)$ consisting of C^r -maps from I to itself. For our purposes, r will always be 0, 1, or 2.

Lemma 11.1. The map $(f, x) \mapsto (f, f(x))$ from $C^r(I, I) \times I$ to itself is

smooth of class C^r . Hence its n -fold iterate $(f, x) \mapsto (f, f^n(x))$ is also smooth of class C^r .

For a proof, the reader is referred to Abraham and Robbin, p. 25. ■

We must study piecewise-monotone mappings from I to itself. To simplify the discussion, we will only consider the case where the set $\{c_1, \dots, c_{\ell-1}\}$ of turning points is fixed. For example, if I is the interval $[0, \ell]$ we could take $c_1 = 1, \dots, c_{\ell-1} = \ell - 1$. We will assume that some such choice of turning points has been made. Also choose a sign $\varepsilon(I_1) = \pm 1$.

Definition. Let $P^r = P^r(I, c_1, \dots, c_{\ell-1}, \varepsilon(I_1))$ denote the convex subset of $C^r(I, I)$ consisting of piecewise-monotone maps f which are strictly monotone on each interval $I_j = [c_{j-1}, c_j]$, being monotone increasing or monotone decreasing according as $(-1)^{j+1}$ equals $\varepsilon(I_1)$ or $-\varepsilon(I_1)$.

We are interested in the continuity properties of the kneading increment $\nu_i(f)$ of §4 as f varies in P^r . Here is a first result. Choose some turning point c_i .

Lemma 11.2. If f belongs to the space P^0 of piecewise-monotone maps and satisfies the condition that none of the forward images $f(c_i), f^2(c_i), f^3(c_i), \dots$ is a turning point, then the correspondence

$$g \mapsto \nu_i(g)$$

from P^0 to the module $V[[t]]$ of formal power series is continuous at f .

In other words, for each N there exists $\varepsilon > 0$ so that

$$\nu_i(g) \equiv \nu_i(f) \pmod{t^N}$$

whenever g in P^0 satisfies $\|g - f\|_0 < \varepsilon$. The proof, using Lemma 11.1, is clear. ■

Next, let us look at the simplest case where a discontinuity occurs. Suppose that there exists one and only one integer $p \geq 1$ so that $f^p(c_i)$ is a turning point c_j , with $i \neq j$. Let N be a pre-assigned large integer. Suppose, to fix our ideas, that f^p has a local minimum at c_i .

Lemma 11.3. With these hypotheses, every $g \in P^0$ which is sufficiently close to f satisfies either

$$\nu_i(g) \equiv \nu_i(f) \pmod{t^N}$$

or

$$\nu_i(g) \equiv \nu_i(f) - 2t^p \nu_j(f) \pmod{t^N}$$

according as $f^p(c_i) \geq c_j$ or $f^p(c_i) < c_j$.

The proof is straightforward. ■

The case of a local maximum is completely analogous. The only difference is that the inequality symbols \geq and $>$ must be replaced by \leq and $<$ respectively.

The behavior of the kneading determinant $D(f)$ can now be described as follows.

Corollary 11.4. For $f \in P^0$ if each of the turning points $c_1, \dots, c_{\ell-1}$ satisfies the hypothesis of 11.2 or 11.3, then the correspondence

$$g \mapsto D(g)$$

from P^0 to $Z[[t]]$ is continuous at f .

Proof. This follows immediately, since the discontinuity of 11.3 corresponds to an elementary row operation in the kneading matrix $[N_{ij}]$, and hence does not change the determinant. ■

A more complicated discontinuity occurs when $f^p(c_i) = c_i$. To get a reasonable result in this case we will need to make use of first derivatives.

Lemma 11.5. Suppose that $f^P(c_i) = c_i$, where f is smooth of class C^1 . Then, for every g which is sufficiently close to f in the space P^1 , the real numbers

$$g^P(c_i) - c_i, g^{2P}(c_i) - c_i, g^{3P}(c_i) - c_i, \dots$$

are either all zero or all have the same sign. Furthermore, given $\varepsilon > 0$ there exists $\delta > 0$ so that $|g^m(c_i) - f^m(c_i)| < \varepsilon$ uniformly for all m whenever $\|g - f\|_1 < \delta$.

Proof. To simplify the notation, set $F = f^P$, $G = g^P$. Since the first derivative \dot{F} of F must vanish at c_i , we can choose $\varepsilon > 0$ small enough so that $|\dot{F}(x)| < 1/4$ for $|x - c_i| \leq \varepsilon$. Now choose g close enough to f so that $|\dot{G}(x)| < 1/2$ for $|x - c_i| \leq \varepsilon$, and so that $|G(c_i) - c_i| < \varepsilon/2$. If $G(c_i) = c_i$, the proof is completely straightforward. But if $G(c_i) \neq c_i$, then it follows by an easy induction that

$$|G^{m+1}(c_i) - G^m(c_i)| < |G(c_i) - c_i|/2^m$$

and hence that

$$|G^m(c_i) - G(c_i)| < |G(c_i) - c_i| < \varepsilon/2$$

for all $m \geq 1$. The conclusion then follows easily. ■

Now suppose that the forward images $f(c_i), f^2(c_i), \dots, f^{P-1}(c_i)$ are not turning points, but that $f^P(c_i) = c_i$. Then it is not difficult to see that the i -th kneading increment takes the form

$$\nu_i(f) = \eta(t)/(1-t^P)$$

where $\eta(t)$ is a polynomial of degree p . More precisely,

$$\eta(t) = (I_{i+1} - I_i) + 2tI_{\alpha_1} + \dots + 2t^{P-1}I_{\alpha_{P-1}} + t^P(I_{i+1} + I_i),$$

where the coefficient of t^m is positive or negative according as f^m has a local minimum or a local maximum at c_i , and where

$f^m(c_i) \in I_{\alpha_m}$. Suppose, to fix our ideas, that f^p has a local minimum at c_i .

Corollary 11.6. With these hypotheses, for every g which is sufficiently close to f in the space P^1 we have either

$$\nu_i(g) = \eta(t)/(1-t^p) \quad \text{or} \quad \nu_i(g) = \eta(t)/(1+t^p)$$

according as $g^p(c_i) \geq c_i$ or $g^p(c_i) < c_i$.

The proof, using 11.5, is not difficult and will be left to the reader. ■

If the other rows of the kneading matrix do not change, then it follows that the determinant $D(g,t)$ is equal to $D(f,t)$ in the first case and to $D(f,t)(1-t^p)/(1+t^p)$ in the second. If f^p has a local maximum at c_i , then the situation is completely analogous except that the inequalities are reversed.

Remark. The C^1 -topology is essential here. For example, for the map $f(x) = x^2$ on the interval $[-1,1]$ we can choose a smooth map g , with $\ell = 2$, uniformly close to f so that g maps some small interval $[-\varepsilon, \varepsilon]$ to itself by a quadratic map having any growth number at all in the interval $1 \leq s \leq 2$. Thus there are uncountably many different possibilities for the kneading determinant $D(g)$.

We will next show that a "generic" homotopy between generic elements of the space P^2 involves only the discontinuities which have already been described in 11.4 and 11.6. We will need to work in an open subset of a Banach space, in order to make use of the inverse function theorem. To achieve this, we introduce second derivatives.

Definition. A mapping $f \in C^2(I,I)$ will be called a Morse function if the first derivative $f'(x)$ and the second derivative $f''(x)$ do not vanish simultaneously, and if f' does not vanish at the two boundary points. Let $M \subset P^2$ denote the dense open subset consisting of Morse functions in P^2 which satisfy $f(I) \subset \text{interior}(I)$.

In other words, M consists of all maps $f \in C^2(I, \text{interior } I)$ such that f' vanishes only at the turning points $c_1, \dots, c_{\ell-1}$, and such that $f''(c_i) \neq 0$, this second derivative being alternately positive or negative according as $(-1)^i \varepsilon(I_1)$ is positive or negative.

It is easy to check that M is a convex open set in the Banach space B consisting of all elements $f \in C^2(I, \mathbb{R})$ which satisfy the linear equations $f'(c_1) = \dots = f'(c_{\ell-1}) = 0$.

Define submanifolds $M_n(i, j) \subset M$ as follows. For any $n \geq 1$ and any turning points c_i and c_j , let $M_n(i, j)$ be the set of all maps $f \in M$ such that $f(c_i), f^2(c_i), \dots, f^{n-1}(c_i)$ are not turning points, but $f^n(c_i) = c_j$.

