One-Dimensional Dynamics

by

Welington de Melo and Sebastian van Strien

To Gilza, Hilary
Daniel and David

Preface

One-dimensional dynamics has developed in the last decades into a subject in its own right. Yet, many recent results are inaccessible and have never been brought together. For this reason, we have tried to give a unified account of the subject and complete proofs of many results. To show what results one might expect, the first chapter deals with the theory of circle diffeomorphisms. The remainder of the book is an attempt to develop the analogous theory in the non-invertible case, despite the intrinsic additional difficulties. In this way, we have tried to show that there is a unified theory in one-dimensional dynamics. By reading one or more of the chapters, the reader can quickly reach the frontier of research.

Let us quickly summarize the book. The first chapter deals with circle diffeomorphisms and contains a complete proof of the theorem on the smooth linearizability of circle diffeomorphisms due to M. Herman, J.-C. Yoccoz and others. Chapter II treats the kneading theory of Milnor and Thurston; also included are an exposition on Hofbauer's tower construction and a result on full multimodal families (this last result solves a question posed by J. Milnor). In the third chapter we analyze structural stability and hyperbolic properties of one-dimensional systems with a simplified proof of a result of R. Mañé. This chapter has a section describing the ideas that are used in the theory of rational maps on the Riemann sphere corresponding to those which are used here; it also serves as an introduction to the last chapter. The fourth chapter shows that Denjoy's result for circle diffeomorphisms also holds for non-invertible maps and that the period of the attractors of these maps is bounded. The first part of this result extends work of J. Guckenheimer, J.-C. Yoccoz, A.M. Blokh and M.Yu. Lyubich. The first three sections of Chapter IV give all the distortion tools necessary for the remaining chapters.

From Chapter V onwards, we confine ourselves to unimodal maps. The fifth chapter gives an account of the state-of-the-art concerning ergodicity, Cantor attractors and invariant measures. We shall give a unified treatment of results of A.M. Blokh, P. Collet, J.-P. Eckmann, J. Guckenheimer, S.D. Johnson, M.Yu. Lyubich, G. Keller, M. Martens, M. Misiurewicz, T. Nowicki, S. van Strien and partly of the results of Jakobson and Świątek. Also we give a complete account of the proof of M. Benedicks and L. Carleson of Jakobson's result on invariant measures in families of maps. In the final chapter, a complete proof is given of D. Sullivan's recent results on the universal structures of one-dimensional systems coming from renormalization. This chapter uses ideas from complex analysis, in particular Teichmüller theory. The necessary background for this book is treated in an appendix.

This has resulted in a lengthy volume. However, the reader should not be discouraged by this. The book is organized in such a way that each of the chapters is essentially independent of the others. For this reason, we have defined some concepts twice or more. (For example the notion of wandering interval is defined in both Chapter I and II.)

We have used Chapter II of this text for an introductory course at first year graduate level on the combinatorial properties of piecewise monotone maps. Viviane Baladi used Chapter I and the first part of Chapter II for an 8 week introductory graduate course in dynamical systems. Sections IV.1-IV-3 and V.1-V.6 were used for another course on the metric theory of unimodal maps. A more advanced graduate course on the theory of renormalizations included Sections III.1, IV.1-IV.3 and Chapter VI. Most exercises include substantial hints (some of these hints are even complete proofs).

We should emphasize that we have not tried to give a complete account of all developments in one-dimensional dynamics. Most notably we have only given a short introduction to the combinatorial theory of one-dimensional systems; for more references on this see Section II.11 and recent monographs by Sarkovskii et al. (1989), Alsedà et al. (1990), Block and Coppel (1992) and by Misiurwicz and Nitecki (1991). We have not touched on thermodynamical theory, singularity spectra, decay of correlations and such matters, though some references to these subjects are given.

This book grew from notes written by the first author which were distributed during a course organized by the Brazilian Mathematical Society. The first part of this book retains the same structure as those notes but has been very much expanded. Moreover, the last two chapters are new. In these chapters results are given on ergodic properties of one-dimensional dynamical systems and also a complete account is given of Dennis Sullivan's proof of the renormalization conjectures of Feigenbaum, Coullet and Tresser. Exercises have been added.

We wish to thank Klaas Pieter Hart and Lucio Rodrígues for their help with making TEX-macros and Tineke Mulder, Kitty Cijfer and Ruud Schotting for their practical help. We also wish to thank Viviane Baladi, Thierry Barbot, Christian Bonatti, Karen Brucks, Henk Bruin, Robbert Fokking, Jean-Marc Gambaudo, Ernst Hansen, Toby Hall, Gerhard Keller, Bernd Krauskopf, Ricardo Mañé, Hilary Marland, Marco Martens, Gilza de Melo, Pedro Mendes, Jack Milnor, Tomasz Nowicki, Hiroe Oka, Paulo Sad, Atsuro Sannami, Mitsu Shishikura, Charles Tresser, Masato Tsujii, and Edson Vargas for reading parts of the manuscript. Their comments, in particular those of Viviane Baladi, Henk Bruin, Bernd Krauskopf and Tomasz Nowicki, have helped us to eliminate many mistakes. Duncan Sands' comments on a previous version of the proof of Benedicks and Carleson's result (Section V.6) have been very useful. Conversations with Michael Lyubich were very helpful for us in completing the proof of the Grötzsch Inequality in the proof of Proposition VI.7.1. Frederick Gardiner's comments on the proof of the Coiling Lemma, see Theorem VI.8.2, were also very valuable. We would like to thank Dennis Sullivan for many inspiring discussions and in particular spending many hours with us explaining his renormalization results. The first author would like to thank IHES and the Technical University of Delft for the hospitality during several visits that were essential for the preparation of this book.

Amsterdam, Rio de Janeiro
December 1992

Welington de Melo
Sebastian van Strien

Contents

Chapter 0.

Introduction

This book is about real one-dimensional discrete dynamical systems. We consider continuous maps $f\colon N \to N$ where N is an interval or a circle and – if this leads to better results – we shall often assume that f is smooth. Typical examples we will encounter are maps of the form

$$f : [0,1] \to [0,1] \text{ defined by } f(x) = ax(1-x)$$

or

$$f : \mathbb{R}/\mathbb{Z} \to \mathbb{R}/\mathbb{Z} \text{ defined by } f(x) = x + a\sin(2\pi x) + \alpha \bmod 1.$$

We will consider iterates f^n of f. These are inductively defined by $f^0(x) = x$ and $f^n(x) = f(f^{n-1}(x))$ for $n = 1, 2, \ldots$. Our aim is to describe the orbits $x, f(x), f^2(x), \ldots$ of points x and to describe how these are distributed. Even for the simple maps mentioned above this turns out to be highly non-trivial and also extremely interesting, both from a mathematical and an applied point of view. There are many motivations for studying such dynamical systems. Let us mention a few.

1. Historical. Iterating one-dimensional maps has a very long history. After all, to construct an accurate calendar, the Babylonians had to consider – in our terminology – a rotation of the circle and give a precise estimate for its angle α of rotation based on a piece of its orbit. For this they – and later the Greeks – considered the line $(t, t\alpha)$ in the plane and developed a continued fraction algorithm to estimate its slope α. For references on this we refer to Fowler (1987) and Series (1983). Ever since, continued fractions have played an important role in mathematics and in particular in number theory. It was Poincaré, in the 1880s, who generalized this point of view by considering the dynamics of more general maps of the circle. A detailed understanding of the dynamics of these maps leads to small-divisor problems and plays an important role in the modern theory on celestial mechanics.

Of course, since the 18th century one has another important one-dimensional dynamical system: the iteration scheme to determine the zeros of a function.

These are now referred to as the Newton-Raphson method – for a debate on this attribution see Kollerstrom (1992). Because one may have several zeros it is obvious that something interesting is occuring. Surprisingly, only at the beginning of this century was a more systematic analysis made of this algorithm in the complex plane by Julia and Fatou and more general non-linear iteration problems were considered.

2. Mathematical beauty. As we hope to show in this book, the theory of one-dimensional dynamical systems has a very beautiful structure. It is built-up in several layers: the first one is based on the order structure on an interval or a circle and the Intermediate Value Theorem for continuous maps. This already leads to a surprisingly rich 'combinatorial' theory. The second layer is connected to the affine structure and based on the fact that in one-dimensional space the size (i.e., volume) and the diameter of a connected set are the same: an example of a tool of this type is the Mean Value Theorem. Thirdly, one has the projective group acting on the real line which will lead to the Koebe Principle. These three structures allow the study of the orbit of a configuration of respectively two, three and four points. Finally, an interval or a circle can be embedded in the complex plane where one has the notion of conformal and quasiconformal maps. This leads to ways to define 'geodesic arcs' between such dynamical systems.

Because one has such a good mathematical framework, the general theory of one-dimensional dynamical systems is very well developed. For example, we can completely describe the dynamics which can occur in these systems in a topological sense. This dynamics can be extremely complicated (for example infinitely many periodic points can coexist in a far more complicated way than in the usual horseshoe maps) and yet is completely understood. In a metric sense, the orbit structure is also increasingly understood. Surprisingly, one often has rigidity (or universality as physicists choose to call it): the metric structure of orbits is completely determined by the topological one.

Moreover, the mathematical tools that are used in this theory come from different and beautiful branches of mathematics – number theory, topology, ergodic theory, complex analysis, real analysis, general dynamical systems and foliation theory to name a few.

3. Displays many features of higher-dimensional dynamics. Since Newton it has been common to model 'real life' by ordinary differential equations. Of course, for theoretical and practical reasons these are often discretized. However, there is a deeper relationship between flows and discrete systems: from the beginning of this century it was realized that discrete systems also arise from flows with periodic orbits as the first return map to some codimension-one section. In this way a flow in a three-dimensional space is related to an invertible discrete system in a two-dimensional space. For example, the Hénon map

$$(x, y) \mapsto (1 - ax^2 + y, bx)$$

is considered to be a natural model for the first return map of a three-dimensional flow. If the flow is dissipative then the corresponding invertible system is area contracting. Intuitively, this should mean that the essential dynamics is confined to a one-dimensional subspace; so if the flow can be reasonably modelled by the Hénon map then one may expect that the three-dimensional flow has essential features of the dynamics of the interval map $f(x) = 1 - ax^2$.

Indeed, there is increasing evidence that higher-dimensional dynamical systems can often be 'reduced' to one-dimensional systems. In some cases nothing is lost in this reduction. For example, Smale's solenoidal diffeomorphism (1967) on a full torus $D^2 \times S^1$ can be completely reduced to a degree two map of the circle. Another example is of course the geometric model of the Lorenz equations. The chaotic dynamics of these equations became only understood after it was realized that some first return map is essentially one-dimensional, see Guckenheimer (1976). Similarly, especially through the work of M. Morse it is known that geodesic flow on a hyperbolic surface are very closely related to one-dimensional dynamical systems, see for example the papers in Bedford et al. (1991).

In some other cases, the relationship is less clear but partial results can still be proved. For example, the bifurcation structure of higher-dimensional systems can be understood from the bifurcations of one-dimensional maps, see for example Collet et al. (1980), Levi (1981), Van Strien (1981) and Holmes and Whitley (1984). Another example is Sarkovskii's beautiful theorem (1964) on the periodic orbits of continuous interval maps (which was rediscovered by Li and Yorke (1975) who popularized it under the name 'period three implies chaos'). This result also has two-dimensional analogues, see Boyland (1987), (1988) and Gambaudo et al. (1990). Moreover, dynamical properties of the Hénon maps $(x, y) \mapsto (1 - ax^2 + y, bx)$ can sometimes be completely reduced to those of one-dimensional maps. In Gambaudo et al. (1989) it is shown there are Hénon maps with an infinite number of periodic points which are 'Ω-conjugate' to the Feigenbaum quadratic interval map. Furthermore, a deep result of Benedicks and Carleson (1991) shows that Hénon maps can have 'strange attractors'. Their method of proof is to extend one-dimensional ideas.

Finally, as is not surprising, many results and techniques are very similar to those developed for iterations of rational maps on the Riemann sphere.

4. Relations with other areas of mathematics. As mentioned before, the theory of dynamical systems brings together many ideas from different branches of mathematics. However, dynamical systems also gives something in return! For example, much in differential topology is linked to dynamical systems. To give just a few examples, the work of Smale on the Poincaré Conjecture is based on gradient flows. Several questions in the foliations are related to dynamical systems, see for example Sacksteder (1965). Thurston's work on isotopy classes on surfaces uses a dynamical approach: by iterating closed curves, he constructs maps on branched one-dimensional manifolds and his pseudo-Anosov maps, see for example Casson and Bleiler (1988) and Fathi et al. (1979). We should also mention the theory of Fuchsian groups, see for example Sullivan (1980) (and other essays in this memorial volume). There is also a strong link with statistical mechanics (via symbolic dynamics, Perron-Frobenius-type operators a.s.o.).

Altogether we may conclude that the theory of dynamical systems plays a central role in mathematics!

5. They model 'real' applications. In some cases, one-dimensional dynamical systems accurately model 'real' applications. For example, many 'real life' problems can be modeled by Hamiltonian dynamical systems. Often, because of symmetries, these have many first integrals which implies that the flow takes place on tori. Many of these tori even persist if one perturbes the system because of K.A.M. results. In this way circle diffeomorphisms can arise naturally as a first return map to a smooth flow without singularities on a two-dimensional torus (this happens for example in a forced pendulum or in electrical networks).

Another application is to biological systems. The simplest growth model is that the number of animals x_{n+1} at time $n+1$ only depends on x_n. This leads to studying iterates of one-dimensional maps.

In some other cases, one can empirically see that something should be modeled by a one-dimensional dynamical system. For example, there are chemical reactions (e.g. the Belousov-Zhabotinskii reaction) whose data hints to interval dynamics: measuring some quantity at discrete times $n \in \mathbb{N}$ one gets a sequence x_n of real numbers; one seems to have that x_{n+1} is a continuous, but not an invertible function of x_n, for references see for example Schuster (1989), Cvitanović (1989), or Jackson (1989/1990).

6. They 'explain' experimental patterns. Since Hénon maps are natural models for first return maps of three-dimensional flows, many experimentally observed features are familiar to those who have studied one-dimensional systems. In several fluid experiments, for example the Rayleigh-Bénard experiment, bifurcations are observed which have many similarities with the bifurcations of the quadratic family $f_a(x) = ax(1-x)$, see for example the papers in Cvitanović (1989). In Figure 0.1 we have drawn the well-known bifurcation diagram of the quadratic family.

The bifurcation diagram of the quadratic family is not only qualitatively but also quantitatively ubiquitous. Indeed, let a_n be the smallest parameter for which f_{a_n} has a periodic orbit of period 2^n. Then one can show that a_n converges to some number a_∞ as $n \to \infty$ and that

$$\lim_{n\to\infty} \frac{a_{n+1} - a_n}{a_{n+2} - a_{n+1}}$$

tends to some number $\delta = 4.6692016091029\ldots$. This limit is the same for any smooth unimodal family of maps $\tilde{f}_a$, provided each of these maps also has a non-zero second derivative at its extremum. In fact it is even the same for maps from the Hénon family. So δ is a universal number! Moreover, in experiments one often find similar bifurcation diagrams with the same rate δ of convergence of period doubling.

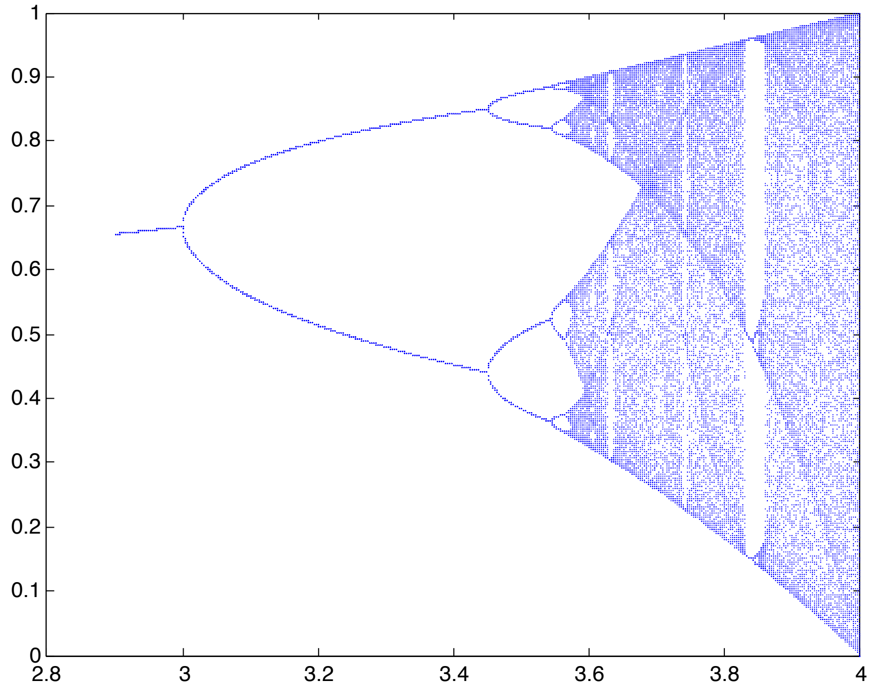

Fig. 1: The bifurcation diagram of $f_a(x) = ax(1-x)$

Similarly, one can prove that the limit

$$\lambda(a,x) = \lim_{n\to\infty} \frac{1}{n} \log |Df_a^n(x)|$$

exists and is equal to some constant $\lambda(a)$ for Lebesgue almost all x. This number $\lambda(a)$ is called the Liapounov number. Experimentally measured Liapounov numbers often also have the same kind of parameter dependence as those in the quadratic family, see for example the papers in Cvitanović (1989).

7. It is fun!

Fig. 0.2. The Liapounov number of $f_a(x) = ax(1-x)$ as a function of a.

In this book we will study one-dimensional dynamics from a mathematical point of view. Questions about one-dimensional dynamics fit into four categories: combinatorial, topological, ergodic and smooth. Let us first describe these briefly.

1. *Combinatorial.* The main distinguishing feature of real one-dimensional dynamics is that an interval or a circle has a natural ordering. This allows the development of a 'combinatorial theory' which describes properties of orbits related to this ordering. Examples of results in this combinatorial theory are Sarkovskii's hierarchy of periodic orbits and the theory of kneading invariants (or rotation numbers).

2. *Topological.* In the topological theory we want to describe the dynamics of these maps from a topological viewpoint. What are the attractors, which maps are conjugate and which are structurally stable?

3. *Ergodic.* In the ergodic theory we want to know the behaviour of typical orbits: to which sets are randomly chosen points attracted?

4. *Smooth.* Finally, one would like to establish smoothness properties of conjugacies between two systems. Surprisingly, two conjugate maps are sometimes 'smoothly' conjugate, simply because they satisfy some suitable combinatorial and topological conditions.

Now we will discuss and survey these theories in more detail.

1.1. The combinatorial theory

One of the main questions in the field of dynamical systems is whether two systems have 'the same behaviour'. Often one says that $x_{n+1} = f(x_n)$ and $y_{n+1} = g(y_n)$ are 'the same' if they are identical up to a coordinate change. This means that there is a homeomorphism $h\colon N \to N$ such that $h \circ f = g \circ h$. In this case $h(f^n(x)) = g^n(h(x))$ so h maps orbits of f onto orbits of g and we say that f and g are *topologically equivalent* or *conjugate*. However, in one-dimensional dynamics it turns out to be useful to first study a weaker notion of equivalence. Therefore, we say that $f, g : N \to N$ are *combinatorially equivalent* if, roughly speaking, there exists an order-preserving surjective 'map' $h : N \to N$ with the property that the image of a point is either again a unique point or a single closed interval such that $h \circ g = f \circ h$. So h may collapse some intervals to points or blow-up some points to intervals. (A precise definition will be given in Section II.3.) For those who are familiar with kneading theory we should remark that two maps without periodic attractors and without wandering intervals are combinatorially equivalent if and only if they have same kneading invariants. As we will see, in Sections I.2 and II.6, maps from a very large class are conjugate if and only if they are combinatorially equivalent.

Poincaré (1880) realized that any circle homeomorphism without periodic points is combinatorially equivalent to a rotation. Moreover, two such homeomorphisms are combinatorially equivalent if and only if their rotation numbers are the same. The corresponding statements for continuous piecewise monotone interval maps were shown much later. Parry (1964) showed that interval maps are combinatorially equivalent (or, in fact, semi-conjugate) to piecewise linear maps. Also in 1964, Sarkovskii's proved his remarkable theorem that a periodic point of period three implies the existence of periodic points of any period. Later, Metropolis, Stein and Stein (1973) and – in their well-known handwritten notes – Milnor and Thurston (1977) proved that two piecewise monotone continuous interval maps are combinatorially equivalent if and only if the orbits of their turning points are ordered in the same way. Moreover, Milnor and Thurston showed that any such map is combinatorially equivalent to a polynomial map and that the topological entropy depends continuously on the map.

In Section I.1 the combinatorial theory in the case of circle homeomorphisms will be discussed. For these maps one can define a 'rotation number'. This is usually done by using some limit related to a real valued lift or, alternatively, by some counting argument. However, this cannot be done in the non-invertible case. So instead we associate to a circle diffeomorphism a symbolic coding of orbits and also a sequence of first return maps. These first return maps define return times p_n, q_n and a_n. Two maps are combinatorially equivalent if and only if the corresponding numbers coincide. As one would expect one can also define the rotation number of the map in terms of these numbers. Indeed, the ratios p_n/q_n are the best approximants of some number $\rho(f)$ and the integers a_n are the coefficients in the continued fraction expansion of $\rho(f)$. As we will show, this definition of $\rho(f)$ coincides with the usual definition of the rotation number of f.

There is another reason for defining the rotation number in this combinatorial way. As we will see, throughout the whole book first return maps and disjointness properties of iterates of intervals play a vital role. So we would like to stress these notions here. And indeed, it turns out that all the well-known properties of these approximants follow simply from the dynamics of these first return maps (and so are proved dynamically rather than using the less intuitive algebraic manipulations). For example, the property that p_n/q_n is the best approximation of $\rho(f)$ is a trivial consequence of the fact that q_n is the return time of some first return map.

As it turns out, for piecewise monotone interval maps a completely similar coding can be made. Moreover, in each reasonable family f_μ of smooth maps we can find a map f_{μ_0} which is combinatorially equivalent to a given topological piecewise monotone map g. However, interval maps can have a more complicated dynamics. For example, such maps can periodically permute smaller and smaller intervals whose orbits shrink down to an invariant Cantor set; such maps are called *infinitely renormalizable*. Feigenbaum maps are one of the many maps with this property.

This and the theorems mentioned above are proved in Chapter II. These results will also explain why the bifurcations of such interval families are ordered in such a universal way.

1.2. The topological theory

Suppose two maps are combinatorially equivalent. When are they also topologically equivalent, i.e., conjugate? As we shall see this question is very closely related to the existence of *wandering intervals* and *periodic attractors*. Here an interval is said to be wandering if all its forward iterates are disjoint and if these iterates do not tend to a periodic orbit. Similarly, a closed forward invariant set A is said to be *attracting* if its basin $B(A) = \{x\,;\ f^n(x) \to A \text{ as } n \to \infty\}$ satisfies the following two properties: i) the closure of $B(A)$ contains intervals, ii) each closed forward invariant proper subset A' of A has a smaller basin of attraction, i.e., $\mathrm{cl}\,(B(A)) \setminus \mathrm{cl}\,(B(A'))$ contains intervals. As it turns out, a topological description of the dynamics of maps is possible once we can show that they have no wandering intervals and that the period of periodic attractors is bounded.

Although the non-existence of wandering intervals is a topological property of a map, it can be shown that these intervals cannot exist for a very large class of smooth maps. Therefore we restrict our attention to maps f with some smoothness properties. The first result in this direction is for smooth circle diffeomorphisms and is due to Denjoy (1932). This proof is given in Section I.2. If f has *critical points* (i.e., points x where $f'(x) = 0$) then the situation is more involved. During the last ten years, contributions to the general result were made by Guckenheimer, Yoccoz, de Melo and Van Strien, Blokh and Lyubich and finally Martens, de Melo and Van Strien. Initially, these results assumed that the map f satisfies some magic condition:

$$Sf = \frac{f'''}{f'} - \frac{3}{2}\left[\frac{f''}{f'}\right]^2 \quad < 0$$

which is usually called the negative Schwarzian condition. However, later it was shown that this assumption is superfluous and not so mysterious after all. Indeed, as we shall see in Chapter IV, the situation is now completely understood: one-dimensional maps which satisfy some smoothness properties and also have no flat critical points have no wandering intervals. The proof uses a mixture of combinatorial and distortion techniques. The *combinatorial* techniques are essentially based on the following considerations:

1. closest returns and first return maps;
2. disjointness of certain orbits of intervals;

The aim of the *distortion* techniques is to give ‘bounds on the shape’ of high iterates of f. Of course if f has *critical points* then f is certainly not almost linear. After all, even $|Df(x)|/|Df(y)|$ is then unbounded. Instead these distortion results give

3. bounded distortion in the absence of critical points;
4. the Koebe Principle bounding the distortion ‘away from the critical values’.

These distortion results will turn out to be closely related to similar distortion results from complex analysis.

As already mentioned, the non-existence of wandering intervals is closely related to a topological description of the dynamics. For example, let $\Omega(f)$ be the *non-wandering set* of f. This is the set of points x such that for any neighbourhood U of x there exists $n > 0$ with $f^n(U) \cap U \neq \emptyset$ and so arbitrarily close to these points are points which come back nearby. (Of course, the notions of 'wandering interval' and 'wandering set' are quite different.) The non-existence of wandering intervals immediately gives that the non-wandering set of a C^2 circle diffeomorphism consists of periodic orbits (all of the same period) or is equal to the circle. In the last case the diffeomorphism is conjugate to a rotation. In the case of interval maps the situation is more complicated. In this case one can often decompose or *renormalize* a map in the sense that there exists an interval I such that some iterate f^n maps I into itself, so that $I, \dots, f^{n-1}(I)$ are disjoint. So if a map is renormalizable then either points always stay outside the orbit of I or they enter eventually. For points which enter eventually, it is more convenient to study the restriction of f^n to I, i.e., to use a 'microscope'. Sometimes this map f^n is again renormalizable. In this way we will prove in Chapter III that interval maps can have three types of topological attractors:

1. a periodic attractor;
2. an invariant *solenoidal* Cantor set (this means that f acts as an adding machine on this set and that f is infinitely often renormalizable);
3. a finite union of intervals containing a dense orbit;

In Chapter III this result will be stated more precisely.

Another aspect of the topological theory deals with the structural stability of maps. As usual we say that a map is *structurally stable* if any nearby map is conjugate to it. As is easy to see, no map is structurally stable in the C^0 topology. Therefore one usually considers C^r maps with the C^r topology where r is at least 1 (and often $r \geq 2$). Circle diffeomorphisms are C^r structurally stable if and only if they are Morse-Smale. This means that each recurrent point is periodic and each periodic point is hyperbolic, i.e., if $f^n(p) = p$ then $|Df^n(p)| \neq 1$. For non-invertible maps the situation is more interesting. These can be structurally stable even if many recurrent points are non-periodic. It turns out that in order to study structural stability one has to introduce a more general notion of hyperbolicity. A forward invariant set K is *hyperbolic* if there exists $C > 0$ and $\lambda > 1$ such that $|Df^n(x)| \geq C\lambda^n$ for each $x \in K$ and each $n \in \mathbb{N}$. Furthermore, a map is called Axiom A if the set of points which are not contained in the basin of a periodic attractor is hyperbolic. It turns out that any Axiom A map, for which the orbits of the critical points are disjoint, is structurally stable. This result is of course the analogue of the higher-dimensional result. In higher dimensions, Axiom A occurs rarely while in one-dimensional dynamics it is rather common. Indeed, as Mañé (1985) has shown, any C^2 map for which each critical point is attracted to a periodic orbit and for which each periodic point is hyperbolic, satisfies the Axiom A property. So a non-uniform condition on each periodic orbit suffices to get hyperbolicity. Even so, it is unknown whether hyperbolicity of the non-wandering set holds for generic maps. These results are proved and discussed in Chapter III including an elementary proof of Mañé's result.

In fact, even if some of the critical points do not tend to a periodic orbit but nonetheless never enter a neighbourhood of the critical points, something similar holds. Such maps are said to have the Misiurewicz property and Misiurewicz (1981) and Van Strien (1990) have shown that one has a kind of non-uniform hyperbolicity for these maps. In the discussion of ergodic properties of maps this will play an important role.

1.3. The ergodic theory

What points have 'typical' orbits? This vague question is related to the ergodicity of the map f. A map is *ergodic* if for any two measurable forward invariant sets X and Y whose intersection has Lebesgue measure zero, either X or Y has zero Lebesgue measure. So this means that one cannot decompose the space into 'visible' forward invariant sets. It is not hard to show that C^2 circle diffeomorphisms without periodic orbits are ergodic. In Blokh and Lyubich (1986), (1989c) and Lyubich (1990) the corresponding result is shown for interval maps. However, even if many orbits are dense it could still be the case that the forward limits of almost all points are contained in a set with zero Lebesgue measure. To study such questions, we say, following Milnor (1985), that a closed forward invariant set A is a *metric attractor* if the basin $B(A)$ of A, i.e., $B(A) = \{x\,;\; f^n(x) \to A \text{ as } n \to \infty\}$ satisfies

1. the measure of $B(A)$ is positive;
2. each closed forward invariant subset A' which is strictly contained in A has a smaller basin of attraction: $B(A) \setminus B(A')$ has positive Lebesgue measure.

Blokh and Lyubich (1986) have shown that interval maps with negative Schwarzian derivative can only have finitely many attractors. From this and Guckenheimer (1979) it follows that, for unimodal maps, the attractor is either a Cantor set, a finite union of intervals or a periodic orbit. If it is a Cantor set then it is equal to $\omega(c)$ and Martens (1990) showed that this last set either contains intervals or has Lebesgue measure zero. Related results were obtained by Guckenheimer and Johnson (1990) and Keller (1990a). Lyubich and Milnor (1991) and Lyubich (1992) completed the classification of these attractors, by showing that an attractor of a smooth unimodal map with a quadratic critical point which is a Cantor set is necessarily solenoidal (and therefore map is infinitely often renormalizable). For these results we refer to Section V.1.

Moreover, one would like to know how orbits are distributed. One way to analyze this is through *invariant measures*. So suppose that a map f has a probability measure μ which is invariant: $\mu(A) = \mu(f^{-1}(A))$ for each measurable set A. If f is ergodic and μ is absolutely continuous then Birkhoff's Ergodic Theorem gives us the asymptotic distribution of orbits: for each open set U with a non-empty intersection with the support of μ one has for Lebesgue almost all x,

$$\lim_{N\to\infty} \frac{\#\{i\,;\ 0 \le i < N,\ f^i(x) \in U\}}{N} = \mu(U) > 0.$$

To give an example, let f be a circle diffeomorphisms without periodic points and U an open interval in S^1. Then

$$\lim_{N\to\infty} \frac{\#\{i\,;\ 0 \le i < N,\ f^i(x) \in U\}}{N}$$

exists and is independent of x. Hence there is an invariant probability measure μ on S^1 such that this limit is equal to $\mu(U)$. For example, if R is a rotation without periodic points then this measure is simply the Lebesgue measure. So, under a rotation, iterates of x are *uniformly distributed.* An equivalent way to express this is to denote the Dirac measure in a point $z \in S^1$ by δ_z and to say that $\frac{1}{n}\sum_{i=0}^{n-1} \delta_{R^i(x)}$ tends to the Lebesgue measure on S^1 in the weak topology.

However, orbits under non-linear circle diffeomorphisms are in general far from evenly distributed. Indeed, even though the orbit of any point is dense there exists a diffeomorphism f as above and a set $K \subset S^1$ of full Lebesgue measure such that

$$\lim_{N\to\infty} \frac{\#\{i\,;\ 0 \le i < N,\ f^i(x) \in K\}}{N} = 0.$$

Constructions of such maps were given by Finzi (1950) and Arnol'd (1961). As will become clear in Sections I.4 and I.5, it is no great surprise that something like this happens. After all, consider the family of diffeomorphisms $f(x) = x + a\sin(2\pi x) + \alpha \bmod 1$ and suppose that for a certain parameter (a,α) the map f has no periodic orbits but still there exists a point p which is 'almost periodic' (of, say, period n). Then of course orbits linger for an extremely long time near $\{p,\dots,f^{n-1}(p)\}$. So orbits are already far from evenly distributed. The examples of Arnol'd are constructed by repeating this infinitely often. On the other hand, Arnol'd (1961) and, in a more general context, Herman (1979), have shown that the invariant probability measure of a circle diffeomorphism whose rotation number is sufficiently irrational is absolutely continuous.

For interval maps the situation is similar: Johnson (1987) has given examples of quadratic maps with similar features to those of Arnol'd mentioned above. Moreover, Hofbauer and Keller (1990a) have even given examples of quadratic maps for which $\frac{1}{n}\sum_{i=0}^{n-1} \delta_{f^i(x)}$ tends to the Dirac measure of a repelling fixed point for almost all x. In other words, most orbits spend most of their time near a repelling point and this repelling fixed point is, from a statistical point of view, an attractor! We will prove this in Section V.5.

On the other hand, many quadratic maps $x \mapsto ax(1-x)$ have invariant probability measures which are absolutely continuous. Misiurewicz (1981), Collet and Eckmann (1983), Keller (1990) and Nowicki and Van Strien (1991a) have given conditions under which such invariant measures exist. All these conditions require some non-uniform type of hyperbolicity. We shall prove these results in Sections V.2-V.4. These conditions are satisfied for a large set of parameters. Indeed, Jakobson (1981) and also Benedicks and Carleson (1991) have shown that these measures exist for a set of parameters of positive Lebesgue measure. This result is proved in Section V.1. Here we will follow Benedicks and Carleson (1991). The ideas used in the proof of this last result also indicate that there might exist weak topological conditions which suffice for the existence of such absolutely continuous invariant measures.

1.4. The smooth theory

If we want to relate two conjugate maps f and g in detail then it would be very useful if they were smoothly conjugate. Surprisingly, there are several theorems stating that if two conjugate maps satisfy some suitable combinatorial and topological conditions – given by their kneading invariants – then the conjugacy between them is necessarily smooth.

The first result in this direction is for the circle. Provided a circle diffeomorphism has a rotation number which is sufficiently irrational then it is smoothly conjugate to a rotation. This result was first proved by Arnol'd (1961) for analytic diffeomorphisms which are close to a rotation. Moser (1966) and Herman (1979) greatly extended these results to smooth circle diffeomorphisms. We will prove this result in Chapter I using the version of Yoccoz (1984a). The basic tool in Herman's result is a sophisticated version of the distortion results used in Denjoy's theory and requires additional smoothness. In fact, any C^3 circle diffeomorphism f without periodic points is almost linear in the sense that there exists a sequence of iterates $q_n \to \infty$ such that $||Df^{q_n} - 1|| \to 0$. This already shows that the rotation is a good model for circle diffeomorphisms. Under suitable additional conditions on the rotation number, one gets that $||Df^k - 1||$ is bounded for all $k \in \mathbb{Z}$ and that f is C^1 conjugate to a rotation. There is a very well developed theory in which necessary and sufficient conditions are given in terms of the smoothness and its rotation number for a circle diffeomorphism to be C^k conjugate to a rotation.

Later, a similar result was discovered for infinitely renormalizable non-invertible maps. Feigenbaum (1979) and Coullet-Tresser (1978) found numerically that the map at the limit of period doubling has the following property: for each such map f with a quadratic critical point the ratios

$$\frac{|f^{2^{n+1}}(c) - c|}{|f^{2^n}(c) - c|}$$

tend to the same number (namely $0.3995\ldots$). Later, computer assisted proofs were given by Lanford (1984), see also Eckmann and Wittwer (1979), Campanino, Epstein and Ruelle (1981) and others. This discovery was the first indication of the existence of *metric rigidity* of the critical orbit. Let us formulate this notion more precisely. Two maps are said to have the same combinatorial type if their critical orbits are ordered in the same way. This order structure we refer to as the *combinatorial critical orbit type*. We say that a combinatorial critical orbit type implies *rigidity* if any two 'sufficiently regular' maps whose critical orbits have this combinatorial type are smoothly equivalent in the sense that there exists a smooth diffeomorphism of the real line that maps one critical orbit onto the other conjugating the maps along these orbits. In Chapter VI we shall prove a recent rigidity result, due to Sullivan (1992). In this result he considers critical orbit types which come from infinitely renormalizable maps of bounded type as will be defined in Section VI.1. For these maps the closure of the critical orbit is a Cantor set. For the proof of this result one defines an operator, called the renormalization operator, on some subset of the space of unimodal interval maps. This operator defines a dynamical system on an infinite-dimensional space and the rigidity theorems follow from a complete description of its dynamics: it has precisely the same dynamics as Smale's horseshoe. The well-known Feigenbaum map (which is the fixed point under the usually studied period doubling operator) corresponds to one of the fixed points of Smale's horseshoe.

One of the main differences between the circle diffeomorphism case and the non-invertible case, is that in the former the rotation is clearly the right 'model' map. In the latter, no well-understood model can be easily given. It is for this reason that Sullivan had to use complexifications of the maps and tools from Teichmüller theory to obtain the model by a contraction argument. We shall explain these ideas at length in the last chapter.

Recently, it was shown by Lyubich and Milnor that some other, non-renormalizable maps have universal structures, see Section VI.10. A smooth theory like the one which already exists for circle diffeomorphisms seems to emerge for non-invertible maps. One-dimensional dynamics is becoming a mature subject!

Chapter I.

Circle Diffeomorphisms

This chapter is devoted to the study of invertible one-dimensional dynamical systems. In the later chapters we shall see that, although the non-invertible case is quite different, many techniques and theorems for analyzing these non-invertible systems find their roots in the invertible case. One of our aims in this book is to emphasize these similarities.

The circle S^1 is defined to be the quotient space of the real line $\mathbb{R}$ by the group of translations by integers: $S^1 = \mathbb{R}/\mathbb{Z}$. Let $\pi\colon \mathbb{R} \to S^1$ be some covering map. Notice that $\pi(x) = \pi(y)$ if and only if $x - y \in \mathbb{Z}$ and π maps the open interval $(0,1)$ diffeomorphically onto $S^1 \setminus \{\pi(0)\}$. A map $f\colon S^1 \to S^1$ is said to be a C^r, $r \geq 0$, diffeomorphism if and only if there exists a C^r diffeomorphism $\hat{f}\colon \mathbb{R} \to \mathbb{R}$ such that $\pi \circ \hat{f} = f \circ \pi$. We say that $\hat{f}$ is a lift of f. Notice that if $\hat{g}$ is another lift of the same homeomorphism f then there exists an integer $n \in \mathbb{Z}$ such that $\hat{f}(x) - \hat{g}(x) = n$ for all $x \in \mathbb{R}$. A homeomorphism $\hat{f}\colon \mathbb{R} \to \mathbb{R}$ is a lift of some circle homeomorphism if and only if $\hat{f}(x+n) = \hat{f}(x) \pm n$ for every $x \in \mathbb{R}$. Hence, if $\hat{f}$ preserves orientation then $\hat{f} = Id + \phi$ where Id is the identity map and ϕ is periodic of period 1. In particular each translation $T_\alpha(x) = x + \alpha$ is a lift of a rotation R_α of the circle.

In Section 1 the dynamics of circle homeomorphisms will be studied from a combinatorial point of view. Here the only structure to be considered is the circular order of S^1. This theory goes back to Poincaré (1885), who introduced the important dynamical invariant called the *rotation number*. He proved that the rotation number of a homeomorphism is irrational if and only if it has no periodic points and moreover that a homeomorphism with an irrational rotation number is combinatorially equivalent to a rotation with the same rotation number. By that we mean that each orbit of the homeomorphism lies in the circle in the same order as an orbit of the corresponding rotation. More precisely, if $f\colon S^1 \to S^1$ is a homeomorphism without periodic points then there exists a rotation R such that the mapping $h\colon O_f(x) \to O_R(x)$ defined by $h(f^n(x)) = R^n(x)$ for all $n \in \mathbb{N}$ is order preserving. (Here $O_g(x)$ is the forward orbit of x under the map $g : S^1 \to S^1$.)

From Poincaré's theorem it follows, because h is order preserving and $O_R(x)$

is dense in S^1, that h extends continuously to a monotone map of the circle satisfying $h \circ f = R \circ h$. Since h is monotone, the inverse image of each point is either a point or a closed interval, and it maps orbits of f onto orbits of R. Moreover, h is strictly monotone if and only if $O_f(x)$ is dense. If h is not strictly monotone (and we will see in Section 2 that this can happen even if f is a C^1 diffeomorphism) we say that h is a semi-conjugacy between f and R.

The proof we present of the above theorem differs slightly from the usual one. We use the technique of symbolic dynamics, because in this way the treatment of the combinatorial aspects of invertible and non-invertible one-dimensional dynamical systems becomes more unified. Moreover it contains an introduction to the techniques of renormalization that will be discussed in Chapter VI. The use of coding for analyzing the dynamics of circle homeomorphisms and related problems has a long history, see Siegel et al. (1992) and Series (1985).

In Section 2 we will develop the theory of Denjoy (circa 1930) which describes the dynamics of diffeomorphisms from a topological point of view. We will need more smoothness and the main theorem states that a C^2 diffeomorphism without periodic points is in fact topologically conjugate to a rotation. The main ingredient is the control of the distortion of iterates of the original map when restricted to certain intervals of the circle. The distortion of a map, as defined in Section 2, is a measure of its non-linearity. We prove, following Denjoy, that all iterates of the original map, when restricted to some special intervals have bounded distortion. This technique will also play an important role later on in Chapter IV when the results of this chapter are generalized to non-invertible one-dimensional dynamical systems.

In Section 3 we will discuss the dynamics of circle diffeomorphisms under a more quantitative point of view. The main result states that a C^3 diffeomorphism whose rotation number satisfies some arithmetic condition, is C^1 conjugate to a rotation. The proof is based on a much finer control of the distortion of the iterates and for that a new technical tool is needed: the Schwarzian derivative. The above theorem is an important instance of a rigidity result: requiring that two diffeomorphisms be topologically conjugate, implies, under some arithmetical conditions, that the conjugacies have some smoothness properties. As we will see in Section 3, this is also related to ergodic properties of the diffeomorphism. A topological hypothesis, a condition on the rotation number, implies the existence of a nice invariant measure for the diffeomorphism. This is related to certain situations which occur for non-invertible dynamical systems and which will be discussed in later chapters.

In Section 4 one parameter families of circle diffeomorphisms will be discussed. It will be shown that for many families the rotation number is locally constant (as a function of the parameter) precisely when it is rational. This is called frequency locking and the regions in the parameter space which correspond to rational rotation numbers are called Arnol'd tongues. Using these families, it will be shown in Section 5 that the conjugacy between analytic circle diffeomorphisms is often not even absolutely continuous. This shows that one cannot drop the conditions on the rotation numbers which were imposed in

Section 3. Finally, in Section 6 it will be shown that for these families of diffeomorphisms, although there are many parameters for which one cannot hope to get smooth conjugacies with rotations, the set of parameters for which one does have smooth conjugacies has large measure.

1 The Combinatorial Theory of Poincaré

We consider the circle S^1 as the quotient space of the real line by the group of translations by integers: $S^1 = \mathbb{R}/\mathbb{Z}$ and we consider the circular ordering on S^1. Let $\pi\colon \mathbb{R} \to S^1$ be the quotient map. In S^1 we consider the metric and the orientation induced from the metric and orientation of the real line via π (hence the distance between any two points is at most $\frac{1}{2}$). We are interested in analyzing the dynamics of homeomorphisms $f\colon S^1 \to S^1$. The simplest homeomorphisms are rotations: those are the orientation preserving isometries of the circle. If x is a periodic point of period n of a rotation f then any other point y is also a periodic point of the same period. Indeed, the length of the positively oriented arc of S^1 from x to y is equal to the length of the positively oriented arc between $f^n(x) = x$ and $f^n(y)$. Hence $f^n(y) = y$.

If a rotation f does not have a periodic point then the orbit $O_f(x) = \{f^n(x)\,;\ n \in \mathbb{Z}\}$ is dense in the circle. Let us prove this: if F denotes the closure of $O_f(x)$ then F is a closed invariant set, i.e., $f(F) = F = f^{-1}(F)$ (the closure of an invariant set is invariant). Its complement, $A = S^1 \setminus F$, is open and also invariant (the complement of an invariant set is also invariant). If A is not empty and A_0 is a connected component of A then, for each n, $f^n(A_0)$ is also a connected component of A. The intervals $\{f^n(A_0)\,;\ n \in \mathbb{Z}\}$ cannot all be disjoint because they all have the same length. Consequently, there exists $m \in \mathbb{Z}$ such that f^m maps one of the intervals $f^n(A_0)$ homeomorphically onto itself. But this implies that f^m has a fixed point in the closure of $f^n(A_0)$. Since f does not have periodic points, this contradiction proves that $A = \emptyset$ and that all orbits are dense in the circle.

In this section we will discuss a result of Poincaré which distinguishes two cases.

1. If $f\colon S^1 \to S^1$ is a homeomorphism which has a periodic point then its dynamics turns out to be trivial: any orbit is asymptotic to a periodic orbit and (if f preserves orientation) any two periodic orbits have the same period.

2. If f does not have a periodic point then Poincaré's result asserts that there exists a rotation $g\colon S^1 \to S^1$ such that any orbit of f has the same order as any orbit of g. More precisely, the map $h\colon O_f(x) \to O_g(y)$ defined by $h(f^n(x)) = g^n(y), n \in \mathbb{Z}$ is monotone: if $u, v, w \in O_f(x)$ and v lies between u and w in the positive orientation of the circle then $h(v)$ lies between $h(u)$ and $h(w)$. From this fact it follows easily, since any orbit of g is dense in the circle, that h extends continuously to a monotone map $h\colon S^1 \to S^1$ satisfying the equation: $h \circ f = g \circ h$. We say that h is a

semi-conjugacy between f and g. From the above equation it follows that $h \circ f^n = g^n \circ h$, that is, h sends orbits of f into orbits of g. In general h is not a conjugacy because the inverse image of some point may be an interval.

We start with a very simple observation on the dynamics of an invertible continuous map of an interval. Let J be a closed interval and $f\colon J \to J$ be a continuous injective map. If f is orientation preserving (i.e., monotone increasing) then any orbit of f is asymptotic to a fixed point. In order to prove this, take $x \in J$ and let us first consider the case that $f(x) > x$. Then $f^2(x) = f(f(x)) > f(x)$ since f is monotone increasing. By induction we get $f^n(x) = f(f^{n-1}(x)) > f^{n-1}(x)$. Hence the sequence $f^n(x)$ is monotone increasing and converges to $y = \sup\{f^n(x)\}$. By the continuity of f, we get $f(y) = f(\lim_{n\to\infty} f^n(x)) = \lim_{n\to\infty} f^{n+1}(x) = y$. Therefore the ω-limit set of the orbit of x is the fixed point y. Similarly, if $f(x) < x$ then the sequence $f^n(x)$ is decreasing and also converges to a fixed point of f. If $f(J) = J$ we can apply the argument above to f^{-1} instead of f, and conclude that the α-limit set of any orbit of f is also a fixed point.

If f is orientation reversing (i.e., monotone decreasing), then the second iterate $f^2 = f \circ f$ is monotone increasing. Hence, using the previous argument for f^2, we get that any orbit of f is asymptotic to either the fixed point of f or to a periodic point of period two.

Let us now consider circle homeomorphisms. If a homeomorphism $f\colon S^1 \to S^1$ has a periodic point y of period k then y is a fixed point for the homeomorphism f^k. If f preserves orientation then f^k, restricted to the invariant interval $J = S^1 \setminus \{y\}$ is a monotone increasing map. Hence, as before, the ω-limit set of any orbit of f is a periodic orbit of period k because, under iteration by f^k, any point goes to a fixed point of f^k.

So consider the case that f reverses orientation. We claim that f has in this case precisely two fixed points. Let us prove this claim. Take x such that $f(x) \neq x$. Then take two open arcs $A = (p, x)$ and $B = (x, q)$ in S^1 starting at x and going in opposite directions; choose A and B maximal such that $f(A) \cap A = \emptyset$ and $f(B) \cap B = \emptyset$. Since f is orientation reversing, $f(A)$ and A are contained in one component of $S^1 \setminus \{x, f(x)\}$ and $f(B)$ and B in the other one. Moreover, by maximality $f(p) = p$ and $f(q) = q$. From this the claim easily follows. Since f^2 preserves orientation, we get as before that any point is asymptotic to one of the fixed points or to a periodic point of period two. Thus we have a complete description of the dynamics of a homeomorphism if it has a periodic point.

To describe the dynamics of a circle homeomorphism without periodic points (which are necessarily orientation preserving) we are going to introduce a powerful topological invariant called the rotation number which was defined for the first time by Poincaré (1881-1886). One way to define this number would be to take a point $c \in S^1$ and to take the positively oriented (half-open) arc $L = [c, f(c))$ connecting c to $f(c)$. One can show that for each $x \in S^1$ the

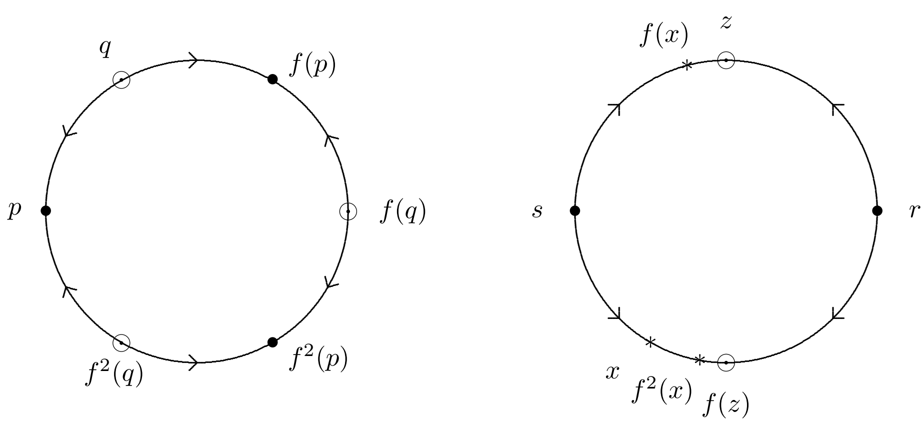

Fig. 1.1: The typical dynamics of a homeomorphism $f\colon S^1 \to S^1$ with periodic points: $\omega(x)$ is a periodic orbit for each $x \in S^1$. On the left an orientation preserving homeomorphism with two periodic points p and q of period three. On the right the dynamics of an orientation reversing homeomorphism with two fixed points r and s and a periodic point z of period two.

fraction $\#\{i\,;\ f^i(x) \in L, 0 \le i \le n-1\}/n$ has a limit as n tends to infinity, and that this limit does not depend on c and x; we have taken L half-open because otherwise this statement is for example false if f is a rotation $f(x) = x+\alpha \bmod 1$ when α rational. This limit is called the *rotation number* of f. However, this approach will not be feasible in the non-invertible case and therefore, in order to unify the treatment in the invertible and the non-invertible case, we will introduce the rotation number in a more combinatorial and somewhat indirect way. In this way, we will also immediately get much more precise information about the orbit of points. Furthermore, using the idea of first return maps we will see in an extremely natural way that certain orbits of intervals are disjoint and cover the circle, see Statement (1.6) below. This kind of disjointness will play a crucial role throughout this book. As the reader will notice, we shall derive the usual algebraic properties about continued fraction expansions from dynamical properties rather than from algebraic considerations.

More precisely, just as we will do in the non-invertible case later on, we will code orbits of a circle homeomorphism f using symbolic dynamics: to each point $x \in S^1$ we will associate a sequence of symbols $\{L, c, R\}$. This sequence will be called the itinerary $\underline{i}_f(x)$ of x. It will turn out that the itinerary of one point completely determines the combinatorial type of the homeomorphism. Because the maps we will deal with in the remainder of this chapter are orientation preserving it is quite easy to get much more information about these itineraries. In fact, these itineraries will turn out to be determined by a sequence of integers $a_1, a_2, a_3, \ldots$. This sequence of integers will be defined as follows. Suppose that

f has no fixed points. Take a point $c \in S^1$ and let $J' = (c, f(c))$ be the positively oriented open arc in S^1 connecting c to $f(c)$. Now define $a(f) \geq 0$ to be the smallest integer such that $J' \cap f^{a(f)+1}(J') \neq \emptyset$. Then the closure of $J' \cup f(J') \cup \cdots \cup f^i(J')$ is connected and $f^{i-1}(J')$ and $f^i(J')$ are adjacent for each $i = 1, \ldots, a(f)+1$; the closures of the intervals $J, \ldots, f^{a(f)+1}(J')$ together cover S^1. If $f^{a(f)+1}(J') = J'$ then f has a periodic point of period $a(f)+1$ and the procedure stops. If $f^{a(f)+1}(J') \neq J'$ then we define $J_1 = f^{a(f)+1}(J') \cup J'$ and we let $f_1 : J_1 \to J_1$ be the first return map of f to J_1 (this return map is defined by $f_1(x) = f^{k(x)}(x)$ where $k(x) \geq 1$ is the minimal integer such that $f^{k(x)}(x) \in J_1$). If $a(f) \geq 1$ then J_1 is an interval and we will see below that $f_1 : J_1 \to J_1$ can again be considered as a circle map by identifying the boundary points of J_1. Therefore we can apply a similar procedure again to $f_1 : J_1 \to J_1$ (except now we will let J_1' be the clockwise oriented arc connecting c to a boundary point of J_1). If f has no periodic points then, as we will show below, we get in this way a nested sequence of intervals $J_n \subset S^1$ containing c, a sequence of return maps $f_n : J_n \to J_n$, and a sequence of integers $a(f_n) \in \mathbb{N}$ for $n = 1, \ldots$. If f does has a periodic point then this procedure stops; even in this case we shall be able to use the finite sequence of integers $a(f_n)$. This (possibly finite) sequence of integers $a(f_n)$ will then determine a_n and the continued fraction expansion of the rotation number of f.

One advantage of this procedure is that it also allows us to get detailed insight into the orbits of these intervals J_n under the original homeomorphism f. Throughout the remainder of this chapter, this will turn out to be one of the main ingredients for most results.

If we take an interval J of S^1 then the first return map of f to J is in general not continuous. However, in some cases this return map can be regarded as a circle homeomorphism again. To make this precise we will now show how to identify a homeomorphism $g : S^1 \to S^1$ with an interval map $f : [0,1] \to [0,1]$. Fix the orientation and the metric in the circle induced from the real line by the quotient map $\pi : \mathbb{R} \to S^1 = \mathbb{R}/\mathbb{Z}$. Let $g : S^1 \to S^1$ be an orientation preserving homeomorphism without fixed points. Let $\hat{g} : \mathbb{R} \to \mathbb{R}$ be a lift of g. Since g has no fixed points, there is a unique point $c \in (0,1)$ for which $\hat{g}(c) \in \mathbb{Z}$. So there is a unique map $f : [0,1] \to [0,1]$ defined by

$$f(x) = \begin{cases} \hat{g}(x) \text{mod } 1 & \text{for } x \in [0,1] \setminus \{c\} \\ 0 & \text{for } x = c. \end{cases}$$

(we could equally well have defined $f(c) = 1$). Since $\pi \circ f^n = g^n \circ \pi$, orbits of f are mapped by π onto orbits of g. Clearly, f is a continuous map except at the point $c = c(f)$. In fact, $\lim_{t \uparrow c} f(t) = 1$ and $\lim_{t \downarrow c} f(t) = 0$, see Figure 1.2. Moreover, $f(0) = f(1)$. For simplicity we will write $f(c^-)$ for $\lim_{t \uparrow c} f(t)$ and $f(c^+)$ for $\lim_{t \downarrow c} f(t)$. Conversely, any map of $f : [0,1] \to [0,1]$ such that $f(0) = f(1)$ for which there exists a unique point $c \in (0,1)$ such that f is continuous and monotone increasing on $[0,c)$ and on $(c,1]$ and such that $f(c^-) = 1$, $f(c^+) = 0$, defines a homeomorphism of the circle. Hence the space of orientation preserving homeomorphisms of the circle without fixed points can be identified with the space $\mathcal{S}(J)$ of maps $f\colon J \to J$ (where J is a closed interval) without fixed points satisfying the following properties:

1. f has a unique point of discontinuity $c(f)$, this point belongs to the interior of J and $\lim_{x\downarrow c(f)} f(x) = f(c(f))$ is equal to the left endpoint of J and $\lim_{x\uparrow c(f)} f(x)$ is equal to the right endpoint of J;
2. f is monotone increasing in each component of $J \setminus \{c(f)\}$;
3. f maps both boundary points of J to a single point in the interior of J.

In this identification, the set of rotations of the circle corresponds to the set $\mathcal{I}(J)$ of maps in $\mathcal{S}(J)$ which are piecewise linear with slope equal to one. From now on we shall treat maps $f \in \mathcal{S}(J)$ as if they were circle maps by identifying both endpoints of J. After this identification the image under f of c simply becomes one point (and f becomes again a homeomorphism on the circle). We will also show that the return maps to certain intervals J are maps in $\mathcal{S}(J)$, and therefore can again be considered as circle homeomorphisms.

Let us introduce the symbolic dynamics associated to f. Let $\Sigma = \{L, c, R\}^{\mathbb{N}}$, namely, an element of Σ is a sequence $\underline{x} = (x_0, x_1, \dots, x_n, \dots)$ where each $x_n \in \{L, c, R\}$. Let $\sigma\colon \Sigma \to \Sigma$ be the shift map $\sigma\underline{x} = \underline{y}$ where $y_i = x_{i+1}$.

Definition. Let $f \in \mathcal{S}(J)$ and $x \in J$. The *itinerary* of x with respect to f is the sequence $\underline{i}_f(x) = (i_0(x), i_1(x), \dots)$ where $i_j(x) = L$ if $f^j(x) < c(f)$, $i_j(x) = R$ if $f^j(x) > c(f)$ and $i_j(x) = c$ if $f^j(x) = c(f)$. Furthermore, we consider the following (lexicographical) ordering in the space Σ: if $\underline{x}, \underline{y} \in \Sigma$ then $\underline{x} \prec \underline{y}$ if and only if there exists an integer k such that $x_i = y_i$ for $i < k$ and $x_k < y_k$. Here we consider the symbols L, c, R ordered as $L < c < R$. In Σ we consider the topology defined by the metric $d(\underline{x}, \underline{y}) = \sum_{i=0}^{\infty} \frac{1}{2^i} d(x_i, y_i)$ where $d(x_i, y_i) = 1$ if $x_i \neq y_i$ and $d(x_i, x_i) = 0$. This topology corresponds to the one induced by the ordering $\prec$. With this topology Σ is a compact totally disconnected metric space and the shift transformation $\sigma\colon \Sigma \to \Sigma$,

$$\sigma(x_0, x_1, x_2, \dots) = (x_1, x_2, \dots)$$

is continuous.

Lemma 1.1. *Let $f \in \mathcal{S}(J)$. Then:*

1. *if $x < y$ then $\underline{i}_f(x) \preceq \underline{i}_f(y)$;*
2. *if $\underline{i}_f(x) \prec \underline{i}_f(y)$ then $x < y$;*
3. *$\underline{i}_f(f^i(x)) = \sigma^i(\underline{i}_f(x))$ for all $i \geq 1$.*

Proof. Assume $x < y$, suppose $\underline{i}_f(x) \neq \underline{i}_f(y)$ and let k be such that $i_j(x) = i_j(y)$ for all $j = 0, 1, \dots, k-1$ and $i_k(x) \neq i_k(y)$. Then $f^j(x)$ and $f^j(y)$ cannot be in distinct components of $J \setminus \{c\}$ for $j = 0, 1, \dots, k-1$. Hence the map f^k

is continuous in the interval (x, y) and, consequently, it is monotone increasing. Therefore $f^k(x) < f^k(y)$. This proves Statement 1. The easy proofs of Statements 2 and 3 are left to the reader. $\square$

For convenience we shall use the following notation: letting $\underline{x} = (x_0, x_1, \dots)$, $\underline{y} = (y_0, y_1, \dots) \in \Sigma$ we define $\underline{x}_n = (x_0, x_1, \dots, x_{n-1})$,

$$\underline{x}_n \cdot \underline{y}_m = (x_0, x_1, \dots, x_{n-1}, y_0, y_1, \dots, y_{m-1})$$

and

$$\underline{x}_n \cdot \underline{y} = (x_0, x_1, \dots, x_{n-1}, y_0, y_1, \dots).$$

If $f \in \mathcal{S}(J)$, we define $K^+(f), K^-(f) \in \Sigma$ to be

$$K^+(f) = \lim_{x \downarrow c} \underline{i}_f(x)$$

and

$$K^-(f) = \lim_{x \uparrow c} \underline{i}_f(x).$$

By Lemma 1.1, these limits exist and if c is non-periodic then

$$K^+(f) = (R, L) \cdot \underline{i}_f(f^2(c)) \text{ and } K^-(f) = (L, R) \cdot \underline{i}_f(f^2(c)).$$

Definition. We say that $f, \tilde{f} \in \mathcal{S}(J)$ are *combinatorially equivalent* if the orbit of $c(f)$ by f has the same order as the orbit of $c(\tilde{f})$ by $\tilde{f}$, namely, the map $h(f^n(c(f))) = \tilde{f}^n(c(\tilde{f}))$ for $n \in \mathbb{Z}$, is strictly order preserving. (We impose that h is strictly order preserving to make sure that this notion is an equivalence relation.)

Now we will prove that two maps f and $\tilde{f}$ in $\mathcal{S}(J)$ without periodic points are combinatorially equivalent if and only if $K^+(f) = K^+(\tilde{f})$. This result implies that the sequence $K^+(g)$ is also well defined for a circle homeomorphism $g\colon S^1 \to S^1$ without periodic points: although g corresponds to many interval maps, the sequence K^+ we obtain will not depend on this choice. Indeed, for any $x, y \in S^1$, the map h defined by $h(g^n(x)) = g^n(y)$ for all $n \in \mathbb{Z}$ strictly respects the circular order on S^1. It follows that any two interval maps f and $\tilde{f}$ in $\mathcal{S}(J)$ corresponding to $g\colon S^1 \to S^1$ are combinatorially equivalent. Hence, from the next lemma, $K^+(f) = K^+(\tilde{f})$. So $K^+(g)$ is well-defined.

Proposition 1.1. *Let $f, \tilde{f} \in \mathcal{S}(J)$ and assume that these maps have no periodic orbits. Then f and $\tilde{f}$ are combinatorially equivalent if and only if $K^+(f) = K^+(\tilde{f})$.*

Proof. If f and $\tilde{f}$ are combinatorially equivalent then clearly $K^+(f) = K^+(\tilde{f})$. So assume that $K^+(f) = K^+(\tilde{f})$. Write $c = c(f)$ and $\tilde{c} = c(\tilde{f})$. Then define $h(x) = \sup\{y\,;\ \underline{i}_{\tilde{f}}(y) \preceq \underline{i}_f(x)\}$.

Since $K^+(f) = K^+(\tilde f)$ and since the maps $f, \tilde f$ have no periodic orbits, the formula above the previous definition implies that $\underline{i}_f(c) = \underline{i}_{\tilde f}(\tilde c)$. Moreover, if the sequences $\underline{i}_f(x)$ and $\underline{i}_{\tilde f}(y)$ in Σ start with the symbol c then $x = c$ and $y = \tilde c$. Combining all this gives $h(c) = \tilde c$. It follows from this and the previous lemma that h is order preserving and that $h \circ f = \tilde f \circ h$. Moreover, $h(f^n(c)) = \tilde f^n(h(c)) = \tilde f^n(\tilde c)$ for $n \in \mathbb{Z}$. Since $\tilde f$ has no periodic points, this equation implies that h injective on the full orbit of c. This completes the proof of the lemma. □

Everything we have done so far will also work in the non-invertible case. We will now relate $K^+(f)$ to an inductively defined sequence of integers a_n: we will describe the dynamics of $f \in \mathcal{S}(J)$ via the first return maps to a sequence of intervals around the discontinuity point $c = c(f)$. To each of these return maps an integer a_n will be associated. $K^+(f)$ will turn out to be determined by these integers. In this way we will be able to tell which sequences $\{L, c, R\}^{\mathbb{N}}$ are of the form $K^+(f)$ for some $f \in \mathcal{S}(J)$. More precisely, if $I \subset J$ is a closed interval such that the forward f-orbit through any point of I intersects I, we define the *first return map* $\mathcal{R}(f)\colon I \to I$ of f to I as $\mathcal{R}(f)(x) = f^k(x)$ where $k = k(x) = \min\{i > 0\,;\ f^i(x) \in I\}$. Notice that if $J \supset I_1 \supset I_2$, $r_1\colon I_1 \to I_1$ is the first return map of f to I_1 and $r_2\colon I_2 \to I_2$ is the first return map of r_1 to I_2 then r_2 is also the first return map of f to I_2.

So let $f \in \mathcal{S}(J)$ and denote the interiors of the components of $J \setminus \{c(f)\}$ by J' and J''. From the conditions above it follows easily that one of the components, say J', is mapped into the other component J'' and that $f(J'')$ contains J'. (It is possible that $f(J') = J''$ but in this case f permutes J' and J'' and f has a periodic point of period 2. This situation happens for example when $f(x) = x + 1/2 \bmod 1$.)

Lemma 1.2. *Let $f \in \mathcal{S}(J)$ have no fixed points and let $c = c(f)$, J' and J'' be as above. Let $a(f)$ be the smallest integer such that J' and $f^{a(f)+1}(J')$ have a point in common and let $J(f)$ be the closure of $f^{a(f)+1}(J') \cup J'$. Then*

1. $a(f)$ is the smallest integer such that the closure of $J' \cup f(J') \cup \cdots \cup f^{a(f)+1}(J')$ covers the circle;

2. if $f^{a(f)}(J')$ contains c in its closure then $f^{a(f)+1}(J') = J' = J(f)$ and $f^{a(f)+1}(c) = c$; in this case the first return map $\mathcal{R}(f)$ to $J(f)$ is equal to $f^{a(f)+1}$ and has fixed points in $\partial J(f)$.

3. otherwise, $J(f)$ strictly contains J' and the first return map $\mathcal{R}(f)$ of f to $J(f)$ is contained in $\mathcal{S}(J(f))$; furthermore, $\mathcal{R}(f)$ maps the non-empty interval $J'' \cap J(f)$ into $J' = J' \cap J(f)$; tis return map coincides with $(f|J'')^{a(f)} \circ (f|J')$ in J' and with $f|J''$ in $J'' \cap J(f)$;

Proof. Since f does not have a fixed point and f is monotone increasing on J'',

$$a'(f) = \max\{k \in \mathbb{N}\,;\ f^i(J') \subset J'' \text{ for all } i = 1, \ldots, k\}$$

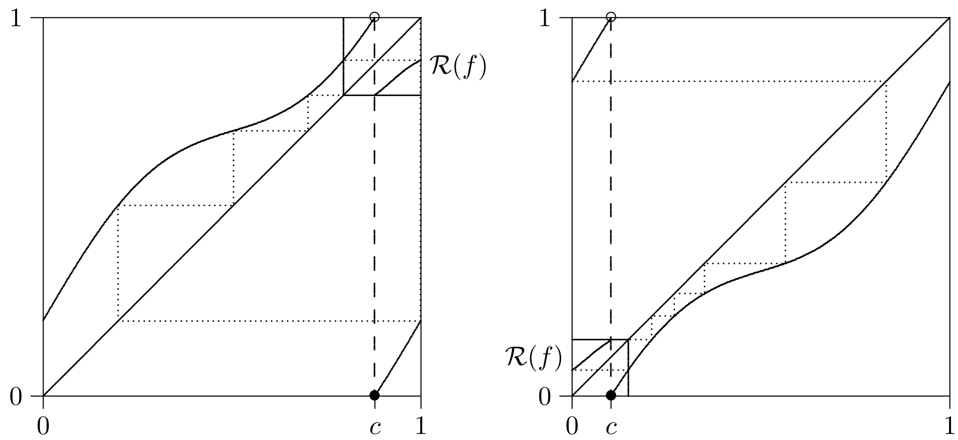

Fig. 1.2: $a(f)$, J_1 and the map $\phi_1\colon J_1 \to J_1$ from below when J' is to the left respectively to the right of J''.

is finite, see Figure 1.2. Moreover, $f(J')$ has a common endpoint with J'', the intervals $f(J'), \dots, f^{a'(f)+1}(J')$ lie ordered, each of these intervals has one common boundary point with the next one, and $f^{a'(f)+1}(J')$ has a non-empty intersection with J'. From this description it follows that $a(f) = a'(f)$ and that $J(f) = f^{a(f)+1}(J') \cup J'$ is an interval. If c is contained in the closure of $f^{a(f)}(J')$ then c is an end point of this interval and $f^{a(f)+1}$ maps J' onto itself and $f^{a(f)+1}(c) = c$. If c is not contained in the closure of $f^{a(f)}(J')$ then $f^{a(f)+1}(J')$ contains c in its interior. From the previous description $a(f)+1$ is the smallest integer such that $f^{a(f)+1}(x) \in J(f)$ for each $x \in J' = J' \cap J(f)$. Now consider $J'' \cap J(f)$. In order to be more explicit we assume that J'' is to the left of J'; if J'' is to the right of J' then simply interchange the words left and right in what follows. In this case, the image of $J'' \cap J(f)$ under f is equal to $J' \setminus f^{a(f)+1}(J')$, see Figure 1.2. It follows that f maps $J'' \cap J(f)$ into J' and that its image contains the right endpoint of J'. Therefore the return map $\mathcal{R}(f)$ has all stated properties.

□

In the next example we will explicitly determine $\mathcal{R}(R_\alpha)$ when R_α is a rotation of the circle.

Example. Let $R_\alpha\colon [0,1] \to [0,1]$ be the rotation defined by $R_\alpha(x) = x + \alpha \bmod 1$. This is the simplest type of map in $\mathcal{I}(J) \subset \mathcal{S}(J)$. Let us first consider the case that $\alpha \in (0,1/2]$. Then one has $c = 1-\alpha$, $J' = (1-\alpha, 1)$, $J'' = (0, 1-\alpha)$, $J(R_\alpha) = [(n-1)\alpha, 1]$ where n is the largest integer with $n\alpha \le 1$ and $a(R_\alpha) = n-1$. If we denote by $[x]$ the largest integer which is less than or equal to x this gives $a(R_\alpha) = \left[\frac{1}{\alpha}\right] - 1$ when $\alpha \in (0,1/2]$. Furthermore, $\mathcal{R}(R_\alpha)$ is equal to $x + n\alpha$ for $x \in [1-\alpha, 1]$ and equal to $x+\alpha$ for $x \in [(n-1)\alpha, 1-\alpha)$. Taking an orientation preserving linear scaling $h\colon [0,1] \to J(R_\alpha)$ gives that $h^{-1} \circ \mathcal{R}(R_\alpha) \circ h = R_{\alpha'}$ where $\alpha' = \frac{\alpha}{1-n\alpha+\alpha} = \frac{\alpha}{1-a(R_\alpha)\alpha}$. Since $\alpha \le 1 - n\alpha + \alpha < 2\alpha$ we have that $\alpha' \in (1/2, 1]$ (note that R_1 is of course equal to $R_0 = id$). So if we let $G\colon (0,1] \to [0,1)$ be the *Gauss map* defined by

$$G(x) = \frac{1}{x} - \left[\frac{1}{x}\right]$$

then

$$\left.\begin{aligned} \mathcal{R}(R_\alpha) &= R_{\alpha'} \quad \text{where } \alpha' = \frac{1}{1+G(\alpha)} \in (1/2,1], \\ a(f) &= \left[\frac{1}{\alpha}\right] - 1 \end{aligned}\right\} \quad \text{when } \alpha \in (0,1/2].$$

Similarly, when $\alpha \in (1/2,1)$, $J(R_\alpha) = [0, 1-n(1-\alpha)]$ where $a(R_\alpha) = n-1$ and therefore $a(R_\alpha) = \left[\frac{1}{1-\alpha}\right] - 1$. Scaling the return map gives $h^{-1} \circ \mathcal{R}(R_\alpha) \circ h = R_{\alpha'}$ where $\alpha' = \frac{1-n(1-\alpha)}{1-\alpha+1-n(1-\alpha)} \in (0,1/2)$. In other words, we get

$$\left.\begin{aligned} \mathcal{R}(R_\alpha) &= R_{\alpha'} \quad \text{where } \alpha' = \frac{G(1-\alpha)}{1+G(1-\alpha)} \in (0,1/2), \\ a(f) &= \left[\frac{1}{1-\alpha}\right] - 1 \end{aligned}\right\} \quad \text{for } \alpha \in (1/2,1).$$

Note that if $\alpha \in (0,1/2]$ then $\alpha' \in (1/2,1]$ and if $\alpha \in (1/2,1)$ then $\alpha' \in (0,1/2)$. Moreover, if $\alpha' = 1 = 0 \bmod 1$, then $\mathcal{R}(R_\alpha)\colon J(R_\alpha) \to J(R_\alpha)$ has fixed points and we cannot apply the previous lemma again ($\alpha' = 1$ happens precisely when $\alpha = \frac{1}{n}$ for some integer $n \geq 2$).

Let us now repeat the procedure from Lemma 1.2 inductively. So define $J_0 = J$, $\phi_0\colon J_0 \to J_0$, $\phi_0 = f$, we let $a_1 = \infty$ if f has fixed points and otherwise we define

$$a_1 = \begin{cases} a(f) + 1 & \text{if } J' \text{ is to the right of } J'' \\ 1 & \text{if } J' \text{ is to the left of } J'', \end{cases}$$

$$J_1 = \begin{cases} J(\phi_0) & \text{if } J' \text{ is to the right of } J'' \\ J & \text{if } J' \text{ is to the left of } J'', \end{cases}$$

and

$$\phi_1 = \begin{cases} \mathcal{R}(f) & \text{if } J' \text{ is to the right of } J'' \\ f & \text{if } J' \text{ is to the left of } J''. \end{cases}$$

(The reason we distinguish between these two cases is in order to make sure that ϕ_1 always maps the left component of $J_1 \setminus \{c\}$ into the right component. Furthermore, the definition of a_1 is slightly different from the formulas for $a_2, a_3, \ldots$ below because later we will show that these numbers determine the continued fraction expansion of the 'rotation number' of f. If we defined $a_1 = a(f)$ we would not get the rotation number but only a closely related number.)

Now suppose that $n \geq 2$ and that $J_1, \ldots, J_{n-1}$, $\phi_1, \ldots, \phi_{n-1}$ are defined, and that $\phi_{n-1}\colon J_{n-1} \to J_{n-1}$ has no fixed points. Then define the interval J_n, the return map ϕ_n to J_n, and the integer a_n inductively by

$$J_n = J(\phi_{n-1}), \quad \phi_n = \mathcal{R}(\phi_{n-1})\colon J_n \to J_n,$$

$$a_n = a(\phi_{n-1}).$$

On the other hand, if $\phi_{n-1}\colon J_{n-1} \to J_{n-1}$ has fixed points then we let $a_n = \infty$ and we stop the inductive definition. In other words, if J' is to the right of J'' we have

$$\begin{aligned} a_1 &= a(f) + 1, & \phi_1 &= \mathcal{R}(f), \\ a_n &= a(\mathcal{R}^{n-1}(f)), & \phi_n &= \mathcal{R}^n(f) \text{ for all } n = 2, 3, \ldots \end{aligned} \tag{1.1}$$

and, otherwise,

$$\begin{aligned} a_1 &= 1, & \phi_1 &= f, \\ a_n &= a(\mathcal{R}^{n-2}(f)), & \phi_n &= \mathcal{R}^{n-1}(f) \text{ for all } n = 2, 3, \ldots. \end{aligned} \tag{1.2}$$

In Chapter VI we shall call a similarly defined map the *n-th renormalization* of f.

Of course, ϕ_n is the first return map of f to J_n. In particular, if f has no periodic points then ϕ_n has no fixed points for each integer n and therefore the construction never stops. If J' is to the left of J'' then $a_1 = 1$, $J_1 = J$. It follows that ϕ_1 always maps the left component of $J_1 \setminus \{c\}$ into the right component. Let J'_n be the interior of the left component of $J_n \setminus \{c\}$ if n is odd and of the right component if n is even. Denote the interior of the other component of $J_n \setminus \{c\}$ by J''_n. From the previous lemma, the role of the right and left component of $J_n \setminus \{c\}$ is interchanged in each step of the induction. More precisely,

$$J'_n = J''_{n-1} \cap J_n \text{ and } J''_n = J'_{n-1} \cap J_n = J'_{n-1}. \tag{1.3}$$

Consequently, we get by induction that ϕ_n maps J'_n into J''_n for all $n \geq 1$ for which ϕ_n is defined. Also, $\phi_1|J'_1 = f$ and $\phi_1|J''_1 = f^{a_1}$. From (1.3) and the previous lemma one has that

$$\phi_n|J''_n = (\phi_{n-1}|J''_{n-1})^{a(\phi_{n-1})} \circ (\phi_{n-1}|J'_{n-1})$$

and

$$\phi_n|J'_n = \phi_{n-1}|J''_{n-1}.$$

Therefore we get by induction that

$$\phi_n|J'_n = f^{q_{n-1}} \text{ and } \phi_n|J''_n = f^{q_n}$$

where q_n is defined inductively by

$$\begin{aligned} q_0 = 1 \ &, \ q_1 = a_1 \\ q_{n+1} = q_{n-1} &+ a_{n+1} q_n \text{ for } n \geq 1. \end{aligned} \tag{1.4}$$

From the fact that ϕ_n is a first return map and from Lemma 1.2 it follows that if f has a periodic point of period $N > 1$ then $\phi_n\colon J_n \to J_n$ has a periodic point of period $< N$. (At least one point and not all of the points of the periodic orbit are contained in the domain of the return map.) In particular, if f has a periodic point there exists some finite $n \in \mathbb{N}$ for which ϕ_n will have fixed points: the process stops and $a_n = \infty$.

Since $\phi_n : J_n \to J_n$ is a map in $\mathcal{S}(J_n)$, we get that $J_n = [f^{q_{n-1}}(c), f^{q_n}(c)]$ (and this interval contains c) and that $J_n' = (c, f^{q_n}(c))$ and $J_n'' = (f^{q_{n-1}}(c), c)$. Here and also in the remainder of this book we use the convention that (a,b) is the smallest open interval containing a and b in its boundary irrespective of whether $a < b$ or $b < a$. Similarly, $[a,b]$ is the corresponding closed interval. Furthermore, we get for $n \geq 1$, $0 \leq j \leq q_{n+1}$ that

$$f^j(c) \in J_n \text{ if and only if } j = q_{n-1} + iq_n \tag{1.5}$$

for some $i \in \{0, \ldots, a_{n+1}\}$. So in this sense the iterates $f^{q_n}(c)$ are *closest returns* to c and it makes sense to call the intervals J_n' and J_n'' *closest return intervals*. These closest returns are drawn in Figure 1.3. We claim that

$$\text{the union of } \bigcup_{i=0}^{q_{n-1}-1} f^i(J_n') \text{ and } \bigcup_{i=0}^{q_n-1} f^i(J_n'') \text{ "tiles" the interval.} \tag{1.6}$$

More precisely, the closure of this set covers the interval (or circle) and all the intervals in the union are disjoint. As we will see this result is fundamental for all metric results on circle diffeomorphisms. The proof of this result is surprisingly natural in our setting: this is because first return maps usually give a lot of disjointness. So let us prove (1.6). The disjointness can be seen as follows. The first return map of $J_n' \cup J_n''$ to itself is equal to $f^{q_{n-1}}$ on J_n' and equal to f^{q_n} on J_n''. Furthermore, the images of the two intervals J_n' and J_n'' under this first return map are disjoint. But then the orbits of these intervals up to the return time must also be disjoint. The fact that the closure of the union of these intervals cover the circle is also easy to see. Indeed, take $x \in J$ and let $k \geq 0$ be minimal so that $f^{-k}(x)$ is contained in the closure of $J_n' \cup J_n''$. Such an integer k certainly exists because the union of $J_n', \ldots, f^{q_{n+1}}(J_n')$ covers the interval. So there are two possibilities: if $f^{-k}(x)$ is in the closure of J_n' then because J_n' returns within time q_{n-1} to the closure of $J_n' \cup J_n''$ one has $k < q_{n-1}$ and therefore x is in the closure of $f^k(J_n')$; in the other case, x is in the closure of $f^k(J_n'')$ for some $k < q_n$. So in both cases the result follows.

Because some corollaries of (1.6) are fundamental in future sections of this chapter we will state them in a lemma.

Lemma 1.3. *Let $f \in \mathcal{S}(J)$, $x \in J$, $I_{n-1}(x) = (x, f^{q_{n-1}}(x))$ and $\hat{I}_{n-1}(x) = I_{n-1}(f^{-q_{n-1}}(x)) \cup I_{n-1}(x) \cup \{x\} = (f^{-q_{n-1}}(x), f^{q_{n-1}}(x))$. Then $\hat{I}_{n-1}$ is an interval and the intervals*

$$I_{n-1}(x), f(I_{n-1}(x)), \ldots, f^{q_n-1}(I_{n-1}(x))$$

are pairwise disjoint. Furthermore, the intervals

$$\hat{I}_{n-1}(x), f(\hat{I}_{n-1}(x)), \ldots, f^{q_n-1}(\hat{I}_{n-1}(x))$$

cover J and each point is contained in at most two of these intervals.

Proof. Let $y = f^{-q_n}(x)$. Of course, the choice of c is arbitrary. So (1.6) holds if we replace J'_n and J''_n by $I_n(x) = (x, f^{q_n}(x))$ and $I_{n-1}(x) = (f^{q_{n-1}}(x), x)$ or, alternatively, by $I_n(y) = (y, f^{q_n}(y))$ and $I_{n-1}(y) = (f^{q_{n-1}}(y), y)$. Hence,

$$J'_n \cup J''_n \subset \hat{I}_n(x) \subset I_{n-1}(x) \cup I_{n-1}(y).$$

The lemma follows from this: by (1.6) the first $q(n)$ iterates of $J'_n \cup J''_n$ certainly cover J; moreover, again by (1.6), the first $q(n)$ iterates of $I_{n-1}(x)$ and also of $I_{n-1}(y)$ are disjoint. □

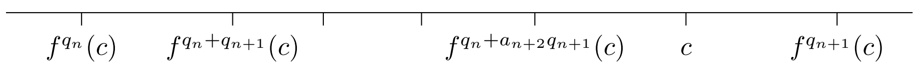

Fig. 1.3: The position of $f^{q_n}(c)$, $f^{q_{n+1}}(c)$, $f^{q_n+iq_{n+1}}(c)$ for $i = 0, \dots, a_{n+2}$.

Of course, Statements (1.5), (1.6) and Lemma 1.3 also hold for the original circle homeomorphism $g\colon S^1 \to S^1$. In this case $I_{n-1}(x)$ is the segment in S^1 bounded by x and $g^{q_{n-1}}(x)$ which does not contain $g(x)$.

If f is equal to a rotation R_α then one can make more quantitative statements. We shall do this in the next example.

Example. Let us consider a rotation $R = R_\alpha$. We shall give an algorithm to determine α and show that α is necessarily irrational if $a_n < \infty$ for all n. Let $\theta_0 = \alpha$, i.e., the length of the interval $(c, R(c^-)) = (c, 1)$ and let θ_n be the length of $J'_n = (c, R^{q_n}(c^+))$ for $n \geq 1$. Since $R^n(c) = c + n\alpha \bmod 1$, this gives

$$\theta_n = \inf_{p \in \mathbb{Z}} |q_n\alpha - p|.$$

From the previous description J'_{n+1} is equal to $J''_n \setminus [\phi_n(J'_n) \cup \cdots \cup \phi_n^{a_{n+1}}(J'_n)]$ where $J'_{n-1} = J''_n$ and the union is disjoint. It follows that

$$\begin{aligned} 1 &= a_1\theta_0 + \theta_1, \\ \theta_{n-1} &= a_{n+1}\theta_n + \theta_{n+1}, n \geq 1 \end{aligned} \tag{1.7}$$

and $\theta_n \downarrow 0$ as $n \to \infty$. From (1.6) the two collections

$$J'_n, R(J'_n), \dots, R^{q_{n-1}-1}(J'_n)$$

and

$$J''_n, R(J''_n), \dots, R^{q_n-1}(J''_n)$$

together 'tile' the interval (or circle). Since the intervals from the first and second collection have length respectively θ_n and θ_{n-1}, this gives

$$q_{n-1}\theta_n + q_n\theta_{n-1} = 1 \text{ for each } n \geq 1. \tag{1.8}$$

Now let $p_0, p_1, p_2, \dots$ be the sequence of integers so that

$$q_n\alpha - p_n = (-1)^n\theta_n \tag{1.9}$$

(since $\theta_n = \inf_{p\in\mathbb{Z}} |q_n\alpha - p|$ and J'_{n+1} and J'_n are on opposite sides of c, this is possible). Then $p_0 = 0$, $p_1 = 1$ and since $\theta_{n-1} = a_{n+1}\theta_n + \theta_{n+1}$ and $q_{n+1} = a_{n+1}q_n + q_{n-1}$,

$$\begin{aligned} p_{n+1} &= q_{n+1}\alpha - (-1)^{n+1}\theta_{n+1} \\ &= [a_{n+1}q_n + q_{n-1}]\alpha - (-1)^{n+1}[\theta_{n-1} - a_{n+1}\theta_n] \\ &= a_{n+1}p_n + p_{n-1} \end{aligned} \tag{1.10}$$

for all $n \geq 1$. From the 'tiling equation' $q_{n+1}\theta_n + q_n\theta_{n+1} = 1$,

$$\begin{aligned} &q_{n+1}p_n - q_np_{n+1} \\ &\quad = q_{n+1}[q_n\alpha - (-1)^n\theta_n] - q_n[q_{n+1}\alpha - (-1)^{n+1}\theta_{n+1}] \\ &\quad = (-1)^{n+1}[q_{n+1}\theta_n + q_n\theta_{n+1}] = (-1)^{n+1}. \end{aligned} \tag{1.11}$$

In particular, p_n and q_n are coprime as long as p_n and q_n are not both one, but then $n = 1$. Moreover, $|\frac{p_n}{q_n} - \frac{p_{n+1}}{q_{n+1}}| = \frac{1}{q_nq_{n+1}}$,

$$\frac{p_2}{q_2} < \frac{p_4}{q_4} < \frac{p_6}{q_6} < \cdots < \frac{p_7}{q_7} < \frac{p_5}{q_5} < \frac{p_3}{q_3} < \frac{p_1}{q_1} \tag{1.12}$$

and $\frac{p_n}{q_n}$ converges. This limit is equal to α because $\frac{p_n}{q_n} = \alpha + (-1)^{n+1}\frac{\theta_n}{q_n}$ and $q_n \to \infty$. Furthermore, $0 < |\alpha - \frac{p_n}{q_n}| < |\frac{p_{n+1}}{q_{n+1}} - \frac{p_n}{q_n}| = \frac{1}{q_nq_{n+1}} < \frac{1}{q_n^2}$. The angle of rotation α is irrational because otherwise $\alpha = k/l$ where k and l are integers and then the last inequality would imply that $|q_nk - p_nl| < l/q_n$ and therefore for n sufficiently large we would have $|q_nk - p_nl| < 1$. Since $q_nk - p_nl$ is an integer and p_n and q_n are coprime, this is impossible. Since $|R^{q_{n-1}}(c) - c| > |R^{q_n}(c) - c|$ and $R^j(c) \in J_n = (R^{q_{n-1}}(c), R^{q_n}(c)]$ for $j > 0$ implies $j \geq q_n$, it follows that if $j > 0$ and $|R^j(c) - c| \leq |R^{q_n}(c) - c|$ then $j \geq q_n$. Hence

$$|\alpha - \frac{p}{q}| > |\alpha - \frac{p_n}{q_n}| \text{ for all } p, q \in \mathbb{N} \text{ with } 0 < q < q_n \tag{1.13}$$

(therefore the ratios $\frac{p_n}{q_n}$ are said to be the *best rational approximations* for the irrational number α).

Let us now give a formula for a_n. Rescaling J_n in an orientation preserving way to $[0,1]$, one gets that $\phi_n(R_\alpha)\colon J_n \to J_n$ becomes equal to some rotation $R_{\alpha(n)}$. When $\alpha \in (1/2, 1)$, then by definition $a_1(\alpha) = 1 = \left[\frac{1}{\alpha}\right]$, $J_1 = J$ and $\phi_1 = f$ and therefore $\alpha(1) = \alpha$ which turns out be equal to $\frac{1}{1+G(\alpha)}$ in this case (so we do not need to use the calculations from the previous example for α' in this case). If $\alpha \in (0, 1/2]$ then $a_1(\alpha) = a(R_\alpha) + 1 = \left[\frac{1}{\alpha}\right]$ and from the previous example, $\alpha(1) = \frac{1}{1+G(\alpha)} \in (1/2, 1]$. Note that in both cases the formulas for $a_1(\alpha)$ and $\alpha(1)$ become the same (this is the reason we defined $J_1 = J$ when J' is to the left of J''). If this last number is equal to 1 then $\phi_1\colon J_1 \to J_1$ has fixed points and the procedure stops. If the procedure can be repeated then we have in both cases $\alpha(1) \in (1/2,1)$ and we can determine $\alpha(n)$ inductively. As we remarked at the end of the previous example, because of $\alpha(1) \in (1/2,1)$, we get $\alpha(n) \in (0,1/2)$ for n even and $\alpha(n) \in (1/2,1)$ for n odd (as long as $\alpha(k) \neq 1$ for $k = 1,\dots,n$). From the previous example we therefore get the following recursive formulas for α: $\alpha(n+1)$ is equal to $\frac{G(1-\alpha(n))}{1+G(1-\alpha(n))}$ when n is odd and equal to $\frac{1}{1+G(\alpha(n))}$ when n is even. As we will see in Exercise 1.1 below, from this one gets that $\alpha(n)$ is equal to $\frac{1}{1+G^n(\alpha)} \in (1/2,1)$ when n is odd and equal to $\frac{G^n(\alpha)}{1+G^n(\alpha)} \in (0,1/2)$ when n is even, as long as $G^k(\alpha) \neq 0$ for $k = 1,\dots,n$. From this it follows quite easily, see Exercise 1.1, that

$$a_n(\alpha) = \left[\frac{1}{G^{n-1}(\alpha)}\right]. \tag{1.14}$$

Note also that if $b_1, b_2, \dots$ is an arbitrary sequence of positive integers then there exists a rotation R_α such that $a_n(\alpha) = b_n$ for all $n \in \mathbb{N}$. Indeed, the set of α for which $a_1(\alpha) = [1/\alpha]$ is equal to b_1 is an interval $E_{b_1} = (\frac{1}{b_1+1}, \frac{1}{b_1}]$ and the Gauss map $G\colon (0,1] \to [0,1)$ sends this interval onto $[0,1)$. It follows that for $n \geq 2$,

$$\{G^{n-1}(\alpha)\,;\ \alpha \in (0,1)\} = \{G^{n-1}(\alpha)\,;\ \alpha \in E_{b_1}\}$$

and so

$$\{a_n(\alpha)\,;\ \alpha \in (0,1)\} = \{a_n(\alpha)\,;\ \alpha \in E_{b_1}\}.$$

So fixing a_1 does not restrict the choice one has for a_n for $n \geq 2$. By induction one gets a nested sequence of intervals $E_{b_1} \supset E_{b_1,b_2} \supset E_{b_1,\dots,b_n}$ such that $a_i(\alpha) = b_i$ for $i \leq n$ and $\alpha \in E_{b_1,\dots,b_n}$. Hence for $\alpha \in \cap_{n\geq 0} E_{b_1,\dots,b_n}$ one has $a_n(\alpha) = b_n$ for all $n \geq 1$. As we have seen, α is necessarily irrational.

As we shall show now, these numbers a_n are in fact what is called the coefficients of the continued fraction expansion of α. Indeed, let $a_1, a_2, \dots, a_n$ be a sequence of finite integers $a_n \geq 1$ and define the rational number

$$[0; a_1, a_2, \dots, a_n] = \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cfrac{1}{a_3 + \cfrac{1}{\ddots + \frac{1}{a_n}}}}}.$$

Then for $n \geq 1$,

$$\frac{p_n}{q_n} = [0; a_1, a_2, \dots, a_n]. \tag{1.15}$$

This can be seen by first showing by induction that

$$[0; a_1, a_2, \dots, a_n, x] = \frac{xp_n + p_{n-1}}{xq_n + q_{n-1}} \tag{1.16}$$

for each positive real number x. For $n = 1$ this is trivial. Furthermore, if (1.16) holds for n then using $p_{n+1} = a_{n+1}p_n + p_{n-1}$, $q_{n+1} = a_{n+1}q_n + q_{n-1}$ we find

$$\begin{aligned}[0; a_1, a_2, \ldots, a_n, a_{n+1}, x] &= [0; a_1, a_2, \ldots, a_n, a_{n+1} + \frac{1}{x}] \\ &= \frac{(a_{n+1} + 1/x)\, p_n + p_{n-1}}{(a_{n+1} + 1/x)\, q_n + q_{n-1}} \\ &= \frac{x p_{n+1} + p_n}{x q_{n+1} + q_n}.\end{aligned}$$

So (1.16) and therefore (1.15) follows by induction. It follows that

$$\alpha = \lim_{n\to\infty} [0; a_1, a_2, \ldots a_n] = [0; a_1, a_2, a_3, \ldots]$$

$$= \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cfrac{1}{a_3 + \cfrac{1}{\ddots}}}}.$$

The numbers $\frac{p_n}{q_n}$ are called the *convergents* of α. Similarly, if $a_n = \infty$ then $\alpha = [0; a_1, \ldots, a_{n-1}]$.