Lemma 11.7. Each $M_n(i, j)$ is a C^2 -smooth codimension 1 submanifold of M . Furthermore, each intersection of two distinct submanifolds $M_{n_1}(i_1, j_1)$ and $M_{n_2}(i_2, j_2)$ is transverse, and hence is a codimension 2 submanifold of M whenever it is non-vacuous.

By definition, a subset S of a Banach manifold M is a C^r -smooth codimension k submanifold if for every point s of S there exists a local chart mapping a neighborhood of s in the pair (M, S) by a C^r -diffeomorphism onto an open subset of some pair $(\mathbb{R}^k \times B', 0 \times B')$ where B' is a Banach space.

Proof of 11.7. We will make use of Lemma 11.1 together with the inverse function theorem. Fixing some point $x_0 \in I$, consider the C^2 -map $\varphi : M \rightarrow I$ defined by $\varphi(f) = f^n(x_0)$. If we set $x_k = f^k(x_0)$, then computation shows that the derivative $D\varphi(f)$ is the linear map

$$\Delta f \mapsto \sum_{i=1}^n \Delta f(x_{i-1}) f'(x_i) f'(x_{i+1}) \dots f'(x_{n-1})$$

from the ambient Banach space B to \mathbb{R} .

Now let us specialize to the case $x_0 = c_i$. If f belongs to the set $M_n(i, j) \subset M$, note that $D\varphi(f) \neq 0$. In fact, choosing

$\Delta f \in B$ so that $\Delta f(x)$ vanishes except for x in a small neighborhood of c_1 , it follows easily that $D\varphi(f)(\Delta f) \neq 0$. Hence, using the inverse function theorem for Banach spaces, it follows that $M_n(i,j)$ is a codimension 1 submanifold (compare Lang, p. 15, Corollary 25).

The proof that intersections are transverse is completely analogous. ■

Remark. The manifold $M_n(i,j)$ is not a closed subset of M when $n \geq 2$. However any limit point must belong to some 2-fold intersection of the form $M_p(i,h) \cap M_q(h,k)$ with $p+q \leq n$. Note that $i \neq h$. This follows almost by definition when $i \neq j$, and by 11.5 when $i = j$.

Definition. Let $M^{(1)}$ denote the union of all of these codimension 1 submanifolds $M_n(i,j)$, and let $M^{(2)}$ denote the union of the codimension 2 submanifolds $M_{n_1}(i_1,j_1) \cap M_{n_2}(i_2,j_2)$.

Lemma 11.8. Any two points f_0 and f_1 in $M - M^{(2)}$ can be joined by a path in $M - M^{(2)}$ which intersects each manifold $M_n(i,j)$ only finitely often.

Proof. Let H_0 denote the set consisting of all C^2 -paths from f_0 to f_1 in the convex open set M , topologized as a subset of the Banach space $C^2(I \times I, \mathbb{R})$. Define subsets $H_0 \supset H_1 \supset H_2 \supset \dots$ inductively as follows. Let H_n be the set of all paths belonging to H_{n-1} which

- (a) intersect the manifold $M_n(i,j)$ transversally for all i and j , and
- (b) are disjoint from $M_n(i,j) \cap M_{n_1}(i_1,j_1)$ for all $n_1 \leq n$ and all $i \neq i_1, j, \text{ and } j_1$.

It is not difficult to check inductively that each H_n is a dense open subset of the complete metric space \bar{H}_0 . Hence the intersection of the H_n is non-vacuous, and contains the required homotopy

from f_0 to f_1 . ■

Finally we will need the following. Fix some large integer N .

Lemma 11.9. Any map $f \in P^0$ can be uniformly approximated by a map in $M - M^{(1)}$ which has the same kneading matrix modulo t^N .

Proof. First suppose that the various forward images $f^p(c_i)$ with $1 \leq p < N$, $1 \leq i \leq \ell - 1$ are not turning points of f . Then the proof is straightforward: First uniformly approximate f by an element g belonging to the subspace $M \subset P^2$ of Morse functions. Then approximate g by an element h in the subspace $M - M^{(1)}$. If the approximations are sufficiently close, then it is easy to check that $\nu_i(f) \equiv \nu_i(g) \equiv \nu_i(h)$ modulo t^N (compare 11.2).

Now consider the more difficult case where, for example, $f^p(c_1) = c_j$. To fix our ideas, suppose that f has a local minimum at c_1 . Let g coincide with f except in a very small neighborhood of c_1 and suppose that $g(x) > f(x)$ within this neighborhood. If the approximation is sufficiently close, then one can check that the forward images $g(c_1), \dots, g^p(c_1), \dots, g^{N-1}(c_1)$ are not turning points, and that the kneading invariants of g are congruent to those of f modulo t^N . Making analogous alterations at $c_2, \dots, c_{\ell-1}$, and then proceeding as above, the proof can easily be completed. ■

We are now ready to prove our main theorem.

Proof of 10.1. Recall that the space M , consisting of all C^2 -smooth Morse functions $f: I \rightarrow \text{interior}(I)$ with prescribed turning points and prescribed sign $\varepsilon(I_1) = \pm 1$ for the derivative $f'(c_0)$ at the leftmost point, is a convex open subset of a suitable Banach space.

It is first necessary to check the formula $D(t) = Z(t)^{-1}$ for just one arbitrarily chosen function f_0 in M . For example, let f_0 map the entire interval I into the interior of a lap I_j with $\varepsilon(I_j) = +1$. Then all of the interesting part of the kneading matrix

$[N_{pq}(f_0)]$ is concentrated in the j -th column. Cancelling this j -th column, it follows easily that

$$D(f_0, t) = 1/(1-t).$$

On the other hand, there are no periodic orbits at all of negative type, so the formula 10.2 takes the form $Z(f_0, t) = 1 - t$, as required.

Now consider a generic homotopy f_u between maps f_0 and f_1 , as constructed in 11.8. We will show that the product

$$Z(f_u, t)D(f_u, t) \in \mathbb{Z}[[t]]$$

varies continuously with the parameter $u \in [0, 1]$ and hence is constant.

Fixing some parameter value u , recall from 11.8 that there at most one turning point c_i such that some forward image $f_u^p(c_i)$ is a turning point c_j . There are two cases to consider.

Case 1. Suppose that no turning point is periodic, so that $f_u^k(c_i) \neq c_i$ for $k \geq 1$ and for all i . Then every fixed point of f_u is an interior point of a lap of f_u^k . Evidently the number of fixed points which occur in laps on which f_u^k is decreasing remains constant under small deformations of f_u . In other words the correspondence

$$g \mapsto Z(g, t)$$

from P^0 to the ring of formal power series is continuous at f_u . Together with 11.4, it follows that the correspondence

$$g \mapsto Z(g, t)D(g, t)$$

is continuous at f_u also.

Case 2. Suppose that just one turning point c_i is periodic, say $f_u^p(c_i) = c_i$ with p minimal. Suppose, to fix our ideas, that f_u^p has a local minimum at c_i as illustrated in Figure 11.10. Then

the graph

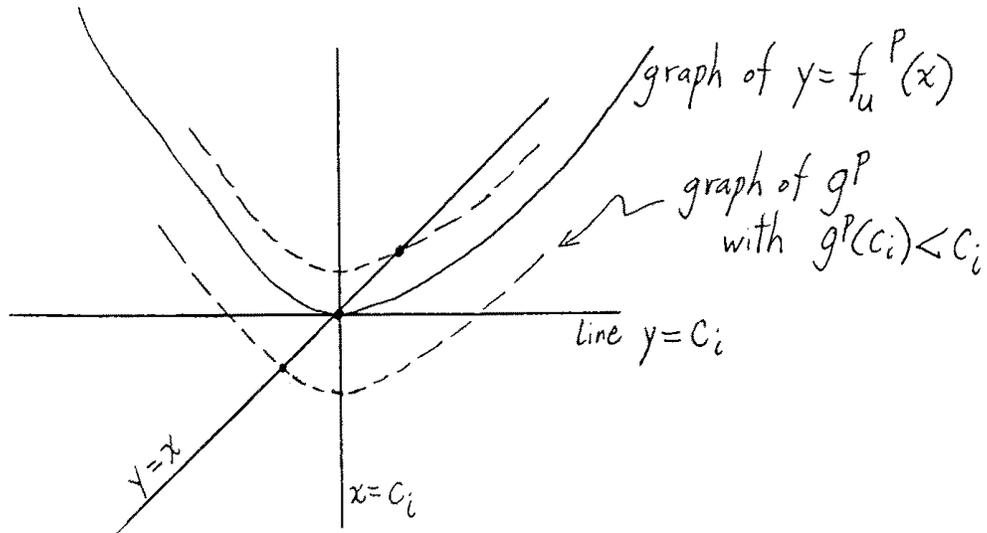


Figure 11.10

the graph of f_u^P intersects the diagonal transversally at c_i . For any g which is sufficiently close to f_u in the space P^1 , it follows that the graph of g^P intersects the diagonal transversally somewhere near c_i . The resulting fixed point of g^P will be of negative type if and only if $g^P(c_i) < c_i$. In this case there is a new factor of $(1+t^P)/(1-t^P)$ in the zeta function $Z(g,t)$. There is no such new factor if $g^P(c_i) \geq c_i$.

Comparing Lemma 11.6, it follows that the correspondence

$$g \mapsto Z(g,t)D(g,t)$$

from the space P^1 to $\tilde{Z}[[t]]$ is continuous at f_u .

Since these are the only possibilities allowed by 11.8, this proves that the function $u \mapsto Z(f_u,t)D(f_u,t)$ from $[0,1]$ to the

totally disconnected space $\tilde{Z}[[t]]$ is continuous and hence constant. Therefore

$$Z(f_1, t)D(f_1, t) = Z(f_0, t)D(f_0, t) = 1.$$

This completes the proof of 10.1 for any function $f_1 \in M$ which does not belong to the exceptional subset $M^{(1)}$, so that 11.8 applies.

Now for any piecewise-monotone map $f \in P^0$, according to Lemma 11.9, we can approximate by a map $f_1 \in M - M^{(1)}$ which has the same kneading determinant modulo t^N . Here N can be any preassigned large integer. Inspecting the proof, it is not difficult to show that the congruence

$$Z(f_1, t) \equiv Z(f, t) \pmod{t^N}$$

will also be satisfied. Thus

$$Z(f, t) \equiv Z(f_1, t) = D(f_1, t)^{-1} \equiv D(f, t)^{-1} \pmod{t^N}$$

for every N , and hence $Z(f, t) = D(f, t)^{-1}$ as required. ■

12. An intermediate value theorem

This section will continue to study the way in which various invariants of the mapping f change under smooth deformation of f . We concentrate on the case $\ell = 2$, but also study continuity properties of the growth number $s(f)$ for arbitrary fixed lap number.

Let us first ask which kneading determinants $D(t) = 1 \pm t \pm t^2 \pm \dots$ can actually occur, for maps with $\ell = 2$. As in §3, we order the additive group $\tilde{Z}[[t]]$ of formal power series by setting $\alpha = a_0 + a_1 t + \dots > 0$ whenever $a_0 = \dots = a_{n-1} = 0$ but $a_n > 0$ for some $n \geq 0$. It is convenient to set $|\alpha| = \pm \alpha$, taking the plus sign if and only if $\alpha > 0$.

Definition. A formal power series $D(t) = 1 + D_1 t + D_2 t^2 + \dots$ with

coefficients $D_i = \pm 1$ will be called admissible if

$$D(t) \leq |D_n + D_{n+1}t + D_{n+2}t^2 + \dots|$$

for every $n \geq 1$.

To illustrate this definition, note that a series of the form $1 + t + (\text{higher terms})$ is admissible if and only if it is equal to $(1-t)^{-1} = 1 + t + t^2 + \dots$.

Theorem 12.1. A given power series $D(t) = 1 \pm t \pm t^2 \pm \dots$ actually occurs as the kneading determinant of some map f with $\ell(f) = 2$ if and only if this series is admissible.

The proof will show that every such admissible power series actually occurs as kneading determinant for some quadratic mapping

$$x \mapsto bx(1-x)$$

from the unit interval to itself, with $1 \leq b \leq 4$. More generally, consider a one-parameter family of C^1 -smooth maps $g_b : I \rightarrow I$, depending C^1 -smoothly on the parameter b for $b_0 \leq b \leq b_1$, all with the single turning point c_1 .

Intermediate Value Theorem 12.2. Any admissible power series $1 \pm t \pm t^2 \pm \dots$ which lies between the kneading determinants of g_{b_0} and g_{b_1} must actually occur as the kneading determinant of g_b for some parameter value between b_0 and b_1 .

Proof. Let $D(b;t) \in \mathbb{Z}[[t]]$ be the kneading determinant associated with the parameter value b , and let $A(t)$ be the given admissible series. Suppose, to fix our ideas, that $D(b_0;t) > A(t) > D(b_1;t)$. Let b' be the supremum of parameter values $b < b_1$ for which $D(b;t) \geq A(t)$. Thus there are parameter values b^- and b^+ arbitrarily close to b' with $D(b^-;t) \geq A(t) > D(b^+;t)$. We will show that $D(b';t)$ must be equal to $A(t)$.

To simplify notations, let $f = g_{b'}$. First suppose that none of the iterated images $f(c_1), f^2(c_1), \dots$ is equal to the turning point

c_1 . Then according to 11.2 or 11.4 the correspondence $b \mapsto D(b;t)$ is continuous at b' , and it follows immediately that $D(b';t) = A(t)$.

Now suppose that $f^p(c_1) = c_1$, with $p \geq 1$ minimal. Then the formal power series $D(b;t)$ is periodic of period p , having the form $\varphi(t)/(1-t^p)$ for some polynomial $\varphi(t) = 1 + D_1 t + \dots + D_{p-1} t^{p-1}$ of degree $p-1$ (compare §4.5). Furthermore, for every b sufficiently close to b' we have

$$D(b;t) = \varphi(t)/(1+t^p)$$

by §11.6. Thus

$$\varphi(t)/(1-t^p) \geq A(t) > \varphi(t)/(1+t^p)$$

by the definition of b' . However, it is not difficult to check that the only admissible series satisfying these inequalities is $\varphi(t)/(1-t^p)$ itself. This completes the proof of 12.2. ■

Proof of 12.1. In one direction, this follows immediately by applying 12.2 to the one parameter family of maps $x \mapsto bx(1-x)$. In fact, for $b = 1$ we obtain the largest possible admissible kneading determinant $1 + t + t^2 + \dots$ while for $b = 4$ we obtain the smallest possible admissible kneading determinant $1 - t - t^2 - \dots$. (For further discussion of such quadratic maps, see §13.)

For the proof in the other direction, we must study the kneading determinant $D(t) = \sum D_k t^k$ associated with an arbitrary map f with $\ell(f) = 2$. Suppose, to fix our ideas, that f attains its minimum at the unique turning point c_1 . Then $f(c_1) \leq f(x)$ for all x , and in particular,

$$f(c_1) \leq f^{n+1}(c_1)$$

for all $n \geq 1$. Applying the invariant coordinate function θ of §3, it follows that $\theta(f(c_1)) \leq \theta(f^{n+1}(c_1))$. Similarly, taking the limit as $x \rightarrow c_1$, $x > c_1$, we have

$$\theta(f(c_1+)) \leq \theta(f^{n+1}(c_1+)).$$

Recall that $\theta(x)$ is a formal power series in t , of the form $I_{i_0} \pm I_{i_1} + \dots$, provided that x is not a pre turning point. Since there are only two formal basis elements $I_1 < I_2$, corresponding to the two laps, we can simplify the situation by applying the linear transformation

$$I_1 \mapsto -1, I_2 \mapsto +1$$

of §3.3. Then $\theta(x)$ will map to a formal power series in $\mathbb{Z}[[t]]$ of the form

$$\bar{\theta}(x) = \pm 1 \pm t \pm t^2 \pm \dots,$$

which still depends monotonely on x , so that

$$(*) \quad \bar{\theta}(f(c_1+)) \leq \bar{\theta}(f^{n+1}(c_1+)).$$

With this notation, it is not difficult to check that $\bar{\theta}(c_1+)$ is equal to the kneading determinant $1 + D_1 t + D_2 t^2 + \dots$. Similarly,

$$\bar{\theta}(f(c_1+)) = D_1 + D_2 t + D_3 t^2 + \dots,$$

and more generally

$$\bar{\theta}(f^{n+1}(c_1+)) = D_n (D_{n+1} + D_{n+2} t + \dots).$$

Now, multiplying the inequality (*) by t and adding $+1$, we obtain

$$1 + D_1 t + D_2 t^2 + \dots \leq D_n (D_n + D_{n+1} t + \dots) = |D_n + D_{n+1} t + \dots|,$$

as required. ■

Next let us study continuity properties of the growth number $s(f)$. We consider piecewise monotone C^1 -maps from I to itself with fixed turning points $c_1, \dots, c_{\ell-1}$.