One can also reverse this construction. Rather than associating a sequence of integers a_n to a circle homeomorphism we can also do the opposite: as we have shown above, for each sequence of integers $a_n \geq 1$, there exists an *irrational* rotation R_α with $a_n(\alpha) = a_n$. The rational numbers $p_n/q_n = [0; a_1, a_2, \ldots, a_n]$ converge to α in the way described above.

Usually one defines coefficients a_n of the continued fraction expansion of a number α by the expression (1.14). Because continued fraction expansion and rotations of the circle are so closely connected it seems in this book more natural to define a_n dynamically.

Exercise 1.1. Give a proof of the formulas for $\alpha(n)$ and a_n in the previous example. (Hint: use induction, $G(\frac{x}{1+x}) = G(x)$ and $[x + 1] = [x] + 1$; furthermore, distinguish between n even and n odd or in other words between $\alpha_n \in (0, 1/2)$ and $\alpha_n \notin (0, 1/2)$. Finally use the result of the first example that for $n \geq 2$, $a_n = \left[\frac{1}{\alpha_{n-1}}\right] - 1$ when n is even and $a_n = \left[\frac{1}{1-\alpha_{n-1}}\right] - 1$ when n is odd.)

From all this it follows that the itineraries K^+ and K^- of f are of a very special form. As before we will use the following notation related to sequences $\underline{x} = (x_0, x_1, \ldots)$, $\underline{y} = (y_0, y_1, \ldots) \in \Sigma$. Let $\underline{x}_n = (x_0, x_1, \ldots, x_{n-1})$ and

$$\underline{x}_n \cdot \underline{y}_m = (x_0, x_1, \ldots, x_{n-1}, y_0, y_1, \ldots, y_{m-1}).$$

Similarly define $\underline{x}_n^1 = \underline{x}_n$ and by induction for $m \geq 1$, $\underline{x}_n^m = \underline{x}_n \cdot \underline{x}_n^{m-1}$.

Lemma 1.4. *Let $f \in \mathcal{S}(J)$ and assume that f has no periodic points. Then for $n \geq 1$,*

$$K^+_{q_{2n+2}} = K^+_{q_{2n}} \cdot (K^+_{q_{2n+1}})^{a_{2n+2}}$$
$$K^+_{q_{2n+1}} = K^+_{q_{2n-1}} \cdot (K^-_{q_{2n}})^{a_{2n+1}}$$

The first equality also holds for $n = 0$ but then one has to replace the first two symbols RR by RL.

Proof. In order to prove the first equality it suffices to prove that

$$\underline{i}_f(f^{q_{2n}+i\cdot q_{2n+1}}(c))_{q_{2n+1}} = K^+_{q_{2n+1}},$$

for each $i = 0, \ldots, a_{2n+2} - 1$. From the description in the previous lemma, $f^i(c), f^i(f^{q_{2n}}(c)) \notin (f^{q_{2n+1}}(c), f^{q_{2n}}(c))$ for $i = 0, \ldots, q_{2n+1} - 1$ and therefore $f^i(c, f^{q_{2n}}(c)] \cap (f^{q_{2n+1}}(c), f^{q_{2n}}(c)) = \emptyset$ for all $i = 0, \ldots, q_{2n+1} - 1$. In particular, $c \notin f^i\,(c, f^{q_{2n}}(c))$ for $i = 0, \ldots, q_{2n+1} - 1$. Since $f^{q_{2n}}(c) > c$, it follows that $(\underline{i}_f(x))_{q_{2n+1}} = K^+_{q_{2n+1}}$ for all $x \in (c, f^{q_{2n}}(c)]$. Because $f^{q_{2n}+i\cdot q_{2n+1}}(c^+) \in (c, f^{q_{2n}}(c)]$ for $i = 0, \ldots, a_{2n+2}$, the first equality follows except when $q_{2n} = 1$ and $i = 0$ because then $f(c)$ is ambiguous. This can only hold if $n = 0$ and then the modified equality easily follows. The second equality follows similarly because $(\underline{i}_f(x))_{q_{2n}} = K^-_{q_{2n}}$ for all $x \in [f^{q_{2n-1}}(x), c)$. (More on the sequences $K^{\pm}(f)$ can be found in Gambaudo (1987) and Gambaudo et al. (1984).) □

Remark. From this lemma it follows that one can construct itineraries K^+ and K^- which correspond to a circle homeomorphism by using certain substitution rules, see e.g. Gambaudo (1987). From this lemma it also follows that each periodic orbit of a circle homeomorphism has 'a finite degree of complexity'. Here we say that the periodic point p has the first degree of complexity if $\underline{i}_f(p) = (L^n \cdot R)^\infty$ or $\underline{i}_f(p) = (R^n \cdot L)^\infty$. It has the second degree of complexity if it is of the form $\underline{i}_f(p) = ((L^{n\pm1}\cdot R)^m \cdot (L^n\cdot R))^\infty$ or $\underline{i}_f(p) = ((R^{n\pm1}\cdot L)^m \cdot (R^n \cdot L))^\infty$. In the same way one can define in general the degree of complexity of a periodic orbit. The rotation numbers corresponding to periodic orbits of the same degree of complexity are on the same level in the Farey tree. (The Farey tree orders all rational numbers in $[0,1]$ with increasing denominators. At the top level are the numbers $0/1$ and $1/1$ and on the second level $1/2$. All other rationals can be generated from these in a unique way by using that the largest rational between p/q and p'/q' is $(p+q)/(p'+q')$. So this also determines the level of a rational number in the tree.) More on this can be found in Gambaudo (1987), Gambaudo et al. (1984), see also Mira (1986).

Proposition 1.2. *If f and $\tilde{f}$ have no periodic points then $K^+(f) = K^+(\tilde{f})$ if and only if $a_i(f) = a_i(\tilde{f})$ for all $i \geq 0$.*

Proof. If $a_i(f) = a_i(\tilde{f})$ for all $i \geq 0$, then, by Lemma 1.4, we have that $K^+(f) = K^+(\tilde{f})$. The converse is also easy to verify. □

We can now state and prove the main result of this section easily from the results we obtained so far.

Theorem 1.1. (Poincaré) *If $f \in \mathcal{S}(J)$ does not have periodic points then there exist a unique rotation $R \in \mathcal{I}([0,1])$ and a continuous and surjective monotone map $h\colon J \to [0,1]$ such that*

$$h \circ f = R \circ h.$$

So h is a semi-conjugacy between f and R. Writing $R(x) = x + \alpha \bmod 1$ then α is irrational and equal to the rotation number of f.

Proof. Since f has no periodic points, $K^+(f) = (a_i(f))$ is defined for all $i \geq 0$. From the previous example, there exists a rotation R such that $a_i(f) = a_i(R)$ for all $i \geq 0$. From Proposition 1.2 we have that $K^+(f) = K^+(R)$. For convenience write $c_f = c(f)$ and $c_R = c(R)$. By Proposition 1.1, the map $h\colon O(c_f) \to O(c_R)$ defined by $h(f^n(c_f)) = R^n(c_R)$ for all $n \in \mathbb{Z}$ is monotone strictly increasing. R has no periodic points because otherwise $R^n(c_R)$ and therefore $f^n(c_f)$ would tend to some set consisting of finitely many points. This would imply that f has a periodic point, a contradiction. Hence R has no periodic points, and as we have seen in the beginning of this section this implies that $O(c_R)$ is dense in $[0,1]$. It follows that h extends continuously to the closure F of $O^+(c_f)$. Let $I = (x, y)$ be a connected component of $J \setminus F$. Since h is monotone on F, we have $h(x) \leq h(y)$. We claim that $h(x) = h(y)$. Indeed, if $h(x) < h(y)$ then, since the forward orbit of c_R is dense in $[0,1]$, there exists an integer n such that $h(x) < R^n(x) < h(y)$. On the other hand, we must have either $f^n(c_f) \leq x$ or $f^n(c_f) \geq y$ because $I \cap F = \emptyset$. Since $h(f^n(c_f)) = R^n(c_R)$, we get a contradiction because h is monotone. This proves that $h(x) = h(y)$. Therefore we can extend h continuously to I by setting $h(z) = h(x)$ for all $z \in I$. Hence h extends continuously to a monotone map from J to $[0,1]$ which clearly satisfies the condition of the theorem. □

Now we will give the definition of the rotation number of a circle homeomorphism.

Definition. Let $g \in \mathcal{S}(J)$. Then let a_n be defined as in (1.1) and (1.2). The *rotation number* of g is the real number defined by the continued fraction

$$\rho(g) = [0; a_1, a_2, \ldots, a_{n-1}, \ldots]$$

if all integers $a_n \geq 1$ are finite and

$$\rho(g) = [0; a_1, a_2, \ldots, a_{n-1}]$$

if $a_n = \infty$ (this can only occur if g has periodic points; if $a_1 = \infty$ then g has a fixed point and then we define $\rho(g) = 0$).

Lemma 1.5. *$\rho(g)$ is rational if and only if g has periodic points. Moreover, the map $\mathcal{S}(J) \ni g \mapsto \rho(g)$ is continuous.*

Proof. If g has no periodic points then the integers a_n defined below Lemma 1.2 always exist for each $n \geq 1$ and as we have seen this implies that $\rho(g) = [0, a_1, a_2, \dots]$ is irrational. Also we have seen that if g has periodic points then $a_n = \infty$ for some n and then $\rho(g)$ is rational. Furthermore, it follows immediately from the definitions that for each $N < \infty$ the numbers $a_n(g)$ and $a_n(\tilde{g})$ corresponding to $g, \tilde{g} \in \mathcal{S}(J)$ agree for $n = 1, \dots, N$ provided g and $\tilde{g}$ are sufficiently nearby in the C^0 topology. In particular, the map $\mathcal{S}(J) \ni g \mapsto \rho(g)$ is continuous. □

Let us relate the above construction with the classical definition of rotation number. If $g\colon S^1 \to S^1$ is a continuous map we consider a lift of g to the universal cover, namely, a map $\hat{g}\colon \mathbb{R} \to \mathbb{R}$ such that $\pi \circ \hat{g} = g \circ \pi$. Notice that there exists a unique lift $\hat{g}$ such that $\hat{g}(0) \in [0, 1)$. If g is a rotation, then its lift $\hat{g}$ is just the translation $T_\alpha(x) = x + \alpha$ where $\alpha \in [0, 1)$ is the rotation number. If $\hat{h}\colon \mathbb{R} \to \mathbb{R}$ is the lift of the semi-conjugacy $h\colon S^1 \to S^1$ of Theorem 1.1, then clearly $\hat{h} \circ \hat{g} = T_\alpha \circ \hat{h}$. Therefore $\hat{h} \circ \hat{g}^n = T_\alpha^n \circ \hat{h}$. Since $\hat{h}$ is the lift of h, we must have $\hat{h} = Id + \phi$ where Id is the identity map and ϕ is a periodic function of period 1. Thus $\hat{g}^n(x) + \phi(\hat{g}^n(x)) = x + \phi(x) + n\alpha$. Since ϕ is bounded, we get from the last equation that the limit of $\dfrac{\hat{g}^n(x) - x}{n}$ as $n \to \infty$ exists and is equal to α for every $x \in S^1$. Therefore an alternative (and more conventional) definition of the rotation number of g is

$$\rho(g) = \lim_{n \to \infty} \frac{\hat{g}^n(x) - x}{n}.$$

A comparison between different numerical algorithms for determining the rotation number of a circle diffeomorphism is made in Bruin (1992a).

Let us finish this section by showing that a homeomorphism of S^1 without periodic points has a unique minimal set. (A *minimal* set is a non-empty compact invariant set which has no non-empty, compact, invariant proper subset.) Indeed,

Proposition 1.3. *If $g\colon S^1 \to S^1$ has no periodic points then there exists a set K such that $\alpha(z) = \omega(z) = K$ for each $z \in S^1$. If K has interior points then $K = S^1$. If K has no interior points then K is perfect and totally disconnected, i.e., K is a Cantor set.*

Proof. Take $z \in S^1$ and let $K = \omega(z)$. Since S^1 is compact, K is non-empty, compact and completely invariant. Take $y \in S^1 \setminus K$. The orbit of y has at most one point in each connected component of $S^1 \setminus K$; otherwise one of the connected components of $S^1 \setminus K$ would be mapped by some iterate of g into itself, and g would have a periodic point. It follows that $\omega(y), \alpha(y) \subset \omega(z) = K$ for each $y \in S^1$ and similarly $\omega(y), \alpha(y) \subset \alpha(z)$. Since this holds for any $y, z \in S^1$ one gets that $\omega(y)$ does not depend on $y \in S^1$ and that $\alpha(y) = \omega(y) = K$. In particular, K does not contain any smaller non-empty closed invariant sets.

Since g is invertible and from the definition of $\omega(x)$, no point of K is isolated. Moreover, if K has an interior point then each point of K is an interior point. $\square$

Note that each orbit of an irrational rotation is dense in S^1. Therefore g is conjugate to an irrational rotation if and only if its minimal set K is equal to S^1. If g is conjugate to an irrational rotation there is no interval J such that all of its iterates $g^i(J)$, $i \geq 0$ are disjoint. Moreover, if g is not conjugate to such a rotation then $K \neq S^1$ and for each connected component J of $S^1 \setminus K$ the iterates $\{g^i(J)\}_{i\in\mathbb{N}}$ are pairwise disjoint. So g is conjugate to a rotation if and only if there exists no interval J such that $J, g(J), g^2(J), \ldots$ are pairwise disjoint. The existence of such intervals will be the main topic of the next section.

Exercise 1.2. Prove that for each homeomorphism $g\colon S^1 \to S^1$ there exists a probability measure μ on S^1 which is invariant under g, i.e., for which $\mu(g^{-1}(A)) = \mu(A)$ for any measurable set A. (Hint: take the Lebesgue measure λ on S^1 and define

$$\lambda_n = \frac{\lambda + g_*\lambda + \cdots + g_*^{n-1}\lambda}{n},$$

where $g_*\mu$ is the measure defined by $g_*\mu(A) = \mu(g^{-1}(A))$. Using the compactness of the space of probability measures on S^1, show that a subsequence of λ_n has a limit and that this limit measure is invariant.)

Exercise 1.3. Let $g\colon S^1 \to S^1$ be a homeomorphism without periodic points. Show that the support of any invariant probability measure μ of g is equal to the minimal set of g. Moreover, show that there exists at most one (and therefore precisely one) such invariant measure. (Hint: show that for each interval I and each $x \in S^1$, $\mu(I) = \lim_{n\to\infty} \#\{0 \leq i \leq n-1;\ g^i(x) \in I\}/n$.) Furthermore, let L be the clockwise oriented arc connecting x to $g(x)$: show that

$$\rho(g) = \mu(L).$$

Exercise 1.4. Let $g\colon S^1 \to S^1$ be a homeomorphism with irrational rotation number $\rho(g)$. Using the invariant measure from the previous exercise show that there is a continuous monotone map $h\colon S^1 \to S^1$ of degree one such that $h \circ g = R_\rho \circ h$. (Hint: let μ be the invariant measure of g, take a point x and choose $h(x)$ arbitrary. Then define h so that the Lebesgue measure of the segment $(h(x), h(y))$ is equal to $\mu(x, y)$.)

Exercise 1.5. Let $g\colon S^1 \to S^1$ be a non-invertible continuous map of the circle of degree one. In this exercise we shall see that one can apply some of the previous ideas to these non-invertible maps. Let $\hat{g}\colon \mathbb{R} \to \mathbb{R}$ be a lift of g. Show that the set of limits of

$$\frac{\hat{g}^n(x) - x}{n}$$

where $n \to \infty$ and $x \in S^1$ is a closed interval. This set is called the *rotation interval of* g. (Hint: associate to g a family of continuous order preserving circle maps $g_t\colon S^1 \to S^1$ of degree one such that i) for each $t \in [0,1]$, g_t is constant on some interval, ii) g_t coincides with g outside the regions where it is locally constant, and iii) for the lift $\hat{g}_t$ of g_t one has $\hat{g}_0(x) \leq \hat{g}(x) \leq \hat{g}_1(x)$ for all $x \in \mathbb{R}$, see Figure 1.4. Usually $\hat{g}_0$ is called the lower map and $\hat{g}_1$ the upper map. Since these maps g_t are monotone, one gets as for circle homeomorphisms that $\rho(g_t)$ exists. Show that $t \mapsto \rho(g_t)$ is increasing and that

$$\frac{\hat{g}_0^n(x) - x}{n} \leq \frac{\hat{g}^n(x) - x}{n} \leq \frac{\hat{g}_1^n(x) - x}{n}.$$

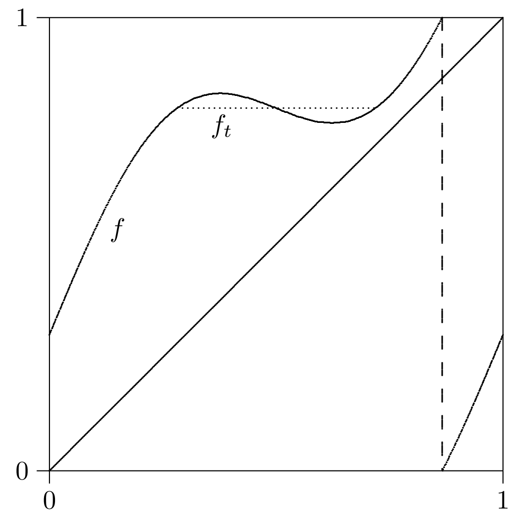

Fig. 1.4: The maps $f, f_t\colon [0,1] \to [0,1]$ corresponding to $g, g_t\colon S^1 \to S^1$ from Exercise 1.5.

In particular, the set of limits of $\frac{\hat{g}^n(x)-x}{n}$ is contained in the interval $[\rho(g_0), \rho(g_1)]$. Moreover, for each ρ in this interval, there exists $t \in [0,1]$ such that $\rho(g_t) = \rho$. Show that one can choose $x \in \mathbb{R}$ so that for all $n \in \mathbb{N}$, $g_t^n(x)$ is outside the region where g_t is locally constant. From this it follows that $g^n(x) = g_t^n(x)$ and therefore that $\frac{\hat{g}^n(x)-x}{n}$ converges to ρ.) Next show that for each ρ in the rotation interval there exists a point $x \in S^1$ such that its g-iterates are ordered in the circle in the same way as its R_ρ iterates, i.e., the map defined by $h(g^n(x)) = R_\rho^n(x)$ is injective and preserves the ordering of S^1. In particular, g has infinitely many periodic points if its rotation interval is non-trivial. (Hint: use the maps g_t defined above.)

2 The Topological Theory of Denjoy

In this section we are going to discuss a theorem proved by Denjoy (1932), stating that any C^2 circle diffeomorphism without periodic points is topologically equivalent to a rotation. More precisely, Poincaré's theorem from the previous section states that any homeomorphism of the circle without periodic points is semi-conjugate to a rotation; Denjoy's theorem strengthens this result by asserting that this semi-conjugacy is in fact a conjugacy provided both the homeomorphism and its inverse are sufficiently smooth. The main analytical ingredient of the proof is the control of the distortion on the restriction of iterates of the map to some intervals. Furthermore, examples in this section will be given which show that some such smoothness conditions are really necessary.

We will present two proofs for this theorem. The first one, which is the original proof of Denjoy, uses even weaker hypotheses: f is a C^1 diffeomorphism whose derivative is a function of bounded variation. The second proof was obtained by Schwartz (1963), which, although it requires a stronger smoothness hypothesis, uses almost no dynamical properties of circle diffeomorphisms and, consequently, can be used in more general situations as we will see.

In Section 1, we proved that if f is a circle diffeomorphism without periodic points then there exists a semi-conjugacy h between f and a rotation g. If h is not a conjugacy, then there exists a point x of the circle whose inverse image by h is an interval J. Since $h \circ f = g \circ h$, we have that $h(f^n(J)) = g^n(x)$. It follows that the intervals $J, f(J), f^2(J), \ldots$ are pairwise disjoint. (As we pointed out below the proof of Lemma 1.2, such an interval J exists if and only if the unique minimal set K of f is not equal to S^1.) This disjointness motivates the following definition.

Definition. We say that J is a *wandering interval* of the map f if:

- 1. the intervals $J, f(J), \ldots$ are pairwise disjoint;
- 2. the ω-limit set of J is not equal to a single periodic orbit.

A homeomorphism $f\colon S^1 \to S^1$ with periodic points does not have wandering intervals. Indeed, if J would be a wandering interval then the forward iterates of J were all disjoint. Therefore J could not contain any periodic point and it would follow as in the previous section that all points in J are asymptotic to one single periodic orbit; this would contradict condition 2) of the definition of a wandering interval. Moreover, it follows from the theorem of Poincaré, proved in Section 1, that a homeomorphism f without periodic points is conjugate to a rotation if and only if f does not have a wandering interval. It therefore suffices to prove that a C^1 diffeomorphism which has no periodic points and whose derivative is a function of bounded variation does not have a wandering interval. For that we will need some analytical tools.

Definition. Let $g\colon N \to N$ be a C^1 map where N is either the circle S^1 or the interval $[0,1]$. If $T \subset N$ is an interval such that $Dg(x) \neq 0$ for every $x \in T$, we define the *distortion* of g in T as:

$$\operatorname{Dist}(g, T) = \sup_{x,y \in T} \log \frac{|Dg(x)|}{|Dg(y)|}$$

Here $|Dg(x)|$ denotes the norm or absolute value of the derivative of g in x.

Notice that an affine map has distortion equal to zero. Given an interval J we shall denote its length by $|J|$.

Lemma 2.1. *Let $f\colon N \to N$ and T be an interval such that the restriction of f^n to T is a C^1 diffeomorphism. Then*

$$\operatorname{Dist}(f^n, T) \leq \sum_{i=0}^{n-1} \operatorname{Dist}(f, f^i(T)). \tag{2.1}$$

Proof. By the chain rule,

$$\log \frac{|Df^n(x)|}{|Df^n(y)|} = \sum_{i=0}^{n-1} \log \frac{|Df(f^i(x))|}{|Df(f^i(y))|}.$$

Since $f^i(x), f^i(y) \in f^i(T)$, we have that

$$\log \frac{|Df^n(x)|}{|Df^n(y)|} \leq \sum_{i=0}^{n-1} \mathrm{Dist}\,(f, f^i(T))$$

and this proves the lemma. □

Corollary 2.1. *Let $f\colon N \to N$ be a C^1 map such that $Df(x) \neq 0$ for all $x \in N$ and such that $x \mapsto \log|Df(x)|$ has Lipschitz constant C. Then for any interval $T \subset N$,*

$$\mathrm{Dist}\,(f^n, T) \leq C \sum_{i=0}^{n-1} |f^i(T)|.$$

In particular, if the intervals $T, f(T), \ldots, f^{n-1}(T)$ are pairwise disjoint, then

$$\mathrm{Dist}\,(f^n, T) \leq C\,\mathrm{diam}(N).$$

Proof. Proof The distortion of f on an interval J is bounded by $C \cdot |J|$. Therefore $\mathrm{Dist}\,(f^n, T) \leq \sum_{i=0}^{n-1} \mathrm{Dist}\,(f, f^i(T)) \leq C \sum_{i=0}^{n-1} |f^i(T)|$. □

Corollary 2.2. *Let $f\colon N \to N$ be a C^1 map such that the map $x \mapsto \log|Df(x)|$ has a variation which is bounded by C. Then there for any interval $T \subset N$ such that $T, f(T), \ldots, f^{n-1}(T)$ are pairwise disjoint intervals,*

$$\mathrm{Dist}\,(f^n, T) \leq C\,\mathrm{diam}(N).$$

Proof. The distortion of a map f on an interval J is bounded by the variation of $x \mapsto \log|Df(x)|$ on this interval. Hence, as the intervals are disjoint, the distortion of f^n on the interval T is bounded by the total variation of $x \mapsto \log|Df(x)|$. □

Corollary 2 gives us a control of the distortion of the restriction of the n-th iterate to some intervals whenever we have disjointness of the first n iterates of this interval. As we have seen in Lemma 1.3 of the previous section such disjointness often holds for circle diffeomorphisms f which have no periodic points. Indeed, let q_n be the n-th convergent of the rotation number of f then for any point x the first q_n iterates of the interval $[x, f^{-q_n}(x)]$ are all disjoint. If J is a wandering interval then all iterates of J are disjoint and it follows that the first q_n iterates $T, f(T), \ldots, f^{q_n-1}(T)$ of $T = [J, f^{-q_n}(J)]$ are all disjoint. Here T is the smallest subinterval of $S^1 \setminus f^{q_n}(J)$ which contains J and $f^{-q_n}(J)$. From this we get the following

Theorem 2.1. (Denjoy) *If $f\colon S^1 \to S^1$ is a C^1 diffeomorphism and its derivative is a function of bounded variation then f does not have a wandering interval.*

Proof. Let V be the total variation of $\log|Df|$ on S^1. Suppose, by contradiction, that there exists a wandering interval J. Since $\sum_{n=-\infty}^{\infty}|f^n(J)| \le 1$, the length of the iterates of J tends to zero. Let q_n be as in the previous section and as in Lemma 1.3 let $T = [f^{-q_n}(J), J]$ be the smallest interval containing J and $f^{-q_n}(J)$ which does not contain $f(J)$. The distortion of f^{q_n} in T is, as we have already seen, bounded by the sum of the variations of $\log|Df|$ in the q_n first iterates of T. Since, as we have seen in Lemma 1.3, these intervals are disjoint, we have that the distortion of f^{q_n} restricted to T is bounded by V. On the other hand, by the Mean Value Theorem, there exist points $x \in J$ and $y \in f^{-q_n}(J)$ such that $|Df^{q_n}(x)| = \frac{|f^{q_n}(J)|}{|J|}$ and $|Df^{q_n}(y)| = \frac{|J|}{|f^{-q_n}(J)|}$. Thus $\frac{|J||J|}{|f^{q_n}(J)||f^{-q_n}(J)|} \le \exp(\mathrm{Dist}\,(f^{q_n}, T)) \le \exp(V)$. This is not possible for n large because $|f^k(J)| \to 0$ as $|k| \to \infty$. □

Remark. Note that the previous results do not apply to C^∞ homeomorphisms $f\colon S^1 \to S^1$ with a critical point c (this is a point such that $Df(c) = 0$). Indeed, in this case $\log Df$ is unbounded. Later on, in Chapter IV, we will develop tools which also apply to maps with critical points. The proof is, however, much more elaborate. The proof above uses that Df exists almost everywhere, that the Mean Value Theorem can be applied, and that $\log|Df|$ can be extended to a map with bounded variation. In particular, Denjoy's theorem is valid for homeomorphisms of S^1 which are piecewise linear. Later on, in Chapter IV we shall see that this theorem also holds under slightly different conditions (that $\log Df$ satisfies the Zygmund condition).

The next theorem shows that the non-existence of wandering intervals also holds for piecewise monotone maps $f\colon N \to N$ as long as one requires more smoothness of f. We should emphasize that by definition a piecewise monotone map has a finite number of turning points.

Theorem 2.2. **(Schwartz)** *Let N be a compact interval and $f\colon N \to N$ be a continuous mapping satisfying the following conditions: i) f is piecewise monotone and piecewise C^1 and ii) the mapping $x \mapsto \log|Df(x)|$ extends to a Lipschitz function on N. Then f does not have wandering intervals.*

Proof. It follows from (ii) that there exists a constant $C > 0$ such that $\mathrm{Dist}\,(f, T) < C|T|$ for every interval $T \subset N$. Hence, by Lemma 2.1, we have that for all intervals T and all integers n,

$$\mathrm{Dist}\,(f^n, T) \le \sum_{i=0}^{n-1} C|f^i(T)|.$$

Suppose, by contradiction, that f has a wandering interval J. Let $J_0 \supset J$ be a maximal wandering interval containing J, in the sense that there is no wandering interval that properly contains J_0. Inductively, we define J_n as a maximal wandering interval that contains the interior of $f(J_{n-1})$. We claim that the intervals we have defined are pairwise disjoint. Indeed, suppose, by contradiction, that there exist integers $0 \le n < m$ such that $J_n \cap J_m \neq \emptyset$. Since J_m contains the interior of $f^{m-n}(J_n)$, it follows that $f^{m-n}(J_m) \cap J_m \supset f^{m-n}(J_m \cap J_n) \neq \emptyset$. This is a contradiction since J_m is a wandering interval and therefore the claim is proved. By definition, f has at most a finite number, say n_0, of turning points (i.e, local maxima and minima). Hence, there exists n_1 such that if $n \ge n_1$ then $f^n(J_0)$ does not contain a turning point of f. In particular, f^k is a homeomorphism in J_{n_1} for every $k > 0$. Hence, if $m > n \ge n_1$ then $f^{m-n}(J_n) = J_m$. Since the intervals $\{J_n\}_{n\ge 0}$ are pairwise disjoint, there exists $n_2 > n_1$ such that $\sum_{j\ge 0} |f^j(J_{n_2})| \le 1$. Take $0 < \delta < (1/2)\exp(-2C)$ and let $I = J_{n_2}$. Now choose an interval T which strictly contains I such that $|T| \le (1+\delta)|I|$. Let us prove by induction that $|f^i(T)| \le 2|f^i(I)|$ for $i = 0, 1, \ldots, n-1$. For $n = 1$ this statement is obvious. So assume it holds for some $n > 1$. As we have seen in Corollary 1, the induction assumption implies that the distortion of f^n in T is bounded by $2C$. Moreover, since the restriction of f^n to I is a diffeomorphism, there exists by the Mean Value Theorem a point $x_n \in I$ such that $Df^n(x_n) = \frac{|f^n(I)|}{|I|}$. Therefore $|Df^n(y)| \le \exp(2C)\frac{|f^n(I)|}{|I|}$ for every $y \in T$. Hence,

$$|f^n(T)| \le |f^n(I)| + \exp(2C)\frac{|f^n(I)|}{|I|}|T \setminus I| \le (1 + \delta\exp(2C))|f^n(I)| \le 2|f^n(I)|.$$

This completes the induction step. It follows that $\lim_{n\to\infty} |f^n(T)| = 0$.

We claim that if T is an interval such that $\inf_{n\to\infty} |f^n(T)| = 0$ then T is either a wandering interval or $\omega(T)$ is a periodic orbit. This statement is called the Contraction Principle. Clearly, because $|f^n(T)| \to 0$ it suffices to prove this principle: since I is a maximal wandering interval and T strictly contains I neither of these possibilities can hold. This contradicts the existence of a wandering interval and completes the proof of the theorem.

So let us prove the claim. Let $\mathcal{I} = \cup_{n\ge 0} f^n(\text{int}\,(T))$. Clearly $\mathcal{I}$ is forward invariant. First suppose that there exists a component U of $\mathcal{I}$ and $s > 0$ such that $f^s(U) \cap U \neq \emptyset$. Since $\mathcal{I}$ is forward invariant this implies $f^s(U) \subset U$. If U is an interval which contains a fixed point p of $f^s\colon U \to U$ in its interior, then some iterate of T contains this fixed point of f^s in its closure and since $\inf_{k\ge 0} |f^k(T)| = 0$ this fixed point of f^s attracts T. So we are finished in this case. Otherwise, $\text{cl}\,(U)$ contains in its boundary an attracting fixed point p of $f^s\colon \text{cl}\,(U) \to \text{cl}\,(U)$. If $f^s(U) \neq U$ then for every $x \in \text{cl}\,(U)$ one gets $f^{ks}(x) \to p$ as $k \to \infty$. If $f^s(U) = U$ then the boundary point $\{q\} = \partial U \setminus \{p\}$ is a repelling periodic point and $\inf_{k\ge 0} |f^k(T)| = 0$ implies that no iterate of T contains q in its closure. Since every point in $\text{int}\,(U)$ is asymptotic to p under iterates of f^s this implies that $f^{ks}(x) \to p$ for $x \in T$ as $k \to \infty$. Again the result follows.

Now assume that for every component U of $\mathcal{I}$ one has $f^s(U) \cap U = \emptyset$ for all $s \ge 1$. Since $\mathcal{I}$ is forward invariant and this holds for each component, this implies that $f^n(U) \cap f^m(U) = \emptyset$ for all $n > m \ge 0$. It follows that U and therefore I is a wandering interval (or asymptotic to a periodic orbit). $\square$

Remark. We should emphasize that this proof also shows the following more abstract statement: suppose that the Lipschitz constant of $\log|Df|$ is bounded by $C<\infty$. Then there exists $\delta>0$ so that if for some $n\in\mathbb{N}$ and some interval J, $f^n|J$ is a diffeomorphism and $\sum_{i=0}^{n-1}|f^i(J)|\le 1$ then for any interval $T\supset J$ with $|T|\le(1+\delta)|J|$ one has

$$|f^i(T)|\le 2\cdot|f^i(J)|, \text{ for all } i=1,\dots,n-1,$$

$$\exp(-2C)\frac{|f^n(J)|}{|J|}\le|Df^n(x)|\le\exp(2C)\frac{|f^n(J)|}{|J|},$$

for all $x\in T$. So if one has a bound on $\sum_{i=0}^{n-1}|f^i(J)|$ one gets an estimate on the non-linearity of f^n on an interval T which is definitely larger than J. Note the difference with Corollaries 1 and 2 above: there bounds on $\sum_{i=0}^{n-1}|f^i(T)|$ or disjointness of $T,\dots,f^{n-1}(T)$ are required (so assumptions on $f^i(T)$ rather than on $f^i(J)$ are made).

Let us show that this result of Schwartz can be applied to non-invertible maps, and can be used to obtain results about flows on surfaces.

Corollary 2.3. *Let $f\colon S^1\to S^1$ be a C^2 mapping such that $Df(x)\ne 0$ for every $x\in S^1$. Then f does not have a wandering interval.*

Proof. Since the derivative of f is not zero in every point, we have that f is a covering map of degree d, $|d|\ge 1$ (and can be homotopied in $\mathbb{R}/\mathbb{Z}$ to a map of the form $z\to d\cdot z \bmod 1$; this means that there is a family f_t of covering maps depending continuously on t such that $f_0=f$ and $f_1(z)=d\cdot z$). Suppose that f has a wandering interval J. Let $x\in\text{interior}(J)$ and $\phi\colon S^1\setminus\{x\}\to(0,1)$ be an isometry. Then the mapping g which is defined by $\phi\circ f\circ\phi^{-1}$ in $(0,1)$ and $g(0)=g(1)=\phi(f(x))$ is C^2 in $[0,1]\setminus\phi(f^{-1}(x))$. Since $g^n(\phi(f(J)))=\phi(f^n(f(J)))$, we have that $I=\phi(f(J))$ is a wandering interval for g and $\cup_{k\ge0}g^k(I)$ does not intersect the $|d|$ intervals $U=\phi(f^{-1}(J))\supset\phi(f^{-1}(x))$. Therefore we can modify g in U, in order to obtain a mapping $h\colon[0,1]\to[0,1]$ which coincides with g outside of U and satisfies the conditions i) and ii) of the previous theorem, see Figure 2.1. Hence, I is a wandering interval of h and this is a contradiction. □

Corollary 2.4. *Let $f\colon N\to N$ be a C^2 map with a finite number of critical points where N is either the circle or a compact interval. If f has a wandering interval J then the ω-limit set of J contains a critical point.*

Proof. If f has a wandering interval which does not accumulate at critical points, we can, as in the proof of the previous corollary, modify f in order to get a mapping which satisfies the conditions of Theorem 2.2 and which has a wandering interval. □

Both Denjoy and Schwartz had vector fields on surfaces in mind when they proved their results. Denjoy used his theorem to prove the following result for

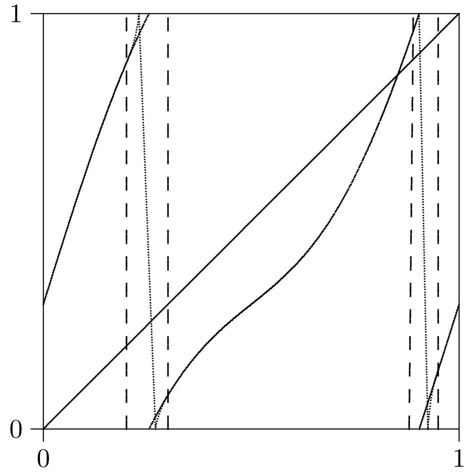

Fig. 2.1: This modification of the map $f\colon S^1 \to S^1$ makes it into an interval map with 'sharp' turning points.

C^2 vector fields on a torus, $S^1 \times S^1$. If such a vector field has no singularities and no periodic orbits then each orbit of this vector field is dense in the torus. This is done by showing that such a vector field has a global cross section; the first return map to this cross section defines a C^2 circle diffeomorphisms. Schwartz generalized this last result to general compact manifolds:

Corollary 2.5. (Schwartz) *Let X be a C^2 vector field in a two-dimensional compact manifold M. Then every minimal set K of X is trivial. More precisely, there are only three possibilities:*

- *1. K is a singularity;*
- *2. K a periodic orbit;*
- *3. K is equal to M and M is the torus.*

Proof. By contradiction let K be a non trivial minimal set of X, i.e., K is a closed, invariant subset, K does not contain a closed invariant proper subset, and K is not the torus, a closed orbit or a singularity of X. In particular, K is equal to the orbit closure of each of its points. Take a point of K and a circle Σ through this point transversal to the vector field X. (For a proof that such a curve exists see, for example, Palis and de Melo (1982, pp.144). Let $D \subset \Sigma$ be the domain of definition of the first return map $\Phi\colon D \to \Sigma$. We have that D is a union of open intervals. Since D contains $K \cap \Sigma$ and K is compact, only a finite number of connected components of D cover K. Therefore we can modify Φ in the complement of these connected components in order to obtain a C^2 circle map $g\colon \Sigma \to \Sigma$ which satisfies the conditions of Theorem 2.2 and which coincides with Φ in a neighbourhood of V of $K \cap \Sigma$. We may assume that V

consists of a finite union of intervals and therefore that there exists at most a finite number of connected components of $\Sigma \setminus K$ which are not contained in V. It follows that we can take a connected component J of $\Sigma \setminus K$ such that $g^n(J) \subset V$ and therefore $g^n|J = \Phi^n|J$ for all $n \geq 0$. Since Φ^n is invertible, both endpoints of $\Phi^n(J)$ are contained in K and because these endpoints are not periodic, it follows that the intervals $\{\Phi^n(J)\}_{n\geq 0}$ are pairwise disjoint and no point of J is asymptotic to a periodic orbit. Hence, J is a wandering interval of Φ and therefore of g. This is in contradiction with Theorem 2.2. □

The theorem below, also due to Denjoy (1932), shows that it is definitely not enough to assume that f is C^1 in order to get the non-existence of wandering intervals. In other words, this result states that there exists a C^1 diffeomorphism of the circle which has no periodic points and whose orbits are non-dense in the circle. In fact, similar counter-examples were first constructed by Bohl in 1911, see the paper by Rosenberg in Denjoy (1975).

Theorem 2.3. *For every irrational number α, there exists a C^1 diffeomorphism f with rotation number equal to α which has a wandering interval.*

Proof. For each integer $n \in \mathbb{Z}$, choose a positive number λ_n such that

$$\sum_{-\infty}^{\infty} \lambda_n = 1 \text{ and} \tag{2.2}$$

$$\frac{\lambda_{n+1}}{\lambda_n} \to 1 \text{ as } |n| \to \infty. \tag{2.3}$$

Let R be the rotation with rotation number equal to α and let $x \in S^1$, $x_n = R^n(x)$ for $n \in \mathbb{Z}$. We claim that there exists a family $\{I_n\,;\ n \in \mathbb{Z}\}$ of disjoint open intervals in S^1 which are ordered in the same way in S^1 as $\{x_n = R^n(x)\,;\ n \in \mathbb{Z}\}$, such that $|I_n| = \lambda_n$, for all $n \in \mathbb{Z}$ and such that $\cup_{n=-\infty}^{\infty} I_n$ is dense in S^1. For this we shall use a diagonal argument. Indeed, because $\sum_{j=-n}^{n} \lambda_j < 1$, there exists for each $n \in \mathbb{N}$ disjoint closed intervals $T_{n,j}$ where $|j| \leq n$ which have the same order in the circle as x_j, $|j| \leq n$ such that $|T_{n,j}| = \lambda_j$ for $|j| \leq n$ and such that each of the components of $S^1 \setminus \cup_{|j|\leq n} T_{n,j}$ has equal length. Next, since S^1 is compact, for each $N \in \mathbb{N}$ we can take a sequence $n_1(N), n_2(N), n_3(N), \dots$ with $n_i(N) \to \infty$ as $i \to \infty$ such that for each $|j| \leq N$, $T_{n_i(N),j}$ converges to an interval T_j^N as $i \to \infty$. Of course for each $N \in \mathbb{N}$, we can choose the sequence $n_1(N), n_2(N), n_3(N), \dots$ to be a subsequence of the sequence $n_1(N-1), n_2(N-1), n_3(N-1), \dots$. If we do this, then for fixed j the intervals T_j^N are all the same for $N \geq |j|$ and so if we define $I_j = T_j^j$ we obtain a family $\{I_j\,;\ j \in \mathbb{Z}\}$ satisfying the claim. Let $A = \cup_{n=-\infty}^{\infty} I_n$. From the claim and (2.2) we have that A is an open, dense subset of the circle and has full measure.

From (2.3), there exists, for each $n \in \mathbb{Z}$, a C^∞ diffeomorphism $f_n \colon I_n \to I_{n+1}$ such that the derivative of f_n is equal to one in the boundary of I_n and both the maximum and the minimum of the derivative of f_n tend to 1 as $|n| \to \infty$. Let $f\colon A \to A$ be defined by $f(x) = f_n(x)$ if $x \in I_n$ and $h\colon A \to S^1$, $h(I_n) = x_n$. Clearly $h \circ f = R \circ h$ and both f and h are monotone maps. Therefore, as both A and $O_R(x)$ are dense in the circle, h and f extend continuously to the circle and h is a semi-conjugacy between f and R. In particular, f has rotation number α.

It remains to prove that f is a C^1 diffeomorphism. For that, it is clearly enough to prove that f is differentiable at each point $y \in S^1 \setminus A$ and that its derivative at y is equal to one. Let y and z be nearby points and let $[y,z]$ denote the smallest arc of the circle connecting y to z. To prove that f is differentiable at y and its derivative is equal to one we have to show that

$$\lim_{z \downarrow y} \frac{|f([y,z])|}{|[y,z]|} = 1 \tag{2.4}$$

and similarly that the corresponding left-sided limit is equal to one. Since the derivative of $f_n|\partial I_n$ is equal to one, (2.4) holds whenever a right-sided neighbourhood of y is completely contained in one of the intervals I_n. Therefore we may assume that each right-sided neighbourhood of y intersects an infinite number of the intervals I_n. The length of $[y,z]$ can be estimated from above and from below by summing all intervals I_n entirely contained in $[y,z]$ plus (for the estimate from above) a piece of the interval I_n that contains z, if any. More precisely, let $J(z) = \{n\,;\ I_n \cap (y,z) \neq \emptyset\}$ and $J'(z) = \{n\,;\ I_n \subset (y,z)\}$. Because $\cup I_n$ has full measure in S^1,

$$\sum_{n \in J'(z)} \lambda_n \leq |[y,z]| \leq \sum_{n \in J(z)} \lambda_n.$$

For the same reason, because I_n is contained in $[y,z]$ if and only if I_{n+1} is contained in $f([y,z])$,

$$\sum_{n \in J'(z)} \lambda_{n+1} \leq |f([y,z])| \leq \sum_{n \in J(z)} \lambda_{n+1}.$$

Hence,

$$\frac{\sum_{n \in J'(z)} \lambda_{n+1}}{\sum_{n \in J(z)} \lambda_n} \leq \frac{|f([y,z])|}{|[y,z]|} \leq \frac{\sum_{n \in J(z)} \lambda_{n+1}}{\sum_{n \in J'(z)} \lambda_n}. \tag{2.5}$$

By construction $J'(z) \subset J(z)$ and $\#(J(z) \setminus J'(z)) \leq 2$ and $\inf\{|n|\,;\ n \in J(z)\}, \inf\{|n|\,;\ n \in J'(z)\} \to \infty$ as $z \to y$. Therefore, because $\lambda_n \to 0$ and $\lambda_{n+1}/\lambda_n \to 1$ as $n \to \infty$, the outer terms in (2.5) tend to one as $z \downarrow y$. This proves (2.4) and completes the proof of this theorem. □

Remarks.

1. It is not hard to find diffeomorphisms $f_n\colon I_n \to I_{n+1}$ as in the proof of Theorem 2.3 such that $||Df_n - 1|| \leq 2\frac{\lambda_{n+1}}{\lambda_n}$. It follows that $||Df - 1|| \leq 2\sup_{n\in\mathbb{Z}} \frac{\lambda_{n+1}}{\lambda_n}$. In particular, in every C^1 neighbourhood of R_α there exists a C^1 diffeomorphism

f with a wandering interval. In Section X of Herman (1979) this statement is generalized as follows: for each irrational number α there exists a dense subset R in the space of C^1 diffeomorphisms of the circle with rotation number α such that each diffeomorphism in R has a wandering interval.

2. We should also note that even if $\epsilon > 0$ is small and f is a $C^{2-\epsilon}$ diffeomorphism, $\log|Df|$ need not have bounded variation. Here f is said to be $C^{2-\epsilon}$ if Df satisfies the Hölder condition that

$$\sup_{x \neq y} \frac{|Df(x) - Df(y)|}{|x-y|^{1-\epsilon}} < \infty.$$

In particular, a $C^{2-\epsilon}$ diffeomorphism need not satisfy the conditions of Theorem 2.1. And indeed, in Section X of Herman (1979) there are examples of $C^{2-\epsilon}$ diffeomorphisms with (an arbitrary) irrational rotation number and having a wandering interval.

3. For any Cantor set K of the circle it is easy to construct a homeomorphism f of the circle such that the minimal set of f coincides with K. It is not so clear what restrictions one has to take on the Cantor set K in order for there to be a diffeomorphism with a certain smoothness which has K as a minimal set, see however McDuff (1981).

4. As we noted before we shall generalize the theorem of Denjoy to smooth maps satisfying some weak regularity conditions. Now if we study flows on surfaces one gets return maps which can have discontinuities (due to the presence of saddles). From Levitt (1983), see Exercise 2.2 below, it is known that in general such maps can have wandering intervals. For special piecewise continuous maps one can sometimes exclude the existence of wandering intervals, see Berry and Mestel (1991).

Exercise 2.1. Show that the diffeomorphism from the previous theorem can also be chosen so that its minimal set K has positive Lebesgue measure. (Hint: Choose the numbers $\lambda_n > 0$ so that $\rho^{-1} := \sum_{-\infty}^{\infty} \lambda_n < 1$. Furthermore, choose the gaps in the construction of the intervals $T_{n,j}$, $|j| \leq n$ so that the component of $S^1 \setminus \cup_{|j| \leq n} T_{n,j}$ to the right of $T_{n,j}$ has length equal to $\rho \cdot \lambda_n$. Then $A = \cup I_n$ has Lebesgue measure equal to $\rho^{-1} < 1$ and

$$\sum_{n \in J'(z)} \rho\lambda_n \leq |[y,z]| \leq \sum_{n \in J(z)} \rho\lambda_n.$$

From this it follows as before that the map f is a C^1 diffeomorphism.)

Exercise 2.2. In this exercise we shall show that there exists a piecewise affine interval exchange transformation on $[0,1]$ which has wandering intervals. This result is due to G. Levitt (1983) and is based on the uniqueness of ergodic measures for interval exchange transformations, see Sataev (1975), Keynes and Newton (1976) and Keanes (1977) and also Cornfield et al. (1982). More precisely, $f\colon [0,1] \to [0,1]$ is called an interval exchange transformation if $[0,1]$ is the union of a finite number of intervals $I_1, \ldots, I_r$ such that f acts as a translation on each of these intervals and such that the closure of $f(I_1) \cup \cdots \cup f(I_r)$ together covers $[0,1]$ again. The results we just mentioned state that there are interval exchange transformations f which are transitive (so all orbits are dense) and with two different ergodic probability measures μ and ν (notice that the Lebesgue measure is certainly invariant under f). It is easy to see that this implies that $\mu(I_i)$ and $\nu(I_i)$ cannot be equal for all i. In particular, we may assume that

$$\mu(I_1) > \nu(I_1) \text{ and } \nu(I_2) > \mu(I_2).$$

The construction goes in a few steps. 1) Assume that $\mu_1 > \nu_1 > 0$ and $\nu_2 > \mu_2 > 0$. Show that there exists positive numbers a, b such that

$$a^{\mu_1} b^{\mu_2} < 1 < a^{\nu_1} b^{\nu_2}.$$

(Hint: choose for example t so that $\frac{\mu_2}{\mu_1} < -t < \frac{\nu_2}{\nu_1}$ and let $a = e^t$ and $b = e$.) 2) Let x be a typical point with respect to μ for f and similarly let y be typical for ν and f^{-1}. By this we mean that

$$\frac{\#\{0 \le k \le n-1\,;\ f^k(x) \in I_i\}}{n} \to \mu(I_i),$$

$$(*) \qquad \frac{\#\{-n+1 \le k \le 0\,;\ f^k(y) \in I_i\}}{n} \to \nu(I_i)$$

as $n \to \infty$ for $i = 1, \dots, r$. That this is possible follows from the Birkhoff ergodic theorem, see the Appendix. Now we replace (glue) in the position of the point $f^k(x)$, $k \ge 0$ an interval J_k whose length is defined as follows. Choose the length of J_0 arbitrarily and let for $k \ge 0$

$$|J_{k+1}| = \begin{cases} a \cdot |J_k| & \text{if } f^k(x) \in I_1 \\ b \cdot |J_k| & \text{if } f^k(x) \in I_2 \\ |J_k| & \text{otherwise,} \end{cases}$$

where a and b are defined as in step 1 of this exercise. Similarly, define for $k \ge 0$ an intervals J'_{-k} such that J'_{-k} is glued in the position of $f^{-k}(y)$ whose length is defined as follows. The intervals J'_0 and J_0 have equal length and, furthermore,

$$|J'_{-k}| = \begin{cases} a^{-1} \cdot |J'_{-k+1}| & \text{if } f^{-k}(y) \in I_1 \\ b^{-1} \cdot |J'_{-k+1}| & \text{if } f^{-k}(y) \in I_2 \\ |J'_{-k+1}| & \text{otherwise.} \end{cases}$$

Show that $\sum_{k \ge 0} |J_k| + |J'_{-k}|$ is bounded. This implies that the new interval $\hat{I}$ with these intervals $\cup_{k\ge 0} J_k \cup J'_{-k}$ glued in is again compact interval. (Hint: use $(*)$ and the definition of a and b.) 3) Let $\hat{I}$ be the interval from above and define $\hat{f} \colon \hat{I} \to \hat{I}$ so that for $k > 0$ $\hat{f}$ is an affine map from J_k onto J_{k+1} and from J'_{-k} onto J'_{-k+1}. Moreover, make sure that $\hat{f}$ maps J'_0 affinely onto J_0. Check that $\hat{f}$ is piecewise affine and exchanges $r+4$ intervals. (Hint: since the points $f^k(x)$, $f^{-k}(y)$, $k \ge 0$ are dense, it follows that $\hat{f}$ is well defined, Moreover, it is continuous on the intervals corresponding to the original intervals I_i except on the boundary of the intervals J'_0 and J_0. So the resulting map g has four more discontinuities. From the definition of the length of J_k one has that

$$D\hat{f}(z) = \begin{cases} a & \text{if } z \in I_1 \\ b & \text{if } z \in I_2 \\ 1 & \text{otherwise,} \end{cases}$$

At least if $z \notin J_0 \cup J_0'$; in that case the formula is slightly different. Hence, $\hat{f}$ is piecewise affine.) 4) All iterates of J_0 under $\hat{f}$ are disjoint by construction and, since f has no periodic points, $\hat{f}$ has no periodic points either.

2.1 The Denjoy Inequality

Let $\frac{p_n}{q_n}$ be the convergents of $\alpha \in (0,1)$ as in Section 1. In this section we shall show that if $\log Df$ has bounded variation then $||Df^{q_n}||$ is bounded from below and from above.

Theorem 2.4. **(Denjoy Inequality)** *Assume that* $f\colon S^1 \to S^1$ *is a* C^1 *diffeomorphism and that* $x \to \log|Df(x)|$ *has bounded variation. Then* $||\log Df^{q_n}||$ *is bounded.*

In other words, the derivative of the n-th renormalization $\mathcal{R}^n(f)$ of f from Section 1 is uniformly bounded and bounded away from zero for all $n \in \mathbb{N}$. In the next section – in Lemma 3.4 – we will see that one can find improved bounds provided f is C^3. In order to show that $|\log Df^{q_n}|$ is bounded we will use the tools of the first part of this section for bounding the non-linearity of f^{q_n}. This gives estimates for the non-linearity of f^i on an interval T provided f is C^2 and the first q_n iterates of the interval T are disjoint. In Section 1 it is shown how we can choose T: let $I_n = [x, f^{q_n}(x)]$ be the interval bounded by x and $f^{q_n}(x)$ which does not contain $f(x)$. Similarly define $J_n(x) = [f^{-q_n}(x), f^{q_n}(x)] = I_n(f^{-q_n}(x)) \cup I_n(x)$ in $S^1 \setminus \{f(x)\}$. Then Lemma 1.3 states that the intervals $I_n(x), f(I_n(x)), \ldots, f^{q_{n+1}-1}(I_n(x))$ are pairwise disjoint and that the intervals $J_n(x), f(J_n(x)), \ldots, f^{q_{n+1}-1}(J_n(x))$ cover the circle.

Proof of the Denjoy inequality. First we claim that there exists a constant C which is independent of n such that

$$\frac{1}{C} < \frac{|[x, f^{q_n}(x)]|}{|[f^{-q_n}(x), x]|} = \frac{|I_n(x)|}{|I_n(f^{-q_n}(x))|} < C$$

for every $x \in S^1$. In order to see this let $K_0 = I_n(f^{-q_n}(x)) = [f^{-q_n}(x), x]$, $K_1 = I_n(x) = f^{q_n}(K_0)$ and $K_2 = I_n(f^{q_n}(x)) = f^{q_n}(K_1)$. These intervals have disjoint interiors. We must find a uniform upper and lower bound for $\frac{|K_1|}{|K_0|}$. Since the iterates of K_j up to $q_{n+1} - 1$ are pairwise disjoint, we have that:

$$\sum_{i=0}^{q_{n+1}-1} |f^i(K_0 \cup K_1 \cup K_2)| < 3.$$

From Lemma 2.1, we get a constant $C > 1$ such that

$$\frac{1}{C} < \frac{|Df^k(t)|}{|Df^k(y)|} < C \tag{2.6}$$

for all $t, y \in K_0 \cup K_1 \cup K_2$, and for all $0 \le k < q_{n+1}$. As $f^{q_n}(K_0) = K_1$ and $f^{q_n}(K_1) = K_2$ we have, by (2.6) that

$$\frac{1}{C}\frac{|K_2|}{|K_1|} < \frac{|K_1|}{|K_0|} < C\frac{|K_2|}{|K_1|}.$$

Thus let $A := \frac{|K_1|}{|K_0|}$. Then $\frac{|K_2|}{|K_1|} > \frac{1}{C}A$ and therefore

$$|K_0| + |K_1| = \frac{1}{A}|K_1| + |K_1| < (\frac{1}{A}+1)\frac{C}{A}|K_2|.$$

From the above inequality and (2.6) we have that

$$|f^i(K_0 \cup K_1)| < C(\frac{1}{A}+1)\frac{C}{A}|f^i(K_2)|$$

for $i = 0, \ldots, q_{n+1}-1$. Since the first $q_{n+1}-1$ iterates of $K_0 \cup K_1$ cover the circle and because the first q_{n+1} iterates of $I_n(x)$ are disjoint,

$$1 \le \sum_{i=0}^{q_{n+1}-1} |f^i(K_0 \cup K_1)| < C(\frac{1}{A}+1)\frac{C}{A}.$$

Therefore $A \le 2C^2$ and this proves that

$$\frac{|K_1|}{|K_0|} = A \le 2C^2 = C.$$

Similarly we can show that $\frac{|K_1|}{|K_0|} > \frac{1}{C}$. This completes the proof of the claim.

As $f^{q_n}(I_n(f^{-q_n}(x)) = I_n(x)$ we get from (2.6), taking $k = q_n$,

$$\frac{1}{C}\frac{|I_n(x)|}{|I_n(f^{-q_n}(x))|} < |Df^{q_n}(x)| < C\frac{|I_n(x)|}{|I_n(f^{-q_n}(x))|}.$$

Hence, the proof follows from the claim. □

2.2 Ergodicity

Now we will show that if $f\colon S^1 \to S^1$ is a C^1 diffeomorphism such that $\log Df$ has bounded variation then every measurable invariant set Λ either has zero or full Lebesgue measure. As before, if $A \subset S^1$ is measurable denote the Lebesgue measure of A by $|A|$. A map $f\colon S^1 \to S^1$ is said to be *ergodic with respect to the Lebesgue measure* if any measurable set X for which $f^{-1}(X) = X$ has either zero or full Lebesgue measure. In Chapter V we shall discuss this notion in detail and see that this notion is extremely useful.

Theorem 2.5. *Let $f\colon S^1 \to S^1$ be a C^1 diffeomorphism such that $x \to \log|Df(x)|$ has bounded variation. Then f is ergodic.*

Proof. Suppose that the Lebesgue measure of Λ is positive. Then there exists a density point of Λ, i.e., a point x such that for each sequence of neighbourhoods I_n of x with $|I_n| \to 0$ as $n \to \infty$,

$$\frac{|\Lambda \cap I_n|}{|I_n|} \to 1, \text{ as } n \to \infty \tag{2.7}$$

(see the appendix for more details on this). Take the neighbourhood $J_n(x)$ of x bounded by $f^{-q_n}(x)$ and $f^{q_n}(x)$. As we have seen in Lemma 1.3, each point of S^1 is contained in at least one and at most two of the first $q_{n+1} - 1$ iterates of $J_n(x)$. From this disjointness and from Lemma 2.1, there exists $C < \infty$ such that

$$\frac{1}{C} \leq \frac{|Df^i(y)|}{|Df^i(z)|} \leq C,$$

for all $i = 0, \ldots, q_{n+1} - 1$ and each $y, z \in J_n(x)$. In particular, since Λ is invariant,

$$\frac{|f^i(J_n(x)) \setminus \Lambda|}{|f^i(J_n(x))|} \leq C \cdot \frac{|J_n(x) \setminus \Lambda|}{|J_n(x)|}. \tag{2.8}$$

Since the intervals $J_n(x), f(J_n(x)), \ldots, f^{q_{n+1}-1}(J_n(x))$ cover S^1,

$$|S^1 \setminus \Lambda| \leq \sum_{i=0}^{q_{n+1}-1} |f^i(J_n(x) \setminus \Lambda)|$$

and applying equation (2.8) gives,

$$|S^1 \setminus \Lambda| \leq C \cdot \frac{|J_n(x) \setminus \Lambda|}{|J_n(x)|} \cdot \sum_{i=0}^{q_{n+1}-1} |f^i(J_n(x)|. \tag{2.9}$$

Since each point of S^1 is contained in at most three of the intervals $J_n(x)$, $f(J_n(x)), \ldots, f^{q_{n+1}-1}(J_n(x))$, $\sum_{i=0}^{q_{n+1}-1} |f^i(J_n(x)| \leq 3$ and therefore from (2.9),

$$|S^1 \setminus \Lambda| \leq 3 \cdot C \cdot \frac{|J_n(x) \setminus \Lambda|}{|J_n(x)|}.$$

From (2.7), this last ratio tends to zero and it follows that $S^1 \setminus \Lambda$ has zero Lebesgue measure. □

Exercise 2.3. Show that the proofs of Denjoy's Theorem 2.1, Denjoy's inequality in §2.2 and the result of this section work equally well for piecewise linear homeomorphisms $f: S^1 \to S^1$. (In §3 we shall see that an improved Denjoy inequality holds if f is C^3.)

3 Smooth Conjugacy Results

We have seen in Section 2 that a C^2 diffeomorphism f without periodic points is conjugate to an irrational rotation. Moreover, as we saw in the exercises at the end of Section 1, an irrational rotation is uniquely ergodic (i.e., it has a unique invariant probability measure). Therefore f also has a unique invariant probability measure μ which is defined by $\mu(A) = \lambda(h(A))$, where h is the conjugacy between f and the rotation g and λ is the Lebesgue measure, which is the only probability measure invariant by the rotation g. Consequently, if $\phi\colon S^1 \to \mathbb{R}$ is a continuous function, then the Birkhoff Ergodic Theorem implies that the sums $\frac{1}{n}\sum_{i=0}^{n-1} \phi\circ f^i$ converge uniformly to a constant $\int \phi\, d\mu$, see Section 5. From this it follows easily that if I is an interval then the frequency with which the orbit of a point x visits I (or the sejour time of x in I) coincides with the measure of I, i.e., $\mu(I) = \lim_{n\to\infty} \frac{1}{n}\#\{i\,;\ 0 \le i \le n \text{ such that } f^i(x) \in I\}$. In the case of a rigid rotation this sejour time is therefore equal to the Lebesgue measure of the interval. If we want to get more quantitative information on the sejour time of orbits of f we are led to analyze the regularity of the conjugacy h. For example, if h is a diffeomorphism, the sejour time of an orbit in the interval I is comparable to the Lebesgue measure of I since this is comparable to the Lebesgue measure of $h(I)$ which is the sejour time in $h(I)$ of the orbit of g through $h(x)$.

In Arnol'd (1961) there are examples of analytic diffeomorphisms of the circle, without periodic points, for which the conjugacy to a rotation is not even absolutely continuous (we will present this example in Section 5 of this chapter). The orbits of such a diffeomorphism, although dense in the circle, spend more time in some regions than in others. This is due to the arithmetic properties of the rotation numbers, which, in the case of these examples, are irrational numbers which are very well approximated by rational numbers (Liouville numbers). In the same article, Arnol'd proved that if the rotation number α of an analytic diffeomorphism f satisfies a Diophantine condition (i.e., $|\alpha - \frac{p}{q}| > \frac{K}{q^{2+\beta}}$ for every rational number $\frac{p}{q}$, where K and β are positive constants) and if, furthermore, f is close enough to a rotation, then the conjugacy h is analytic. The proof of this theorem uses a method of approximation to solve the functional equation $h\circ f = R_\alpha \circ h$ which is a modification of Newton's method. This method, which is based on the very rapid convergence of the Newton method, was suggested by Kolmogorov to face problems with *small divisors*. Later, Moser extended the above theorem to the smooth category, see Moser (1966).
This result has a local nature: it is important that f is near to a rotation. M. Herman (1979) proved the global result which had been conjectured by Arnol'd: there exists a subset $A \subset (0,1)$ with Lebesgue measure 1, such that if f is a C^∞ diffeomorphism with rotation number in A then it is C^∞ conjugate to a rotation. Herman's result also applies if one has only a finite amount of differentiability. However, even for very good rotation numbers the conjugacy is in general less differentiable then the diffeomorphism. This loss of differentiability is typical in this type of problem.

Let us say that f is $C^{k-\delta}$ where $k \ge 1$ is an integer and $\delta \in (0,1)$, if f is C^{k-1} and its $k-1$-th derivative satisfies a Hölder condition:

$$\frac{|D^{k-1}f(x) - D^{k-1}f(y)|}{|x-y|^{1-\delta}} < \text{constant}.$$

The best result on these smooth linearizations is

Theorem 3.1. **(Katznelson and Ornstein)** *Let $f\colon S^1 \to S^1$ be a C^k, k a positive real number, diffeomorphism whose rotation number α satisfies the Diophantine condition:*

$$|\alpha - \frac{p}{q}| > \frac{C}{q^{2+\beta}} \text{ for all } \frac{p}{q} \in \mathbb{Q}.$$

Then, if $\beta + 2 < k$, the homeomorphism h which conjugates f with the rotation R_α is of class $C^{k-1-\beta-\epsilon}$ for all $\epsilon > 0$.

This result is due to Katznelson and Ornstein (1989a) and sharpens earlier results due to Herman (1979) and Yoccoz (1984a).

Here we will present an extension of Herman's theorem due to Yoccoz (1984a). This theorem guarantees the differentiability of the conjugacy whenever the rotation number satisfies a Diophantine condition. As we will see in Section 5, without this condition the conjugacy is in general not even absolutely continuous.

Theorem 3.2. **(Herman and Yoccoz)** *Let f be a C^k circle diffeomorphism, $k \geq 3$. Suppose the rotation number α of f satisfies the Diophantine condition:*

$$|\alpha - \frac{p}{q}| > \frac{K}{q^{2+\beta}} \text{ for all } \frac{p}{q} \in \mathbb{Q}\,,$$

where K and β are positive constants. Then, if $k > 2\beta+1$, there exists a C^1 diffeomorphism which is a conjugacy h between f and a rotation R_α. Furthermore, h is $C^{k-1-\beta-\epsilon}$ for every $\epsilon > 0$.

Khanin and Sinai (1987), (1989) and Stark (1988) gave different proofs of part of Theorem 3.2 using renormalization techniques. One of the main points in all the proofs of these results is that certain derivatives cancel. A new analytical tool plays a fundamental role in the proof of this result: this is the Schwarzian derivative which we will define below.

The proof of these theorems is quite intricate. To simplify the exposition of Theorem 3.2 we will prove only that a C^3 diffeomorphism whose rotation number satisfies a Diophantine condition is C^1 conjugate to a rotation and we refer to Yoccoz (1984a) for the complete proof. Along the way, we shall obtain rigidity results which do not require conditions on the rotation number, see Steps 3 and 4 of the proof below.

In order to simplify the notation we will use the letter C for all constants that will appear in our estimates that do not depend on n or on a point in the circle. So in each (of the finite) steps of the proof, C may be bigger than the previous constant C and so we will sometimes write expressions like $C = 2C$.

The proof of Theorem 3.2 will be subdivided in several steps.

Step 1: It is enough to show that derivatives stay bounded

First we prove the following

Proposition 3.1. *Let f be a C^1 diffeomorphism of the circle. Then f is C^1 linearizable if and only if the sequence $||\log Df^i||$ is bounded. (Here $||\cdot||$ denotes the supremum norm.)*

The proof of this proposition will be given in the next two lemmas. First note that if h is C^1, R_α a rotation then it easy to see that $h \circ f = R_\alpha \circ h$ is equivalent to the following cocycle condition: $\log Dh \circ f - \log Dh = -\log Df$. A general lemma of Gottschalk and Hedlund gives necessary and sufficient conditions for solving this kind of cocycle condition.

Lemma 3.1. (Gottschalk-Hedlund) *Let X be a compact metric space and $f\colon X \to X$ be a minimal homeomorphism (i.e., every orbit of f is dense in X). If $g\colon X \to \mathbb{R}$ is a continuous function then the following statements are equivalent:*
1. there exists a continuous function $\phi\colon X \to R$ such that

$$\phi \circ f - \phi = g;$$

2. there exists $x_0 \in X$ such that

$$\sup_{n\in\mathbb{N}} |\sum_{i=0}^{n} g(f^i(x_0))| < \infty.$$

Proof. Let us first show that 1) implies 2). In fact, $g(f^i(x_0)) = \phi(f^{i+1}(x_0)) - \phi(f^i(x_0))$. Thus $\sum_{i=0}^n g(f^i(x_0)) = \phi(f^{n+1}(x_0)) - \phi(x_0)$ and therefore for each n, $|\sum_{i=0}^n g(f^i(x_0))| < 2\sup|\phi| < \infty$.

Now we prove that 2) implies 1). Let $F\colon X \times R \to X \times R$ be the homeomorphism defined by $F(x,t) = (f(x), t + g(x))$. By induction we get $F^n(x,t) = (f^n(x), t + \sum_{i=0}^{n-1} g(f^i(x)))$. Let $M \subset X \times \mathbb{R}$ be the closure of the orbit of the point $(x_0, 0)$. It is clear that $F(M) \subset M$. From the assumption the second component of $F^n(x_0, 0)$ is bounded and therefore M is compact. Let $N \subset M$ be a minimal set of F (i.e., N is a closed set with $F(N) = N$ such that no non-trivial closed subset of it is invariant under F). We have the following properties: a) $\pi_1(N) = X$, where π_1 is the projection onto the first factor, because f is minimal; b) N can be written as the graph of a function $X \to \mathbb{R}$. Indeed, suppose, by contradiction, that there exist $(x,y) \in N$ and $(x, y+\lambda) \in N$. Let $T_\lambda\colon X \times R \to X \times R$, $T_\lambda(x,t) = (x, t+\lambda)$. We have that $T_\lambda \circ F = F \circ T_\lambda$. Hence $T_\lambda(N) = T_\lambda(F(N)) = F(T_\lambda(N))$, i.e., $T_\lambda(N)$ is also invariant by F. As $(x, y+\lambda) \in N \cap T_\lambda(N)$ and N is minimal, we get that $N \subset T_\lambda(N)$. From the compactness of N it follows that the last inclusion is impossible for $\lambda \neq 0$, and this proves Property b). Let $\phi\colon X \to \mathbb{R}$

be the function whose graph is equal to N. We have that ϕ is continuous since N is compact. As $F(x, \phi(x)) = (f(x), \phi(x) + g(x)) \in N$, we have that $(f(x), \phi(x) + g(x)) = (f(x), \phi(f(x)))$, and this proves the lemma. □

Lemma 3.2. *If $f\colon S^1 \to S^1$ is a C^1 diffeomorphism without periodic points and is such that the sequence of functions $\{||\log Df^i||\}$ is bounded then f is C^1 - linearizable.*

Proof. As $\log Df^n = \sum_{i=0}^{n-1} \log Df \circ f^i$ and $\sup_n ||\log Df^n|| < \infty$, we have, from Lemma 3.1, that there exists a continuous function $\psi\colon S^1 \to \mathbb{R}$ such that

$$\log Df = \psi - \psi \circ f.$$

If $\hat{f}$ is a lift of f we get

$$(*) \qquad \log D\hat{f} = \hat{\psi} - \hat{\psi} \circ \hat{f},$$

where $\hat{\psi}$ is a lift of ψ. Because equation $(*)$ remains valid if we add a constant to $\hat{\psi}$, we may assume that $\int_0^1 e^{\hat{\psi}} dx = 1$. Let

$$\hat{h}(x) = \int_0^x e^{\hat{\psi}(t)} dt$$

We have that $\hat{h}$ is a diffeomorphism of the real line and $\hat{h}(x+1) = \hat{h}(x) + 1$. Hence $\hat{h}$ is a lift of a diffeomorphism $h\colon S^1 \to S^1$. From $(*)$ we get that

$$(D\hat{h} \circ \hat{f}) \times D\hat{f} = D\hat{h}.$$

Thus $\hat{h}(\hat{f}(x)) = \hat{h}(x) + \lambda$ for some constant λ. Let $R_\lambda\colon S^1 \to R^1$ be the rotation over λ and $\hat{R}_\lambda$ be the lift to $\mathbb{R}$ of this rotation. Then the last inequality implies $\hat{h} \circ \hat{f} = \hat{R}_\lambda \circ \hat{h}$. □

Now it is very easy to complete the proof of Proposition 3.1.

Proof of Proposition 3.1. If f is C^1 linearizable, $f = h \circ R_\alpha \circ h^{-1}$ then $\log Df^n(x) = \log Dh(R_\alpha^n(h^{-1}(x)) + \log Dh^{-1}(x)$ and therefore $||\log Df^n||$ is bounded. This proves the 'only if' part of Proposition 3.1. The 'if' part was proved in Lemma 3.2. □

By Proposition 3.1, in order to construct C^1 linearizations, it suffices to show that $||\log Df^i||$ is bounded. (The Denjoy inequality from §2.a merely states that $||\log Df^{q_n}||$ is bounded.) Later on, in Section 5 of this chapter, we will show that not every analytic diffeomorphism without periodic points satisfies this condition, but we will show in the remainder of this section that this condition is satisfied for all C^3 diffeomorphisms which have special rotation numbers.

Step 2: It is enough to show that the length of the intervals of closest return do not vary too much

Let $\frac{p_n}{q_n}$ be the convergents of $\alpha \in (0,1)$ as in section 1. These numbers p_n and q_n were characterized by the following two properties:

$$\mathrm{dist}(x, R_\alpha^{q_n}(x)) = \min\{\mathrm{dist}(x, R_\alpha^i(x))\,;\ 0 \le i \le q_n\},$$

$$0 < \alpha_n = \mathrm{dist}(x, R_\alpha^{q_n}(x)) = (-1)^n(q_n\alpha - p_n) < 1.$$

(Note that these numbers are independent of x because R_α is an isometry.) As before let

$$I_n(x) = [x, f^{q_n}(x)]$$

be the interval in $S^1 \setminus \{f(x)\}$ connecting x and $f^{q_n}(x)$. We refer to this interval as the *interval of closest return*. Moreover, we introduce the following notation: Let $m_n(x)$ denote the length of the interval $I_n(x)$,

$$m_n = \min_{x\in S^1} m_n(x) \text{ and } M_n = \max_{x \in S^1} m_n(x).$$

From the next proposition it follows that in order to show that $||\log Df^i||$ is bounded it suffices to estimate $\frac{M_n}{m_n}$ from above and below.

Proposition 3.2. *Let* $0 \le i < q_{n+1}$. *Then:*

$$\frac{1}{C}\frac{m_n}{M_n} < |Df^i(x)| < C\frac{M_n}{m_n}.$$

Proof of Proposition 3.2. Since $f^i(I_n(x)) = I_n(f^i(x))$, from the Mean Value Theorem, there exists $z \in I_n(x)$ such that $|Df^i(z)| = \frac{m_n(f^i(x))}{m_n(x)}$. Thus $\frac{m_n}{M_n} < |Df^i(z)| < \frac{M_n}{m_n}$. On the other hand, since the first $i-1$ iterates of $I_n(x)$ are pairwise disjoint, we get $\frac{1}{C}|Df^i(z)| < |Df^i(x)| < C|Df^i(z)|$. Combining these inequalities completes the proof. □

So it suffices to show that there is a constant such that

$$\frac{1}{C} \le \frac{m_n(x)}{m_n(y)} \le C$$

for all $x, y \in S^1$ and all n. Now we want to show that the function $m_n(x)$ does not vary too much.

Step 3: An estimate on the variation of the length of the intervals of closest return and an improved version of the Denjoy Inequality

In this step the main purpose is to estimate $|m_n(x) - m_n(y)|$ in terms of M_n. In all the estimates we have obtained so far we have used only that the diffeomorphism f is of class C^2. Now we will need that f is C^3. The purpose of this step is to prove the following

Proposition 3.3. *Let f be C^3. Then*

$$|m_{n+1}(x) - m_{n+1}(y)| \leq C\{M_n^{\frac{1}{2}} m_{n+1}(x) + M_n m_n(x)\}$$

for each $y \in I_n(x)$.

Let us explain why this proposition is useful. Rewriting the inequality in this proposition one gets

$$\frac{M_{n+1}}{m_{n+1}} \leq 1 + C\{M_n^{\frac{1}{2}} + M_n \frac{M_n}{m_{n+1}}\}.$$

As we will see in the next step, $m_n \leq M_n$ tends exponentially fast to zero independently of the rotation number of f. In Step 5 we will see that a Diophantine condition on the rotation number suffices to get a bound on the growth rate of $\frac{M_n}{m_{n+1}}$. Together with the previous inequality this will give that M_n/m_n can be estimated from above for these rotation numbers.

In the proof of Proposition 3.3 we need to estimate the derivative $D^2 \log Df$ of the non-linearity $Nf = D \log Df$ of f. Related to this derivative is the differential operator defined below, which needs derivatives up to the order 3, and which plays a fundamental role in many metric results in one-dimensional dynamics. In Chapter III and IV we will go deeper into some of the properties of the Schwarzian derivative. In particular, in Section 1 of Chapter IV we will show that this derivative is related to the hyperbolic Poincaré metric and show its connections with cross-ratios.

Definition. Let g be a C^3 function such that $Dg > 0$. The *Schwarzian derivative* of g is the differential operator defined by:

$$Sg = D^2 \log Dg - \frac{1}{2}(D \log Dg)^2 = \frac{D^3 g}{Dg} - \frac{3}{2}(\frac{D^2 g}{Dg})^2.$$

Remark. 1. From the above definition, the following formula for the Schwarzian derivative of the composition of two functions: $S(g \circ f) = (Sg \circ f) \times (Df)^2 + Sf$ follows immediately. From this we get: $Sf^n(x) = \sum_{i=0}^{n-1} Sf(f^i(x)) \times (Df^i(x))^2$. 2. As we have seen before, the operator $Ng = D \log Dg$ plays an important role in the study of the distortion of functions. Indeed,

$$Ng = D \log Dg = \frac{D^2 g}{Dg}$$

and so if g is a diffeomorphism on (x, y) then

$$\log \frac{Dg(y)}{Dg(x)} = \int_x^y \frac{D^2 g(t)}{Dg(t)} \, dt = \int_x^y Ng(t) \, dt.$$

This operator Ng is related with the Schwarzian derivative in the following way:

$$Sg = D(Ng) - \frac{1}{2}[Ng]^2.$$

We will study the Schwarzian derivative in much greater detail in Chapter IV.

In the next lemma it will be shown that the non-linearity of f^k, i.e., $Nf^k = D\log Df^k$ can be estimated by M_n and m_n. In this lemma essential use is made of the disjointness properties from Lemma 1.3. More specifically we should note that it is only because one has such nice 'tiling' properties that one gets the estimate (3.2). (Without this one could only get $|D\log Df^k(x)| \le C\frac{M_n^{\frac{1}{2}}}{m_n}$ which gives in general a much weaker estimate.) So let us state this lemma.

Lemma 3.3. *If* $x \in S^1$ *and* $0 \le k \le q_{n+1}$ *then:*

$$|D\log Df^k(x)| \le C\frac{M_n^{\frac{1}{2}}}{m_n(x)}; \tag{3.1}$$

$$|D^2\log Df^k(x)| \le C\frac{M_n}{(m_n(x))^2}. \tag{3.2}$$

Remark. This lemma is the only place where the assumption that f is C^3 is explicitly used. One can prove (3.1) and another version of (3.2) also if f is C^{2+z} (this is the class of C^2 diffeomorphisms for which D^2f satisfies a Zygmund condition; this condition will be defined in Section IV.2.a). Therefore, to get a C^1 conjugacy in Theorem 3.1 it is enough that the diffeomorphism belongs to this class.

Proof. Let us prove that

$$|Sf^k(x)| \le C\frac{M_n}{(m_n(x))^2}. \tag{3.3}$$

Since $Sf^k = \sum_{i=0}^{k-1}\left[Sf(f^i(x))\right]\,(Df^i(x))^2$ and $|Sf| \le C$, we have that

$$|Sf^k(x)| \le \;C\,(\max_{0\le i\le k-1}|Df^i(x)|)\;\sum_{i=0}^{k-1}|Df^i(x)|.$$

By Lemma 1.3, the intervals $I_n(x), \dots, f^{q_{n+1}-1}(x)$ are pairwise disjoint and therefore, from Corollary 1 of Lemma 2.1, we get that for $0 \le k \le q_{n+1}$,

$$|Df^i(x)| \le C\frac{|f^i(I_n(x))|}{|I_n(x)|}.$$

Again because these intervals are disjoint we get for $0 \le k \le q_{n+1}$,

$$\sum_{i=0}^{k-1}|Df^i(x)| \le \frac{C}{|I_n(x)|}\sum_{i=0}^{k-1}|f^i(I_n(x))| < \frac{C}{|I_n(x)|} = \frac{C}{m_n(x)}.$$

Hence

$$|Sf^k(x)| \le C\frac{M_n}{(m_n(x))^2}.$$

This proves (3.3).

Now let us prove (3.1). Let $x_0 \in S^1$ be such that $x \mapsto |D \log Df^k(x)|$ is maximal in x_0. Then, $D^2 \log Df^k(x_0) = 0$ and therefore

$$Sf^k(x_0) = -\frac{1}{2}(D \log Df^k(x_0))^2.$$

Thus

$$|D \log Df^k(x_0)|^2 \leq 2 \cdot |Sf^k(x_0)| \leq C \frac{M_n}{(m_n(x_0))^2}.$$

Hence

$$||D \log Df^k|| = |D \log Df^k(x_0)| \leq C \frac{M_n^{\frac{1}{2}}}{m_n}. \tag{3.4}$$

In order to prove (3.1) let $z \in S^1$ be such that $m_n(z) = m_n$. Again by Lemma 1.3, there exist $t \in I_n(z) \cup I_n(f^{-q_n}(z))$ and $0 \leq i < q_{n+1}$ such that $x = f^i(t)$. Therefore

$$D \log Df^{k+i}(t) = \left(D \log Df^k(x)\right) Df^i(t) + D \log Df^i(t)$$

and so

$$|D \log Df^k(x)| \leq \frac{|D \log Df^{k+i}(t)| + |D \log Df^k(t)|}{|Df^i(t)|}.$$

Using (3.4), the chain rule

$$D \log Df^{q_n+i}(x) = \left(D \log Df^i(f^{q_n}(x))\right) Df^{q_n}(x) + D \log Df^{q_n}(x),$$

and the Denjoy inequality we get that formula (3.4) even holds for $k = 0, 1, 2, \ldots, 2q_{n+1}$ (with a different constant C). Therefore the last inequality and (3.4) give

$$|D \log Df^k(x)| \leq \frac{1}{|Df^i(t)|} \times C \frac{M_n^{\frac{1}{2}}}{m_n}. \tag{3.5}$$

On the other hand, there exists $y \in I_n(t)$ such that $|Df^i(y)| = \frac{|I_n(x)|}{|I_n(t)|} = \frac{m_n(x)}{m_n(t)}$. Since the intervals $I_n(t), \ldots, f^{i-1}(I_n(t))$ are pairwise disjoint, we get as before

$$|Df^i(t)|^{-1} \leq C|Df^i(y)|^{-1} = C \frac{m_n(t)}{m_n(x)}. \tag{3.6}$$

Because $t \in I_n(z) \cup I_n(f^{-q_n}(z)) = I_n(z) \cup f^{-q_n}(I_n(z))$, one has $|t - z| \leq m_n(z) + m_n(z)||Df^{-q_n}|| \leq Cm_n(z)$ (here we use the Denjoy inequality). Using this, $Dm_n(x) = \pm(Df^{q_n}(x) - 1)$ and the Mean Value Theorem implies that $m_n(t) = m_n(z) + Dm_n(\hat{z})(t - z) \leq m_n(z) + ||Df^{q_n} - 1|| \cdot |t - z| \leq Cm_n(z)$ for all $t \in I_n(z) \cup I_n(f^{-q_n}(z))$. Using this in (3.6) gives $|Df^i(t)|^{-1} \leq C \frac{m_n}{m_n(x)}$. From this and (3.5) we get (3.1). Finally, (3.1) and (3.3) imply

$$|D^2 \log Df^n(x)| \leq |Sf^n(x)| + \frac{1}{2}|D \log Df^n(x)|^2 \leq C \frac{M_n}{(m_n(x))^2}$$

and this completes the proof of Lemma 3.3. □

Using the previous lemma we can improve Denjoy's inequality substantially. The contents of Denjoy's inequality was that $||\log Df^{q_n}||$ was bounded; in the next lemma we see that these numbers even tend to zero.

Lemma 3.4. (An improved version of Denjoy's Inequality)

$$||Df^{q_n} - 1|| \le CM_n^{\frac{1}{2}}.$$

In other words, the derivative of the n-th renormalization $\mathcal{R}^n(f)$ of f goes quite rapidly to one as n tends to infinity.

Proof. Let $x \in S^1$. Choose $z \in S^1$ so that $m_n(z) = m_n$. We have that $Df^{q_n}(z) - 1 = 0$ (since $x \to m_n(x)$ takes its minimum at $x = z$ and $Dm_n(x) = \pm(Df^{q_n}(x) - 1)$). As the intervals $f^j\{I_n(z) \cup I_n(f^{-q_n}(z))\}, 0 \le j < q_{n+1}$ cover the circle, there exists $t \in I_n(z) \cup I_n(f^{-q_n}(z))$ such that $f^i(t) = x$ for some $0 \le i < q_{n+1}$. Hence

$$\log Df^{q_n}(x) = \log Df^{q_n+i}(t) - \log Df^i(t) =$$

$$= \log Df^{q_n}(t) + [\log Df^i(f^{q_n}(t)) - \log Df^i(t)].$$

On the other hand,

$$|\log Df^i(f^{q_n}(t)) - \log Df^i(t)| \le ||D \log Df^i|| \cdot |I_n(t)|,$$

and

$$|\log Df^{q_n}(t)| = |\log Df^{q_n}(t) - \log Df^{q_n}(z)| \le$$

$$\le ||D \log Df^{q_n}|| \cdot |I_n(z) \cup I_n(f^{-q_n}(z))|,$$

since $\log Df^{q_n}(z) = 0$. As the intervals $I_n(t)$ and $I_n(z) \cup I_n(f^{-q_n}(z))$ are contained in the interval $[f^{-q_n}(z), f^{2q_n}(z)]$ whose length is smaller or equal to Cm_n, we get from (3.1),

$$|\log Df^{q_n}(x)| \le Cm_n \left\{||D \log Df^i|| + ||D \log Df^{q_n}||\right\} \le CM_n^{\frac{1}{2}}.$$

Clearly the lemma follows. □

In the next two lemmas we will estimate the variation of $|m_{n+1}(y) - m_{n+1}(x)|$ for $x \in S^1$ and $y \in I_n(x)$. If m_{n+1} is monotone on the interval $I_n(z)$ for some $z \in S^1$ then it will be possible to estimate $|m_{n+1}(y) - m_{n+1}(x)|$ in terms of $|m_{n+1}(f^{q_n}(x)) - m_{n+1}(x)|$ and to use inequality (3.1). On the other hand, if there exists no such z then one gets local minima of $\log Df^{q_{n+1}}$ on $I_n(x)$ and one can combine this information with inequality (3.2). In this way we get estimates in both cases.

Lemma 3.5. *Suppose that m_{n+1} is monotone on the interval $I_n(z)$ for some $z \in S^1$. Then, for every $x \in S^1$ and $y \in I_n(x)$ we have*

$$\left|\frac{m_{n+1}(y)}{m_{n+1}(x)} - 1\right| \le CM_n^{\frac{1}{2}}.$$

Proof. By Lemma 3.4, we have $||Df^{q_n} - 1|| \leq CM_n^{\frac{1}{2}}$. Since $I_{n+1}(f^{q_n}(x)) = f^{q_n}(I_{n+1}(x))$ we have, from the Mean Value Theorem

$$m_{n+1}(f^{q_n}(x)) = Df^{q_n}(\xi)m_{n+1}(x)$$

for some $\xi \in I_{n+1}(x)$. Thus $\frac{m_{n+1}(f^{q_n}(x))}{m_{n+1}(x)} - 1 = Df^{q_n}(\xi) - 1$. Therefore

$$|\frac{m_{n+1}(f^{q_n}(x))}{m_{n+1}(x)} - 1| < CM_n^{\frac{1}{2}}$$

for each $x \in S^1$. As m_{n+1} is monotone on $I_n(z)$ this inequality implies $|\frac{m_{n+1}(t)}{m_{n+1}(z)} - 1| < CM_n^{\frac{1}{2}}$ for every $t \in [z, f^{q_n}(z)]$ and, since $f^{q_n}(I_{n+1}(t)) = I_{n+1}(f^{q_n}(t))$, by using the previous inequality again, that

$$|\frac{m_{n+1}(t)}{m_{n+1}(z)} - 1| < CM_n^{\frac{1}{2}}$$

for every $t \in [f^{-2q_n}(z), f^{q_n}(z)]$. In the same way, $|\frac{m_{n+1}(z)}{m_{n+1}(t)} - 1| < CM_n^{\frac{1}{2}}$ for every $t \in [f^{-2q_n}(z), f^{q_n}(z)]$. Thus

$$|\frac{m_{n+1}(t')}{m_{n+1}(t)} - 1| < CM_n^{\frac{1}{2}}$$

for every $t, t' \in [f^{-2q_n}(z), f^{q_n}(z)]$. Take now $x \in S^1$ and $y \in I_n(x)$. By the second statement of the second lemma in §2.a, there exist $t, t' \in [f^{-2q_n}(z), f^{q_n}(z)]$ and $0 \leq j < q_{n+1}$ such that $f^j(t) = x$ and $f^j(t') = y$. Thus $m_{n+1}(x) = m_{n+1}(t)Df^j(\xi)$ and $m_{n+1}(y) = m_{n+1}(t')Df^j(\xi')$ with $\xi, \xi' \in [f^{-2q_n}(z), f^{q_n}(z)]$. Hence $|\xi - \xi'| < Cm_n(z)$. From this and (3.1),

$$|\log Df^j(\xi) - \log Df^j(\xi')| = |D\log Df^j(\bar{\xi})| \cdot |\xi - \xi'| \leq C\frac{M_n^{\frac{1}{2}}}{m_n(\bar{\xi})}m_n(z).$$

As $I_n(z) \subset [f^{-3q_n}(\bar{\xi}), f^{3q_n}(\bar{\xi})]$, we have that $m_n(z) < Cm_n(\bar{\xi})$. Therefore

$$|\log Df^j(\xi) - \log Df^j(\xi')| \leq CM_n^{\frac{1}{2}}.$$

Hence, combining all this,

$$|\frac{m_{n+1}(y)}{m_{n+1}(x)} - 1| = |\frac{m_{n+1}(t')Df^j(\xi')}{m_{n+1}(t)Df^j(\xi)} - 1| \leq CM_n^{\frac{1}{2}}. \quad \square$$

Lemma 3.6. *If $t \to m_{n+1}(t)$ is not monotone on every interval of the type $I_n(z)$, $z \in S^1$, then, for every $x \in S^1$ and $y \in I_n(x)$,*

$$|m_{n+1}(y) - m_{n+1}(x)| \leq CM_n m_n(x).$$

Proof. By the hypothesis, the intervals $I_n(f^{-q_n}(x)), I_n(x)$ both contain a zero of $\log Df^{q_{n+1}}$. Thus by the Mean Value Theorem, $D\log Df^{q_{n+1}}$ vanishes in some point $\xi \in [f^{-q_n}(x), f^{q_n}(x)]$. By inequality (3.2) of Lemma 3.3,

$$|D\log Df^{q_{n+1}}(t)| = |D\log Df^{q_{n+1}}(t) - D\log Df^{q_{n+1}}(\xi)| =$$

$$= |D^2\log Df^{q_{n+1}}(\bar{\xi})|\,|t-\xi| \leq C\frac{M_n}{(m_n(\bar{\xi}))^2}\,|t-\xi|$$

for every $t \in I_n(x)$. Now

$$|t-\xi| \leq m_n(f^{-q_n}(x)) + m_n(x) \leq m_n(x)||Df^{-q_n}|| + m_n(x) \leq Cm_n(x)$$

and

$$m_n(t) = m_n(x) - Dm_n(\hat{t})(t-x) \geq m_n(x) - ||Df^{q_n} - 1|| \cdot |t-x| \geq Cm_n(x)$$

for every $t \in [f^{-q_n}(x), f^{q_n}(x)]$. Therefore

$$|D\log Df^{q_{n+1}}(t)| \leq C\frac{M_n}{m_n(x)}$$

for every $t \in I_n(x)$. Since $\log Df^{q_{n+1}}(\bar{t}) = 0$ for some $\bar{t} \in I_n(x)$, we get, using again the Mean Value Theorem,

$$|\log Df^{q_{n+1}}(t)| < CM_n,\ \forall t \in I_n(x).$$

Finally,

$$|m_{n+1}(y) - m_{n+1}(x)| = |Dm_{n+1}(\bar{x})|\,|x-y| = |Df^{q_{n+1}}(\bar{x}) - 1|\,|x-y|$$

$$< C|\log Df^{q_{n+1}}(\bar{x})|\,m_n(x) < CM_nm_n(x). \quad \square$$

Proof of Proposition 3.3. The proof of Proposition 3.3 follows immediately by combining Lemmas 3.5 and 3.6. $\square$

Step 4: Comparing the length of the closest return intervals for the diffeomorphism with the corresponding numbers for the rotation

By Step 2 it suffices to prove that the sequence $\frac{M_n}{m_n}$ is bounded. In order to achieve this we will need to make an assumption on the rotation number of f. More precisely, let α_n be the length of the interval $[x, R_\alpha^{q_n}(x)] \subset S^1 \setminus \{R_\alpha(x)\}$. Then $\alpha_1 > \alpha_2 > \cdots$ and, since R_α is a rotation, α_n does not depend on x and, by the definition of the convergents, $\alpha_n = (-1)^n(q_n\alpha - p_n)$. We have also that

$$\alpha_n = a_{n+2}\alpha_{n+1} + \alpha_{n+2}$$

where $\alpha = [0; a_1, a_2, \ldots]$ is the continued fraction expansion of α. As $a_n \geq 1$ we have that $\alpha_{n+2} \leq \frac{1}{2}\alpha_n$. Thus the sequence α_n decreases exponentially. To analyze the behaviour of the sequences M_n and m_n, we are going to compare them with the sequence α_n.