Lemma 12.3. As g tends to f in the C^1 -topology, keeping the lap number $\ell(g)$ and the turning points $c_1, \dots, c_{\ell-1}$ fixed, the growth number $s(g)$ tends to $s(f)$.

Compare Guckenheimer, 1979.

Here it is essential to use the C^1 -topology. Furthermore, it is essential to keep the lap number fixed, or at least bounded. For Misiurewicz and Szlenk, and independently Newhouse, have given examples to show that the correspondence $g \mapsto s(g)$ is not continuous if the number of turning points is allowed to jump to an arbitrarily large value, even if we use the C^r -topology with r large.

Proof of 12.3. First let us prove continuity at f under the hypothesis that none of the turning points c_i is periodic under f .

We will use the notation $D(g,t)$ for the kneading determinant of g . Let t be a complex variable ranging over the disk $|t| \leq r$ of some radius $r < 1$. Then we will show that the uniform norm

$$\sup_{|t| \leq r} |D(g,t) - D(f,t)|$$

tends to zero as g tends to f in the space P^0 of §11, or, in other words, as g tends uniformly to f with fixed turning points.

Setting $D(g,t) = \sum D_n(g)t^n$, first note that the integer coefficients $D_n(g)$ are uniformly bounded,

$$|D_n(g)| \leq A_n \quad \text{for all } g \in P^0,$$

where A_n is some polynomial function of n . For example, one can take $A_n = 2^{\ell-1}(n+1)(n+2)\dots(n+\ell-1)$.

To verify this estimate, it is convenient to introduce two special notations. We will write $\sum a_n t^n \ll \sum b_n t^n$ whenever $a_n \leq b_n$ for all n , and will use the notation $\|\sum a_n t^n\|$ for the power series $\sum |a_n| t^n$. Recall that the product $(1-t)D(g,t)$ is defined to be the determinant of a certain $(\ell-1) \times (\ell-1)$ kneading matrix. Each entry $\sum a_n t^n$ of this kneading matrix satisfies $|a_n| \leq 2$; or, in other words, $\|\sum a_n t^n\| \ll 2/(1-t)$. Adding up $(\ell-1)!$ monomials to form the determinant, we obtain

$$\|(1-t)D(g,t)\| \ll (\ell-1)!(1/(1-t))^{\ell-1},$$

and therefore

$$\|D(g,t)\| \ll (\ell-1)!2^{\ell-1}(1-t)^{\ell}.$$

Expanding $(1-t)^{-\ell}$ by the binomial theorem, this yields the required upper bound for $|D_n(g)|$.

Since A_n is a polynomial in n , it follows that the comparison series $\sum A_n t^n$ converges uniformly throughout the disk $|t| \leq r < 1$. Hence for each $\varepsilon < 0$ there exists an N so that the sum from N to infinity satisfies

$$(1) \quad \left| \sum_N^{\infty} D_n(g) t^n \right| \leq \sum_N^{\infty} A_n r^n < \varepsilon$$

uniformly throughout the disk $|t| \leq r$, and uniformly for all g .

But if g is sufficiently close to f in the space P^0 , then

$$(2) \quad D_n(g) = D_n(f) \quad \text{for } n < N.$$

This follows using 10.1, or can be verified directly by the methods of §11. Combining (1) and (2) it follows that

$$|D(g,t) - D(f,t)| < 2\varepsilon$$

for $|t| \leq r$. In other words, the analytic function $D(g,t)$ tends uniformly to $D(f,t)$ throughout the disk of radius r as g tends to f in the C^0 -topology.

Using 6.3 and the fact that the zeros of a complex analytic function vary continuously with the function, it follows that $s(g)$ tends to $s(f)$ as g tends to f .

Now consider the more general case where one or more of the turning points c_i may be periodic under f . We must now use the C^1 -topology. If $f^p(c_i) = c_i$, then for g arbitrarily close to f in the space P^1 the function $D(g,t)$ may uniformly approximate not $D(f,t)$ but rather the product $D(f,t)(1-t^p)/(1+t^p)$ (compare §11.6).

If k of the $\ell - 1$ turning points are periodic, then there may be as many as k of these additional factors of the form $(1-t^p)/(1+t^p)$. However, these additional factors are all smooth non-zero analytic functions throughout the disk $|t| \leq r$, so that they do not affect the location of zeros. The argument now proceeds just as before. ■

One immediate corollary is that every number in the interval $[1, \ell]$ occurs as growth number $s(f)$ for some map f with lap number ℓ . For we can choose a smooth one parameter family of maps in the space P^1 leading from a map with $s = 1$ to a map with $s = \ell$.

For the case $\ell = 2$, it follows that there are uncountably many distinct admissible kneading invariants $D(t)$.

13. More on quadratic maps

Let us take a further look at the archetypal one parameter family of maps

$$y \mapsto by(1-y)$$

from the unit interval to itself, or the family of linearly conjugate maps $f_a(x) = (x^2 - a)/2$, where $a = b^2 - 2b$. The growth number $s(f_a)$, as a function of the parameter a , is plotted in Figure 13.3, and tabulated in 13.5. We have noted earlier that the behavior of the iterates of the function f_a becomes more and more complicated as the parameter a increases. Here is a precise statement.

Theorem 13.1. The kneading invariant $D(f_1) \in Z[[t]]$ is monotone decreasing as a function of the parameter a .

Corollary 13.2. For each fixed n , the number of laps $\ell(f_a^n)$ is monotone increasing as a function of a . Hence the growth number $s(f_a)$ is also monotone increasing as a function of a .

These statements were first proved by Douady, Hubbard and

Sullivan (unpublished). The proofs given below were suggested by their argument. Another closely related statement, which follows from the Douady-Hubbard-Sullivan argument but will not be proved here, is the following. For each fixed k , the number of periodic points of f_1 of period k is monotone increasing as a function of a (see Jonker, 1982, for a weaker form of this result).

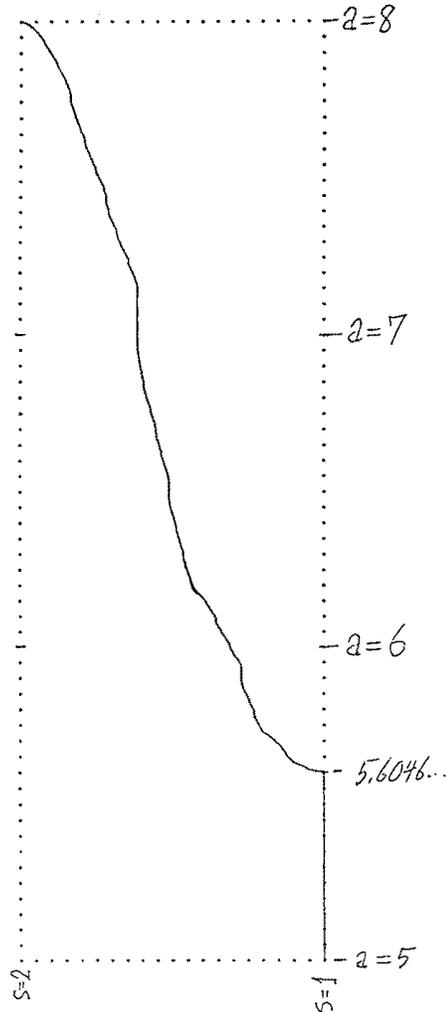


Figure 13.3. Graph of the growth number $s(f_a)$ as a function of the parameter a , where $f_a(x) = (x^2 - a)/2$. This function is continuous and monotone. Its intervals of constancy are everywhere dense. For example $s(f_a) = (1 + \sqrt{5})/2$ for $7 \leq a \leq 7.161\dots$ and $s(f_a) = \sqrt{(1 + \sqrt{5})}/2$ for $5.89\dots \leq a \leq 5.94$.

Before beginning of the proof of 13.1, let us briefly discuss other theorems and problems concerning quadratic maps. We have seen in Sections 11 and 12 that every strictly periodic admissible kneading invariant corresponds to an entire interval of parameter values a . However, it turns out that these open intervals become extremely small as the period increases. In fact, according to Jakobson [1981], we have the following remarkable statement: The set of parameter values a for which the kneading invariant $D(f_a)$ is nonperiodic has positive Lebesgue measure.

It is conjectured that this set of positive measure does not contain any interval, or equivalently that the complementary set, consisting of all parameter values for which $D(f_a)$ is strictly periodic, is everywhere dense in the interval $[-1,8]$. Still another equivalent conjecture would be that each admissible nonperiodic kneading invariant corresponds to a unique parameter value a . In the special case of an eventually periodic (but not periodic) kneading invariant, this would not be difficult to prove by the methods described below; but we have no idea how to handle the general case.