Let $g\colon S^1 \to S^1$ be a diffeomorphism without periodic points. Take a lift $\hat{g}\colon R \to R$ of g such that $\hat{g} = Id + \hat{\phi}$, with $\hat{\phi}$ one-periodic and $0 < \hat{\phi}(x) < 1$ for every $x \in \mathbb{R}$. As we have seen in Section 1,

$$\rho(g) = \lim_{k\to\infty} \frac{1}{k}(\hat{g}^k(x) - x) = \lim_{k\to\infty} \frac{1}{k} \sum_{i=0}^{k-1} \hat{\phi}(\hat{g}^i(x))$$

Let $\pi\colon \mathbb{R} \to S^1$ be the canonical covering map and define $\phi\colon S^1 \to \mathbb{R}$ such that $\phi(\pi(x)) = \hat{\phi}(x)$ for each $x \in \mathbb{R}$. Clearly $\phi(x)$ is the length of the arc in the circle which connects x to $g(x)$ and which is positively oriented. Hence

$$\rho(g) = \lim_{k\to\infty} \frac{1}{k} \sum_{i=0}^{k-1} \phi(g^i(x)) = \int \phi \, d\mu$$

where μ is the probability measure on the circle invariant by g. Applying this to $g = f^{q_n}$ and, taking into account that the length of the oriented arc of the circle between x and f^{q_n} is equal to either $m_n(x)$ or $1 - m_n(x)$ according to the parity of n, we have

$$\begin{aligned}\int_{S^1} m_n(x)\, d\mu(x) &= \min\{\rho(f^{q_n}), 1 - \rho(f^{q_n})\} \\ &= \min\{\rho(R_\alpha^{q_n}), 1 - \rho(R_\alpha^{q_n})\} = \alpha_n.\end{aligned}$$

Thus

$$M_n \geq \alpha_n \geq m_n.$$

As we have in Step 2, to complete the proof of the Theorem 3.2, it is sufficient to prove that the sequence $\frac{M_n}{m_n}$ is bounded and therefore it is enough to prove that $\frac{M_n}{\alpha_n}$ and $\frac{\alpha_n}{m_n}$ are bounded. In this section we will show that $\frac{M_{n+1}}{\alpha_{n+1}}$ can be estimated from above in terms of $\frac{M_n}{\alpha_n}$. In the next step, this inequality and the Diophantine condition will then imply that $\frac{M_n}{\alpha_n}$ and $\frac{\alpha_n}{m_n}$ are both bounded.

Proposition 3.4. *For n sufficiently large,*

$$M_{n+1} \leq M_n \frac{\frac{\alpha_{n+1}}{\alpha_n} + CM_n}{1 - CM_n^{\frac{1}{2}}},$$

$$m_{n+1} \geq m_n \frac{\frac{\alpha_{n+1}}{\alpha_n} - CM_n}{1 + CM_n^{\frac{1}{2}}}.$$

(Notice that by Denjoy's theorem $M_n \to \infty$ as $n \to \infty$ so the above inequalities make sense for large n.)

Let us first motivate this inequality. Since $\alpha_{n+2} \leq \frac{1}{2}\alpha_n$, we have from the first of these inequalities that $M_{n+2} \leq \frac{2}{3}M_n$ for n sufficiently large. Now rewrite the first inequality as

$$\frac{M_{n+1}}{\alpha_{n+1}} \leq \frac{M_n}{\alpha_n} \frac{1 + \frac{\alpha_n}{\alpha_{n+1}} CM_n}{1 - CM_n^{\frac{1}{2}}}.$$

Therefore if $\frac{\alpha_{n+1}}{\alpha_n}$ does not grow too fast as n goes to infinity then we expect the sequence $\frac{M_{n+1}}{\alpha_{n+1}}$ to be bounded.

In order to prove Proposition 3.4 we will need two lemmas. The first of these is crucial in comparing the locally defined numbers $m_n(x)$ with the corresponding numbers α_n for the rotation. It shows that the ratio of the lengths of some dynamically defined intervals are determined by the rotation number. In other words, this lemma gives an example of a rigidity result.

Lemma 3.7. *For every $x \in S^1$, there exist $y \in I_n(x)$ and $z \in I_{n+1}(x)$ such that*

$$\frac{m_{n+1}(y)}{m_n(z)} = \frac{\alpha_{n+1}}{\alpha_n}.$$

Proof. Proof Of course the Lebesgue measure λ is invariant under R_α. Let h is the conjugacy between f and R_α and $\mu = h_\star\lambda$ be the measure defined by $h_\star\lambda(A) = \lambda(h^{-1}(A))$. Clearly μ is invariant under f. Let $\omega\colon S^1 \to R$ be a map with a single discontinuity at $f(x)$ such that $\pi \circ \omega$ is the identity map of the circle. We have

$$m_n(t) = |\omega f^{q_n}(t) - \omega(t)|.$$

By the invariance of the measure we have

$$\int_{I_n(x)} \omega \circ f^{q_{n+1}}\, d\mu = \int_{I_n(f^{q_{n+1}}(x))} \omega\, d\mu$$

and

$$\int_{I_{n+1}(x)} \omega \circ f^{q_n}\, d\mu = \int_{I_{n+1}(f^{q_n}(x))} \omega\, d\mu.$$

Because $I_n(f^{q_{n+1}}(x)) \setminus I_n(x) = I_{n+1}(x) \setminus I_{n+1}(f^{q_n}(x))$ one gets therefore

$$\int_{I_n(x)} (\omega \circ f^{q_{n+1}} - \omega)\, d\mu = \int_{I_n(f^{q_{n+1}}(x))\setminus I_n(x)} \omega\, d\mu =$$

$$= \int_{I_{n+1}(x)\setminus I_{n+1}(f^{q_n}(x))} \omega\, d\mu = -\int_{I_{n+1}(x)} (\omega \circ f^{q_n} - \omega)\, d\mu.$$

Applying the Mean Value Theorem to the last equality, one obtains
$y \in I_n(x)$ and $z \in I_{n+1}(x)$ such that

$$m_{n+1}(y) \int_{I_n(x)} d\mu = m_n(z) \int_{I_{n+1}(x)} d\mu.$$

Since

$$\int_{I_n(x)} d\mu = \int_{[h(x), R_\alpha(h(x))]} d\lambda = \alpha_n,$$

we have that

$$m_{n+1}(y)\alpha_n = m_n(z)\alpha_{n+1}$$

and this proves the lemma. □

Lemma 3.8. *For every $x \in S^1$ we have*

$$|m_{n+1}(x) - \frac{\alpha_{n+1}}{\alpha_n} m_n(x)| \leq C[M_n m_n(x) + M_n^{\frac{1}{2}} m_{n+1}(x)].$$

Proof. Let $x \in S^1$. Take $y \in I_n(x)$ and $z \in I_{n+1}(x)$ as in Lemma 3.7 such that $m_{n+1}(y) = \frac{\alpha_{n+1}}{\alpha_n} m_n(z)$. By Proposition 3.3, we have

$$|m_{n+1}(x) - m_{n+1}(y)| \leq C\{M_n^{\frac{1}{2}} m_{n+1}(x) + M_n m_n(x)\}.$$

On the other hand, as $|Df^{q_n}(t) - 1| \leq CM_n^{\frac{1}{2}}$, we have that

$$|m_n(z) - m_n(x)| \leq CM_n^{\frac{1}{2}} |[x, z]| \leq CM_n^{\frac{1}{2}} m_{n+1}(x).$$

Thus

$$|m_{n+1}(x) - \frac{\alpha_{n+1}}{\alpha_n} m_n(x)| \leq$$
$$\leq |m_{n+1}(x) - m_{n+1}(y)| + |\frac{\alpha_{n+1}}{\alpha_n} m_n(z) - \frac{\alpha_{n+1}}{\alpha_n} m_n(x)| \leq$$
$$\leq C\{M_n^{\frac{1}{2}} m_{n+1}(x) + M_n m_n(x)\} + CM_n^{\frac{1}{2}} m_{n+1}(x) =$$
$$= C\{M_n^{\frac{1}{2}} m_{n+1}(x) + M_n m_n(x)\}. \quad \square$$

Proof of Proposition 3.4. Follows immediately from Lemmas 3.7 and 3.8. □

Step 5: For appropriate rotation numbers the variation of the length of the intervals of closest return is bounded; the proof of Theorem 3.2

Notice that the Diophantine condition on the rotation number was not used until now. In this section we will use this condition and Proposition 3.4 to prove

Lemma 3.9. *If the rotation number of f satisfies a Diophantine condition of order β, $0 < \beta < 1$, then the sequences $\frac{M_n}{\alpha_n}$ and $\frac{\alpha_n}{m_n}$ are bounded.*

Proof. By Denjoy's theorem, $M_n \to 0$ as $n \to \infty$. Since, by Proposition 3.4,

$$M_{n+1} \leq M_n \frac{\frac{\alpha_{n+1}}{\alpha_n} + CM_n}{1 - CM_n^{\frac{1}{2}}}$$

we have that, given $\epsilon > 0$,

$$M_{n+1} \leq M_n((1 + \epsilon)\frac{\alpha_{n+1}}{\alpha_n} + \epsilon)$$

for every n big enough. Hence

$$M_{n+2} \leq M_n((1 + \epsilon)^2 \frac{\alpha_{n+2}}{\alpha_n} + O(\epsilon)).$$

Since $\frac{\alpha_{n+2}}{\alpha_n} < \frac{1}{2}$, we have that

$$M_{n+2} \leq \frac{2}{3} M_n$$

for n big enough. Therefore the sequence $\{M_n\}$ decreases even exponentially.

From the Diophantine condition we get $|\alpha - \frac{p}{q}| > \frac{K}{q^{2+\beta}}$ for some $\beta \in (0,1)$. Furthermore, by (1.5) in §1, $|\alpha - \frac{p_n}{q_n}| < \frac{1}{q_n q_{n+1}}$. Hence

$$\frac{K}{q_n^{1+\beta}} < \alpha_n = (-1)^n (q_n \alpha - p_n) < \frac{1}{q_{n+1}}$$

and therefore

$$\alpha_n > \alpha_{n+1} > K \alpha_n^{1+\beta}.$$

Since $0 < \beta < 1$, there exists $\theta \in (0,1)$ such that $(1+\beta+\theta)(1+\theta) < 2-\theta$ and, since $\alpha_n \leq M_n \to 0$, there exists n_0 such that if $n \geq n_0$ then

$$M_{n+2} \leq \frac{2}{3} M_n,$$
$$C M_n^{\frac{1}{2}} \leq \frac{1}{2},$$
$$\alpha_{n+1} \geq \alpha_n^{1+\beta+\theta}.$$

Since the sequence M_n^θ converges at least geometrically to zero, there exists $A > 1$ such that

$$\prod_{n=n_0}^{\infty} \frac{1 + M_n^\theta}{1 - C M_n^{\frac{1}{2}}} < A. \tag{3.7}$$

Claim: there exists n_1 such that if $n \geq n_1$ then

$$M_{n+1} \leq M_n \frac{\alpha_{n+1}}{\alpha_n} \frac{1 + M_n^\theta}{1 - C M_n^{\frac{1}{2}}}. \tag{3.8}$$

Before proving this claim let us finish the proof of the lemma. From equations (3.7) and (3.8), we have

$$\frac{M_n}{\alpha_n} \leq A \frac{M_{n_1}}{\alpha_{n_1}}$$

for every $n \geq \max\{n_0, n_1\}$. This shows that

$$\frac{M_n}{\alpha_n} < C \text{ for all } n \geq 0. \tag{3.9}$$

From the second inequality of Proposition 3.4, we get

$$m_{n+1} \geq m_n \frac{\alpha_{n+1}}{\alpha_n} \frac{1 - C \frac{M_n}{\alpha_{n+1}} \alpha_n}{1 + C M_n^{\frac{1}{2}}}$$

and therefore, using (3.9),

$$\frac{\alpha_{n+1}}{m_{n+1}} \leq \frac{\alpha_n}{m_n} \frac{1 + C \alpha_n^{\frac{1}{2}}}{1 - C \frac{\alpha_n^2}{\alpha_{n+1}}}.$$

Since, by the Diophantine condition, $\alpha_n^2(\alpha_{n+1})^{-1} \le C\alpha_n^{1-\beta}$, and α_n goes to zero exponentially fast, we have that the product

$$\prod_{n=0}^{\infty} \frac{1 + C\alpha_n^{\frac{1}{2}}}{1 - C\frac{\alpha_n^2}{\alpha_{n+1}}}$$

converges and, consequently, the sequence $\frac{\alpha_n}{m_n}$ is also bounded.

It remains to prove the inequality (3.8). By Proposition 3.4,

$$M_{n+1} \le M_n \frac{\alpha_{n+1}}{\alpha_n} \frac{1 + \frac{\alpha_n}{\alpha_{n+1}} C M_n}{1 - C M_n^{\frac{1}{2}}}. \tag{3.10}$$

Therefore it is enough to prove that

$$\frac{\alpha_n}{\alpha_{n+1}} C M_n \le M_n^{\theta} \tag{3.11}$$

for n big enough. We may assume that A satisfies $A^2 > A^{1-\theta} > 4C$. Take n_1 such that $M_{n_1} < \frac{1}{A^2}$. Let

$$r_{n_1} = \frac{1}{A^2}$$

and

$$\kappa_{n_1} = \frac{\log(A^2 M_{n_1})}{\log(\alpha_{n_1})} = \frac{\log(M_{n_1}) - \log(r_{n_1})}{\log(\alpha_{n_1})}.$$

Since $r_{n_1} > M_{n_1} \ge \alpha_{n_1}$, we have that $0 < \kappa_{n_1} < 1$. Since $A > 1$, also $r_{n_1} < 1$.

For $n \ge n_1$ define

$$r_{n+1} = r_n \frac{1 + M_n^{\theta}}{1 - C M_n^{\frac{1}{2}}} \text{ and } \kappa_{n+1} = \kappa_n \tag{3.12a}$$

if $C M_n \frac{\alpha_n}{\alpha_{n+1}} \le M_n^{\theta}$ and otherwise

$$r_{n+1} = r_n, \text{ and } \kappa_{n+1} = (1+\theta)\kappa_n. \tag{3.12b}$$

Let us prove by induction that

$$M_k \le r_k \alpha_k^{\kappa_k} \tag{3.13}$$

for all $k \ge n_1$. For $k = n_1$ equation (3.13) holds (with equality) by definition of κ_{n_1}. Assume that we have proved by induction that (3.13) holds for $n_1 \le k \le n$. Notice that $r_n < 1$ for all $n \ge n_1$ since

$$r_n \le \prod_{i=n_1}^{\infty} \frac{1 + M_i^{\theta}}{1 - C M_i^{\frac{1}{2}}} r_{n_1} \le \frac{1}{A} < 1.$$

Therefore (3.13) implies that $\kappa_k < 1$ because

$$\alpha_k^{\kappa_k} > r_k \alpha_k^{\kappa_k} \ge M_k \ge \alpha_k.$$

We have two cases to consider. If $CM_n \frac{\alpha_n}{\alpha_{n+1}} \leq M_n^\theta$ then from (3.10),

$$M_{n+1} \leq M_n \frac{\alpha_{n+1}}{\alpha_n} \frac{1 + M_n^\theta}{1 - CM_n^{\frac{1}{2}}}.$$

Thus, using the induction assumption and that $\kappa_{n+1} = \kappa_n \in (0, 1)$ in this case,

$$M_{n+1} \leq \frac{\alpha_{n+1}}{\alpha_n} \frac{1 + M_n^\theta}{1 - CM_n^{\frac{1}{2}}} r_n \alpha_n^{\kappa_n} \leq r_{n+1} \alpha_{n+1} \alpha_n^{\kappa_n - 1} \leq r_{n+1} \alpha_{n+1}^{\kappa_{n+1}}$$

which shows that (3.13) remains valid for $k = n + 1$. If on the other hand $CM_n \frac{\alpha_n}{\alpha_{n+1}} > M_n^\theta$, then (3.10) and $CM_n^{\frac{1}{2}} \leq \frac{1}{2}$ imply $M_{n+1} \leq 4CM_n^{2-\theta}$. Using the induction hypothesis, this gives

$$M_{n+1} \leq 4C[r_n \alpha_n^{\kappa_n}]^{2-\theta}.$$

Since $4Cr_n^{1-\theta} \leq A^{1-\theta} r_n^{1-\theta} < 1$, this yields

$$M_{n+1} \leq r_n [\alpha_n^{\kappa_n}]^{2-\theta}. \tag{3.14}$$

Since $(1 + \beta + \theta)(1 + \theta) < 2 - \theta$, and, since $\alpha_n > \alpha_{n+1} \geq \alpha_n^{1+\beta+\theta}$, one has

$$[\alpha_n^{\kappa_n}]^{2-\theta} \leq \alpha_{n+1}^{(1+\beta+\theta)^{-1}(2-\theta)\kappa_n} \leq \alpha_{n+1}^{(1+\theta)\kappa_n} = \alpha_{n+1}^{\kappa_{n+1}}.$$

Using this in (3.14),

$$M_{n+1} \leq r_{n+1} \alpha_{n+1}^{\kappa_{n+1}}$$

and (3.13) holds also in this case. Thus as we saw above, $\kappa_n < 1$ for every $n \geq n_1$. But since $\kappa_{n+1} \geq \kappa_n$ for all $n \geq n_1$, the inequality

$$CM_n \frac{\alpha_n}{\alpha_{n+1}} > M_n^\theta$$

can only occur a finite number of times, because from the definition in (3.12b) otherwise $\kappa_{n+1} = (1 + \theta)\kappa_n$ infinitely often, contradicting $\kappa_n < 1$ for every $n \geq n_1$. This completes the proof of the lemma and also of Theorem 3.2. □

4 Families of Circle Diffeomorphisms; Arnol'd Tongues

Following Arnol'd and Herman we will study in this section families of circle diffeomorphisms. So let $f \colon S^1 \to S^1$ be a circle homeomorphism and define the rotation function, $S^1 \ni \alpha \mapsto \rho(\alpha) := \rho(R_\alpha \circ f)$. It will turn out that the function $\rho \colon S^1 \to S^1$ is continuous and assumes every irrational value in $S^1 = \mathbb{R}/\mathbb{Z}$ exactly once. Under a mild additional condition it will be shown that for each rational number $\frac{p}{q}$, $\rho^{-1}(\frac{p}{q})$ has a non-empty interior. Finally it will be shown that for $0 < |a| < \frac{1}{2\pi}$, the map $f \colon S^1 \to S^1$ defined by $f(t) = t + a\sin(2\pi t)$ is an analytic diffeomorphism satisfying this additional condition. Consequently, for such maps f, the function $\alpha \to \rho(R_\alpha \circ f)$ is locally constant at each rational value! This phenomenon is called *phase locking*.

For convenience of notation let $f_\alpha = R_\alpha \circ f$, and let $\hat{f}$ and $\hat{f}_\alpha$ be the lifts of f respectively f_α such that $\hat{f}_\alpha = \hat{f} + \alpha$.

Lemma 4.1. *Let $f\colon S^1 \to S^1$ be a homeomorphism without periodic points. Then $\rho(R_\alpha \circ f) > \rho(f)$ if $\alpha > 0$.*

Proof. Suppose $f\colon S^1 \to S^1$ is a homeomorphism without periodic points. By Zorn's Lemma, there exists a closed f-invariant subset $K \subset S^1$ which is minimal, i.e., K does not contain any compact, non-empty, proper f-invariant subset. By minimality, every orbit in K is dense in K. Therefore, if $x \in K$, there exists a sequence $n_i \to \infty$ with $f^{n_i}(x) \to x$. So choose x so that it is accumulated from both sides by other points in K (this is possible because otherwise K would be countable). Let $\pi\colon \mathbb{R} \to S^1$ be the canonical projection and take $\hat{x} \in \mathbb{R}$ such that $\pi(\hat{x}) = x$. Therefore there exists a sequence p_i of positive integers such that $\hat{f}^{n_i}(\hat{x}) - \hat{x} - p_i$ tends to zero from both sides. By taking a subsequence we may assume that $\hat{f}^{n_i}(\hat{x}) < \hat{x} + p_i$ for all i. So take $\alpha > 0$. We claim that $\hat{f}^n_\alpha(\hat{x}) \geq \hat{f}^n(\hat{x}) + \alpha$ for all $n \in \mathbb{N}$. Indeed, this is true for $n = 1$ and assuming, by induction, it is true for $n-1$ we get $\hat{f}^n_\alpha(\hat{x}) = \hat{f}_\alpha(\hat{f}^{n-1}_\alpha(\hat{x})) \geq \hat{f}_\alpha(\hat{f}^{n-1}(\hat{x})) = \hat{f}^n(\hat{x}) + \alpha$ and the claim is proved. This, $\hat{f}^n_0(\hat{x}) < \hat{x} + p_i$ and the Intermediate Value Theorem implies that there exists $0 \leq \alpha_i \leq \hat{x} + p_i - \hat{f}^{n_i}(\hat{x})$ such that $(\hat{f}_{\alpha_i})^{n_i}(\hat{x}) = \hat{x} + p_i$. Hence x is a periodic point of f_{α_i}. □It follows that

$\alpha \to \rho(\alpha)$ is increasing and strictly increasing if $\rho(\alpha)$ is irrational.

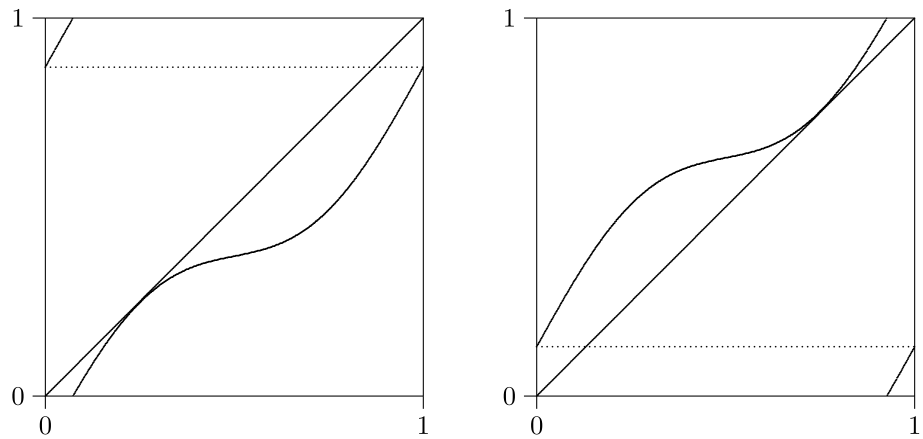

Fig. 4.1: The graph of f^s_α when $\alpha \in K^{\pm}_{r/s}$

Lemma 4.2. *Assume that $f_\alpha^m \neq id$ for all $\alpha \in S^1$ and all $m \in \mathbb{N}$. Then $\rho^{-1}(\frac{r}{s})$ has a non-empty interior for each rational number $\frac{r}{s}$. In particular, the set*

$$\{\alpha\, ;\ \rho(f_\alpha) \text{ is irrational}\}$$

is nowhere dense in S^1.

Proof. Let $K_{r/s}$ the set of α such that $\rho(f_\alpha) = \frac{r}{s}$. From the continuity and the monotonicity of $\alpha \mapsto \rho(\alpha)$ it follows that this set is non-empty. Equivalently,

$$K_{r/s} = \{\alpha\, ;\ \hat{f}_\alpha^s(x) = x + r \text{ at some point } x\}.$$

Similarly, define

$$K_{r/s}^+ = \{\alpha \in K_{r/s}\, ;\ \hat{f}_\alpha^s(x) \geq x + r \text{ for all } x\},$$

$$K_{r/s}^- = \{\alpha \in K_{r/s}\, ;\ \hat{f}_\alpha^s(x) \leq x + r \text{ for all } x\}.$$

The proof in the previous lemma also implies that these sets are non-empty and they both consist of single points. Since $\hat{f}_\alpha^s(x) \not\equiv x + r$ for all α, $K_{r/s}^+ \neq K_{r/s}^-$. Therefore $K_{r/s}$ is a non-trivial interval and the lemma follows. $\square$

If $\alpha \in K_{r/s}$ then f_α has periodic points of period s. If $\alpha \in K_{r/s}^\pm$ then each of these periodic points of f_α attracts from one side and is repelling from the other side, see Figure 4.1. If $\alpha \in \text{interior}(K_{r/s})$ then at least one of the periodic points of f is attracting from both sides.

Lemma 4.3. *Consider S^1 as the unit circle in the complex plane. If $f\colon S^1 \to S^1$ is an analytic diffeomorphism which has an entire extension to the complex plane which is not affine, then there exists no integer n such that $f^n \equiv id$ on S^1.*

Proof. If $f^n(x) = x$ for all $x \in S^1$ then $f^n(x) = x$ for all x in the complex plane. Therefore $f^{n-1} \circ f = id$ and f is an entire biholomorphic transformation. This implies that f is of the form $f(z) = cz+d$. This contradicts the assumptions. $\square$

In particular, for each $0 < |a| < \frac{1}{2\pi}$, the map $f\colon S^1 \to S^1$ defined by $f_a(t) = t + a\sin(2\pi t)$ is an analytic diffeomorphism and satisfies the conditions of Lemmas 4.2 and 4.3. In particular, for each such a, the function $\alpha \to \rho(R_\alpha \circ f_a)$ is monotone, locally constant at each α for which $\rho(R_\alpha \circ f_a)$ is rational, and non-constant at each irrational value. A function with these properties is often called a *devil's staircase*.

figure 4.2: The function $\alpha \to \rho(R_\alpha \circ f_a)$.

figure4.3: The boundary of the set $\{(\alpha, a)\,;\ \rho(f_{\alpha,a}) = \text{constant}\}$. In the vertical direction the parameter a is drawn. The tongues corresponding to rotation numbers 1/2, 4/7, 3/5, 5/8, 2/3, 5/7, 3/4, 4/5, 5/6, 6/7, 7/8 and 1 are given.

Furthermore, consider the two-parameter family

$$f_{\alpha,a} = R_\alpha \circ f_a.$$

From Lemma 4.2 for each irrational number ρ, the set $\{(\alpha, a)\,;\ \rho(f_{\alpha,a}) = \rho\}$ is the graph of a continuous function. For ρ rational this set has a non-empty interior, and is bounded by two continuous curves, see Figure 4.3. The wedges between these two curves are commonly referred to as Arnol'd tongues. Although $\{\alpha\,;\ \rho(f_{\alpha,a})$ is irrational$\}$ (where a is some fixed number as before) is nowhere dense in S^1, this set has positive Lebesgue measure. This follows from the result of M. Herman which we will discuss in the Section 6.

Exercise 4.1. In this exercise we will show that for any $r \geq 1$ the set of Morse-Smale diffeomorphisms is open and dense in the space of all C^r diffeomorphisms. Here a diffeomorphism $g\colon S^1 \to S^1$ is called a *Morse-Smale* if g has only a finite number of periodic points and if each of these periodic points is hyperbolic (a periodic point p of g is called *hyperbolic* if $g^n(p) = p$ implies $|Dg^n(p)| \neq 1$). (In higher dimensions a Morse-Smale diffeomorphism has to satisfy some other properties.) We will show this statement in a few steps.
i) Show that Lemma 4.1 implies that a C^r diffeomorphism $f\colon S^1 \to S^1$ can be approximated in the C^r topology by a diffeomorphism $f_1\colon S^1 \to S^1$ with a periodic point.
ii) Let p be a periodic point of f_1 and assume it has period n. Show that f_1 can be approximated in the C^r topology by a C^r diffeomorphism $f_2\colon S^1 \to S^1$ such that p is again a periodic point of f_2 of period n but with $Df_2^n(p) \neq 1$. (Hint: simply change f_1 in a neighbourhood U of p such that $U \cap O(p) = \{p\}$ and so that $f_2(p) = f_1(p)$ and $Df_2(p) \neq Df_1(p)$ by using bump functions.)
iii) Show that f_2 can be approximated by a diffeomorphism f_3 having p as a periodic point of period n and such that all periodic points of f_3 are hyperbolic. (Hint: since f_2 has a periodic point of period n, all periodic points of f_2 have period equal to n (unless f_2 reverses orientation in which case the situation is simpler). Now repeat the previous construction. Note that it may happen that f_2^n is the identity on some interval.) Steps i)-iii) imply the density of Morse-Smale diffeomorphisms. iv) Let f be a C^r Morse-Smale diffeomorphism, where as before $r \geq 1$. Show that any diffeomorphism g sufficiently close in the C^r topology to f has only hyperbolic periodic points and that the number of periodic points of f and g is the same. (Hint: use the implicit function theorem.)

Exercise 4.2. Let $f, \tilde{f}\colon S^1 \to S^1$ be two Morse-Smale diffeomorphisms with the same number, say k, of periodic points and with the same period. Show that these diffeomorphisms are conjugate.
i) Show that the number of periodic points of a Morse-Smale diffeomorphism $g\colon S^1 \to S^1$ is even. (Hint: between each two attracting periodic points there is a repelling periodic point.)

ii) Let p be an attracting periodic point of g. Show that there exists two half-open intervals I_1 and I_2 in the basin $B(p)$ of p with the property that each orbit in $B(p)$ intersects one of these intervals exactly once. The set $I_1 \cup I_2$ is called a *fundamental domain* of p. (Hint: let q_1 and q_2 be the neighbouring periodic points, i.e., the periodic point so that the arc (q_1, q_2) contains no other periodic points except p. Let $x_i \in (q_i, p) \subset (q_1, q_2)$ and define $I_i = [x_i, g(x_i))$. The union $\cup_{i=1,2} \cup_{n \in \mathbb{Z}} f^n(I_1)$ coincides with the basin of p.)
iii) Now construct a conjugacy between f and $\tilde{f}$. (Hint: Let $I_1, \ldots, I_k$ be the fundamental domains for f and $\tilde{I}, \ldots, \tilde{I}_k$ the fundamental domains for $\tilde{f}$. Take any homeomorphism $h_* \colon S^1 \to S^1$ which sends I_i homeomorphically to $\tilde{I}_i$. For each non-periodic point x there exists a unique integer $n(x)$ such that $f^n(x) \in I_i$. Next define $h(x) = \tilde{f}^{-n(x)} \circ h_* \circ f^{n(x)}(x)$. Show that h extends to a homeomorphism on S^1. By definition one immediately has $\tilde{f} \circ h = h \circ f$.)
Combining this and the previous exercise one has that each Morse-Smale diffeomorphism is C^r *structurally stable*, i.e., each such diffeomorphism is conjugate to any C^r nearby diffeomorphism.

Exercise 4.3. Consider the two-parameter family $f_{\alpha,a}$ for $a \geq 0$ from before. Show that the wedge $\rho(f_{\alpha,a})$ at $(0,0)$ is bounded by the curves $\alpha = \pm|a|$. Similarly, by looking at the second and third iterate of $f_{\alpha,a}$, show that the wedges in $(1/2, 0)$ and $(1/3, 0)$ are bounded by respectively

$$\alpha = 1/2 + \pm(\pi/2)a^2 + O(a^3)$$

and

$$\alpha = 1/3 + (\sqrt{3}\pi/6)a^2 \pm (\sqrt{7}\pi/6)a^3 + O(a^4).$$

See also Arnol'd (1961).

5 Counter-Examples to Smooth Linearizability

In this section we assume that $f \colon S^1 \to S^1$ is analytic and satisfies $f^n \not\equiv id$ for all $n \in \mathbb{Z}$. Furthermore, let $f_\alpha = R_\alpha \circ f$ and

$$\mathcal{R} = \{\alpha \in S^1\, ;\ \rho(f_\alpha) \text{ is irrational}\}.$$

From the previous section it follows that we could take here the family,

$$f_\alpha(t) = t + a\sin(2\pi t) + \alpha \bmod 1,$$

where $0 < |a| < \frac{1}{2\pi}$. Moreover, the results from the previous section imply that $\mathcal{R}$ is perfect (has no isolated points) and totally disconnected (has no interior points) and that its closure has the same properties. In other words, the closure of $\mathcal{R}$ is a complete metric space and this closure is just the union of $\mathcal{R}$ with a countable set. Therefore $\mathcal{R}$ has the Baire property: the countable intersection of a collection of open and dense subsets in $\mathcal{R}$ is again dense in $\mathcal{R}$.

In general an analytic diffeomorphism without periodic points is not C^1 conjugate to a rotation, see Finzi (1950). In the theorem below, due to Arnol'd, it will be shown that, even when a diffeomorphism f as above is analytic, there exists a dense set of parameters $\alpha \in \mathcal{R}$ such that the conjugacy between f_α and the rotation is not even absolutely continuous. So in general a conjugacy sends a set of positive Lebesgue measure into a set of zero Lebesgue measure. In particular, these maps do not have an absolutely continuous invariant probability measure.

Theorem 5.1. *For f and f_α as above there exists a dense set of α's in $\mathcal{R}$ such that there exists no absolutely continuous conjugacy between f_α and a rotation. (Note that for $\alpha \in \mathcal{R}$, f_α is certainly conjugate to a rotation.)*

Corollary 5.1. *There exists a dense set of α's in $\mathcal{R}$ such that there exists no C^1 conjugacy between f_α and a rotation.*

Proof of Corollary. Although the corollary follows immediately from Theorem 5.1 let us give an independent proof. Notice that if h is a C^1 conjugacy between f and R_ρ, then $f^n = h \circ R_\rho^n \circ h^{-1}$ for all $n \in \mathbb{Z}$. Since h is C^1, $\sup_{n\in\mathbb{Z}} \log ||Df^n||$ is finite. So let us show that there exist many parameters α such that $\sup_{n\in\mathbb{Z}} \log ||Df_\alpha^n||$ is infinite. This is done as follows. Take the open set

$$U_k = \{\alpha \in \mathcal{R}\,;\ \sup_{n\in\mathbb{Z}} \log ||Df_\alpha^n|| > k\}.$$

Let us show that this set is dense in $\mathcal{R}$. So take $\alpha \in \mathcal{R}$. From Lemma 4.2, arbitrarily close to α there exists a number $\alpha' \in K^{\pm}_{r/s}$. Since $f_{\alpha'}$ has periodic points which are attracting (from one-side), $\sup_{n\in\mathbb{Z}} \log ||Df_{\alpha'}^n|| = \infty$. In particular, for any α'' sufficiently close to α' one has $\sup_{n\in\mathbb{Z}} \log ||Df_{\alpha''}^n|| > k$. From Lemma 4.2 one can choose such a parameter α'' so that it is contained in $\mathcal{R}$ and so that it is arbitrarily close to α'. Combining this gives that U_k is dense in $\mathcal{R}$. From the Baire property one gets that $\cap_{k\geq 0} U_k$ is dense in $\mathcal{R}$. Since $\sup_{n\in\mathbb{Z}} \log ||Df_\alpha^n|| = \infty$ for $\alpha \in \cap_{k\geq 0} U_k$, f_α is not C^1 conjugate to a rotation. □

Note that the diffeomorphisms which were constructed in the previous result are near to diffeomorphisms with a one-sided attractor O. This forces orbits to spend most of their time near O. Such behaviour is sometimes called *intermittency.* We will elaborate on this in the next exercise.

Exercise 5.1. Let f and f_α as above. For each $\alpha' \in K^{+}_{r/s}$, the diffeomorphism $f_{\alpha'}$ has a one-sided periodic attractor O (this set consists of one or possibly of a finite number of periodic orbits). Show that for any neighbourhood U of O and any $x \in S^1$, there exists $\epsilon > 0$ such that for any $\alpha \in (\alpha', \alpha' + \epsilon) \cap \mathcal{R}$ any limit measure μ of the measures

$$\frac{1}{n}\sum_{i=0}^{n-1} \delta_{f_\alpha^i(x)}$$

has the property that $\mu(U) \geq 1/2$. (Hint: For any neighbourhood U, the cardinality of $\{0 \leq k < n\,;\ f^k(x) \in U\}$ is at least $n/2$ for n sufficiently large provided $\epsilon > 0$ is sufficiently small.) The idea of the proof of Theorem 5.1 is to combine this with a Baire argument.

Proof of Theorem 5.1. Let V_k be the open subset of $\mathcal{R}$ of parameters α for which there exist open sets A with $|A| < \frac{1}{k}$ (where $|A|$ denotes the Lebesgue measure of A) and an integer N with $f_\alpha^N(S^1 \setminus A) \subset A$. We claim that for each k the set V_k is dense.

Before proving this claim let us show that the theorem follows from this claim. So let $V = \cap_{k\geq 0} V_k$. Then the Baire property implies that V is dense in $\mathcal{R}$. For $\alpha \in V$, there exists a sequence of sets A_k and integers N_k such that i) $f_\alpha^{N_k}(S^1 \setminus A_k) \subset A_k$ and ii) $|A_k| < \frac{1}{k}$. Let us show that there exists no absolutely continuous homeomorphism h such that $h \circ f_\alpha \circ h^{-1}$ is a rotation. Since rotations preserve Lebesgue measure, for any homeomorphism h such that $h \circ f_\alpha \circ h^{-1}$ is a rotation one has that the measure of $h(A_k)$ is equal to the measure of $h(f_\alpha^{N_k}(A_k))$. From this and i),

$$1 - |h(A_k)| = 1 - |h(f_\alpha^{N_k} A_k)| = |h(f_\alpha^{N_k}(S^1 \setminus A_k))| \leq |h(A_k)|.$$

This implies that $|h(A_k)| \geq \frac{1}{2}$. Since $|A_k| \leq \frac{1}{k}$, the conjugacy h is definitely not absolutely continuous!

So we only need to prove that V_k is dense in $\mathcal{R}$. Take $\alpha \in \mathcal{R}$. From Lemma 4.2, arbitrarily close to α there exists a number $\alpha' \in K^{\pm}_{r/s}$. Since f is analytic, $f_{\alpha'}$ has a finite number of periodic points $x_1, \dots, x_r$ all of which are fixed points of $f_{\alpha'}^s$ and attracting from one-side (and repelling from the other side). It follows that for each $x \in S^1$, $f_{\alpha'}^n(x) \to \{x_1, \dots, x_r\}$ as $n \to \infty$. In particular, if we let A_k be an open neighbourhood of $\{x_1, \dots, x_r\}$ with Lebesgue measure $< \frac{1}{k}$, then there exists $N(k) < \infty$ such that for each $x \in S^1 \setminus A_k$, $f_{\alpha'}^N(x) \in A_k$ and therefore $f_{\alpha'}^N(S^1 \setminus A_k) \subset A_k$. By continuity and, since A_k is open, it follows that for each α'' sufficiently near α', one also has

$$f_{\alpha''}^N(S^1 \setminus A_k) \subset A_k.$$

From Lemma 4.2 one can choose $\alpha'' \in \mathcal{R}$ arbitrarily close to α'. Hence $\alpha'' \in V_k$. The density of V_k in $\mathcal{R}$ follows. □

Remark. As we have seen in one of the exercises in §1, every circle homeomorphism without periodic points has precisely one invariant probability measure. In particular, if f is conjugate to the rotation R and has an invariant probability measure μ then $h_*\mu$ is the Lebesgue measure. Here h is the conjugacy between f and R, $h \circ f = R \circ h$, and $h_*\mu$ is the measure defined by $h_*\mu(A) = \mu(h^{-1}(A))$ for all measurable sets A. From the previous result it follows that in general h is not absolutely continuous, and therefore that μ is not absolutely continuous with respect to the Lebesgue measure. Even if f has no finite invariant measure which is absolutely continuous with respect to the Lebesgue measure, it may still have an σ-finite invariant measure which is absolutely continuous with respect to the Lebesgue measure. This is an invariant measure μ such that S^1 it the countable union of intervals I_i, such that $\mu(I_i) < \infty$ and such that μ is absolutely continuous, i.e., there exists a L^1-function $\phi\colon S^1 \to [0,\infty]$ such that

$$\mu(A) = \int_A \phi \, d\mu.$$

Equivalently, letting $K = \phi^{-1}(0,\infty)$ one has that $\mu(K) > 0$ and that

$$\phi \circ f \cdot Df = \phi \text{ for } \mu \text{ almost all points.}$$

It is not known whether all diffeomorphisms without periodic points from the family $f(t) = t + a\sin(2\pi t) + \alpha$, $0 < |a| < \frac{1}{2\pi}$, have such an invariant σ-finite measure. However, there are examples of C^∞ diffeomorphisms which have no σ-finite absolutely continuous invariant measure and also of such diffeomorphisms which have no finite but do have an σ-finite absolutely continuous measure, see Katznelson (1977). In the next exercise an example due to Herman (1979) of a continuous, piecewise linear circle homeomorphism without invariant absolutely continuous σ-finite measures is given.

Exercise 5.2. Show that in general an analytic diffeomorphism f without periodic points is not quasi symmetrically conjugate to a rotation. Here we say that a homeomorphism $h\colon S^1 \to S^1$ is *quasisymmetric* if there exists $K < \infty$ such that for each two intervals I_1 and I_2 with a common end-point such that $|I_1| = |I_2|$ one has $\frac{|h(I_1)|}{|h(I_2)|} \le K$. (This concept plays an important role in the last chapter of this book because of its connections with quasiconformal maps on the Riemann sphere.) (Hint: use the same type of diffeomorphism as in the proof of Theorem 5.1 and choose intervals $I_1 \cup I_2$ on one side of a point which is almost a one-sided fixed point.) (Recently, Yoccoz has shown that any two analytic homeomorphisms with a unique critical point and with the same rotation number are quasi-symmetrically conjugate. Later on we shall see that quasisymmetry plays an important role in non-invertible one-dimensional systems.)

Exercise 5.3. In this exercise we shall show, following Herman (1979), that there exist continuous, piecewise linear circle homeomorphisms without invariant absolutely continuous σ-finite measures. Define a continuous map $F\colon [0,1] \to [0,1]$ with slope λ on an interval of the form $[0,a]$ and slope $\frac{1}{\lambda}$ on $[a,1]$, and such that $F(0) = 0$. (Clearly F is completely determined and $a = 1/(\lambda+1)$.) Now define $f\colon S^1 \to S^1$ so that $f(x) = F(x) + b \bmod 1$, where $b \in \mathbb{R}$ is chosen so that f has no periodic points. Suppose by contradiction that f has an absolutely continuous σ-finite invariant measure μ, which is absolutely continuous with respect to the Lebesgue measure. Let ϕ be its density and let $K = \phi^{-1}(0,\infty)$ (where $\mu(K) > 0$). a) Show that the invariance of μ implies that

$$\phi \circ f \cdot Df = \phi \text{ for } \mu \text{ almost all points.}$$

b) Let

$$g(x) = \begin{cases} \exp\left(2\pi i \frac{\log \phi(x)}{2\log\lambda}\right) & \text{when } x \in K, \\ 0 & \text{when } x \notin K. \end{cases}$$

Using step a) show that $g \circ f(x) = -g(x)$ for $x \in K$. In particular $g \circ f^2 = g$ on F. c) Show that g is strictly positive on a set of positive Lebesgue measure. (Use that μ is absolutely continuous and that $g(x) > 0$ for $x \in K$ and $\mu(K) > 0$.) d) As we remarked in exercise 1 in §2.b, f and also f^2 is ergodic. Show that this implies that g is μ-almost everywhere constant. (Hint: use $g \circ f^2 = g$ and deduce from this that for each $c \in \mathbb{R}$ the set $\{x\,;\ g(x) \geq c\}$ is f^2-invariant. From the ergodicity this set has either Lebesgue measure zero or full Lebesgue measure. Since this holds for each $c \in \mathbb{R}$, g must be constant.) e) Show that g cannot be almost everywhere constant. This contradicts part d) of this exercise.

6 Frequency of Smooth Linearizability in Families

Following Herman (1977), we will show in this section that for smooth families $[0,1] \ni t \to f_t$ of smooth diffeomorphisms such that $\rho(f_0) \neq \rho(f_1)$, the set of parameters for which f_t is C^1 linearizable has positive Lebesgue measure.

Theorem 6.1. *Let $[0,1] \ni t \to f_t$ be a family of C^3 diffeomorphisms depending C^1 on t such that $\rho(f_0) \neq \rho(f_1)$. Then the set of parameters for which f_t is C^1 linearizable has positive Lebesgue measure.*

First we need the following lemma. This lemma states that the rotation number depends Lipschitz on the perturbation at smoothly linearizable maps.

Lemma 6.1. *Assume that $t \to f_t$ is C^1 and that $h \circ f_{t_0} \circ h^{-1} = R_\alpha$ where h is a C^1 conjugacy. Then*

$$\frac{|\rho(f_t) - \rho(f_{t_0})|}{|t - t_0|} \leq ||Dh|| \cdot ||\frac{\partial}{\partial t} f_t||.$$

Proof. First notice that $|\rho(f) - \rho(R_\alpha)| \leq ||f - R_\alpha||$. Indeed, let $\epsilon = ||f - R_\alpha||$; then the lift $\hat{f}$ of f can be chosen such that $x + \alpha - \epsilon \leq \hat{f}(x) \leq x + \alpha + \epsilon$ for all x. Hence

$$\alpha - \epsilon \leq \frac{\hat{f}^n(x) - x}{n} \leq \alpha + \epsilon$$

and $|\rho(f) - \rho(R_\alpha)| \leq ||f - R_\alpha||$ follows. Hence

$$|\rho(f_t) - \rho(f_{t_0})| = |\rho(h \circ f_t \circ h^{-1}) - \rho(R_\alpha)| \leq$$

$$\leq ||h \circ f_t \circ h^{-1} - R_\alpha|| \leq ||\frac{\partial}{\partial t} h \circ f_t \circ h^{-1}|| \times |t - t_0| \leq$$

$$\leq ||Dh|| \cdot ||\frac{\partial}{\partial t} f_t|| \times |t - t_0|. \quad \square$$

Lemma 6.2. *Let f be a C^3 diffeomorphism such that $\alpha = \rho(f)$ satisfies the Diophantine condition:*

$$|\alpha - \frac{p}{q}| > \frac{K}{q^{2+\beta}} \text{ for all } \frac{p}{q} \in \mathbb{Q}$$

where K and β are positive constants. Then f is C^1 conjugate to a rotation and $||Dh||$ can be estimated from above in terms of $||\log Df||$, $||Sf||$ and K and β.