The proof of 13.1 will be based on the following statement.

Lemma 13.4. Consider two different real quadratic mappings for which the critical point is contained in a periodic orbit of period p . If these two have the same kneading invariant, then they are linearly conjugate to each other.

Proof. Since the cases $p = 1, 2$ are easy to verify, we may assume that $p \geq 3$. The proof will be based on complex variable methods, and in particular, on Teichmüller theory. Let M_p be the complex $(p-2)$ -dimensional manifold consisting of all p -tuples (z_1, \dots, z_p) of distinct complex numbers, well defined up to a simultaneous linear substitution

$$z_j \mapsto \alpha z_j + \beta$$

with $a \neq 0$. Then the universal covering \tilde{M}_p can be identified with the Teichmüller space consisting of all marked Riemann surfaces of genus zero with $p+1$ punctures (see for example Abikoff).

Consider any complex quadratic mapping q whose critical point c is periodic of period p . Then the finite orbit $\{z_j = q^j(c)\}$ determines a point (z_1, \dots, z_p) in the manifold M_p . It is not difficult to check that two different quadratic maps, each having such a superattractive periodic orbit, determine the same point of M_p if and only if they are linearly conjugate.

Let us define a many-valued, locally biholomorphic mapping F from M_p into itself by the formula

$$F(z_1, \dots, z_p) = (\pm\sqrt{z_2 - z_1}, \dots, \pm\sqrt{z_p - z_1}, 0).$$

Then the fixed points of F are precisely the points (z_1, \dots, z_p) associated with quadratic maps having a superattractive period p orbit. For if $z_j = q^j(c)$, then setting $q(z) = (z-c)^2/a^2 + z_1$ we will have $q^{-1}(w) = \pm a\sqrt{w - z_1} + c$, hence

$$z_j = \pm a\sqrt{z_{j+1} - z_1} + c \quad \text{for } j < p,$$

and $z_p = a \cdot 0 + c$, as required.

Now let us choose a basepoint b in the universal covering manifold \tilde{M}_p which lies over such a fixed point of F . Evidently F lifts to a unique holomorphic mapping \tilde{F} of \tilde{M}_p into itself which fixes the point b . The key step of our proof is the following observation. If we give the manifold \tilde{M}_p the Teichmüller metric, then this map \tilde{F} from \tilde{M}_p into itself must be strictly distance decreasing. Hence \tilde{F} can have no other fixed points.

If we assume this last statement for a moment, then the proof of 13.4 can be easily completed as follows. Let q be a real quadratic map with a superattractive orbit of period p . Let $\Delta \subset M_p$ be the open $(p-2)$ -simplex consisting of all p -tuples (x_1, \dots, x_p) of dis-

tinct real numbers, well defined up to a simultaneous linear substitution, which are ordered in the same way as the base p -tuple $(q(c), \dots, q^p(c) = c)$. We may identify Δ with a subset of the Teichmüller space \tilde{M}_p , containing the basepoint b . Then \tilde{F} maps Δ into itself. If there were another real quadratic map q' having the same kneading invariant as q , and also having a superattractive periodic orbit, then it would determine another fixed point of \tilde{F} in Δ . Since there is only one such fixed point, it follows that q' and q must be linearly conjugate to each other.

In order to prove that \tilde{F} reduces Teichmüller distance, we must first recall the definition of this distance. First consider a diffeomorphism $f : z \mapsto z'$ between two Riemannian 2-manifolds. If we fix some tangent direction at a point of the first manifold, then the mapping stretches distances (at this point and in this direction) by a factor of $ds'/ds > 0$. The ratio $K(z) = \sup(ds'/ds)/\inf(ds'/ds)$, taking the \sup and the \inf over all directions through some given point z , is called the dilatation of f at z . A map f , which is a diffeomorphism except possibly at a finite number of points, is called admissible (or quasi-conformal) if $K(z)$ is bounded, so that the number $K[f] = \sup_z K(z)$ is finite. The smallest possible value of $\log K[f]$ as f varies over a homotopy class of maps is called the Teichmüller distance associated with this homotopy class. In the hyperbolic case, a basic theorem of Teichmüller asserts that the minimum dilatation is always realized by one and only one mapping which has the following description (see for example Abikof, p. 59). There is an essentially unique quadratic differential $\phi(z)dz^2$, which is holomorphic except possibly for simple poles at the puncture points, so that f stretches distances by a maximum factor in the positive directions associated with this quadratic differential, and by a minimum factor in the orthogonal negative directions, the ratio of these

factors being exactly $K[f]$ everywhere.

For our purposes, the two Riemann surfaces should be the complex plane \mathcal{C} , punctured at the infinite point and either at the points z_1, \dots, z_p or at corresponding points z'_1, \dots, z'_p . The mapping F assigns to each of these punctured Riemann surfaces a new punctured Riemann surface which can be described geometrically as follows. First form the 2-fold covering, branched only at the critical value z_1 (or z'_1) and at the infinite point. Each of the remaining puncture points has two distinct representatives in this 2-fold covering. We must fill in one of these two puncture points, but keep the other one. (There is a choice here, so that F is not a single-valued function.) Finally, we must renumber the finite puncture points cyclically.

Let us start with the unique admissible map f from $\mathcal{C} - \{z_1, \dots, z_p\}$ to $\mathcal{C} - \{z'_1, \dots, z'_p\}$, within a specified homotopy class, which minimized the maximum dilatation. Now pass to 2-fold branched covering surfaces, as described above. Evidently f lifts to an admissible map \tilde{f} between these branched coverings, with the same constant dilatation $K[f]$. This proves that \tilde{F} can never increase distance.

To see that \tilde{F} must actually decrease Teichmüller distance, we must show that the lifted map \tilde{f} cannot be the unique best possible map described in Teichmüller's theorem. Evidently the map \tilde{f} is associated with a quadratic differential, obtained by lifting the given quadratic differential on $\mathcal{C} - \{z_1, \dots, z_p\}$ to the 2-fold covering. Recall that every quadratic differential on a surface of genus zero must have at least four poles. In general, these poles must be counted with the appropriate multiplicity, but in our case only simple poles can occur. Hence there must be at least two poles at the non-branch points z_2, \dots, z_p . Each of these will be covered by

two poles on the covering surface, one at a puncture point and one at an ordinary point. Thus this quadratic differential on the covering surface has poles at ordinary (nonpuncture) points, which proves that its associated mapping is not the best possible mapping which minimizes the maximal dilatation. Hence the Teichmüller distance between these two marked covering surfaces must be strictly less than the Teichmüller distance between the two original surfaces. ■

Proof of Theorem 13.1. As the parameter a increases from -1 to 8 , we must show that kneading invariant $D(f_a)$ decreases monotonely from $1+t+t^2+\dots$ to $1-t-t^2-\dots$. Suppose to the contrary that $D(f_a) < D(f_b)$ for some $a < b$. Since $D(f_a) \neq D(f_b)$, there must clearly exist some parameter value c between a and b so that f_c has a superattractive periodic orbit. Then either $D(f_c) < D(f_b)$ or $D(f_c) > D(f_a)$ or both. In the first case, according to the intermediate value theorem 12.2, there exists a parameter value $d > b > c$ with $D(f_d) = D(f_c)$. Furthermore, the proof constructs this parameter value d so that f_d also has a superattractive periodic orbit. But this is impossible by 13.4. Since the argument in the second case $D(f_c) > D(f_a)$ is completely analogous, this completes the proof of 13.1. ■

Proof of Corollary 13.2. If two quadratic maps f and g satisfy $D(f) > D(g)$, using the usual total ordering of the formal power series ring $Z[[t]]$, then we assert that $\ell(f^n) < \ell(g^n)$ for every n . For, as h varies continuously from f to g , the number $\ell(h^n)$ can change only as we pass through a map having a superattractive orbit of period less than n . Using 11.6 and 5.8, we see that $\ell(h^n)$ can only increase. Evidently, the corollary follows. ■

To conclude this section, here is a table showing some actual kneading invariants, together with the associated parameter values a and growth numbers s . In order to make a short list, we have singled

out those admissible kneading invariants which are periodic, with period $p \leq 6$.

Table 13.5. List of admissible kneading determinants of period $p \leq 6$.

parameter a in the interval	coefficients of periodic kneading determinant	period p	growth number s
[-1 , 0]	+ + + + + + + + ...	1	
(0 , 4]	+ - + - + - + - ...	2	1
(4 , 5.242...]	+ - - + + - - + ...	4	

[5.898..., 5.904...]	+ - - + - + + - ...	6	1.272..

[6.497..., 6.501...]	+ - - + - + - - ...	5	1.512..

[7 , 7.019...]	+ - - + - - + - ...	3	1.618..
(7.019..., 7.091...]	+ - - - + + + - ...	6	

[7.442..., 7.443...]	+ - - - + + - - ...	5	1.722..

[7.629..., 7.629...]	+ - - - + - + - ...	6	1.792..

[7.762..., 7.763...]	+ - - - + - - - ...	4	1.839..

[7.867..., 7.867...]	+ - - - - + + - ...	6	1.883..

[7.941..., 7.941...]	+ - - - - + - - ...	5	1.927..

[7.985..., 7.985...]	+ - - - - - + - ...	6	1.965..

This table is to be interpreted as follows: An entry of say "+ - - + - - + - ..." of period 3 stands for the periodic series $(1-t-t^2)/(1-t^3)$. Note that there are infinitely many other periodic admissible series with period $p > 6$ in every one of the starred gaps in this table. (The precise number of admissible series of period p

is given by the expression $(2p)^{-1} \sum \mu(d) 2^{p/d}$, to be summed over all squarefree odd divisors of p . Here $\mu(d) = \pm 1$ is the Moebius function, equal to $+1$ or -1 according as the number of primes dividing the squarefree number d is even or odd. This expression grows exponentially with p .)

Note that the lengths of the associated parameter intervals tend to zero very rapidly for large p . Thus, in several cases the three decimal places listed are not enough to distinguish the two endpoints of an interval.

14. Constructing new functions out of old

This section will briefly describe several ways of constructing interesting maps f from an interval to itself. Proofs will be largely left to the reader. We will use the notation $D(f,t)$ to indicate the dependence of the determinant $D(t)$ on the map f , and the notation $L(f,t)$ for the power series $\sum \ell(f^k) t^{k-1}$.

Example 14.1 Juxtaposition. Suppose that $f : I \rightarrow I$ maps the left half of I to itself by a map f_1 , and maps the right half to itself by a map f_2 (compare Figure 7.5). Then the lap number $\ell(f^k)$ is evidently equal to $\ell(f_1^k) + \ell(f_2^k) - 1$, hence

$$L(f,t) = (L(f_1,t) + L(f_2,t) - (1-t)^{-1}).$$

It follows that the growth number $s(f)$ is equal to the maximum of the two numbers $s(f_1)$ and $s(f_2)$. Similarly, inspection shows that $\hat{n}(f^k) = \hat{n}(f_1^k) + \hat{n}(f_2^k) - 1$ and $N(f^k) = N(f_1^k) + N(f_2^k) + 1$, hence

$$\hat{\zeta}(f,t) = \hat{\zeta}(f_1,t) \hat{\zeta}(f_2,t) (1-t),$$

$$Z(f,t) = Z(f_1,t) Z(f_2,t) / (1-t).$$

The corresponding formula

$$D(f, t) = D(f_1, t)D(f_2, t)(1-t)$$

for the kneading determinant then follows from 10.1. Alternatively, inspection shows that the kneading matrix of f takes the form

A	* · · · *	O
O	* · · ·	B

Deleting the common column and applying 4.2, we obtain a different proof of the same formula for $D(f, t)$

Example 14.2 Iteration. Given any f , we can think of the 2-fold iterate $f^2 : I \rightarrow I$ as a mapping in its own right. Evidently $s(f^2) = s(f)^2$. In order to describe the power series $tL(f^2, t) = \sum \ell(f^{2k})t^k$, it is convenient to substitute t^2 in place of t . Then inspection shows that

$$t^2L(f^2, t^2) = \sum \ell(f^{2k})t^{2k}$$

is equal to the average $(tL(f, t) - tL(f, -t))/2$. Similarly, a completely formal argument shows that

$$\hat{\zeta}(f^2, t^2) = \hat{\zeta}(f, t)\hat{\zeta}(f, -t), \quad \text{and} \quad Z(f^2, t^2) = Z(f, t)Z(f, -t).$$

Using 10.1, it follows that

$$D(f^2, t^2) = D(f, t)D(f, -t).$$

As an example, starting with the quadratic map $f(x) = (x^2 - 7)/2$, we obtain a quadratic map $f^2(x) = (x^4 - 14x^2 + 21)/8$. Since $D(f, t) = (1-t-t^2)/(1-t^3)$, it follows by multiplying out this formula that $D(f^2, t^2) = (1-3t^2+t^4)/(1-t^6)$, or in other words $D(f^2, t) = (1-3t+t^2)/(1-t^3)$.

Similarly, for any iterate f^k , an analogous argument based on 10.1 shows that

$$D(f^k, t^k) = D(f, \omega_1 t) \dots D(f, \omega_k t),$$

where $\omega_1, \dots, \omega_k$ are the k -th roots of unity. This computation can be expressed in a more perspicuous form if $D(f, t)$ happens to be a rational function of t . In this case we can write

$$D(f, t) = \prod_j (1 - a_j t)^{m_j},$$

where the a_j may be complex. It follows easily that

$$(14.3) \quad D(f^k, t) = \prod_j (1 - a_j^k t)^{m_j}.$$

As an example, the function f illustrated in Figure 1.5 has rational kneading determinant $D(f, t) = (1 - 3t)/(1 - t^2)$. Using 14.3, it follows that

$$D(f^k, t) = \begin{cases} (1 - 3^k t)/(1 - t^2) & \text{for } k \text{ odd} \\ (1 - 3^k t)/(1 - t)^2 & \text{for } k \text{ even.} \end{cases}$$

Example 14.4 Period Doubling. Let us start with a function

$f : [0, 1] \rightarrow [0, 1]$ which maps both boundary points to zero. Define an associated map $g : [-2, 2] \rightarrow [-2, 2]$, as illustrated in Figure 14.5, by setting

$$g(x) = \begin{cases} 2x+2 & \text{for } x \in [-2, -1] \\ f(-x) & \text{for } x \in [-1, 0] \\ -x & \text{for } x \in [0, 2] \end{cases} .$$

Inspection shows that $g^2(x) = f(x)$ for $x \in [0, 1]$. The periodic points of g can be described very precisely as follows.

- (1) There are just two fixed points, at -2 and 0 .
- (2) Each periodic point $x \neq 0$ of f with period p gives rise to periodic points $\pm x$ of g of period $2p$. There are no other periodic points.

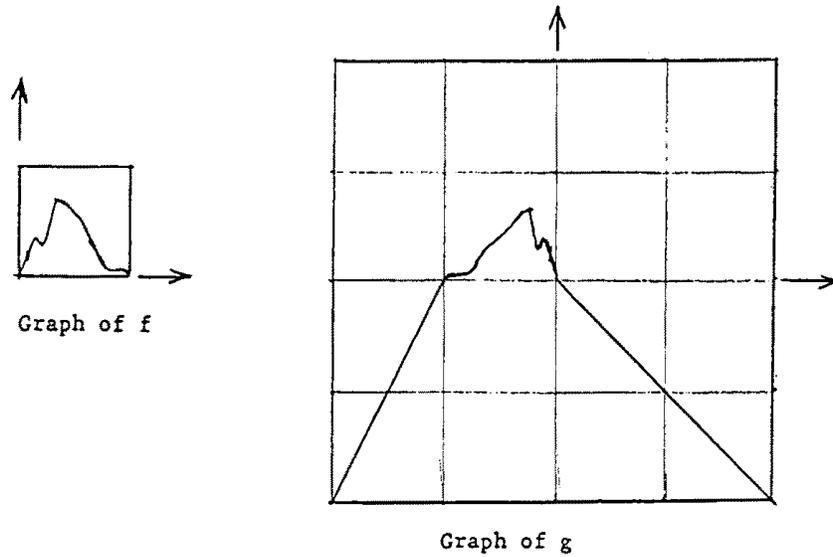


Figure 14.5

It follows easily that

$$\hat{\zeta}(g, t) = \hat{\zeta}(f, t^2)(1-t)/(1-t) \quad \text{and} \quad Z(g, t) = Z(f, t^2)/(1-t).$$

Using 10.1 or by direct computation, we find that

$$D(g, t) = (1-t)D(f, t^2).$$

It follows that $s(g) = s(f)^{1/2}$.