Proof. That there exists a C^1 conjugacy is the contents of Theorem 3.1. So we need to check that an upper-bound of $||Dh||$ can be given which only depends on K, β, $||\log Df||$ and $||Sf||$. So let us go through the proof of Theorem 3.1. The first part of the proof of Theorem 3.1 was to show that f is C^1 conjugate to a rotation if $\sup \log ||Df^i||$ is bounded. Since h is a C^1 and conjugates f to a rotation, one has

$$\log Dh - \log Dh \circ f = \log Df.$$

Since there exists a point $x \in S^1$ such that $Dh(x) = 1$, it follows from the continuity of Dh and from $\log Dh \circ f^n(x) = -\sum_{i=0}^{n} \log Df(f^i(x) = -\log Df^n(x)$ that

$$||\log Dh|| \leq \sup_n ||\log Df^i||.$$

One can easily check that the constants C appearing in Steps 2-4 only depend on $||\log Df||$ and $||Sf||$. Finally the upper-bound obtained for M_n/m_n in Step 5 only depends on C, K, and β. □

Proof of Theorem 6.1 Fix $K > 0$ and $\beta > 0$ and let D be the set of α's in $(\rho(f_0), \rho(f_1))$ such that

$$|\alpha - \frac{p}{q}| > \frac{K}{q^{2+\beta}} \text{ for all } \frac{p}{q} \in \mathbb{Q}.$$

It is well known that the set D has positive Lebesgue measure, see the exercise below. Let $A = \{t \in (0,1)\,;\ \rho(f_t) \in D\}$. Since $t \to \rho(f_t)$ is continuous, one has $\rho(A) = D$. From the previous lemma there exists $C_1 < \infty$ such that for each $t \in A$, the C^1 conjugacy h_t between f_t and a rotation satisfies $||Dh_t|| \leq C_1$. Let $C_2 = \sup ||\frac{\partial}{\partial t} f_t||$. From Lemma 6.1 one gets that $\rho\colon A \ni t \mapsto \rho(f_t) \in D$ has Lipschitz constant $C_1 \cdot C_2$. It follows that

$$0 < |D| = |\rho(A)| \leq C_1 \cdot C_2 \cdot |A|.$$

(Here $|X|$ a.s.o. denotes the Lebesgue measure of a set $X \subset S^1$). It follows that $|A| > 0$, and, since for each $t \in A$ the diffeomorphism f_t is C^1 linearizable, the theorem follows. □

Exercise 6.1. Let $K > 0$ and $\tau > 1$ and let

$$\Omega_{K,\tau} = \{\alpha\,;\ |q\alpha \bmod 1| > \frac{K}{q^\tau} \text{ for all } q \in \mathbb{Z}\}.$$

Show that the set of $\Omega_{K,\tau}$ has positive Lebesgue measure. Moreover, $\Omega_\tau = \cup_{K>0}\Omega_{K,\tau}$ has full Lebesgue measure. (Hint: for $r > 0$ and for fixed $q \in \mathbb{Z}$ the set

$$\left\{\alpha\,;\; -r \le \alpha \le r,\, |q\alpha \bmod 1| < \frac{K}{q^\tau}\right\}$$

has at most Lebesgue measure $cqr\frac{K}{q^{1+\tau}}$. Hence $[-r,r] \setminus \Omega_{K,\tau}$ has at most Lebesgue measure $\sum_{q\in\mathbb{Z}} cqr\frac{K}{q^{1+\tau}} \le c'rK$. Hence $[-r,r] \setminus \Omega_\tau$ has Lebesgue measure zero.)

7 Some Historical Comments and Further Remarks

The best reference on circle diffeomorphism is no doubt M. Herman's thesis (1979). Results on ergodic properties of flows without singularities on a torus can be found in Furstenberg (1961) and also Chapter XVI of Cornfield et al. (1982). If a torus flow has singularities the situation becomes more complicated. In one of the simplest cases the return map is a continuous circle map which is constant on some arcs. This case was first studied by Cherry (1938) and generalized in Martens et al. (1990). For more on flows on surfaces see for example Aranson and Grines (1986), Cornfield et al. (1982) and Godbillon (1983).

The results on the rotation intervals of non-invertible circle homeomorphisms mentioned in the exercises at the end of Section 1 are due to several people. For this we refer to Newhouse et al. (1983) (already circulated in 1977 as a preprint), and for example Bernhardt (1982), Boyland (1985), Chenciner et al. (1984), Misiurewicz (1986), Barkmeijer (1988) and Alsedà and Mañosas (1990).

Denjoy's result that C^2 diffeomorphisms without periodic points do not have wandering intervals is sharp in many ways. Hall (1981) has shown that there exists a C^∞ homeomorphism (with one critical point) without periodic points and which has wandering intervals. Yoccoz (1984b) has shown that analytic homeomorphisms of the circle without periodic points have no wandering intervals. This last result is known in much greater generality now, see Chapter IV.

For the proof of the result that any analytic diffeomorphism which is near to a rotation and has a 'good rotation number' is analytically conjugate to a rotation, see Arnol'd (1965), (1983). In Moser (1990) the problem of simultaneously linearizing commuting diffeomorphisms is solved when they are near rotations. The corresponding global problem (when the diffeomorphisms are not near rotations) is still open. In Chapter VI we shall study similar rigidity results for renormalizable interval maps. Here Lanford proved the local version of the rigidity conjectures of Feigenbaum and later Sullivan proved the global results.

As we have seen in Section 6, for families of smooth circle diffeomorphisms with non-constant rotation number the set of parameter values for which the rotation number is irrational has positive Lebesgue measure. Later it was shown that the analogous result is false for circle homeomorphisms with critical points.

Indeed, in Boyd (1985), families of monotone degree one circle maps, which are constant on some interval and expanding elsewhere, were considered. It was shown for these families that the set of parameters corresponding to rational rotation numbers has full measure. This result was greatly generalized by Świątek (1988), (1989) who proved that for a rather general class of families of maps the parameter for which the rotation number is rational has again full measure, see also Veerman (1989), Veerman and Tangerman (1990a) and Tangerman and Veerman (1991). Moreover, Świątek (1989) shows that the orbit of the critical point of a critical circle map (a smooth map with some critical points) satisfies some very specific scaling laws. Khanin (1990) shows that the conjugacy between a critical circle map and a rotation is definitely not absolutely continuous if its coefficients in the continued fraction expansion of its rotation number are unbounded. Moreover, Herman and Yoccoz have shown that the conjugacy between two analytic homeomorphisms with a unique critical point and the same rotation numbers is quasisymmmetric, see also Świątek (1990). Extending ideas which will be discussed in the last chapter, Faria (1992), has obtained an important result which leads to rigidity for these critical circle maps. We will come back to this in the last chapter.

There are several numerical results about the sizes of the rational regions in Arnol'd tongues. For some numerically observed scalings of the sizes of the 'steps' $\{\alpha\,;\ \rho(R_\alpha \circ f_a) \text{ is rational}\}$, see for example Cvitanović and Söderberg (1988). In this paper it is also conjectured that if p/q and p'/q' are Farey neighbours then the largest step between the two steps which correspond to p/q and p'/q' is the step corresponding to the rotation number $(p+p')/(q+q')$. This last number is the rational number between p/q and p'/q' with the largest denominator (and the number following these rationals in the Farey tree). Some of these numerical observations have recently been confirmed for families of diffeomorphisms and C^∞ homeomorphisms of the circle by Graczyk (1991a), Jonker (1991) and Graczyk and Świątek (1991). Moreover, let f_a be a nice one-parameter family of circle diffeomorphisms. The set of parameter values a for which f_a has a rotation number which is either rational or satisfies some Diophantine condition has Lebesgue measure zero, see Tsujii (1992e). Graczyk (1992) has strengthened this result and shown that this set even has Hausdorff measure zero.

Chapter II.

The Combinatorics of Endomorphisms

In this chapter we will discuss endomorphisms of the circle and of the interval from a combinatorial point of view. The aim is to develop an analogue to the topological description of circle homeomorphisms given in Section I.1. As in that section, the main ingredient here is symbolic dynamics and the structure to be considered is the order structure of the interval or of the circle.

In Section 1, we will show that non-invertible maps have a much richer dynamics than invertible maps, by proving Sarkovskii's remarkable result that the existence of some periodic points implies the existence of many others. For example, if such a map has a periodic orbit of period three then it has periodic orbits of each period. In Section 2, we will describe the dynamics of the simplest non-invertible dynamical systems: covering maps of the circle. It will be proven that any covering map of the circle of degree $d > 1$ is combinatorially equivalent to a unique model, namely, the expanding map $z \mapsto z^d$. In Section 3, we will develop the combinatorial theory of Milnor and Thurston (1977) for maps with a finite number of turning points (a point c is called a *turning point* of an interval map $f\colon I \to I$ if the map has a local extremum at c and if c is in the interior of I). In this theory, a point x is coded by associating to it a sequence of symbols $i_n(x)$, $n = 0, 1, \dots$. Here $i_n(x)$ is an element from a finite list of symbols and depends only on the position of $f^n(x)$ in relationship to the position of the turning points of f. The main result states that if the forward orbits of the critical points of two endomorphisms are ordered in the same way then the endomorphisms are combinatorially equivalent. This notion of combinatorial equivalence will be defined in Section 3. We use this notion here because if two endomorphisms are combinatorially equivalent then not only their kneading invariants are the same but also the dynamics of points attracted to periodic points is 'the same'. In other words, up to some non-essential features, such endomorphisms are conjugate. In Section 4, we will consider families of maps f_μ with l turning points and show when such a family is full. Such a family is full if given a map g with l turning points, there exists a parameter μ such that f_μ

and g are 'essentially' combinatorially equivalent. It turns out that the fullness of families can be proved by solving a certain fixed point problem (using what we call the Thurston map). In Section 5, we will apply some of these results to families and give a first introduction to the theory of renormalization of maps.

Using the results of Section 3 and some important analytical results of Singer and Guckenheimer on the dynamics of quadratic maps, in Section 6 we will prove that any unimodal map is semi-conjugate to a quadratic map. Therefore the quadratic family $f_\mu(x) = \mu x(1-x)$, $\mu \in [0,4]$, plays the same role in the theory of unimodal maps as the rotations for circle diffeomorphisms. In fact, quoting results from Chapter IV on the non-existence of wandering intervals, we prove a corresponding result for families of multimodal maps.

In Section 7, we will introduce an important invariant, the topological entropy, which is a measure of the dynamical complexity of a map and relate this invariant with the growth of the 'lap number' of interval maps. Sections 8 and 9 contain a further development of the Milnor-Thurston theory. Using the combinatorial tools of Section 3 it will be shown that any interval map with positive topological entropy is semi-conjugate to a piecewise linear map with constant slope and the same topological entropy and also that the topological entropy depends continuously on the map. Finally, in Section 10, we will deal again with the quadratic family Q_μ and show that the kneading invariant, and therefore the topological entropy, of Q_μ increases with the parameter μ.

Only Sections 2 to 6 are relevant to the remainder of the book; the reader could skip the other sections.

1 The Theorem of Sarkovskii

As we have seen in the last chapter, the existence of a periodic point for an invertible one-dimensional dynamical system makes the dynamics extremely simple. We will see now that this is not true for non-invertible maps: unlike the case of diffeomorphisms of the circle, periodic points of different period may coexist for non-invertible maps. In fact, as a corollary of the theorem we will prove in this section that the existence of a periodic point of period three for a continuous interval map implies the existence of periodic points of every period.

The result below was proved by Sarkovskii in 1964, and has since been rediscovered by several authors.

Definition. Consider the following ordering on the set of natural numbers, called the *Sarkovskii ordering*:

$$3 \succ 5 \succ 7 \succ \cdots \succ 2n+1 \succ \cdots \succ 6 \succ 10 \succ 14 \succ \cdots \succ 2 \times (2n+1) \succ \cdots$$

$$\succ 2^m \times 3 \succ 2^m \times 5 \succ 2^m \times 7 \succ \cdots \succ 2^m \times (2n+1) \succ \cdots$$

$$\succ \cdots \succ 2^n \succ \cdots \succ 2 \succ 1\,.$$

Theorem 1.1. **(Sarkovskii)**
Let $f\colon [0,1] \to [0,1]$ be a continuous map having a periodic point of period n. If $n \succ m$ in the Sarkovskii ordering then f has a periodic point of period m.

In the proof we will use some ingredients of symbolic dynamics and we will follow the exposition of Block et al. (1980).

Lemma 1.1. *Let $f\colon I \to I$ be a continuous map where I is an interval.*

a) If $J \subset I$ is an interval such that $f(J) \supset J$ then f has a fixed point in the closure of J.

b) If $\{I_i \subset I\,;\ i = 0, 1, 2, \dots\}$ is a family of closed intervals such that $f(I_i) \supset I_{i+1}$ then there exist a nested and decreasing sequence of intervals J_n in I_0 such that $f^n(J_n) = I_n$. In particular, there exists $x \in I_0$ such that $f^i(x) \in I_i$ for every $i \geq 0$.

Proof. a) Let $a < b$ be the boundary points of J. Since $f(J) \supset J$, there exist $z, w \in J$ such that $f(z) \leq a$ and $f(w) \geq b$. So if $g(x) = f(x) - x$ then $g(z) = f(z) - z \leq f(z) - a \leq 0$ and $g(w) \geq 0$. By the Intermediate Value Theorem, there exists x in the interval bounded by z and w such that $g(x) = 0$ and this proves a).

b) As $f(I_0) \supset I_1$, there exists a closed interval $J_1 \subset I_0$ such that $f(J_1) = I_1$. Suppose, by induction, that there exist closed intervals $J_1 \supset J_2 \supset \cdots \supset J_n$ such that $f^i(J_i) = I_i$ for every $1 \leq i \leq n$. Since $f(I_n) \supset I_{n+1}$, there exists a closed interval $\tilde{I}_n \subset I_n$ such that $f(\tilde{I}_n) = I_{n+1}$. On the other hand, as $f^n(J_n) = I_n \supset \tilde{I}_n$, there exists an interval $J_{n+1} \subset J_n$ such that $f^n(J_{n+1}) = \tilde{I}_n$. Hence $f^{n+1}(J_{n+1}) = f(\tilde{I}_n) = I_{n+1}$. Since the intervals J_i lie nested, there exists $x \in \cap_{n=0}^{\infty} J_n$ and, since $f^n(J_n) = I_n$ for every $n \geq 0$, each such x satisfies the required properties. □

Definition. We say that a collection of closed subintervals $\{I_k\}$ of the interval I forms a partition if the interior of the intervals I_k are pairwise disjoint. Given a partition, the *Markov graph* of $f\colon I \to I$ associated to this partition is the graph whose vertices are the intervals of the partition and the edges are the pairs (I_i, I_k) such that $f(I_i) \supset I_k$. Let us denote such an edge by $I_i \to I_k$. (Of course, for some partitions the number of edges might be zero.)

It follows from the previous lemma that every path in the Markov graph of a partition, is associated to a point whose itinerary is exactly the sequence of vertices in this path. If the path is a closed path, i.e., a path of the type $I_{i_0} \to I_{i_1} \to \cdots \to I_{i_{n-1}} \to I_{i_0}$ then there exists a periodic point $x \in I_{i_0}$ such that $f^j(x) \in I_{i_j}$ and $f^n(x) = x$. However, even if the vertices $I_{i_0}, \dots I_{i_{n-1}}$ are distinct, the period of x may be smaller than n because the intervals of the partition are not disjoint (only the interior of the intervals are pairwise disjoint).

Example. Let $0 < p_1 < p_2 < p_3 < 1$. It is easy to construct a continuous map $f\colon [0,1] \to [0,1]$ such that $f(p_1) = p_2$, $f(p_2) = p_3$ and $f(p_3) = p_1$. Indeed, it is enough to connect the points $(p_1, p_2), (p_2, p_3), (p_3, p_1)$ by a curve in the square $[0,1] \times [0,1]$ transverse to each vertical line, see Figure 1.1. This curve is the graph of a function $f\colon [0,1] \to [0,1]$.

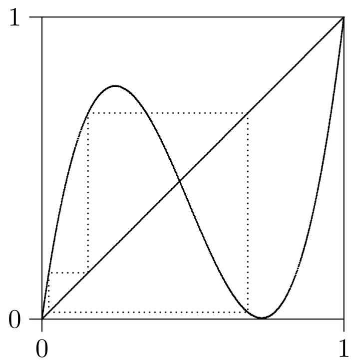

Fig. 1.1: A map $f\colon [0,1] \to [0,1]$ having a periodic point of period three.

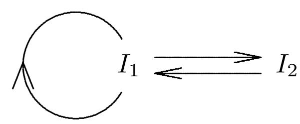

Fig. 1.2: The Markov graph associated to a periodic point of period three.

Consider the partition $I_1 = [p_2, p_3], I_2 = [p_1, p_2]$. Since $f(I_1) \supset I_1 \cup I_2$ and $f(I_2) \supset I_1$, the Markov graph of this partition has a subgraph having two vertices, an edge connecting I_1 to itself, an edge connecting I_1 to I_2 and an edge connecting I_2 to I_1, as in Figure 1.2 (notice that the Markov graph may contain another edge if $f(I_2) \supset I_2$). Let us consider the closed path

$$I_1 \to I_1 \to I_1 \cdots \to I_1 \to I_2 \to I_1$$

with $m > 3$ edges. By Statement b) of Lemma 1.1, there exists an interval $J \subset I_1$ such that $f^i(J) \subset I_1$ for $0 \le i < m-1$, $f^{m-1}(J) \subset I_2$ and $f^m(J) = I_1$. Hence because $J \subset I_1$ and using Statement a) of Lemma 1.1, there exists $x \in J \subset I_1$ such that $f^m(x) = x$. Clearly $f^i(x) \in I_1$ if $0 \le i < m-1$ and $f^{m-1}(x) \in I_2$. We claim that x is a periodic point of period m. So we claim that there is no integer $0 < i < m$ with $f^i(x) = x$. Indeed, if $f^i(x) = x$ with $0 < i < m$ then $f^{m-1}(x) = f^{i-1}(x) \in I_1$ because $x = f^m(x)$. Thus $f^{m-1}(x) \in I_1 \cap I_2 = \{p_2\}$ and therefore $x = f(f^{m-1}(x)) = f(p_2) = p_3$. This is impossible because $f(p_3) = p_1 \notin I_1$. Therefore f has a periodic point of every period > 3. By considering the path $I_2 \to I_1 \to I_2$, we get also a periodic point of period two. With the same argument, we prove that a continuous map $f\colon [0,1] \to [0,1]$ such that $f(p_1) = p_3$, $f(p_3) = p_2$ and $f(p_2) = p_1$ also has periodic points of all periods. Thus we have shown that Theorem 1.1 holds for $n = 3$. From part b) of Lemma 1.1, we get also that, in this example, the itinerary of a point x (i.e., the sequence of intervals $I_{i(1)} I_{i(2)} I_{i(3)} \ldots$ such that $f^k(x) \in I_{i(k)}$ for $k \ge 0$) may be quite arbitrary. In fact, given a list $(n_1, n_2, n_3 \ldots)$ of positive integers, there exists a point whose itinerary is equal to the list $(I_1)^{n_1} I_2 (I_1)^{n_2} I_2 (I_1)^{n_3} I_2 \ldots$, where $(I_1)^{n_j}$ denotes the list made of n_j symbols equal to I_1.

Let us now come to the general case. Let x be a periodic point of period n of a continuous map $f\colon I \to I$. Let $x_0 < x_1 < \cdots < x_{n-1}$ be the points of the orbit of x defining a partition of the interval $J = [x_0, x_{n-1}]$ into $n-1$ closed intervals. We will describe some of the properties of the Markov graph of f associated to this partition.

Lemma 1.2. *There exists a vertex $I_1 = [x_a, x_{a+1}]$ of the Markov graph of f such that $f(I_1) \supset I_1$ and, in fact, $f(x_{a+1}) \leq x_a < x_{a+1} \leq f(x_a)$.*

Proof. Since $O_f(x) = \{x_0, \ldots, x_{n-1}\} \in [x_0, x_{n-1}]$, we have $f(x_0) > x_0$ and $f(x_{n-1}) < x_{n-1}$. Thus there exists an integer $0 < a < n-1$ such that

$$x_a = \max\{x_i\,;\ f(x_i) > x_i\}.$$

Take $I_1 = [x_a, x_{a+1}]$. Since $f(x_a) \geq x_{a+1}$, $f(x_{a+1}) \leq x_a$ and $f(I_1)$ is an interval, we have that $f(I_1) \supset I_1$. □

From now one let I_1 be as in Lemma 1.2.

Lemma 1.3. *Let I_1 be a vertex of the Markov graph such that $f(I_1) \supset I_1$. Then for any vertex K of the Markov graph there exists a path connecting I_1 to K.*

Proof. Let V_i be the set of vertices which are endpoints of some path of size i starting at I_1. Thus $K' \in V_i$ if there exists a path $I_1 \to K_2 \to \cdots \to K_i$ with $K_i = K'$. Hence $I_1 \to I_1 \to K_2 \to \cdots \to K_i$ is a path of size $i+1$ connecting I_1 to K'. Therefore $V_i \subset V_{i+1}$. Let U_i be the set of points contained in some $K' \in V_i$, i.e., $U_i = \cup\{K'\,;\ K' \in V_i\}$. We have that $U_i \subset U_{i+1}$. We claim that if there exists $K' \in V_i$ such that $f(\partial K') \notin U_i$ then $V_{i+1} \neq V_i$. Indeed, if $f(z) \notin U_i$ for some $z \in \partial K'$ then $f(K')$ contains a vertex (i.e., an interval) having $f(z)$ as a boundary point and this vertex is not contained in V_i. This proves the claim. Since the number of vertices is equal to $n-1$, there exists an integer $i \leq n-1$ such that $V_i = V_{i+1}$. From the above claim we conclude that $U_i \cap O_f(x)$ is invariant by f. Therefore $U_i = [x_0, x_n]$ and V_i is the set of all vertices. □

Lemma 1.4. *Suppose there is no vertex, distinct from I_1, that can be connected to I_1 by a path in the Markov graph. Then f maps the elements of the orbit of x that are on the left of the interior of I_1 into the elements that are on the right and vice-versa. Furthermore, in this case the period of x is even and f has a periodic point of period two.*

Proof. If $I_1 = [x_a, x_{a+1}]$ is as in Lemma 1.2 then $f(x_a) \geq x_{a+1}$ and $f(x_{a+1}) \leq x_a$. If there exists $x_i < x_a$ such that $f(x_i) \leq x_a$ then, taking $x_b = \max\{x_i < x_a\,;\ f(x_i) \leq x_a\}$, we have that $f(x_b) \leq x_a$ and $f(x_{b+1}) \geq x_{a+1}$. Hence $f([x_b, x_{b+1}]) \supset I_1$ which is a contradiction. Therefore for every $x_i < x_a$, we have that $f(x_i) \geq x_{a+1}$. Similarly, $f(x_i) \leq x_a$ for every $x_i \geq x_{a+1}$. Hence the period of x is even. Let $J_0 = [x_0, x_a]$ and $J_1 = [x_{a+1}, x_{n-1}]$. Thus $f(J_0) \supset J_1$ and

$f(J_1) \supset J_0$. Therefore there exists $z \in J_0$ such that $f^2(z) = z$ and $f(z) \in J_1$. Therefore z is a periodic point of period two. □

Lemma 1.5. *Assume f has periodic points with odd period. Let $n > 1$ be the smallest such period and x be a periodic point of period n. Then the corresponding Markov graph contains the following paths (see Figure 1.3): i) $I_1 \to I_2 \to \cdots \to I_{n-1} \to I_1 \to I_1$. ii) $I_{n-1} \to I_{2i+1}$ for every i such that $2i + 1 < n$. Furthermore, there is no edge of the type $I_j \to I_{j+k}$ if $k > 1$.*

Proof. By Lemmas 1.2 - 1.4, there exist vertices $I_1, \ldots, I_k$ such that $f(I_k) \supset I_1$ and such that there exists a path $I_1 \to I_2 \to \cdots \to I_k$. Let $k > 1$ be the smallest integer such that there exists a path $I_1 \to \cdots \to I_k \to I_1$. We claim that $k = n - 1$. Indeed, if $k < n - 1$ we conclude, by considering the path $I_1 \to \cdots \to I_k \to I_1$ (if k is odd) or the path $I_1 \to \cdots \to I_k \to I_1 \to I_1$ (if k is even), the existence of a periodic point of odd period smaller than n. This contradicts the hypothesis. By the minimality of $n - 1$, there is no edge of the type $I_j \to I_{j+k}$ if $k > 1$, because, otherwise, we would get a shorter closed path connecting I_1 to itself. Let $I_1 = [x_a, x_{a+1}]$. From Lemma 1.2, $f(x_a) \geq x_{a+1}$ and $f(x_{a+1}) \leq x_a$. Since x is not a periodic point of period two, we must have either $f(x_{a+1}) < x_a$ or $f(x_a) > x_{a+1}$. Suppose the first inequality holds (in the other case the argument is similar). Then $f(x_a) = x_{a+1}$ and $f(x_{a+1}) = x_{a-1}$ since, otherwise, $f(I_1)$ would contain not only I_1 and I_2 but also another vertex and this would contradict the property we have proved. Hence $I_2 = [x_{a-1}, x_a]$: so I_2 is the first interval of the partition which is to the left of I_1. Since $f(x_a) = x_{a+1}$ and $f(x_{a-1}) \geq x_{a+1}$, we must have $f(x_{a-1}) = x_{a+2}$ since otherwise $f(I_2)$ would contain more than one vertex, which again contradicts the above property. Thus $I_3 = [x_{a+1}, x_{a+2}]$ and therefore I_3 is the first interval of the partition which is on the right of I_1. Using this argument repeatedly, we get, by induction, that the intervals with even indices are to the left of I_1 whereas the intervals with odd indices are to the right and that these intervals are ordered in the interval in the following way: $I_{n-1}, \ldots, I_2, I_1, I_3, \ldots, I_{n-2}$. Since for $i = 1, \ldots, \frac{n-3}{2}$, $f(I_{2i})$ contains I_{2i+1} and no other vertex, $f(x_1) = x_{n-1}$ and therefore $f(I_{n-1}) \ni x_{n-1}$. On the other hand, since $f(I_{n-1})$ contains I_1, we get that $f(I_{n-1})$ contains $[x_a, x_{n-1}]$. Therefore $f(I_{n-1})$ contains every vertex which is to the right of I_1 which, as we saw, are the ones with odd index. □

Corollary 1.1. *If f has a periodic point of odd period n then f has periodic points of all periods larger than n as well as all even period smaller than n.*

Proof. If m is an integer larger than n, by looking at the path of length m: $I_1 \to \cdots \to I_1 \to \ldots I_{n-1} \to I_1$ we conclude as before, using Lemma 1.1, the existence of a periodic point of period m. If $m = 2i < n$, the path $I_{n-1} \to I_{n-2i} \to I_{n-2i+1} \to \cdots \to I_{n-1}$ gives a periodic point of period m. □

Lemma 1.6. *If f has a periodic point of even period then f has a periodic point of period two.*

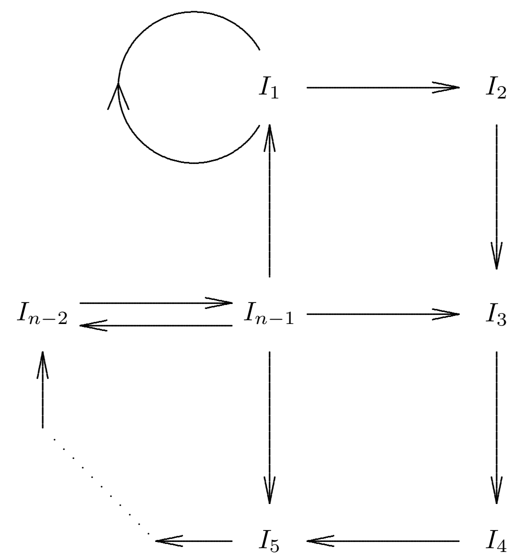

Fig. 1.3: The Markov graph associated to a periodic point of odd period.

Proof. Let $n \geq 2$ be the smallest integer such that f has a periodic point of period n. Suppose, by contradiction, that $n > 2$. By the corollary of Lemma 1.5, n is even since, otherwise, f would have a periodic point of period two. By Lemma 1.4, there exists a vertex I_k such that $f(I_k) \supset I_1$ because f does not have periodic points of period two. As before, let k be the smallest integer such that the Markov graph has a path $I_1 \to I_2 \to \cdots \to I_k \to I_1$. Since f does not have periodic points of period smaller than n, we have, as in the proof of Lemma 1.5, that $k = n-1$ and that there is no edge of the type $I_i \to I_{i+j}$ if $j > 1$. Using the same arguments as in the proof of Lemma 1.5, we conclude that there exist edges connecting I_{n-1} to all vertices with even indices. Hence the path $I_{n-1} \to I_{n-2} \to I_{n-1}$ gives, via Lemma 1.1, a periodic point of period two and the lemma is proved. $\square$

Proof of Theorem 1.1

1) Suppose that f has a periodic point of period $n = 2^k$. If $n \succ m$ then $m = 2^l$ with $l < k$. If $l = 0$ the result is obvious. If $l \neq 0$ then $g = f^{\frac{m}{2}}$ has a periodic point of period 2^{k-l+1}. Therefore, by Lemma 1.6, g has a periodic point of period two. This is a periodic point of period m for f, and the theorem is proved in this case.

2) Let $n = p2^k$ where p is an odd number and $k \geq 0$. If $n \succ m$ we have three cases to consider: a) $m = q2^k$, with $q > p$ odd; b) $m = q2^k$ with q even; c) $m = 2^l$ with $l \leq k$. In the cases a) and b), $g = f^{2^k}$ has a periodic point of period odd equal to p. Since $q > p$, in case a), or q is even in case b), the corollary of Lemma 1.5, gives a periodic point of period q for g. It is easy to see that this point is a periodic point of precisely period m for f. In case c), we get from case b) that f has a periodic point of period 2^{k+1}. Since $l \leq k$, it follows

from 1) that f has a periodic point of period m. This completes the proof of the theorem. □

Example. Figure 1.4 represents the graph of a map $f\colon [0,1] \to [0,1]$ with a periodic point of period five and the corresponding Markov graph. It is easy to see that the map we have drawn does not have a periodic point of period 3. In the same way, we can construct maps having a periodic point of period n but no periodic point whose period is larger than n with respect to the Sarkovskii ordering on $\mathbb{N}$.

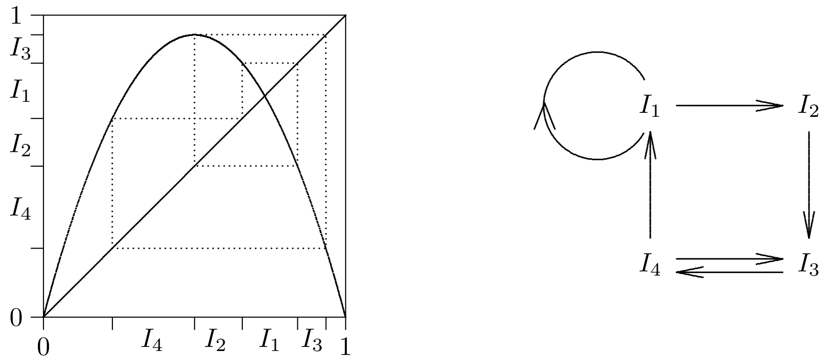

Fig. 1.4: A map $f\colon [0,1] \to [0,1]$ with a periodic point of period 5 and the associated Markov graph.

Exercise 1.1. Let $f\colon S^1 \to S^1$ be a continuous map with a periodic orbit of period 3. Show that if a lift $F\colon \mathbb{R} \to \mathbb{R}$ of f has also a periodic orbit of period 3, then f has periodic orbits of every period. Show that this last condition on the lift F cannot be dropped.

Exercise 1.2. Show that the map $Q_\mu\colon [0,1] \to [0,1]$ defined by $Q_\mu(x) = \mu x(1-x)$ has periodic orbits of each period for every μ sufficiently close to 4. (Hint: show that these maps have a periodic orbit of period 3.)

Exercise 1.3. Show that if a continuous map $f\colon [0,1] \to [0,1]$ has a periodic point p of period 4 such that $p < f(p) < f^2(p) < f^3(p)$, then f has periodic orbits of each period. (Hint: show for example that such a map has also a periodic point of period 3 using the ideas of this section.)

Exercise 1.4. Show that there exist parameters $\mu \in [0,4]$ such that $Q_\mu(x) = \mu x(1-x)$ has periodic orbits of periods 1, 2 and 4 and no other periodic orbits.

2 Covering Maps of the Circle as Dynamical Systems

In this section we will consider covering maps $f\colon S^1 \to S^1$ of degree d with $|d| \geq 2$. By this we mean that f is a surjective local homeomorphism such that the pre-image of each point consists of exactly $|d|$ points. If $d > 0$ then f is orientation preserving and if $d < 0$ it is orientation reversing. If we consider a lift $\hat{f}\colon \mathbb{R} \to \mathbb{R}$ of f to the universal covering space, then $\hat{f}$ is a homeomorphism such that $\hat{f}(x+1) = f(x) + d$ for all $x \in \mathbb{R}$. Conversely any homeomorphism of the real line with this property is a lift of some circle covering map. In particular, for each d the map $\hat{g}_d\colon \mathbb{R} \to \mathbb{R}$, defined by $\hat{g}_d(x) = d \cdot x$ is a lift of a covering map $g_d\colon S^1 \to S^1$. Our goal is to prove the following result of Shub (1969): any covering map of degree d is semi-conjugate to g_d. This result plays the same role in the dynamics of covering maps as the theorem of Poincaré in the dynamics of homeomorphisms as described in Section I.1.

The covering maps g_d satisfy the condition $|Dg_d(x)| = |d| > 1$. This is a special case of the following situation.

Definition. We say that a C^1 map $f\colon S^1 \to S^1$ is *expanding* if there exist constants $C > 0$ and $\lambda > 1$ such that

$$|Df^n(x)| > C\lambda^n$$

for all $n \in \mathbb{N}$ and all $x \in S^1$.

It is not hard to see that every expanding map $f\colon S^1 \to S^1$ is a covering map of degree d with $|d|$ bigger than one.

Theorem 2.1. **(Shub)** *Let $f\colon S^1 \to S^1$ be an expanding C^1 map of degree d. If $g\colon S^1 \to S^1$ is a covering map of degree d, then there exists a (not necessarily strictly) monotone and surjective map $h\colon S^1 \to S^1$ such that $h \circ g = f \circ h$.*

We should emphasize that the conjugacy is not unique in general: if we take $f(z) = 3z \mod 1$ and $h(z) = z + 0.5 \mod 1$ then $h \circ f = f \circ h$. Also notice that, in general, h is not a conjugacy. In fact, if x is a fixed point of f and $\phi\colon S^1 \to S^1$ is a diffeomorphism such that $\phi(x) = x$ and $|D\phi(x)| < |Df(x)|^{-1}$ then x is an attracting fixed point of $g = f \circ \phi$. (That x is an attracting fixed point means that the set $B_g(x) = \{y\,;\ g^n(y) \to x \text{ as } n \to \infty\}$ contains a neighbourhood of x.) Since f has no attracting fixed point, it follows that g is not conjugate to f: the semi-conjugacy of Theorem 2.1 maps $B_g(x)$ onto the full orbit of a fixed point of f. However, if g is also an expanding map then h is a conjugacy. In fact, assume by contradiction that the preimage under h of a point z is an interval I. If z is not a periodic point of f then the intervals $\{g^n(I), n \in \mathbb{N}\}$ are pairwise disjoint since $h(g^n(I)) = f^n(z)$ and if z is periodic then $g^k(I) = I$ where k is the period of z. But both these situations cannot occur because g is expanding. Therefore we get the following consequence of the above theorem.

Corollary 2.1. *Let $End^r(S^1)$, $r \geq 1$, be the space of the C^r endomorphisms of the circle endowed with the C^r topology. Then all expanding maps of $End^r(S^1)$ are structurally stable (i.e., any two nearby expanding maps are conjugate).*

Proof of Theorem 2.1: The idea of the proof is the following pullback argument. We start with a homeomorphism $h_0\colon S^1 \to S^1$; since both g and f have degree d, we can pullback h_0 to a homeomorphism h_1, i.e, $f \circ h_1 = h_0 \circ g$. By induction we construct a sequence h_n of homeomorphisms such that h_n is a pullback of h_{n-1}. Next we prove that the sequence converges to a monotone map h_∞ which is a semiconjugacy between f and g.

It is more convenient to work in the universal covering. So, let $\hat{f}\colon \mathbb{R} \to \mathbb{R}$ and $\hat{g}\colon \mathbb{R} \to \mathbb{R}$ be lifts of f and g. Then $\hat{f}$ and $\hat{g}$ are diffeomorphisms and since f and g are expanding there are constants $K < \infty$ and $\rho < 1$ such that

$$|D\hat{f}^{-n}(x)|, |D\hat{g}^{-n}(x)| \leq K \cdot \rho^n$$

for each $n \geq 0$ and each $x \in \mathbb{R}$. Let $\mathcal{E}$ be the space of continuous monotone maps $\phi\colon \mathbb{R} \to \mathbb{R}$ such that $\phi(x+1) = \phi(x) + 1$. In $\mathcal{E}$ we consider the uniform metric: $d(\phi_1, \phi_2) = \sup_x |\phi_1(x) - \phi_2(x)|$ (the supremum exists because $\phi - Id$ is periodic). Then $\mathcal{E}$ is a complete metric space and each function in $\mathcal{E}$ is a lift of a continuous map of the circle. For each $\phi \in \mathcal{E}$ let $T\phi = \hat{f}^{-1} \circ \phi \circ \hat{g}$. Clearly, $T\phi \in \mathcal{E}$. To prove the theorem we will show that T is a contraction, in the sense that

$$d(T^n\phi_1, T^n\phi_2) \leq K \cdot \rho^n \cdot d(\phi_1, \phi_2).$$

Indeed,

$$\begin{aligned} |\hat{f}^{-n}(\phi_1(\hat{g}^n(x))) - \hat{f}^{-n}(\phi_2(\hat{g}^n(x)))| &\leq K \cdot \rho^n \cdot |\phi_1(\hat{g}^n(x)) - \phi_2(\hat{g}^n(x)| \\ &\leq K \cdot \rho^n \cdot d(\phi_1, \phi_2). \end{aligned}$$

It follows that T has a unique fixed point $\hat{h}$ which means that $\hat{f} \circ \hat{h} = \hat{h} \circ \hat{g}$. Of course $\hat{h}$ is a lift of a monotone map h and $f \circ h = h \circ g$. □

Remarks. 1. The same proof as above gives a similar theorem for expanding maps of compact manifolds of higher dimension, see Shub (1969). 2. Let f and g be C^2 covering maps, f being expanding and g a local C^2 diffeomorphism. If h is the semi-conjugacy given by the theorem and $I = h^{-1}(z)$ is an interval then I is eventually periodic, namely, there are integers n and $p > 0$ such that $g^p(g^n(I)) = g^n(I)$. Indeed, either z is an eventually periodic point of f or it is not. In the first case $f^p(f^n(z)) = f^n(z)$ for some integers n and $p > 0$ and this implies $g^p(g^n(I)) = g^n(I)$ and therefore our assertion. In the second case all the points $z, f(z), \ldots$ are distinct and, since f is expanding, the sequence $f^k(z)$ cannot tend to a periodic orbit. It follows that the intervals $g^k(I) = h^{-1}(f^k(z))$, $k \in \mathbb{N}$ are pairwise disjoint and that $g^k(I)$ also does not tend to a periodic orbit. Hence I is a wandering interval of g. But, as we have proved in Corollary I of Theorem I.2.2, a C^2 covering map cannot have a wandering interval. It follows that the second alternative cannot occur. Thus the assertion is proved. If J is a periodic interval for g, i.e., g is the pre-image of a periodic point of f by h, then g^p maps J diffeomorphically onto itself. Hence every point in J is either periodic or asymptotic to a periodic orbit. Therefore the dynamics of g is obtained from the dynamics of f by blowing up the orbit (both positive and negative) of some periodic points and inserting intervals. In Chapter IV we will prove that the number of (maximal) periodic intervals is finite. 3. Later on, in Exercise

III.2.4, we will see that the structurally stable covering maps form an open and dense set in the space of all covering maps. 4. In Shub and Sullivan (1985)

it is proved that if the conjugacy h from Theorem 2.1 between two expanding C^2 maps is absolutely continuous (in both directions, i.e., for each measurable set A its image $h(A)$ has Lebesgue measure zero if and only if A has zero Lebesgue measure) then it is C^1. Clearly, if h is C^1 then $Dh \circ Df^n = Dg^n \circ Dh$ and therefore eigenvalues of corresponding periodic points of f and g coincide. Therefore in general such conjugacies are certainly not absolutely continuous.

Exercise 2.1. Prove that for an expanding map, the periodic points are dense; each backward orbit is dense and there exist points whose forward orbits are dense. From this and from the previous theorem it easily follows that if a C^2 covering map has a periodic interval then the closure of the complement of the full orbit of all these periodic intervals is an invariant Cantor set.

Exercise 2.2. Show that each expanding map of the circle has periodic orbits of each period. (Hint: use the Markov graphs from the proof of the theorem of Sarkovskii to construct the periodic orbits.)

Exercise 2.3. Show that the conjugacy h from Theorem 2.1 is quasisymmetric provided f and g are expanding $C^{1+\alpha}$ maps with $\alpha \in (0,1)$. This means that there exists $C < \infty$ such that $|\log Df(x) - \log Df(y)| \le C|x-y|^\alpha$ for all $x \neq y$ and similarly for g. (Together with the result of Shub and Sullivan mentioned above, this shows that a quasisymmetric homeomorphism is not necessarily absolutely continuous.) We note that this contrasts with the situation for circle diffeomorphisms see Exercise I.5.2. In Exercise III.6.2 a more general result is proved. Later on, in Chapter VI, we shall see why quasisymmetry is such a useful concept. (Hint: use the 'Naive Distortion Lemma', see Lemma I.2.1. More precisely, assume that there exists $C > 0$ and $\lambda > 1$ such that $|Df^n(x)|, |Dg^n(x)| \ge C\lambda^n$ for each $n > 0$ and let h be the conjugacy between f and g with $h \circ g = f \circ h$. Take two intervals I_1 and I_2 in S^1 with a common endpoint such that $|I_1| = |I_2|$. Since g is expanding, there exists $n > 0$ such that $g^{n+1}(I_1 \cup I_2) \supset S^1$. Let n be the smallest integer with this property. Then there exists $K_1 < \infty$ such that $\sum_{i=0}^{n-1} |g^i(I_1 \cup I_2)| < K_1$. It follows from Lemma 2.1 of Section I.2 that Dg^n has bounded distortion on $I_1 \cup I_2$, i.e., there exists $K < \infty$ such that $|Dg^n(x)|/|Dg^n(y)| < K$ for each $x, y \in I_1 \cup I_2$. Hence $\frac{|g^n(I_1)|}{|g^n(I_2)|} < K$. Furthermore, from the definition of n the length of each of these two intervals is of the order of $|S^1|$, because in one extra iterate these intervals overlap the circle. Therefore by the uniform continuity of h, there exists a universal constant K' such that $\frac{|h \circ g^n(I_1)|}{|h \circ g^n(I_2)|} < K'$. Similarly one has $|Df^n(x)|/|Df^n(y)| < K$ for each $x, y \in h(I_1 \cup I_2)$. This and $f^{-n} \circ h \circ g^n = h$ implies that $\frac{|h(I_1)|}{|h(I_2)|} \le K \cdot K'$ and therefore that h is quasisymmetric.)

3 The Kneading Theory and Combinatorial Equivalence

One of the main questions in the field of dynamical systems is whether two systems are 'the same'. Of course, there are many equivalence relations. In this section we shall introduce the one based on kneading invariants and the slightly stronger one of combinatorial equivalence. We will describe a machinery which can be used to solve this question for interval maps. This will be done through the combinatorial theory developed by Milnor and Thurston (1977) for piecewise monotone, continuous maps of the interval. Parts of this theory date back to Parry (1964) and Metropolis et al. (1973). This theory is the analogue of the Poincaré theory for homeomorphisms of the circle and states that the combinatorial type of such a map is completely determined by the orbits of its turning points. The main ingredient is the use of symbolic dynamics. In the previous chapter we associated to each circle homeomorphism a rotation number. Because the maps in this chapter are not order preserving anymore, the situation is more complicated now. Therefore we shall define some sequences of symbols and show that these sequences completely determine the combinatorial type of such maps. See also Parry (1964), Hofbauer (1981) and Hofbauer and Keller (1982) for the symbolic dynamics of maps that may have some discontinuities. In Section 3.a we shall give some examples and in Section 3.b we will describe a tool, 'the Hofbauer tower construction' which can be used to give a more graphical description of the orbits of the turning points.

Definition. Let I be the compact interval $[0,1]$ and $f\colon I \to I$ be a *piecewise monotone* continuous map. This means that f is continuous and that f has a finite number of turning points, i.e., points in the interior of $[0,1]$ where f has a local extremum. Such a map is called *l-modal* if f has precisely l turning points and if $f(\partial I) \subset \partial I$. More precisely, we assume that f has local extrema at $0 < c_1 < \cdots < c_l < 1$ and that f is strictly monotone in each of the $l+1$ intervals $I_1 = [0, c_1), I_2 = (c_1, c_2), \ldots, I_{l+1} = (c_l, 1]$. In particular, we say that f is *unimodal* if $f(\partial I) \subset \partial I$ and if f has precisely one turning point c. There is no loss of generality in the assumption that $f(\partial I) \subset \partial I$ since any endomorphism of a compact interval can be extended to a bigger interval so that the boundary of the larger interval is mapped into itself.

Often one says that two maps $f, g\colon I \to I$ define 'the same' dynamical system if they are identical up to coordinate change. This means that there is a homeomorphism $h\colon I \to I$ such that

$$\begin{array}{ccc} I & \xrightarrow{h} & I \\ {\scriptstyle f}\downarrow & & \downarrow{\scriptstyle g} \\ I & \xrightarrow{h} & I \end{array}$$

commutes. In this case $h(f^n(x)) = g^n(h(x))$ so h maps orbits of f onto orbits of g and we say that f and g are *topologically conjugate* or simply *conjugate*.

Often two interval maps share many dynamical properties even though they are not quite conjugate. We will show that this happens if two maps f and g 'fold' the interval in the same way. In Figure 3.1 it is shown that – in general – two different maps fold the interval in completely different ways.

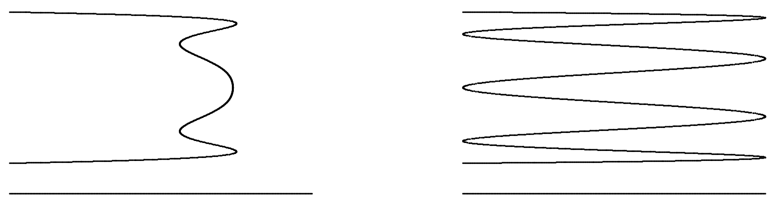

Fig. 3.1: A map $f\colon [0,1] \to [0,1]$ can be represented by a graph but also by a folding map. On the left the folding of $[0,1]$ under f_a^3 is shown for $a = 2$ and on the right for $a = 4$ where $f_a(x) = ax(1-x)$. The segment $[0,1]$ is embedded in the square $[0,1] \times [0,1]$ as shown and then projected into itself.