As an example, if $D(f, t) = (1-t-t^2)/(1-t^3)$, then $D(g, t) = (1-t)(1-t^2-t^4)/(1-t^6)$. (By §12, we can find a quadratic map with this kneading determinant.)

Example 14.6. After making a linear change of coordinates to replace the interval $[-2, 2]$ by $[0, 1]$, we can iterate this construction to obtain a sequence of maps $f_0 = f, f_1 = g, f_2, f_3, \dots$ from $[0, 1]$ to itself, with

$$D(f_1, t) = (1-t)D(f, t^2)$$

$$D(f_2, t) = (1-t)(1-t^2)D(f, t^4),$$

$$D(f_3, t) = (1-t)(1-t^2)(1-t^4)D(f, t^8),$$

and so on. Note that the growth numbers $s(f_i) = s(f)^{1/2^i}$ tend to 1 as $i \rightarrow \infty$. The limit function f_∞ , illustrated in Figure 14.7, is completely independent of f . Its kneading determinant is equal to infinite product

$$D(f_\infty, t) = (1-t)(1-t^2)(1-t^4)(1-t^8)\dots$$

Evidently this infinite product converges, and is nonzero, whenever $|t| < 1$. Hence $s(f_\infty) = 1$. Noteworthy properties of this limit function are the following (compare 1.5).

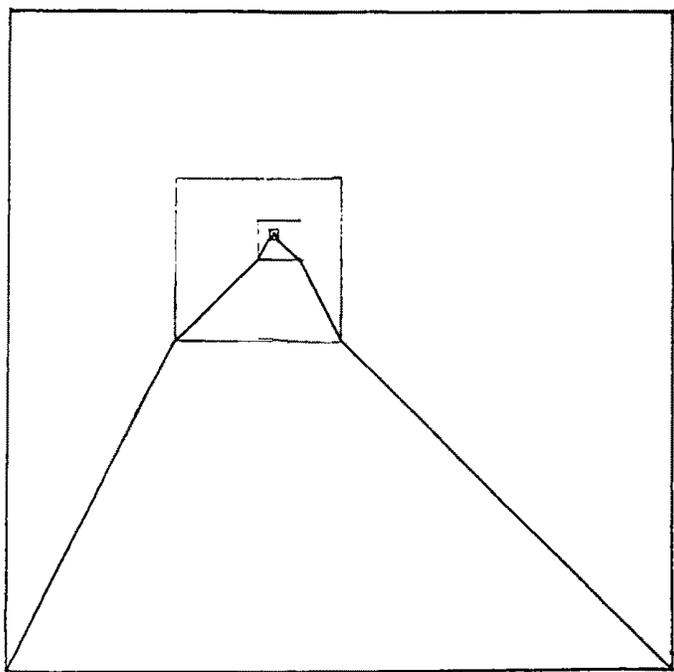


Figure 14.7. Graph of f_∞ .

(1) The reciprocal zeta function $\zeta(f_\omega, t)^{-1} = \hat{\zeta}(f_\omega, t)^{-1}$ is equal to $(1-t)^2(1-t^2)(1-t^4)(1-t^8)\dots$. In other words, there is just one periodic orbit 2^k for each $k \geq 1$, and there is no periodic point at all whose period is not a power of 2 (compare Šarkovskii's theorem in §8).

(2) From 9.7 or by direct computation we see that the sequence of lap numbers $\ell(f_\omega^k)$ grows faster than any polynomial function of k as $k \rightarrow \infty$. Yet this sequence does not have exponential growth since $s = 1$.

(3) Restricting t to the interval $[0, 1]$, the function $D(t) = (1-t)(1-t^2)(1-t^4)\dots$ clearly tends to zero faster than any power of $1-t$ as $t \rightarrow 1$. The same is true for $D(\omega t)$, where ω is any $2k$ -throot of unity. Thus the complex analytic function $D(t)$ has an essential singularity at every dyadic root of unity, and hence at every point of the unit circle. In other words $D(t)$ cannot be extended as an analytic function outside of the disk $|t| < 1$. Similarly, $\zeta(f_\omega, t)$ cannot be extended. The non-differentiability of the map f_ω is of course not a necessary property, but is rather an artifact of this particular method of construction. According to 12, there exists a real number a so that the quadratic map $x \mapsto (x^2 - a)/2$ has precisely the same kneading determinant $D(t) = (1-t)(1-t^2)(1-t^4)\dots$ (compare Feigenbaum). This number $a = 5.6046\dots$ is conjectured to be unique.

Example 14.8 Microimplantation of functions. The construction in 13.4 can be generalized as follows. Suppose that we are given $e : I \rightarrow I$ and $f : J \rightarrow J$ with the following properties. The map e should have just one turning point c_1 , and this turning point should be periodic of period p , so that $e^p(c_1) = c_1$. Furthermore f should map the two end points of J to a single end point. Then we will construct a new map $g : I \rightarrow I$ by implanting a small copy of f into e .

Suppose, to simplify the notation, that $c_1 = 0$. Note that e^{p-1} maps a neighborhood of $e(0)$ homeomorphically onto a neighborhood of 0. It is not difficult to construct an auxiliary map E , also with lap number $\ell = 2$, which coincides with e outside of a small neighborhood of 0, but satisfies either $E^p(x) = |x|$ or $E^p(x) = -|x|$ for all x within some smaller neighborhood $-\delta \leq x \leq \delta$. Evidently E and e have the same kneading matrix.

Construct g as follows. Let $g(x) = E(x)$ for $|x| \geq \delta$, and let g^p map $[-\delta, \delta]$ to itself by a mapping which is linearly conjugate to f . Evidently g is uniquely determined by this description and has the same lap number as f . Each periodic orbit of E which does not intersect the interval $[-\delta, \delta]$ remains periodic for g . Similarly, each periodic orbit of f with period q gives rise to a periodic orbit of g with period pq . It is not difficult to verify that

$$N(g^k) = \begin{cases} N(e^k) & \text{if } k \not\equiv 0 \pmod{p} \\ N(e^k) + N(f^{k/p}) + 1 & \text{if } k \equiv 0 \pmod{p}. \end{cases}$$

Hence

$$Z(g, t) = Z(e, t)Z(f, t^p)/(1-t^p)$$

and

$$(14.9) \quad D(g, t) = (1-t^p)D(e, t)D(f, t^p).$$

(Note that the product $(1-t^p)D(e, t)$ is a polynomial of degree $p-1$ with all coefficients ± 1 , by 4.5.)

It follows that the growth number $s(g)$ is just the maximum of the two numbers $s(e)$ and $s(f)^{1/p}$.

As an example, the quadratic mapping $e(x) = (x^2 - r)/2$ satisfies $e^2(0) = 0$, with $p = 2$ and $(1-t^2)D(e, t) = 1 - t$. In this case we recover the formula

$$D(g, t) = (1-t)D(f, t^2)$$

of 14.4, with $s(g) = \sqrt{s(f)}$.

On the other hand, if the period p is not a power of 2, and if $\ell(f) = \ell(g) = 2$, then according to Bowen and Franks we have

$$s(e) > 2^{1/p} \geq s(f)^{1/p},$$

hence $s(g) = s(e)$. Thus, in this case, we can construct uncountably many distinct admissible kneading determinants, all with the same growth number $s(e)$.

Here is an example with $p = 3$. Let $e(x) = (x^2 - 7.019\dots)/2$ where the constant 7.019... is to be chosen so that $e^3(0) = 0$. In this case, $(1-t^3)D(e,t) = 1-t-t^2$, so the formula becomes

$$D(g,t) = (1-t-t^2)D(f,t^3),$$

with $s(g) = s(e) = (1+\sqrt{5})/2$. Just as in 14.6, this construction can be iterated infinitely often, thus constructing a map with this same growth number whose kneading determinant

$$D = (1-t-t^2)(1-t^3-t^6)(1-t^9-t^{18})\dots$$

has infinitely many zeros within the unit disk. Again by §12, there exists a (presumably unique) real number $a = 7.145\dots$ so that the quadratic map $(x^2 - a)/2$ has this same kneading determinant.

References

- Abikoff, W.: The real analytic theory of Teichmüller space, Lect. Notes Math. 820, Springer, 1980.
- Abraham, R., Robbin, J.: Transversal Mappings and Flows, Benjamin, 1967.
- Adler, R., Konheim, A., McAndrew, M.: Topological entropy, Trans. Amer. Math. Soc. 114, 309-319 (1965).
- Ahlfors, L.: Complex Analysis, McGraw-Hill, 1966.
- Allwright, D.J.: Hypergraphic functions and bifurcations in recurrence relations, SIAM J. Appl. Math. 34, 687-691 (1978).