Definition. We say that two l-modal maps $f, g\colon I \to I$ with turning points $c_1 < \cdots < c_l$ respectively $\tilde{c}_1 < \cdots < \tilde{c}_l$ are *combinatorially equivalent* if there exists an order preserving bijection h such that

$$\begin{array}{ccc} \bigcup_{i=1}^{l} \bigcup_{n\in\mathbb{Z}} f^n(c_i) & \xrightarrow{\;h\;} & \bigcup_{i=1}^{l} \bigcup_{n\in\mathbb{Z}} g^n(\tilde{c}_i) \\ {\scriptstyle f}\big\downarrow & & \big\downarrow{\scriptstyle g} \\ \bigcup_{i=1}^{l} \bigcup_{n\in\mathbb{Z}} f^n(c_i) & \xrightarrow{\;h\;} & \bigcup_{i=1}^{l} \bigcup_{n\in\mathbb{Z}} g^n(\tilde{c}_i) \end{array} \tag{3.1}$$

commutes and $h(c_i) = \tilde{c}_i$ for $i = 1, \ldots, l$. The map h is called the *combinatorial equivalence* between f and g.

We introduce the notion of combinatorial equivalence because it is very close to the usual conjugacy relationship, as will become clear in Theorem 3.1 and its corollary. For those who are already familiar with kneading theory we should remark that two maps without periodic attractors and without wandering intervals are combinatorially equivalent if and only if they have the same kneading invariants (we shall return to this in Theorem 3.2 below). Moreover, if two maps are combinatorially equivalent then their kneading invariants are the same. Because the preimages of turning points are the places where iterates of the map are folded and forward images of turning points determine where these 'fold-points' are mapped, two maps f and g are combinatorially equivalent if and only if the interval I is folded by f^n and g^n in the same way for all $n \in \mathbb{N}$, see Figure 3.1. Below, in Theorem 3.1, we shall see that only the forward iterates of the turning points really matter. Let us analyze this definition in more detail and relate it to the notion of conjugacy. For this we need to introduce the concepts of 'homterval', 'basin' and 'wandering interval'.

Definition. Let us define a *homterval* to be an interval on which f^n is monotone for all $n \geq 0$.

What we want to explain now is that homtervals do not carry too much interesting information. Indeed, homtervals are related to wandering intervals and attracting periodic orbits. Let us first define these notions.

Definition. An interval J is *wandering* if all its iterates $J, f(J), f^2(J), \dots$ are disjoint and if $(f^n(J))_{n\geq 0}$ does not tend to a periodic orbit.

As we will show later, 'smooth' maps do not have wandering intervals. The reason for the second condition in the above definition, is that there are clearly many intervals whose forward orbit consists of disjoint intervals tending to a periodic orbit (one can show that in this case the limiting periodic orbit is attracting).

Definition. Let $O(p)$ be a periodic orbit. This orbit is called *attracting* if

$$B(p) = \{x\,;\ f^k(x) \to O(p) \text{ as } k \to \infty\}$$

contains an open set. The set $B(p)$ is called the *basin* of $O(p)$. The *immediate basin* $B_0(p)$ of $O(p)$ is the union of the components of $B(p)$ which contain points from $O(p)$. If $B_0(p)$ is a neighbourhood of $O(p)$ then this orbit is called a *two-sided attractor* and otherwise a *one-sided attractor*. For later use we will denote by $B(f)$ the union of the basins of periodic attracting orbits and by $B_0(f)$ the union of the immediate basins of periodic attractors.

Lemma 3.1. *Let J be a homterval of $f\colon I \to I$. Then there are two possibilities:*

a) J is a wandering interval;

b) every point in J is contained in the basin of a periodic orbit: some iterate of J is mapped into an interval L such that f^p maps L monotonically into itself for some $p \in \mathbb{N}$.

Proof. Suppose that not all the intervals $J, f(J), \dots$ are disjoint. Then there exist integers $n \geq 0$, $p > 0$ such that the interiors of $f^n(J)$ and $f^{n+p}(J)$ have a non-empty intersection. Hence, for all $k \geq 0$, $f^{n+kp}(J)$ and $f^{n+(k+1)p}(J)$ also intersect and so the closure L of $\cup_{k\geq 0} f^{n+kp}(J)$ is an interval and f^p maps L homeomorphically into itself. So points of J are eventually mapped into fixed points of $f^p|L$ or iterate to some fixed point of $f^p|L$. □

Remarks. 1. In Chapter IV we shall show that maps f which satisfy some mild smoothness conditions do not have wandering intervals. 2. If J is a connected component of the complement of $\bigcup_{i=1}^{l} \bigcup_{n\in\mathbb{Z}} f^n(c_i)$ then there exists no $n \in \mathbb{N}$ for which $f^n(J)$ contains a turning point and therefore f^n is monotone on J for all $n \in \mathbb{N}$. So a combinatorial equivalence is defined on the complement of homtervals. In particular, a combinatorial equivalence defines an order preserving surjective 'map' $h\colon I \to I$ with the property that the image of a point is either again a unique point or a single closed homterval and such that $h\circ g = f\circ h$. So h may collapse some intervals to points or blow-up some points to intervals. If f and g have no homtervals then a combinatorial equivalence h between f and g is in fact a conjugacy between f and g. 3. Note that the basin of a periodic point p of period n is completely invariant, i.e., $f(B(p)) = B(p) = f^{-1}(B(p))$ but that $f(B_0(p)) \subset B_0(p)$ because $B_0(f)$ can contain turning points of f. We also claim that $B(p) = \cup_{n\geq 0} f^{-n}(B_0(p))$. Indeed, if $x \in B(p)$ then $f^{kn}(x)$ tends to a point in $O(p)$ as $k \to \infty$, say p. Hence, either $f^{kn}(x) = p \in B_0(p)$ for k sufficiently big or there exists a (possibly one-sided) neighbourhood U of p such that $f^n(U) \subset U$, $\cap_{k\geq 0} f^{kn}(U) = \{p\}$ and such that $f^{kn}(x) \in U$ for k large. Since this implies that $U \subset B_0(p)$, the claim holds. Note that a l-modal map f is not constant on any interval and therefore the image of a non-trivial interval is also a non-trivial interval. Because for any attracting periodic point p its basin $B(p)$ contains an interval, the immediate basin $B_0(p)$ is a finite union of intervals and f maps $B_0(p)$ into itself. 4. Of course, a monotone map f^n from an interval L into itself can have many fixed points and consequently not all these fixed points need to be attracting. However, every non-periodic point in L is in the basin of an attracting fixed or an attracting periodic orbit of period two of $f^n\colon L \to L$. (This last possibility can only occur when $f^n\colon L \to L$ is orientation reversing.) Moreover, for every $x \in L$ we have $\omega(x) \subset B(p)$ for some (not necessarily attracting) fixed point p of $f^n\colon L \to L$.

How do we know when two maps are combinatorially equivalent? If we look at the definition it would be necessary to determine all images and preimages of c. But this is usually an infinite set. However, the turning points of a map could all be eventually periodic. So it would be much nicer if we could concentrate our attention on forward iterates of the turning points. The next theorem tells us that we are permitted to do so. It also tells us that frequently a combinatorial equivalence between two maps can be extended to a conjugacy.

Theorem 3.1. *Suppose that $f, g\colon I \to I$ are two l-modal maps with turning points $c_1 < \cdots < c_l$ respectively $\tilde{c}_1 < \cdots < \tilde{c}_l$. Assume that the map*

$$h\colon \bigcup_{i=1}^{l} \bigcup_{n\geq 0} f^n(c_i) \to \bigcup_{i=1}^{l} \bigcup_{n\geq 0} g^n(\tilde{c}_i) \tag{3.2}$$

defined by $h(f^n(c_i)) = g^n(\tilde{c}_i)$ is an order preserving bijection. Then the following properties are satisfied.

1. *The maps f and g are combinatorially equivalent.*
2. *If $PT(f)$ denotes the set of periodic turning points of f then the 'conjugacy' h maps $\bigcup_{i=1}^{l} \bigcup_{n\geq 0} f^n(c_i) \bigcap (B_0(f) \cup PT(f))$ into the corresponding set for g.*

3. *Assume that i) f and g have no wandering intervals, ii) there are no intervals consisting of periodic points of constant period, iii) each periodic turning point is attracting and iv) the restriction of the map h from (3.2) to $B_0(f)$, i.e.,*

$$h\colon \bigcup_{i=1}^{l}\bigcup_{n\geq 0} f^n(c_i)\cap B_0(f) \to \bigcup_{i=1}^{l}\bigcup_{n\geq 0} g^n(\tilde{c}_i)\cap B_0(g)$$

extends to a conjugacy from $B_0(f)$ to $B_0(g)$ (here we use assumption iii) to make sure that $B_0(f)\cup PT(f) = B_0(f)$). Then h can be extended to a conjugacy on I.

The assumptions ii) and iii) above are not superfluous: the turning point of each the unimodal maps $Q(x) = 2x(1-x)$, $f(x) = -|1/2-x| + 1/2$ and $g(x) = -\sqrt{1/2}\sqrt{|1/2-x|} + 1/2$ is a fixed point. Therefore they are all combinatorially equivalent and also a map h as in the assumption of the third part of Theorem 3.1 exists trivially. Even so, they are not conjugate because the turning point is respectively an attracting, neutral and repelling fixed point. We should emphasize that any periodic turning point of a C^1 map is necessarily attracting and therefore in this condition iii) can be dispensed with.

We should also remark that we can call an interval consisting of periodic points of constant period an *interval of periodic points* because of the following

Proof. Claim if each point in an interval J is periodic then each point in J must be a fixed point of some iterate f^p of the map. Indeed, let J be a component of the set $\mathrm{Per}(f)$ of periodic points of f. If $x\in J$ has period k then $f^k(J)\cap J\neq\emptyset$. Therefore $f^k(J)\cup J$ is an interval which is contained in $\mathrm{Per}(f)$ and since J is a component of $\mathrm{Per}(f)$ it follows that $f^k(J) = J$. Let us now show that $f^k|J$ is injective. If $x, y\in J$ and $f^k(x) = f^k(y)$ then both $f^k(x)$ and $f^k(y)$ have the same period. Since $x, y\in\mathrm{Per}(f)$ this implies that x and y have the same period $n\in\mathbb{N}$. Hence $x = f^{n-k}(f^k(x)) = f^{n-k}(f^k(y)) = y$. It follows that $f^k\colon J\to J$ is a homeomorphism and therefore that each point of J has either period k or period $2k$.

Before proving this theorem, let us state the following:

Corollary 3.1. *Let f and g be as in the previous theorem and assume that*

1. *the map h from (3.2) is an order preserving bijection;*
2. *the basin of each periodic attractor of f and g contains a turning point and each periodic turning point is an attractor;*
3. *the immediate basins of two periodic attractors have no boundary point in common;*

Then there exists a one-to-one correspondence between periodic attractors of f and g. Moreover, if a periodic attractor of f is one-sided if and only the same holds for the corresponding periodic attractor of g, then the monotone bijection h from (3.2) can be extended to a conjugacy between $f|B_0(f)$ and $g|B_0(g)$. In particular, if f and g have no wandering intervals and have no intervals consisting of periodic points of constant period, then f and g are conjugate.

Proof. Conditions 1, 2 and 3 imply that f and g have the same number of periodic attractors and that each of these attractors can be 'detected' by the orbit of at least one of the turning points. More precisely, let $O(p)$ be a periodic attractor of period n. If f^n is orientation preserving near p then we can take a small open neighbourhood $U(p)$ of p such that f^n sends $U(p)$ orientation preservingly into $U(p)$. Then the set $V(p) = U(p)\setminus f^n(U(p))$ consists of two half-open intervals. If f^n is orientation reversing near p or if f^n has a local extremum at p then take $U(p)$ to be a one-sided neighbourhood of f such that $f^{2n}(U(p)) \subset U(p)$ and let $V(p) = U(p) \setminus f^n(U(p))$. The intervals $V(p)$ form what is called a *fundamental domain* of $O(p)$. Of course one can choose the fundamental neighbourhood $V(p)$ of $O(p)$ so that the forward orbit of each turning point which intersects $B_0(p)\setminus O(p)$ also intersects this neighbourhood. By assumption $O(p)$ attracts at least one turning point c_i and since the map h from (3.2) is order preserving, it follows that $\tilde{c}_i$ is also in the basin of a periodic attractor $O(\tilde{p})$. Moreover, it follows from assumption 3) that this correspondence between periodic attractors of f and g is one-to-one. Therefore we can extend h to a homeomorphism between corresponding fundamental neighbourhoods $V(p)$ and $V(\tilde{p})$. Using $h \circ f = g \circ h$, we get the required extension to a conjugacy between the restriction of f to $B_0(f)$ and the restriction of g to $B_0(g)$. □

Remark. 1. As we shall see in Section 6, the previous result implies that there exists a very large class of smooth piecewise monotone maps S such that $f, g \in S$ are conjugate if and only if they are combinatorially equivalent. 2.

Part of Theorem 3.1 can also be proved using the itineraries defined below, but we prefer to use another technique, the *pullback construction*. One reason we will use this method is because it will play an important role throughout the remainder of this book. Another reason is that the itineraries defined below only give a way to construct conjugacies outside basins of periodic attractors. To extend these one has to use pullback construction. Therefore we prefer to use the pullback construction from the start.

The previous result is the analogue of the theorem of Poincaré for circle homeomorphisms without periodic points. Indeed, consider two l-modal maps f and g without periodic attractors and without wandering intervals. The previous result shows that f and g are conjugate if and only if there exists an order preserving map h from the forward orbits of the turning points of f to the forward orbits of the turning points of g such that $h(c_i) = \tilde{c}_i$ and such that $h \circ f = g \circ f$ on this set. This is the analogue of Poincaré's Theorem I.1.1.

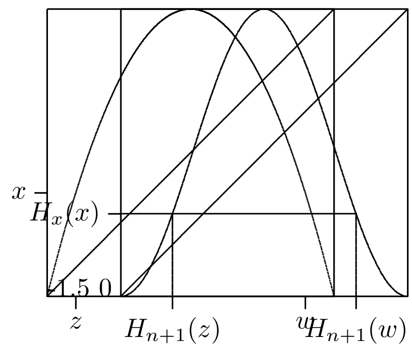

Fig. 3.2: The pulling back construction defines H_{n+1} from H_n by pulling back via f and g.

Proof of Theorem 3.1: Let $I = [0,1]$ and assume that h is as in (3.2). If h extends as in 2) then one can extend h to a homeomorphism $H_0\colon [0,1] \to [0,1]$ in such a way that the restriction of H_0 is a conjugacy between $f\colon B_0(f) \to B_0(f)$ and $g\colon B_0(g) \to B_0(g)$. Now we are ready to define a sequence of maps H_n as follows. Since H_0 maps the f-image of the i-th turning point of f onto the g-image of the i-th turning point of g, consequently g and $H_0 \circ f$ have the same local extremal values and the same number of turning points. Hence there exists a unique homeomorphism $H_1 : [0,1] \to [0,1]$ with

$$g \circ H_1 = H_0 \circ f$$

and $H_1(0) = 0$, $H_1(1) = 1$, see Figure 3.2. Note that H_1 agrees with h on the forward iterates of c_i: for each $n \geq 0$,

$$g \circ H_1(f^n(c_i)) = H_0 \circ f^{n+1}(c_i) = h(f^{n+1}(c_i)) = g^{n+1}(\tilde{c}_i)$$

and by construction $H_1(f^n(c_i))$ and $g^n(\tilde{c}_i)$ lie in the same interval of monotonicity. Hence

$$H_1(f^n(c_i)) = g^n(\tilde{c}_i).$$

We should note that H_1 is a bijective and order preserving map from $f^{-1}(c_i)$ to $g^{-1}(\tilde{c}_i)$. Similarly, one defines inductively H_{n+1} by

$$g \circ H_{n+1} = H_n \circ f.$$

It follows that

$$H_n\colon \bigcup_{k \geq -n} f^k(C(f) \cup B_0(f)) \;\to\; \bigcup_{k \geq -n} g^k(C(g) \cup B_0(g))$$

is an order preserving bijection where $C(f) = \{c_1, \ldots, c_l\}$ and $C(g) = \{\tilde{c}_1, \ldots, \tilde{c}_l\}$. Moreover, H_{n+1} coincides with H_n on this set, and we also have

$$g \circ H_n = H_n \circ f.$$

In particular there exists a unique limit

$$H_* : \bigcup_{n\in\mathbb{Z}} f^n(C(f)\cup B_0(f)) \quad\to\quad \bigcup_{n\in\mathbb{Z}} g^n(C(g)\cup B_0(g))$$

such that $H_*\circ f = g\circ H_*$ on this set and $H_*(c_i)=\tilde{c}_i$. We should emphasize that the maps H_n do not need to have a well defined limit on all of $[0,1]$, see Example 3.1.

If g has no wandering intervals and no intervals consisting of periodic points of the same period, then Lemma 3.1 implies that the set of preimages of turning points and of points which are eventually mapped into $B_0(g)$ is dense in $[0,1]$. It follows that H_* can be extended in a unique way to a monotone and surjective map $H_*\colon [0,1]\to[0,1]$ and so f and g are semi-conjugate. If the same properties hold for f then H_* uniquely extends to a homeomorphism. □

So far we have seen that the combinatorial type of a multimodal map is determined completely by the forward orbits of its turning points and that the notions of combinatorial equivalence and of conjugacy are closely related. In the remainder of this section we will show that there is a very convenient way to describe the orbits of these turning points by using symbolic dynamics. As we shall see it is possible to 'characterize' the dynamics of a map almost completely by a string of symbols. To do this, let us denote by $\mathcal{S}$ the symbols space consisting of the symbols $I_1,\dots,I_{l+1}$ and $c_1,\dots,c_l$ and by $\Sigma=\mathcal{S}^{\mathbb{N}}$ the space of sequences $\underline{x}\colon\mathbb{N}\to\mathcal{S}$, $\underline{x}=(x_0,x_1,\dots)$, where $x_i=\underline{x}(i)$. In Σ we consider the topology defined by the metric $d(\underline{x},\underline{y})=\sum_{i=0}^{\infty}\frac{1}{2^i}d(x_i,y_i)$ where $d(x_i,y_i)=1$ if $x_i\neq y_i$ and $d(x_i,x_i)=0$. With this topology Σ is a compact metric space and the shift transformation $\sigma\colon\Sigma\to\Sigma$,

$$\sigma(x_0,x_1,x_2,\dots)=(x_1,x_2,\dots)$$

is continuous.

Let $\underline{i}\colon I\to\Sigma$ be defined by $\underline{i}(x)=(i_0(x),i_1(x),\dots,i_n(x),\dots)$ where $i_n(x)=I_k$ if $f^n(x)\in I_k$ and $i_n(x)=c_k$ if $f^n(x)=c_k$. The sequence $\underline{i}(x)$ is called the *itinerary* of x under f. The map $\underline{i}$ relates the dynamics of f with the dynamics of the shift transformation: the diagram

$$\begin{array}{ccc} I & \xrightarrow{\ f\ } & I \\ {\scriptstyle \underline{i}}\downarrow & & \downarrow{\scriptstyle \underline{i}} \\ \Sigma & \xrightarrow{\ \sigma\ } & \Sigma \end{array}$$

commutes. Notice that the map $\underline{i}$ is not continuous exactly at the backward iterates of the turning points. But given $n\in\mathbb{N}$ and $x\in I$, there exists $\delta>0$ such that $i_n(y)$ is constant on the interval $(x,x+\delta)$ (but this value is not the same as $i_n(x)$ if $f^n(x)$ is a turning point). It follows that the one-sided limits $\underline{i}(x^+)=\lim_{y\downarrow x}\underline{i}(y)$ and $\underline{i}(x^-)=\lim_{y\uparrow x}\underline{i}(y)$ always exist. Notice that both $\underline{i}(x^+)$ and $\underline{i}(x^-)$ belong to the closed, σ-invariant subspace $\Sigma_0\subset\Sigma$ of sequences with elements in $\mathcal{S}_0=\{I_1,\dots,I_{l+1}\}$. If x is not in the backward orbit of some turning point then $\underline{i}$ is continuous at x and $\underline{i}(x) = \underline{i}(x^+) = \underline{i}(x^-)$. The sequences $\nu_1, \dots, \nu_l$ defined by

$$\nu_i = \underline{i}(c_i^+)$$

are called the *kneading invariants* of f. For simplicity let

$$\nu_0 = \underline{i}(0) \text{ and } \nu_{l+1} = \underline{i}(1).$$

We will see below that they play a role similar to the rotation number. In Section II.8, it will turn out to be useful to rewrite this kneading invariant as a formal power series.

Remark. 1. If f and g are combinatorially equivalent then they have the same kneading invariants. As we will show below, if f and g have the same kneading invariants and both f and g have no periodic attractors, no wandering intervals and no intervals of periodic points then they are also combinatorially equivalent. 2. Even if f and g have the same kneading invariants, they need not be combinatorially equivalent. Indeed, the kneading invariants give no information about the positions of the iterates of the turning points which are contained in homtervals. A very simple example illustrating this is given by the maps $f(x) = 2x(1-x)$ and $g(x) = ax(1-x)$ where $0 < a < 2$. Both these maps have kneading invariant $(I_2, I_1, I_1, \dots)$. However, the turning point of f is a fixed point and the turning point of g is attracted to an attracting fixed point. Moreover, in the multimodal case, iterates of several turning points can land in a fundamental domain inside the basin of a periodic attractor. The kneading invariant gives no information about the relative position of these iterates. In particular, if f and g have the same kneading invariants and they have the same number of periodic attractors with the same orientation, then they are still not necessarily combinatorially equivalent. Similarly, conditions 2 and 3 in the Corollary of Theorem 3.1 are needed to make sure that the dynamics inside the basins of attractors is completely determined by the orbits of the turning points. 3. To any l-modal map $f\colon I \to I$ we can associate a continuous map $\tilde{f}\colon I \to I$ with modality $\leq l$ and without homtervals by simply collapsing all homtervals of f. If f has no homtervals which coincide with an entire interval of monotonicity I_i then the modality of $\tilde{f}$ is again l and it is easy to see that f and $\tilde{f}$ have the same kneading invariants. It follows in particular that kneading invariants give little information on the dynamics within the basins of periodic attractors and also do not detect the presence of wandering intervals which are strictly contained inside a homterval. (If slightly different kneading invariants were used then wandering intervals could sometimes be detected by these invariants, see the next remark.) 4. The kneading invariants we defined above were used in Milnor and Thurston's (1977) paper. Notice that some authors use slightly different kneading invariants. Indeed, if the forward orbit of a turning point c_k does not contain a turning point then $\sigma\nu_k$ coincides with the itinerary of $f(c_k)$. In this case $\sigma\nu_k$ determines ν_k. Therefore, sometimes the alternative invariants $\underline{i}(f(c_i))$ are used instead of those introduced above with limits. For most questions they can be used equally well. However, there is an important difference: the alternative invariant – unlike 'ours' – can detect certain wandering intervals. In particular, the alternative kneading invariants of a map f and its associated map $\tilde{f}$ (as defined in Remark 3 above) are not always equal. This observation was first made in MacKay and Tresser (1988). Indeed, let us illustrate this by considering a bimodal map $f\colon I \to I$ as drawn in Figure 3.3 with turning points c_1 and c_2. It is possible to construct this map so that $f(c_1)$ is in the interior of $I_3 = (c_2, 1]$ and the interval connecting $f(c_1)$ and c_2 is a wandering homterval J. Indeed, first choose a bimodal map $\hat{f}$ such that $\hat{f}(c_1) = c_2$ is not eventually periodic. By 'blowing-up' each of the points of this orbit to a small interval one can modify $\hat{f}$ so that a neighbourhood $\hat{J}$ of c_2 is a wandering interval. Because $\hat{J}$ is wandering, the forward orbit of $\hat{J}$ does not contain c_1 and therefore we can modify $\hat{f}$ near c_1 such that $\hat{J}$ remains a wandering interval. In this way we get our required map. Clearly $\underline{i}_f(f(c_1)) = I_3 \cdot \underline{i}_f(f(c_2))$ and it is not hard to see that the symbols appearing in $\underline{i}_f(f(c_2))$ are not eventually periodic. (This will be explained below Proposition 3.1.) Now any map g with the same 'alternative' kneading invariants as f has the property

$$\underline{i}_g(g(c_1)) = \underline{i}_f(f(c_1)) = I_3 \cdot \underline{i}_f(f(c_2)) = I_3 \cdot \underline{i}_g(g(c_2))$$

and $\underline{i}_g(g(c_2))$ is not eventually periodic. This equality implies that $g(c_1) \neq c_2$ and that each point in the interval $\tilde{J}$ connecting $g(c_1)$ and c_2 has the same itinerary and therefore $\tilde{J}$ is a homterval. Since the symbols in $\underline{i}_g(g(c_2))$ are not eventually periodic $\tilde{J}$ does not tend to a periodic orbit and therefore is a wandering interval. It follows that any map g with the same 'alternative' kneading invariants as f has a wandering interval. As we remarked above, the kneading invariants introduced by Milnor and Thurston – and which we adopted here – do not 'feel' these types of wandering intervals.

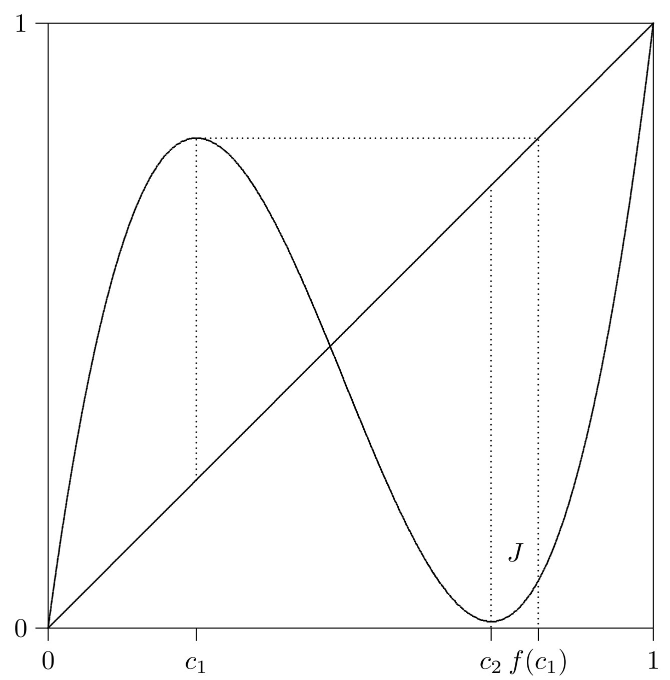

Fig. 3.3: Any bimodal map f such that $\underline{i}_f(f(c_1)) = \underline{i}_f(c_2^+)$, $f(c_1) \neq c_2$ and such that $(c_2, f(c_1))$ is not in the basin of a periodic attractor, has a wandering interval.

Next we introduce an order structure on the space

$$\Sigma = \{I_1, I_2, \ldots, I_{l+1}, c_1, \ldots, c_l\}^{\mathbb{N}}$$

such that the itinerary map $\underline{i}\colon I \to \Sigma$ is order preserving.

Definition. A *signed lexicographical ordering* $\prec$ on Σ is defined follows. Associate to each interval I_j a sign $\epsilon(I_j) = \pm 1$ such that $(-1)^j \epsilon(I_j)$ has a constant sign and define $\epsilon(c_j) = 0$. (Given a l-modal map f we shall always assume that $\epsilon(I_j) = 1$ if and only if f is orientation preserving on I_j.) Then we say that $\underline{x} \prec \underline{y}$ if there exists $n \geq 0$ such that $x_i = y_i$ for $i = 0, \ldots, n-1$ and

$$\left(\prod_{j \in J} \epsilon(x_j)\right) x_n < \left(\prod_{j \in J} \epsilon(y_j)\right) y_n.$$

Here $J = \{j\,;\ 0 \leq j \leq n-1 \text{ and } \epsilon(x_j) \neq 0\}$ if $n \geq 1$ and $J = \emptyset$ if $n = 0$. Moreover, the ordering "$<$" on the symbols $\pm I_k$ and $\pm c_k$ is taken as in the interval: $-I_{l+1} < -c_l < \cdots < -c_1 < -I_1 < 0 < I_1 < c_1 < \cdots < c_l < I_{l+1}$. We also write $\underline{x} \preceq \underline{y}$ if $\underline{x} \prec \underline{y}$ or $\underline{x} = \underline{y}$. (The reason for excluding in J the integers j with $\epsilon(x_j) = 0$ is that $\prec$ would otherwise not define an ordering on Σ. However, the subset $\underline{i}(I)$ of Σ has an additional property: if the first n symbols of $\underline{x}, \underline{y} \in \underline{i}(I)$ coincide and one of these is in $\{c_1, \ldots, c_l\}$ then $\underline{x} = \underline{y}$. So to define the ordering on this subspace $\underline{i}(I)$ one could take equally well $J = \{0, \ldots, n-1\}$.)

This ordering is motivated by the following

Proposition 3.1. *Let $\preceq$ be the lexicographical ordering corresponding to a l-modal map f and let $\underline{i}(x)$ be the itinerary of x under f.*

a) If $x < y$ then $\underline{i}(x) \preceq \underline{i}(y)$.

b) If $\underline{i}(x) \prec \underline{i}(y)$ then $x < y$.

Proof. a) Suppose $x < y$ and $\underline{i}(x) \neq \underline{i}(y)$ and let $n \geq 0$ be such that $i_j(x) = i_j(y)$ for $j = 0, \ldots, n-1$ and $i_n(x) \neq i_n(y)$. Hence there are no turning points in the intervals $[x, y], f([x, y]), \ldots, f^{n-1}([x, y])$ but there is a turning point $c_k \in f^n([x, y])$. Therefore f^n is monotone on the interval $[x, y]$, and is increasing if the product of the signs of the intervals $i_0(x), \ldots, i_{n-1}(x)$ is positive and is otherwise decreasing. If this product is positive then $f^n(x) < f^n(y)$ and, consequently, $i_n(x) < i_n(y)$. Since the product is positive, this implies that $\underline{i}(x) < \underline{i}(y)$. Similarly, if the above product is negative, we have that $f^n(x) > f^n(y)$ which implies again that $\underline{i}(x) < \underline{i}(y)$. This proves a). The proof of b) follows by reversing the argument. □

Let $I(x) = \{y\,;\ \underline{i}(y) = \underline{i}(x)\}$. Proposition 3.1 implies that $I(x)$ is connected. If it is a non-trivial interval then it is a homterval (and no larger interval has this property). In particular, by Lemma 3.1 either each point in this interval is contained in the basin of a periodic orbit or this interval is wandering. If $f^n(x) \in I(x)$ (and $n > 0$ is minimal with this property) then f^n maps $I(x)$ monotonically into itself and so x is either periodic or contained in the basin of a periodic orbit of period n or $2n$. In particular, the kneading invariant ν_i is eventually periodic if and only if the ω-limit of c_i is a periodic orbit.

Next we will show that if f and g have the same kneading invariants, and if they have no periodic attractors and no wandering intervals then they are conjugate. In order to state this result let $\nu_1, \dots, \nu_l$ be the kneading invariants of some l-modal map and as before let $\nu_0 = \underline{i}(0)$ and $\nu_{l+1} = \underline{i}(1)$. Note that ν_0 and ν_{l+1} are determined by the signs of $\epsilon(I_1)$ and $\epsilon(I_{l+1})$ (because $f(\partial I) \subset \partial I$). For example $\nu_0 = I_1^\infty$ if $\epsilon(I_1) = 1$. Next define $\Sigma(\nu_0, \nu_1, \dots, \nu_{l+1})$ to be the sequences $\underline{\alpha} \in \Sigma$ which satisfy the following conditions for each $n \geq 0$ and each $k = 0, 1, \dots, l$:

$$\begin{aligned} &\sigma^n \underline{\alpha} = \underline{i}(c_k) && \text{if } \alpha_n = c_k, \\ &\sigma\nu_k \prec \sigma^{n+1}\underline{\alpha} \prec \sigma\nu_{k+1} && \text{if } \alpha_n = I_{k+1} \text{ and } \epsilon(I_{k+1}) > 0, \\ &\sigma\nu_{k+1} \prec \sigma^{n+1}\underline{\alpha} \prec \sigma\nu_k && \text{if } \alpha_n = I_{k+1} \text{ and } \epsilon(I_{k+1}) < 0. \end{aligned} \tag{3.3}$$

The slightly larger set $\hat{\Sigma}(\nu_0, \dots, \nu_{l+1})$ is defined by (3.3) but with $\prec$ replaced by $\preceq$. The next result tells us that the dynamics of a map f can be 'computed' on a symbolic level. It tells us that it suffices to work with kneading invariants (rather than with the notion of combinatorial equivalence) if we restrict ourselves to maps without periodic attractors and without wandering intervals or if we are not interested in the dynamics on the basins of periodic attractors. It also gives a way to prove the existence of certain orbits of f by constructing them on a symbolic level: for each sequence $\underline{\alpha} \in \Sigma$ as in (3.3) there exists a point which has this sequence as its itinerary. Therefore (3.3) is sometimes referred to as an *admissibility condition*. Jonker (1981) used this method to prove Sarkovskii's Theorem in the unimodal case, see also Collet and Eckmann (1980).

Theorem 3.2. *Let $f \colon I \to I$ be a l-modal map. If f has kneading invariants $\nu_1, \dots, \nu_l \in \Sigma_0$ and ν_0, ν_{l+1} are the kneading invariants in the boundary points of I, then for each $j = 0, \dots, l$, $k = 0, \dots, l$ and $n = 0, 1, 2, \dots$ one has $\nu_j = (I_{j+1}, \dots)$ and if $\sigma^n(\nu_j) = (I_{k+1}, \dots)$ then*

$$\begin{aligned} &\sigma\nu_k \preceq \sigma^{n+1}\nu_j \preceq \sigma\nu_{k+1} && \text{if } \epsilon(I_{k+1}) > 0, \\ &\sigma\nu_{k+1} \preceq \sigma^{n+1}\nu_j \preceq \sigma\nu_k && \text{if } \epsilon(I_{k+1}) < 0. \end{aligned} \tag{3.4}$$

$\underline{i}_f$ maps I into $\hat{\Sigma}(\nu_0, \dots, \nu_{l+1})$ and $\underline{i}_f(I) \supset \Sigma(\nu_0, \dots, \nu_{l+1})$. If f has no wandering intervals then $\underline{i}_f(I) = \Sigma(\nu_0, \dots, \nu_{l+1})$ and if f also has no periodic attractors then one of the inequalities in (3.4) is strict. Furthermore, $\underline{i}_f$ is only constant on wandering intervals, on intervals of periodic points and on intervals which are in the basin of some periodic attractor; therefore $\underline{i}_f \colon I \to \Sigma(\nu_0, \dots, \nu_{l+1})$ is an order preserving bijection if f has no wandering intervals and no periodic attractors. Finally, if f and g have the same kneading invariants, no wandering

intervals, no intervals of periodic points and no periodic attractors then they are conjugate.

Proof. Let us first show that $\underline{i}_f$ maps I into $\Sigma(\nu_0, \dots, \nu_{l+1})$. In fact, if $f^n(x) \in I_{k+1} = (c_k, c_{k+1})$ then $f(c_k) < f^{n+1}(x) < f(c_{k+1})$ if $\epsilon(I_{k+1}) > 0$ and the opposite inequality holds if the sign of I_{k+1} is negative. Therefore $\sigma\nu_k \preceq \underline{i}(f^{n+1}(x)) \preceq \sigma\nu_{k+1}$ if $\epsilon(I_{k+1}) > 0$ and $\sigma\nu_{k+1} \preceq \underline{i}(f^{n+1}(x)) \preceq \sigma\nu_k$ if $\epsilon(I_{k+1}) < 0$. Since $\underline{i}(f^{n+1}(c_j^+)) = \sigma^{n+1}\underline{i}(c_j^+) = \sigma^{n+1}\nu_j$, letting x tend to c_j from the right one gets (3.4). If f has no wandering intervals, no intervals of periodic points and no periodic attractors, $\underline{i}_f$ is injective and consequently one of the inequalities in (3.4) has to be strict. So let us show that $\underline{i}_f(I) \supset \Sigma(\nu_0, \dots, \nu_{l+1})$. So take $\underline{\alpha} \in \Sigma(\nu_0, \dots, \nu_{l+1})$ and suppose, by contradiction, that there is no $x \in I$ such that $\underline{i}_f(x) = \underline{\alpha}$. Then $I = A \cup B$ where $A = \{x \in I\,;\, \underline{i}(x) \prec \underline{\alpha}\}$ and $B = \{x \in I\,;\, \underline{\alpha} \prec \underline{i}(x)\}$. By Proposition 3.1, A and B are intervals. Let $a = \sup A$ and $b = \inf B$. If $a \notin A$ and $b \notin B$ we get a contradiction because $0 \in A$, $1 \in B$ and I is connected. So let us prove that $a \notin A$. The proof that $b \notin B$ goes similarly. In fact, if we assume by contradiction that $a \in A$, then

$$\underline{i}(a) \prec \underline{\alpha} \tag{3.5}$$

and

$$\underline{\alpha} \preceq \underline{i}(a^+). \tag{3.6}$$

Since $x \mapsto \underline{i}(x)$ is continuous unless $f^n(x)$ is a turning point for some n, (3.5) and (3.6) imply that $f^n(a) = c_k$ for some $n \geq 0$ and some k. Let n be the smallest integer with this property. Then $i_j(a) = i_j(a^+) = \alpha_j$ for $j = 0, 1, \dots, n-1$ and $i_n(a) = c_k$. Since $f^n(a) = c_k$ either $f^n(x) \in I_{k+1}$ or $f^n(x) \in I_k$ for each $x > a$ sufficiently close to a. Let us assume we are in the former case. Then $i_n(a^+) = (c_k, c_{k+1}) = I_{k+1}$ and therefore by (3.5) and (3.6), $c_k \leq \alpha_n \leq I_{k+1}$. If $\alpha_n = c_k$ then by definition of the space $\Sigma(\nu_0, \dots, \nu_{l+1})$ we have $\sigma^n(\underline{\alpha}) = \underline{i}(c_k)$ and hence $\underline{i}(a) = \alpha$, a contradiction. Therefore we have $\alpha_n = I_{k+1}$. Moreover, $\sigma^n\underline{i}(a) = \underline{i}(c_k)$ and $\sigma^n\underline{i}(a^+) = \underline{i}(c_k^+) = \nu_k$. So by the definition of the ordering, and by (3.6), $\sigma^n\underline{\alpha} \preceq \sigma^n\underline{i}(a^+) = \nu_k$. This is impossible: if the sign of I_{k+1} is positive then this implies $\sigma^{n+1}\underline{\alpha} \preceq \sigma\nu_k$ and if the sign is negative then $\sigma\nu_k \preceq \sigma^{n+1}\underline{\alpha}$. Since $\sigma^n(\alpha) = (I_{k+1}, \dots)$ both cases are impossible by definition of the class $\Sigma(\nu_0, \dots, \nu_{l+1})$. Hence $\underline{i}_f(I) = \Sigma(\nu_0, \dots, \nu_{l+1})$. It follows that $\underline{i}_f\colon I \to \Sigma(\nu_0, \dots, \nu_{l+1})$ is a well defined order preserving bijection if f has no wandering intervals, no intervals of periodic points and no periodic attractors. From this the last statement of the theorem follows: in that case the monotone bijection $\underline{i}_g^{-1} \circ \underline{i}_f\colon I \to I$ is a conjugacy between f and g. □

Finally, in the remainder of this section we shall characterize the sequences in Σ that can occur as kneading invariants of a map f. Indeed, as we will see in the next theorem, (3.4) gives not only necessary but also sufficient "admissibility" conditions for a collection of sequences $\nu_1, \dots, \nu_l \in \Sigma$ to be the kneading invariants of a l-modal map f. For simplicity, if α is some finite string of symbols from the set $\{I_1,\dots,I_{l+1}\}$ then, as before, $\alpha^\infty \in \{I_1,\dots,I_{l+1}\}^{\mathbb{N}}$ denotes the infinite repetition of this string.

Theorem 3.3. *Consider $\Sigma_0 = \{I_1,\dots,I_{l+1}\}^{\mathbb{N}}$ and let $\prec$ be a signed lexicographical ordering on Σ. Furthermore, let $\nu_1,\dots,\nu_l$ be sequences in Σ_0 of the form $\nu_j = (I_{j+1},\dots)$ and define ν_0 and ν_{l+1} so that*

$$\begin{aligned} \nu_0 = I_1^\infty,\ \nu_{l+1} = I_{l+1}^\infty \quad & \text{if } \epsilon(I_1)=1, \quad \epsilon(I_{l+1})=1 \\ \nu_0 = I_1^\infty,\ \nu_{l+1} = I_{l+1}\cdot I_1^\infty \quad & \text{if } \epsilon(I_1)=1, \quad \epsilon(I_{l+1})=-1 \\ \nu_0 = (I_1\cdot I_{l+1})^\infty,\ \nu_{l+1} = (I_{l+1}\cdot I_1)^\infty \quad & \text{if } \epsilon(I_1)=-1,\ \epsilon(I_{l+1})=-1 \\ \nu_0 = I_1\cdot I_{l+1}^\infty,\ \nu_{l+1} = I_{l+1}^\infty \quad & \text{if } \epsilon(I_1)=-1,\ \epsilon(I_{l+1})=1. \end{aligned}$$

If $\nu_0,\dots,\nu_{l+1}$ satisfies the admissibility conditions (3.4) then there exists a l-modal map $f\colon [0,1]\to[0,1]$ with turning points $c_1,\dots,c_l$ and with kneading invariants $\underline{i}_f(c_i^+)$ equal to ν_i.

Remarks. 1. The proof of Theorem 3.3 shows that the conditions (3.4) are not magical at all: they simply reflect the fact that f is alternatingly order preserving and order reversing on each of the laps of f. More precisely, a simple way to check whether some eventually periodic kneading sequences satisfy the conditions (3.4) goes as follows. For each $k=0,1,\dots$ and each $i=1,\dots,l$, simply embed $\sigma^k(\nu_i)$ in an order preserving fashion into I and mark the points $(\sigma^k(\nu_i),\sigma^{k+1}(\nu_i))$ in the square $I\times I$. If a l-modal graph can be drawn through these points then these conditions are admissible. 2. As explained in Remark

3 above the definition of the lexicographical ordering, we can construct a map f that has no wandering intervals with such admissible kneading invariants. 3.

In the next section we shall strengthen this result: we will show that there also exist polynomial maps with such admissible kneading invariants.

Proof of Theorem 3.3: If $\underline{a},\underline{b}\in\Sigma_0$ then we say that $\underline{a}\sim\underline{b}$ if $\underline{a}=\underline{b}$ or if $\underline{a}=I_{k\pm1}\cdot\sigma\underline{b}$ where $\underline{b}=I_k\cdots$. So for example $\underline{i}(c_k^+)\sim\underline{i}(c_k^-)$. Next choose some $N<\infty$. For $n=0,\dots,N-1$ and $j=0,\dots,l+1$ associate to $\sigma^n\nu_j$ a point p_j^n in the interval $[0,1]$ so that the ordering of these symbols with respect to $\prec$ coincides with the natural ordering on $[0,1]$ and such that $p_j^0=c_j$; here we take $c_0=0$ and $c_{l+1}=1$. So we choose these points so that

$$\begin{aligned} &p_i^n < p_j^m \text{ if } \sigma^n(\nu_i)\prec\sigma^m(\nu_j) \text{ and } \sigma^n(\nu_i)\not\sim\sigma^m(\nu_j) \\ &p_i^n = p_j^m \text{ if } \sigma^n(\nu_i)\sim\sigma^m(\nu_j) \\ &p_i^n > p_j^m \text{ if } \sigma^m(\nu_j)\prec\sigma^n(\nu_i) \text{ and } \sigma^m(\nu_j)\not\sim\sigma^n(\nu_i). \end{aligned}$$

Next let $F_N\colon[0,1]\to[0,1]$ be the piecewise linear map whose graph consists of straight lines connecting the points (p_j^n,p_j^{n+1}) with $j=0,1,\dots,N-1$. We want to show that this map F_N is l-modal. Let us first show that if $\sigma^n\nu_j = (I_k,\dots)$ then $\nu_{k-1} \preceq \sigma^n\nu_j \preceq \nu_k$ and therefore that the corresponding point p_j^n is contained in $[c_k, c_{k+1}]$. This is not difficult to show: from (3.4), $\sigma\nu_{k-1} \preceq \sigma^{n+1}\nu_j \preceq \sigma\nu_k$ if the sign of I_k is positive and $\sigma\nu_k \preceq \sigma^{n+1}\nu_j \preceq \sigma\nu_{k-1}$ otherwise. Moreover, from the definition of the signed lexicographical ordering, one has for any $\alpha, \beta \in \Sigma$ with $\alpha_0 = \beta_0 = I_k$ that

$$\begin{aligned}&[\alpha \prec \beta \text{ implies } \sigma(\alpha) \prec \sigma(\beta)] \text{ if } \epsilon(I_k) > 0,\\ &[\alpha \prec \beta \text{ implies } \sigma(\beta) \prec \sigma(\alpha)] \text{ if } \epsilon(I_k) < 0.\end{aligned} \tag{3.6}$$

Because of all this, $\nu_{k-1} \preceq \sigma^n\nu_j \preceq \nu_k$. It follows that the point p_j^n is contained in $[c_k, c_{k+1}]$. Furthermore, from the requirement that one of the inequalities in (3.4) is strict, $\epsilon(I_k) > 0$ if and only if $\sigma\nu_{k-1} \prec \sigma\nu_k$. Similarly, if $p_j^n < p_i^m$ are in $[c_k, c_{k+1}]$ then, because of the definition of $\sim$, $F_N(p_j^n) = p_j^{n+1} < p_i^{m+1} = F_N(p_i^m)$ if $\epsilon(I_k) > 0$ and $F_N(p_j^n) = p_j^{n+1} > p_i^{m+1} = F_N(p_i^m)$ if $\epsilon(I_k) < 0$. It follows that F_N is order preserving on $[c_k, c_{k+1}]$ if $\epsilon(I_k) > 0$ and order reversing if $\epsilon(I_k) < 0$. Since the sign of I_k alternates, this implies that F_N does have turning points $c_1, \dots, c_l$ and therefore is a l-modal map. Next construct F_{N+1} inductively by choosing new points p_j^N in the interval associated to $\sigma^N\nu_j$, $j = 1, \dots, l$ between the previous points in the right order. It is not hard to do this in such a way that one gets a sequence of maps F_N which converges in the C^0 topology. (For example, keep the old points fixed and choose these new points p_j^N, $j = 1, \dots, l$ so that they are 'equally distributed' over the components of the complement of $\{p_j^n;\ n = 0, \dots, N-1, j = 1, \dots, l\}$. It is easy to show that this forces the maps to converge in the C^0 topology.) Now the kneading invariants of F_N coincide with $\nu_1, \dots, \nu_l$ up to the N-th position. Moreover the position of the first $N-1$ iterates under F_k of c_i is the same for each $k \geq N$ because we do not change the position of the points p_j^n for $n \leq N$. It follows that the limiting map f has the required kneading invariants. □

3.1 Examples

Let us now give a few examples of maps with given kneading invariants.