- Artin, M., Mazur, B.: On periodic points, *Annals of Math.* 81, 82-99 (1965).
- Block, L.: Noncontinuity of topological entropy of maps of the Cantor set and of the interval, *Proc. Amer. Math. Soc.* 50, 388-393 (1975).
- Block, L.: Mappings of the interval with finitely many periodic points have zero entropy, *Proc. Amer. Math. Soc.* 67, 357-360 (1977).
- Block, L.: An example where topological entropy is continuous, *Trans. Amer. Math. Soc.* 231, 201-213 (1977).
- Block, L.: Continuous maps of the interval with finite nonwandering set, *Trans. Amer. Math. Soc.* 240, 221-230 (1978).
- Block, L., Guckenheimer, J., Misiurewicz, M., Young, L.-S.: Periodic points and topological entropy of one dimensional maps, pp. 18-34 of *Global Theory of Dynamical Systems*, ed. Nitecki and Robinson, *Lecture Notes in Math.* 819, Springer 1980.
- Bowen, R.: Topological entropy and Axiom A, pp. 23-41 of *Global Analysis*, ed. Chern and Smale, *Proc. Symp. Pure Math.* 14, A.M.S., 1970.
- Bowen, R.: Entropy for maps of the interval, *Topology* 16, 465-467 (1977).
- Bowen, R.: On Axiom A Diffeomorphisms, *Notes of NSF Regional Conference*, North Dakota State Univ., Fargo, N.D., 1977.
- Bowen, R.: Bernoulli maps of the interval, *Israel J. Math.* 28, 161-168 (1977).
- Bowen, R., Franks, J.: The periodic points of maps of the disk and the interval, *Topology* 15, 337-442 (1976).
- Bowen, R., Lanford, O.E., III: Zeta functions of restrictions of the shift transformation, pp. 43-49 of *Global Analysis*, ed. Chern and Smale, *Proc. Symp. Pure Math.* 14, A.M.S., 1970.
- Collet, P., Eckmann, J.-P.: *Iterated Maps of the Interval as Dynamical Systems*, *Progress in Physics* 1, Birkhauser, 1980.
- Fatou, M.P.: Sur les équations fonctionnelles, *Bull. Soc. Math. France* 47, 161-271 (1919); 48, 33-94, 208-314 (1920).
- Feigenbaum, M.: Quantitative universality for a class of nonlinear transformations, *J. Statist. Phys.* 19 22-52 (1978); 21, 669-706 (1979).
- Feigenbaum, M.: The transition to aperiodic behavior in turbulent systems, *Commun. Math. Phys.* 77, 65-86 (1980).
- Goodman, T.N.T.: Relating topological entropy and measure entropy, *Bull. London Math. Soc.* 3, 176-180 (1971).
- Guckenheimer, J.: Endomorphisms of the Riemann sphere, pp. 95-123 of *Global Analysis*, ed. Chern and Smale, *Proc. Symp. Pure Math.* 14, A.M.S. (1970).

- Guckenheimer, J.: On the bifurcation of maps of the interval, *Invent. Math.* 39, 165-178 (1977).
- Guckenheimer, J.: *Bifurcations of dynamical systems*, C.I.M.E. Lectures, 1978.
- Guckenheimer, J.: Sensitive dependence on initial conditions for one-dimensional maps, *Comm. Math. Phys.* 70, 133-160 (1979).
- Guckenheimer, J.: The growth of topological entropy for one-dimensional maps, pp. 216-223 of *Global Theory of Dynamical Systems*, ed. Nitecki and Robinson, *Lecture Notes in Math.* 819, Springer 1980.
- Jakobson, M.V.: Structure of polynomial mappings on a singular set, *Mat. Sbornik* 77, 105-124 (1968) (= *Math. USSR Sb.* 6, 97-114 (1968)).
- Jakobson, M.V.: On smooth mappings of the circle into itself, *Mat. Sb.* 85, 163-188 (1971) (= *Math. USSR Sb.* 14, 161-185 (1971)).
- Jakobson, M.V.: Absolutely continuous invariant measures for one-parameter families of one-dimensional maps, *Com. Math. Phys.* 81, 39-88 (1981).
- Jonker, L.: Periodic orbits and kneading invariants, *Proc. London Math. Soc.* 39, 428-450 (1979).
- Jonker, L.: A monotonicity theorem for the family $f_a(x) = a - x^2$, *Proc. A.M.S.* 85, 434-436 (1982).
- Jonker, L., Rand, D.: A lower bound for the entropy of certain maps of the unit interval, preprint, Kingston, Ontario, and Univ. of Warwick (1978).
- Jonker, L., Rand, D.: The periodic orbits and entropy of certain maps of the unit interval, *J. London Math. Soc.* (2), 22, 175-181 (1980).
- Jonker, L., Rand, D.: Bifurcations in one dimension; I, The nonwandering set, *Invent. Math.* 62, 347-365 (1981); II, A versal model for bifurcations, *Invent. Math.* 63, 1-15 (1981).
- Julia, G.: Mémoire sur l'iteration des fonctions rationnelles, *J. de Math.* (Liouville), ser. 7, 4, 47-245 (1918).
- Kolata, G.B.: Cascading bifurcations: the mathematics of chaos (news article), *Science* 189, 984-985 (1975).
- Lanford, O.E.: Smooth transformations of intervals, *Séminaire Bourbaki* 563 (1980/81), *Lecture Notes Math.* 901, Springer 1981.
- Lanford, O.E.: A computer-assisted proof of the Feigenbaum conjectures, *Bull. A.M.S.* 6, 427-434 (1982).
- Lang, S.: *Differential Manifolds*, Addison-Wesley, 1972.
- Li, T., Yorke, J.A.: Period three implies chaos, *Amer. Math. Monthly* 82, 985-992 (1975).

- Lorenz, E.: On the prevalence of aperiodicity in simple systems, pp. 53-57 of *Global Analysis*, ed. Grmela and Marsden, Lecture Notes in Math. 755, Springer 1979.
- May, R.M.: Biological populations obeying difference equations: stable points, stable cycles and chaos, *J. Theor. Biol.* 51, 511-524 (1975).
- May, R.M.: Simple mathematical models with very complicated dynamics (review article), *Nature* 261, 459-467 (1976).
- Metropolis, N., Stein, M.L., Stein, P.R.: On finite limit sets for transformations on the unit interval, *Journal of Combinatorial Theory (A)* 15, 25-44 (1973).
- Misiurewicz, M., Szlenk, W.: Entropy of piecewise monotone mappings, *Asterisque* 50, 299-310 (1977) and *Studia Math.* 67, 45-63 (1980).
- Oster, G., Guckenheimer, J.: Bifurcation phenomena in population models, pp. 327-353 of "The Hopf Bifurcation and its Applications," by Marsden and McCracken, Springer 1976.
- Parry, W.: Symbolic dynamics and transformations of the unit interval, *Trans. Amer. Math. Soc.* 122, 368-378 (1966).
- Rothschild, J.: On the computation of topological entropy, Thesis, CUNY, 1971.
- Ruelle, D.: Applications conservant une mesure absolument continue par rapport a dx sur $[0,1]$, *Comm. Math. Phys.* 55, 47-51 (1977).
- Šarkovskii, A.N.: Coexistence of cycles of a continuous map of a line into itself (Russian), *Ukr. mat. Ž.* 16, 61-71 (1964).
- Singer, D.: Stable orbits and bifurcation of maps of the interval, *SIAM J. Appl. Math.* 35, 260-267 (1978).
- Smale, S., Williams, R.: The qualitative analysis of a difference equation of population growth, *J. Mathematical Biology* 3, 1-4 (1976).
- Štefan, P.: A theorem of Šarkovskii on the existence of periodic orbits of continuous endomorphisms of the real line, *Com. Math. Phys.* 54, 237-248 (1977).
- Straffin, P.D., Jr.: Periodic points of continuous functions, *Math. Mag.* 51, 99-105 (1978).
- Targonski, G.: *Topics in Iteration Theory (Studia Math.: Skript 6)*, Vandenhoeck u. Ruprecht, Göttingen 1981.
- Weil, A.: Numbers of solutions of equations in finite fields, *Bull. Amer. Math. Soc.* 55, 497-508 (1949).
- Williams, R.F.: The zeta function in global analysis, pp. 335-339 of *Global Analysis*, ed. Chern and Smale, Proc. Symp. Pure Math. 14, A.M.S. 1970.