Example. Let $0 < c_1 < c_2 < 1$ and I_1, I_2, I_3 be as before. Furthermore, let $\epsilon(I_1) = \epsilon(I_3) = 1$ and $\epsilon(I_2) = -1$ and $\nu_1 = (I_2, I_3, I_2, I_2, I_2, \dots)$, $\nu_2 = (I_3, I_1, I_2, I_2, I_2, \dots)$. It is easy to check that these sequences satisfy the admissibility conditions from the previous lemma. However, a more convenient way to check these conditions is to simply draw a bimodal map which has these kneading invariants. Such a map is drawn in Figure 3.4 on the left with the corresponding kneading invariants. Similarly, let $0 < c < 1$ and $I_1 = [0, c)$ and $I_2 = (c, 1]$. Let $\epsilon(I_1) = -\epsilon(I_2) = 1$ and let $\nu = (I_2, I_2, I_1, I_1)^\infty$ (this means that the block (I_2, I_2, I_1, I_1) is repeated infinitely often). In Figure 3.4 we have drawn on the right a corresponding unimodal map with this kneading invariant.

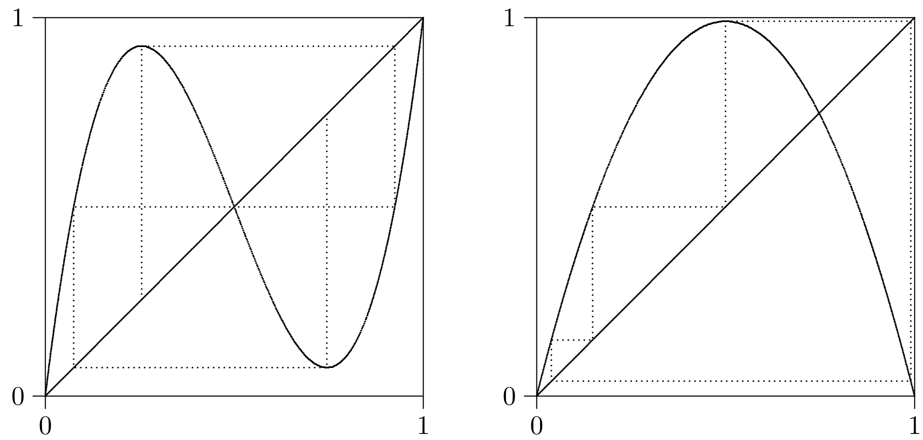

Fig. 3.4: The bimodal and unimodal maps with kneading invariants as in Example 3.1.

Example. Let $f\colon [-1,1] \to [-1,1]$ be the quadratic map $f(x) = -2x^2 + 1$. This map has only one turning point, $c = 0$, and its orbit is $c \to 1 \to -1 \to -1 \dots$. Hence $\underline{i}(-1) = (I_1, I_1, I_1, \dots)$, $\underline{i}(1) = (I_2, I_1, I_1, \dots)$ and $\nu = \underline{i}(c^+) = (I_2, I_2, I_1, I_1, \dots)$. This map is conjugate to the tent map $T\colon [-1,1] \to [-1,1]$ defined by $T(x) = 2x+1, x \le 0$, $T(x) = -2x+1, x > 0$. In fact, let $\phi\colon I \to I$ be the homeomorphism $\phi(x) = \frac{2}{\pi}\sin^{-1} x$. Then $\phi \circ f \circ \phi^{-1}(y) = \frac{2}{\pi}\sin^{-1}(1 - 2\sin^2\frac{\pi}{2}y)) = \frac{2}{\pi}\sin^{-1}(\sin(\frac{\pi}{2}(2y+1)))$. Hence $\phi \circ f \circ \phi^{-1}$ is the tent map T. It is easy to see that the backward orbit of the turning point 0 of T is dense or, indeed, that the backward orbit of any point is dense. (This follows from the fact that for every interval $I \subset [-1,1]$ there exists $n \ge 0$ such that $T^n(I) = [-1,1]$.) Hence the backward orbit of the turning point of f is also dense. Therefore, from Theorem 3.1, it follows that any unimodal map $g\colon I \to I$ with the same kneading invariant as f is semi-conjugate to f.

Example. Let $f\colon I \to I$ be a piecewise linear unimodal map such that there exists an interval J_1 containing the turning point of f such that $J_1, f(J_1), f^2(J_1)$ are disjoint and $f^3\colon J_1 \to J_1$ is again unimodal, see Figure 3.5. Let r_1 be the affine scaling which maps I onto J_1 and let $r(J_1) = J_2$. Since f is affine on each interval $f^i(J_1)$ when $0 < i < 3$, one can modify f on J_1 so that

$$r_1 \circ f \circ r_1^{-1} \text{ equals } f^3 \text{ on } J_1 \setminus J_2.$$

Repeating this infinitely often one gets a map f which is called *infinitely renormalizable*. Another term which is frequently used is that the map is *solenoidal*. This means that there exists a nested sequence of intervals J_n such that $I_n, \dots, f^{3^n-1}(J_n)$ are disjoint and f^{3^n} maps J_n as a unimodal map into J_n. Similarly, one can also construct maps for which the period of J_n is equal to $q(n) = a(1) \cdot a(2) \cdot \dots \cdot a(n)$ where $a(k) \ge 2$. We will come back to these infinitely renormalizable maps in Section 5.

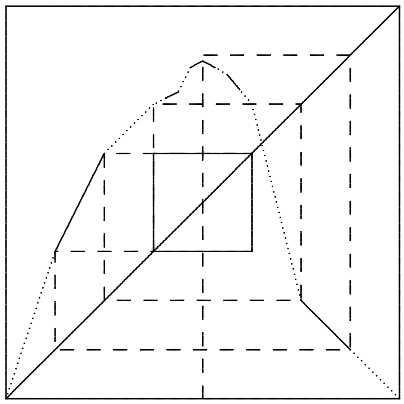

Fig. 3.5: The construction of an infinitely renormalizable map. The first return map to the solid square is a unimodal map with the opposite orientation.

3.2 Hofbauer's Tower Construction

In this section we shall give another way to represent the kneading invariants for unimodal maps. This representation is due to Hofbauer (1979) and is used extensively in the work of Hofbauer and Keller, see also Bruin (1992c). We shall not use this description anywhere in the remainder of this book. However, the ideas of this section are philosophically related to those in Section V.6 and also to the concept of the tableau recently introduced for cubic maps of the Riemann sphere by Branner and Hubbard (1991).

So assume that $f\colon [0,1] \to [0,1]$ is a unimodal map with turning point c, $f(0) = f(1) = 0$ and $f^2(c) < c < f(c)$. (If this last condition does not hold then f has no interesting dynamics.) Then the kneading invariant ν of f is of the form

$$\nu = e_0 e_1 e_2 \dots$$

with $e_i \in \{I_1, I_2\}$, $e_0 = e_1 = I_2$ and $e_2 = I_1$. Let us show that one can write ν in a very nice form. In order to do this define

$$e'_m = \begin{cases} I_1 & \text{if} \quad e_m = I_2, \\ I_2 & \text{if} \quad e_m = I_1. \end{cases}$$

Definition. Take a kneading invariant $\nu(f)$ of a unimodal map f. To such a kneading invariant one can associate in a unique way γ blocks B_i where $\gamma \in \mathbb{N} \cup \{\infty\}$ such that

$$\nu(f) = \begin{cases} I_2 I_2 B_1 B_2 B_3 \dots & \text{if } \gamma = \infty \\ I_2 I_2 B_1 B_2 B_3 \dots B_\gamma & \text{otherwise,} \end{cases}$$

$$(*) \qquad B_i = e_{j+1} e_{j+2} \dots e_{j+m(i)-1} e_{j+m(i)} = e_1 e_2 \dots e_{m(i)-1} e'_{m(i)}$$

for $i < \gamma$ and, if $\gamma < \infty$, the last block B_γ coincides with $e_1e_2\dots$. This is done as follows. First let $m(1)-1$ be largest integer such that

$$B_1 = e_3 \dots e_{3+m(1)-1} = e_1e_2 \dots e_{m(1)-1}.$$

If such a largest integer does not exist then B_1 is an infinite block, $\gamma = 1$ and $m(1) = \infty$. In this case the procedure terminates and $\nu(f) = I_2I_2B_1$; otherwise $\gamma > 1$, $m(1) < \infty$ and B_1 is the finite block such that

$$B_1 = e_3 \dots e_{3+m(1)} = e_1e_2 \dots e'_{m(1)}.$$

Now suppose that $B_1, \dots, B_{k-1}$ are defined such that $I_1I_2B_1 \dots B_{k-1} = e_1 \dots e_j$ and such that $(*)$ holds for $i = 1, \dots, k-1$. Then define B_k such that $m(k)-1$ is the largest integer with

$$B_k = e_{j+1} \dots e_{j+m(k)-1} = e_1e_2 \dots e_{m(k)-1}.$$

If no such integer $m(k)$ exists then we take B_k to be the infinite block and $\gamma = k$; otherwise $(*)$ holds for $i = k$. If for each $i = 1, 2, \dots$ such an integer $m(i)$ exists then we obtain an infinite number of blocks $B_1, B_2, \dots$. Otherwise, we obtain a finite number of blocks $B_1, \dots, B_\gamma$ such that the last one has infinite length and B_γ coincides with $e_1e_2\dots$. It is easy to see that this decomposition of $\nu(f)$ is unique (for example, because $e_2 = I_1$ one has $B_1 = e_2$).

Let $\gamma \in \mathbb{N} \cup \{\infty\}$ be the number from the previous definition and define

$$\mathbb{N}_\gamma = \{n \in \mathbb{N};\ 1 \le n < \gamma\}.$$

We should note that if c is not contained in the basin of periodic attractor then $\gamma = \infty$. In the next lemma we shall show that ν is completely determined by some map $Q\colon \mathbb{N}_\gamma \to \mathbb{N}_\gamma \cup \{0\}$ called the *kneading map.*

Lemma 3.2. *For each f as above there exist $\gamma \in \mathbb{N} \cup \{\infty\}$ and a map $Q\colon \mathbb{N}_\gamma \to \mathbb{N}_\gamma \cup \{0\}$ such that $Q(k) < k$ for each $k \in \mathbb{N}_\gamma$ and with the following properties. Define*

$$S_0 = 1 \text{ and } S_j = S_{j-1} + S_{Q(j)}$$

for $1 \le j < \gamma$. Then, for $1 \le j < \gamma$, the length of the block B_j is equal to some number $S_{Q(j)}$. In fact, $B_1 = e_2$ and

$$B_j = e_{S_{j-1}+1}e_{S_{j-1}+2} \dots e_{S_j-1}e_{S_j} = e_1e_2 \dots e_{S_{Q(j)}-1}e'_{S_{Q(j)}}$$

for $2 \le j < \gamma$ and (if $\gamma < \infty$)

$$B_\gamma = e_{S_{\gamma-1}+1}e_{S_{\gamma-1}+2} \dots = e_1e_2 \dots .$$

We shall prove this lemma by constructing the Hofbauer tower. The orbit of the turning point can be followed extremely well in this tower. The main aim of this construction is to get a Markov extension of the original interval map, but we shall not use this Markov extension anywhere in the remainder of this book. (This extension is said to have the Markov property because, as we will see below, it sends each connected component of its domain to a union of connected components.) We shall, however, give some examples which illustrate these ideas by constructing unimodal maps for which the orbit of the turning point can easily be visualized in the tower construction.

Let us for simplicity assume that c is not periodic. Let $a_k < c$ be so that $[a_k, c]$ is the maximal interval of this form on which f^k is monotone. Of course for $k \geq 2$, a_k is a preimage of c under some iterate of f; for example $a_2 = f^{-1}(c) \cap [0, c]$. Let

$$V_k = f^k[a_k, c].$$

Since a_k is a turning point of f^k, for $k \geq 2$ the interval V_k is of the form

$$V_k = [f^k(c), f^{i_k}(c)]$$

with $1 \leq i_k < k$ and $f^{k-i_k}(a_k) = c$. Here, as before, $[f^k(c), f^{i_k}(c)]$ stands for the segment connecting $f^k(c)$ and $f^{i_k}(c)$ even if $f^k(c)$ is to the right of $f^{i_k}(c)$. So V_k is a one-sided neighbourhood of $f^k(c)$ and for example $V_2 = [f^2(c), f(c)]$ and $i_2 = 1$. One has

$$i_{k+1} = \begin{cases} i_k + 1 & \text{if } c \notin V_k \\ 1 & \text{if } c \in V_k. \end{cases} \tag{3.7}$$

In particular,

$$\text{if } c \notin V_k \text{ then } f(V_k) = V_{k+1}. \tag{3.8}$$

Now let D_k be the closure of the component of $V_k \setminus \{c\}$ which contains $f^k(c)$ in its boundary and E_k the closure of the other component. So $D_k = V_k$ if $c \notin V_k$.

Lemma 3.3.

$$f(D_k) = V_{k+1} \tag{3.9}$$

and

$$\text{if } E_k \neq \emptyset \text{ then } E_k = D_{i_k}. \tag{3.10}$$

Proof. If $c \notin \operatorname{int}(V_k)$ then $D_k = V_k$ and (3.9) follows from (3.8). If $c \in \operatorname{int}(V_k)$ then $D_k = [f^k(c), c]$, $i_{k+1} = 1$, $V_{k+1} = [f^{k+1}(c), f(c)]$ and so again (3.9) holds. Moreover, f^k is monotone on $[a_k, c]$ and $f^{k-i_k}(a_k) = c$. Therefore f^{i_k} is monotone on $[c, f^{k-i_k}(c)]$. Now $a_{i_k} < c$ is chosen so that $[a_{i_k}, c]$ is the largest interval with the property that $f^{i_k}|[a_{i_k}, c]$ is monotone. So either $[a_{i_k}, c]$ contains $[c, f^{k-i_k}(c)]$ or it contains its symmetric $[c, \tau(f^{k-i_k}(c))]$ (where τ is so that $f(\tau(x)) = f(x)$ and $\tau(x) \neq x$ for $x \neq c$). But, since the f-images of intervals which are each others symmetric are the same, one gets in both cases

$$V_{i_k} = f^{i_k}[a_{i_k}, c] \supset f^{i_k}[c, f^{k-i_k}(c)] = [f^{i_k}(c), f^k(c)] = V_k.$$

Since both these intervals have $f^{i_k}(c)$ as a boundary point and since V_k contains c, this implies that the components of $V_{i_k} \setminus \{c\}$ and $V_k \setminus \{c\}$ containing $f^{i_k}(c)$ coincide. In particular, from the definitions of D_{i_k} and E_k, one has (3.10). $\square$

Proof of Lemma 1: Define $S_0 = 1$ and assuming $S_0 < S_1 < S_2 < \cdots < S_j$ are defined we inductively define S_{j+1} as

$$S_{j+1} = \min\{l > S_j\,;\ c \in V_l\}.$$

From (3.7) one has $S_{j+1} - S_j = i_{S_{j+1}}$. Moreover, by (3.10) one has that $V_{i_{S_{j+1}}}$ contains c and so $i_{S_{j+1}}$ is equal to $S_{Q(j+1)}$ for some $Q(j+1) < j+1$. $\square$

Note that S_j are precisely the integers such that $i_{S_j+1} = 1$ and so the places where one can go down in the diagram, see for example Figure 3.6. So define $\hat{D} = \bigcup_{k\geq 2} V_k \times \{k\}$ and $\hat{f}\colon \hat{D} \to \hat{D}$ as follows. Let $\hat{x} = (x, k) \in \hat{D}$. If $x \in D_k$ then we move up in the diagram:

$$\hat{f}(\hat{x}) = (f(x), k+1) \in V_{k+1} \times \{k+1\}.$$

If $x \in E_k$ then $E_k = D_{i_k}$ and so map down to level $i_k + 1$:

$$\hat{f}(\hat{x}) = (f(x), i_k + 1) \in V_{i_k+1} \times \{i_k + 1\}.$$

By (3.9) and (3.10) this map is well defined when $x \neq c$. Moreover, $\hat{f}$ sends each component of $(V_k \setminus \{c\}) \times \{k\}$ monotonically onto a set of the form $V_l \times \{l\}$ for some $l \geq 2$. Therefore $\hat{f}$ is called a Markov map. Examples of these maps are given below.

Remark. It is not too difficult to show that if f is a unimodal map as above then the map $Q\colon \mathbb{N}_\gamma \to \mathbb{N}_\gamma \cup \{0\}$ from Lemma 1 is so that

$$Q(k) < k \text{ for all } k \in \mathbb{N}_\gamma$$

and

$$(Q(j))_{k<j<\gamma} \succeq (Q(Q(Q(k)) + j - k))_{k<j<\gamma}$$

for all $k \in \mathbb{N}_\gamma$ with $Q(k) \geq 1$ (where $\succeq$ denotes the lexicographical ordering on sequences of integers). A map satisfying these inequalities is called a *kneading map.* Building on earlier work of Hofbauer, it is shown in Hofbauer and Keller (1990a) that any kneading map Q also defines the kneading invariant $e_0 e_1 e_2 \ldots$ of a unimodal map: take as before $e_0 = e_1 = I_2$,

$$S_0 = 1 \text{ and } S_k = S_{k-1} + S_{Q(k)}$$

for $1 \leq k < \gamma$ and

$$e_{S_{k-1}+1} e_{S_{k-1}+2} \ldots e_{S_k - 1} e_{S_k} = e_1 e_2 \ldots e_{S_{Q(k)}-1} e'_{S_{Q(k)}}$$

for $2 \leq k < \gamma$. Moreover, if $\gamma < \infty$, $e_{S_{\gamma-1}+1} e_{S_{\gamma-1}+2} \cdots = e_1 e_2 \ldots$.

Using this one can construct many examples of unimodal maps such that the closure of the turning point is a minimal Cantor set. Indeed,

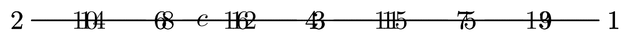

Fig. 3.6: The tower associated to the Feigenbaum map.

Example. The Feigenbaum map: the prototype of the infinitely renormalizable map. In this case we have $\gamma = \infty$ and $Q(j) = j-1$ for all $j \geq 1$. It follows from this that $S_j = S_{j-1} + S_{j-1}$ and since $S_0 = 1$ this gives $S_j = 2^j$. So the interval V_k contains c if and only if k is a power of 2. The tower is drawn below. Since V_{S_k} contains c and $f^{S_k}(c)$ are closest returns,

$$f^2(c) < f^8(c) < f^{32}(c) < \cdots < c < \cdots < f^{16}(c) < f^4(c) < f(c)$$

and

$$\tau(D_{S_{k+1}}) \subset D_{S_k}.$$

From the tower one can check that the 2^j closed intervals $D_{S_j}, \ldots, D_{S_{j+1}-1}$ are disjoint (otherwise one would a contradiction with $Q(j) = j-1$) and each forward iterate of c is contained in the union K_j of these intervals. Moreover, each of these intervals contains precisely two of the 2^{j+1} intervals $D_{S_{j+1}}, \ldots, D_{S_{j+2}-1}$. Since c is contained in D_{S_j} for every j it follows that the forward orbit of c is contained in

$$K = \bigcap_{j \geq 1} \left(D_{S_j} \cup \cdots \cup D_{S_{j+1}-1} \right).$$

Of course K has no isolated points and therefore K is a Cantor set provided it contains no intervals. From the disjointness of $D_{S_j}, \ldots, D_{S_{j+1}-1}$ and since $S_j = 2^j \to \infty$ as $j \to \infty$, it follows that each component of K contains at most one iterate of c. It follows that the closure of the forward orbit of c is a Cantor set (and is equal to $\omega(c)$). If f has no wandering intervals then K is indeed a Cantor set and $K = \omega(c)$.

Example. The Fibonacci map: the prototype of the non-renormalizable map for which the orbit of the turning point is a Cantor set, see Hofbauer and Keller (1990). For this map we have $\gamma = \infty$ and $Q(j) = j-2$ for $j \geq 2$. Since $S_0 = 1$, $Q(1) = 0$ and $S_j = S_{j-1} + S_{Q(j)} = S_{j-1} + S_{j-2}$ the sequence S_j generates the Fibonacci numbers. Again the intervals D_j are decreasing: the intervals $D_{S_{j+1}} \subset D_{S_j}$ are on the same side of c and the interval $D_{S_{j+3}} \subset D_{S_{j+2}}$ on the other side of c. Moreover, it is not hard to show that the first $S_{j+1} - S_j - 1$ forward iterates of D_{S_j} are all disjoint (otherwise one would get a contradiction with $Q(j) = j-2$). So again the closure of the forward orbit of c forms a Cantor set. However, the situation is different from the previous case: here there exists no restrictive interval J containing the turning point. Indeed, if such an interval existed then the smallest such interval (with the same period) would contain forward iterates of c in its boundary. But it is easy to see from the tower in Figure 3.7 that this is impossible. It follows that the corresponding map is non-renormalizable.

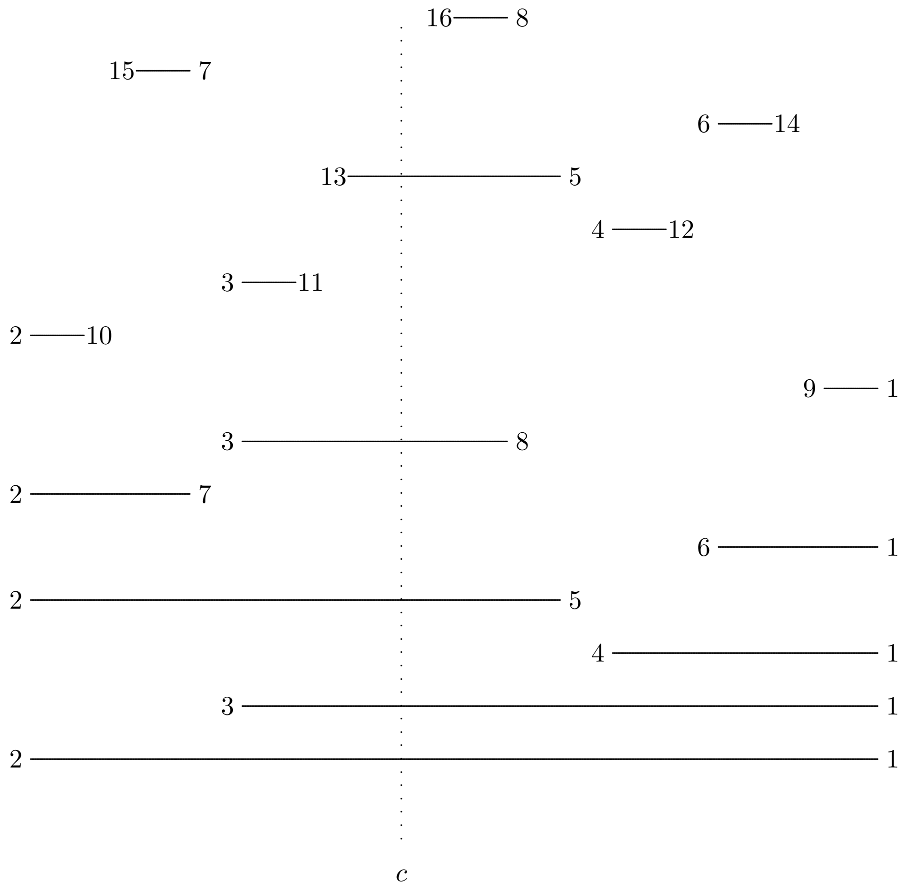

Fig. 3.7: The tower associated to the Fibonacci map.

Exercise 3.1. As in Example 3.2, let $f(x) = -2x^2 + 1$. Show that Theorem 3.2 implies that each sequence in $\Sigma = \{I_1, I_2\}^{\mathbb{N}}$ occurs as the itinerary of some point x in $[-1, 1]$. Using the conjugacy from Example 3.2, show that for each such sequence there exists at most one such point x. Determine γ and the corresponding kneading map Q.

Exercise 3.2. Show that if f is a unimodal map with a periodic attractor then the kneading invariant $\nu(f)$ of f need not be periodic: $\nu(f)$ is of the form $\nu = I_2 \cdot (B \cdot I_1)^\infty$ or of the form $\nu = I_2 \cdot (B \cdot I_2)^\infty$.

Exercise 3.3. Suppose $f\colon [0, 1] \to [0, 1]$ is unimodal and that its turning point has period 4. Determine all possible kneading invariants that f can have. What possible orderings can the orbit of the turning point have? (Hint: check for each periodic sequence $(J_1, J_2, J_3, J_4, \ldots)$ with period 4 that the required compatibility conditions are satisfied.)

Exercise 3.4. Let $\gamma = \infty$ and assume $Q\colon \mathbb{N}_\gamma \to \mathbb{N}_\gamma \cup \{0\}$ is non-decreasing. Show that Q is admissible and therefore corresponds to the kneading map of a unimodal map f.

Exercise 3.5. Suppose that f is unimodal and let Q be the corresponding kneading map. Show that if $Q(n) \to \infty$ as $n \to \infty$ then $\omega(c)$ is a minimal Cantor set.

4 Full Families and Realization of Maps

Full Families and Realization of Maps In this section we will consider families of smooth maps and investigate whether there exists a map in such a family of a given combinatorial type. For convenience we shall make the following assumptions.

Assumptions.

1. Let Δ be some subset of a Euclidean space and for each $\mu \in \Delta$, let $f_\mu\colon I \to I$ be a l-modal map. 2. Assume that f_μ depends continuously on μ in the C^0 topology and that the turning points of f_μ depend continuously on μ. For simplicity, assume $I = [0,1]$ and denote the turning points of f_μ by $0 < c_1(\mu) < \cdots < c_l(\mu) < 1$. By assumption f_μ is strictly monotone on each of the intervals $I_1(\mu) = [0, c_1(\mu)), I_2(\mu) = (c_1(\mu), c_2(\mu)), \dots, I_{l+1}(\mu) = (c_l(\mu), 1]$ and $f_\mu(\partial I) \subset \partial I$. 3. Assume that the Lipschitz norm of f_μ is uniformly bounded, i.e.,

$$\sup_{x \neq y, \mu} \frac{|f_\mu(x) - f_\mu(y)|}{|x-y|} < \infty$$

and that for each $\epsilon > 0$ there exists $\delta > 0$ such that for each $i = 1, \dots, l$ and each $\mu \in \Delta$,

$$|x - c_i(\mu)| < \delta \text{ implies } \frac{|f_\mu(x) - f_\mu(c_i(\mu))|}{|x - c_i(\mu)|} < \epsilon.$$

4. Finally, we require that for each $\delta > 0$, if $\Delta \ni \mu(n) \to \mu$ and if the Lipschitz norm of $f_{\mu(n)}$ tends to zero on some subinterval of $I_i(\mu(n))$ of length δ then the Lipschitz norm of $f_{\mu(n)}|I_i(\mu(n))$ tends to zero (note that Δ need not be closed and therefore f_μ may not be defined).

We say that two l-modal maps $f, g\colon I \to I$ have the same *orientation* if f and g are both increasing or both decreasing on corresponding laps I_i, $i = 1, \dots, l{+}1$. From the assumptions it follows that all maps from a family f_μ as above have the same orientation if Δ is connected. Since successive turning points of f_μ are alternating local maxima and local minima, the map $F\colon \Delta \to I^l$ defined by

$$F(\mu) = (f_\mu(c_1), \dots, f_\mu(c_l))$$

has values in

$$V = \{(v_1, \dots, v_l) \in I^l\,;\ s(-1)^i(v_{i+1} - v_i) < 0 \text{ for } i = 0, \dots, l\}$$

where $s \in \{-,+\}$ depending on the orientation of f_μ and where v_0 and v_{l+1} are respectively the images of the left and right endpoint of I. Here, and below, we have written $f_\mu(c_i)$ instead of $f_\mu(c_i(\mu))$. We assume that F can be extended continuously from the closure of Δ to the closure of V.

In this section we will show that any l-modal map g 'appears' in a reasonable family of smooth maps. So, for example, given a l-modal map $g\colon I \to I$ we want to show that there exists a polynomial map which has essentially the same dynamics as g. Of course one cannot expect that there exists a polynomial map which is conjugate to g. Indeed, g might have wandering intervals and as we shall see polynomial maps do not. Moreover, g might have many 'superfluous' attractors and, as we shall see later, a polynomial map of degree l has at most $l-1$ periodic attractors. So in order to define the notion of a full family we first introduce the notion of two maps being essentially conjugate.

Definition. We call a periodic attractor *essential* if it contains a turning point in its immediate basin. Furthermore, we define an equivalence relation as follows: $x \sim y$ whenever the interval connecting x and y is contained in a union of homtervals which are disjoint from the basins of all essential periodic attractors.

Examples of an interval $[x, y]$ with $x \sim y$ are 1) the closure of a component of the basin of an inessential periodic attractor, 2) a component of a preimage of an interval consisting of points of constant period (some iterate of the map is the identity restricted to this interval) or 3) a wandering interval. In fact, because of Lemma 3.1, equivalence classes are unions of such intervals.

Note that the equivalence classes of this relation are closed intervals and that therefore the quotient map $g/\!\sim\, : I/\!\sim\, \to I/\!\sim$ is a well defined continuous map. However, in general $g/\!\sim$ is not l-modal anymore. For example, if $g(x) = x$ for $x \in [0, 1/2]$ and $g(x) = 1/2 - x$ for $x \in [1/2, 1]$ then all elements of I are equivalent and therefore $I/\!\sim$ consists of only one point. Therefore we shall assume that no equivalence class of $\sim$ contains a lap of g.

Definition. Two l-modal maps f and g are said to be *essentially conjugate* if there exists an order preserving homeomorphism h such that

$$\begin{array}{ccc} I/\!\sim & \xrightarrow{f/\sim} & I/\!\sim \\ h\big\downarrow & & \big\downarrow h \\ I/\!\sim & \xrightarrow{g/\sim} & I/\!\sim \end{array}$$

commutes. Moreover, we say that f and g are *essentially combinatorially equivalent* if the map

$$h\colon \bigcup_{i=1}^{l} \bigcup_{n\geq 0} f^n(c_i)/\!\sim \quad \to \quad \bigcup_{i=1}^{l} \bigcup_{n\geq 0} g^n(\tilde{c}_i)/\!\sim \tag{4.1}$$

defined by $h[f^n(c_i)] = g^n[\tilde{c}_i]$ is an order preserving bijection.

Clearly, if f and g are essentially conjugate then they are essentially combinatorially equivalent. The reverse implication also holds provided each one-sided periodic attractor of f corresponds in a unique way to a one-sided periodic attractor of g, see the Corollary of Theorem 3.1.

Definition. Let Δ be a connected subset Δ of a Euclidean space. We say that a l-modal family f_μ, $\mu \in \Delta$ is *full* if, given a l-modal map $g\colon I \to I$ such that

1. no lap of g is contained in an equivalence class of $\sim$;
2. each periodic turning point is an attractor (this conditions is automatically satisfied if g is C^1);
3. g has the same orientation as maps from the family f_μ,

there exists $\mu' \in \Delta$ such that g and $f_{\mu'}$ are essentially conjugate.

Because of Remark 3 above Theorem 3.2 the kneading invariants of g and $f_{\mu'}$ are the same. However, if g has wandering intervals the alternative 'unusual' kneading invariants from Remark 4 above Theorem 3.2 of g and $f_{\mu'}$ might differ (as was explained in that remark). The reason we impose the first condition on g is that otherwise $g/\sim\colon I/\sim \to I/\sim$ might not have modality l. The second condition excludes a map of the type $g(x) = -s|x - 1/2| + s/2$ with $s = (1 + \sqrt{5})/2$ which has a repelling turning point of period three. We exclude this type of map because no smooth map can be essentially conjugate to g since a periodic turning point of a smooth map is necessarily an essential attractor. It is easy to check that we could drop this assumption if we merely want that f_μ is essentially semi-conjugate to g (this means that there exists an order preserving continuous surjection $h\colon I/\sim \to I/\sim$ such that $h \circ (f/\sim) = (g/\sim) \circ h$). In this case the semi-conjugacy only can 4 the basin a periodic turning point of f_μ (and this happens precisely if the corresponding periodic turning point of g is not attracting).

In this section we will show that many families are full. The appropriate condition for this is given in the following definition.

Definition. We say that $F\colon \Delta \to V$ is *persistently surjective* if the induced homology map

$$F_*\colon H_*(\Delta, F^{-1}(\partial V)) \to H_*(V, \partial V)$$

is surjective. Equivalently, if each interior deformation $F_t\colon \Delta \to V$, $t \in [0,1]$, of F is surjective for each t. Here, an arc of continuous maps $G_t\colon W \to Z$, $t \in [0,1]$, between subsets W and Z of Euclidean spaces is called an *interior deformation* of $G\colon W \to Z$ if G_t extends continuously to a map from $\mathrm{cl}\,(W)$ to $\mathrm{cl}\,(Z)$, if $G_0 = G$, $G_t = G$ on $G^{-1}(\partial Z)$ and if G_t depends continuously on t. (It is not assumed that G_t maps ∂W into ∂Z.)

Definition. A family of l-modal maps $f_\mu\colon I \to I$, $\mu \in \Delta$, satisfying Assumptions 1)-4) from the beginning of this section, is *persistently surjective* if the corresponding map $F\colon \Delta \to V$ is persistently surjective in the sense defined above.

Theorem 4.1. *Assume that f_μ is a l-modal family of maps satisfying assumptions 1)-4) from above. Furthermore, assume that f_μ is persistently surjective, i.e., assume that $F\colon \Delta \to V$ is persistently surjective (this condition is satisfied if F is a homeomorphism). Then f_μ is a full family.*

Before proving the theorem let us state and prove the following

Corollary 4.1. *Let $\underline{a} = (a_0, \dots, a_{l+1}) \in \mathbb{R}^{l+2}$, $P_{\underline{a}}(x) = a_0 + \cdots + a_{l+1}x^{l+1}$ and*

$$\Delta = \{\underline{a} \in \mathbb{R}^{l+2};\ P_{\underline{a}} \text{ is } l - \text{modal}\}.$$

Then the family $P_{\underline{a}}$, $\underline{a} \in \Delta$, is a full family.

Remark. 1. Clearly, in the unimodal case and when Δ is connected, $F\colon \Delta \to V$ is persistently surjective if and only if it is surjective. 2. Even if $F\colon \Delta \to V$ is surjective, $F\colon \Delta \to V$ need not be persistently surjective and f_μ need not be a full family. Indeed, take $g_a(x) = ax(1-x)$ and $\Delta = [1, 2.1] \cup [3.9, 4]$. One can find diffeomorphisms h_a depending smoothly on a such that the map $F\colon \Delta \to V$ associated to $\hat{f}_a = h_a \circ f_a \circ h_a^{-1}$ is surjective. However, there exists no $a \in \Delta$ such that the turning point of $\hat{f}_a$ has period 2 and so $\hat{f}_a$, $a \in \Delta$, is not a full family. In this example Δ is not connected, but in the multimodal case one can give similar examples in which Δ is connected or even simply connected. For this reason we have introduced this stronger surjectivity assumption on F. 3. If $F\colon \Delta \to V$ is a homeomorphism then it is persistently surjective. Indeed, suppose by contradiction that $F_t\colon \Delta \to V$ is an interior deformation of F such that $F_1\colon \Delta \to V$ is not surjective. Then $G_t = F_t \circ F^{-1}\colon V \to V$ satisfies $G_0 = id$. Moreover, G_t can be extended continuously so that it is the identity map restricted to ∂V. Indeed, take $y_n \to y \in \partial V$ and by contradiction assume that for some subsequence $F_tF^{-1}(y_n) \to y' \neq y$. By taking an additional subsequence we may assume that $F^{-1}(y_n)$ converges to some $x \in \partial\Delta$. Then $y_n = FF^{-1}(y_n) \to F(x)$ and therefore $y = F(x)$. Hence $F_tF^{-1}(y_n) \to F_t(x) = F(x) = y$, a contradiction. Since F_1 is not surjective, G_1 is also not surjective and therefore G_1 can be used to define a retract of the simplex onto its boundary. (A retract from a topological space B to a subset A of B is a continuous map $r\colon B \to A$ such that $r|A = id$.) As is well known, see for example Dold (1972) or Massey (1991), this is impossible. 4. From Exercise IV.1.7, it follows that

these polynomial maps $P_{\underline{a}}$ have an important additional property: they have negative Schwarzian derivative (later we shall define this notion). From this and the above Corollary it will follow that polynomial maps model all multimodal maps, see Section 6. 5. Let f_μ be a full l-modal family and let g be a l-modal

map with the same orientation. In general, there exists no parameter μ such that f_μ is combinatorially equivalent to g. This is the reason we had to include the notion of essential combinatorial equivalence in the definition of full families. Let us give an example, due to Milnor, why this stronger equivalence does not always hold for maps as above. Choose a smooth bimodal map $g\colon [0,1] \to [0,1]$ with $g(0) = 0$, $g(1) = 1$ so that its first turning point c_1 maps to a point with itinerary $I_3 I_1 I_1 I_1 I_1 \cdots = I_3(I_1)^\infty$ and its second turning point c_2 maps to a point with itinerary $(I_1)^\infty$. Furthermore, choose g so that 0 is an inessential attracting fixed point, with $g(g(c_1))$ and $g(c_2)$ as distinct points in its immediate basin. For example, g can be a polynomial of degree 4. However, no polynomial P of degree 3 can realize this kneading data. This is because 0 is an inessential fixed point of g. Indeed, in the degree 3 case, since $P(0) \geq 0$, if $P(c_1) > c_1$ and since $P''(x) < 0$ for $x \leq c_1$, we would have $P(x) > x$ for $0 < x \leq c_1$. Hence 0 would be the only point which could have itinerary $(I_1)^\infty$. Since $P(c_1) > c_2$, it would follow that $P(P(c_1)) > P(c_2) \geq 0$, so that $P(c_1)$ can not have the specified itinerary.

Proof of Corollary: Let us first show that Δ is a l-dimensional bounded submanifold of $\mathbb{R}^{2+l}$. By definition a l-modal map sends ∂I into itself. So assuming that $I = [0,1]$ there are several possibilities. Let us take the case that $P_{\underline{a}}(0) = 0$ and $P_{\underline{a}}(1) = 1$ (so l is even). Then $a_0 = 0$ and $a_{l+1} = 1 - a_1 - \cdots - a_l$. For $\underline{a} \in \Delta$ let $c_i(\underline{a})$ be the i-th turning point of $P_{\underline{a}}$. Note that because $P_{\underline{a}}$ is l-modal and this polynomial is of degree $l+1$, all critical points (points where $DP_{\underline{a}}$ is zero) are turning points (the second derivative in these points is non-zero). Let $v_i(\underline{a}) = P_{\underline{a}}(c_i(\underline{a}))$. Then

$$\frac{\partial v_i}{\partial a_j}(\underline{a}) = c_i^j - c_i^{l+1}$$

for $1 \leq i, j \leq l$. Note that $v_i(\underline{a})$ are the components of $F(\underline{a})$. Hence the Jacobian of F is equal to the determinant of the matrix

$$\begin{pmatrix} c_1 - c_1^{l+1} & c_1^2 - c_1^{l+1} & \dots & c_1^l - c_1^{l+1} \\ c_2 - c_2^{l+1} & c_2^2 - c_2^{l+1} & \dots & c_2^l - c_2^{l+1} \\ \vdots & \vdots & \ddots & \vdots \\ c_l - c_l^{l+1} & c_l^2 - c_l^{l+1} & \dots & c_l^l - c_l^{l+1} \end{pmatrix}$$

which is equal to the determinant of

$$\begin{pmatrix} c_1 - c_1^{l+1} & c_1^2 - c_1 & \dots & c_1^l - c_1 \\ c_2 - c_2^{l+1} & c_2^2 - c_2 & \dots & c_2^l - c_2 \\ \vdots & \vdots & \ddots & \vdots \\ c_l - c_l^{l+1} & c_l^2 - c_l & \dots & c_l^l - c_l \end{pmatrix}$$

which is a polynomial Q in $(c_1, \dots, c_l)$ of degree $l + \frac{l(l+1)}{2}$. On the other hand Q vanishes in the spaces $c_i - c_j = 0$, $c_i = 0$ and $c_j = 1$, where $1 \le i, j \le l$. By factoring we get that Q is some polynomial q times

$$\prod_{i=1}^{l} c_i \times \prod_{j=1}^{l} (c_j - 1) \times \prod_{i<j} (c_i - c_j).$$

Comparing the degree of the polynomials we see that q is actually a constant. Hence the map $\underline{a} \mapsto F(\underline{a})$ has non-vanishing Jacobian on Δ and is therefore a local diffeomorphism. So let us show that F is injective on each component of Δ. Since F extends continuously to the boundary, $F(\partial\Delta) \subset \partial V$ and F is a local diffeomorphism, one has that $F^{-1}(y)$ is a finite set for each $y \in V$. Let n be the maximal cardinality of $F^{-1}(y)$ for $y \in V$. Because F is a local diffeomorphism, the set V_0 of $y \in V$ such that $\#F^{-1}(y) = n$ is open. We want to first show that $V_0 = V$; so assume this is not the case. Then take $y_0 \in V_0$ and $y_1 \in V \setminus V_0$ and an arc $\gamma\colon [0,1] \to V$ starting at y_0 and ending at y_1. Let $s \in [0,1]$ be $\min\{t;\ \gamma(t) \in \partial V_0\}$. Because F is a local diffeomorphism, $\gamma\colon [0,s) \to V_0$ has n preimages, say $\gamma_1, \dots, \gamma_n\colon [0,s) \to \Delta$. Because of continuity of F, any limit point of $\gamma_i(t)$ as $t \uparrow s$ belongs to $F^{-1}(s)$ which consists of at most n points and because of $F(\partial\Delta) \subset \partial V$ this gives $F^{-1}(s) \subset V$. So $\gamma_i(t)$ converges to a point $x_i \in V$ as $t \uparrow s$. Because F is a local diffeomorphism x_i cannot be equal to x_j for $i \neq j$, because otherwise $\gamma_i(t) = \gamma_j(t)$ for $t < s$ close to s, a contradiction. So V_0 is closed and therefore $V_0 = V$. This implies that F is a covering map. This means any $y \in V$ has a neighbourhood U such that each component of $F^{-1}(U)$ maps diffeomorphically onto U. To prove this, let $x_1, \dots, x_n$ be the preimages of $y \in V$. Let W_i be a neighbourhood of x_i which is mapped by F diffeomorphically onto $F(W_i)$ and choose these neighbourhoods to be mutually disjoint. Now take $U = \cap F(W_i)$ and let $W_i' = W_i \cap F^{-1}(U)$. So F maps W_i diffeomorphically onto U and since each point of $U \subset V$ has precisely n preimages, $F^{-1}(U) = W_1 \cup \dots \cup W_n$. Hence F is a covering map. Since V is simply connected it follows easily that the restriction of F to each connected component is a homeomorphism. Hence the theorem above implies that $P_{\underline{a}}$, $a \in \Delta$ is a full family. □

The proof of Theorem 4.1, which will occupy the remainder of this section, will use the so-called Thurston map and is based on the fact that there exists no retract from a simplex onto its boundary. There is a widely known alternative way to show that unimodal families are full which is based on a connectedness argument similar to the one used in the proof of Theorem 3.2, see for example Metropolis et al. (1973), Collet and Eckmann (1980) or Van Strien (1987). In the multimodal case the traditional proof does not work because the space of kneading invariants associated to multimodal maps is not ordered. Even in the unimodal case we prefer our proof because it gives an algorithmic way of constructing maps with a certain kneading invariant. This algorithm was used to draw many of the pictures in this book.

Because the proof of this theorem is quite long we have subdivided it into four steps. In Step 1 we shall prove the result in the case that each turning point is eventually periodic under the additional assumption that F is a homeomorphism. In Steps 2, 3 and 4 we shall deal with the general case. Of course, we can collapse each of the components of the basins of inessential periodic attractors of g, the intervals consisting of periodic orbits of constant period and all wandering intervals. Consequently, we may and will assume that g has no inessential periodic attractors, no intervals of periodic points and no wandering intervals. Moreover, by assumption, each periodic turning point of g is attracting. Because of the Corollary below Theorem 3.1, it is enough to show that for each such map g there exists a parameter μ such that the following two properties are satisfied. *Property 1.* The map

$$h\colon \bigcup_{i=1}^{l} \bigcup_{n\geq 0} f_\mu^n(c_i)/\sim \;\to\; \bigcup_{i=1}^{l} \bigcup_{n\geq 0} g^n(\tilde{c}_i) \tag{4.2}$$

defined by $h[f^n(c_i)] = g^n[\tilde{c}_i]$ is an order preserving bijection; here $\sim$ is the (order preserving) equivalence relation from above. (Notice that we do not need to use the equivalence relation on the right hand side of (4.2) because have already collapsed equivalence classes for g. This is why we may say that f_μ and g are *essentially combinatorially equivalent* if (4.2) is satisfied).
Property 2. Whenever c_i, c_j are in the basin of one periodic attractor p of period n, the same holds for $\tilde{c}_i, \tilde{c}_j$ and vice versa; this periodic attractor p is one-sided if and only if the same holds for the corresponding periodic attractor $\tilde{p}$.

Step 1: $F\colon \Delta \to V$ is a homeomorphism and each turning point of g is eventually periodic

Let us first prove the result under the additional assumptions that F is a homeomorphism and that each turning point of g is eventually periodic. In this case the sets from (4.2) have finite cardinality. The main tool in the proof is the pullback argument from the previous section. For simplicity assume that the turning points of g are not mapped into the boundary of I. (If this happens the situation is clear for this turning point.) The parameter μ so that f_μ is combinatorially equivalent to g will be found as a limit of a sequence of parameters $\mu(n)$.

One way to construct these parameters is as follows. Because F is a homeomorphism, there exists a unique parameter $\mu(0)$ such that the values in the critical points of $f_{\mu(0)}$ are the same as those in corresponding critical points of g. Hence there exists a unique orientation preserving homeomorphism h_0 such that $f_{\mu(0)} = g_0 \circ h_0$ where $g_0 = g$. Now write $g_1 = h_0^{-1} \circ g_0 \circ h_0$. Proceeding in this way inductively, we construct homeomorphisms h_n, parameters $\mu(n)$ such that $f_{\mu(n)} = g_n \circ h_n$ and maps $g_{n+1} = h_n^{-1} \circ g_n \circ h_n$ which are topologically conjugate to g. Except in the case that f_μ is a polynomial family (and - as we assumed here - each turning point of g is eventually periodic) it is not known whether this sequence $\mu(n)$ converges, see also Section II.10.

Therefore we shall follow a somewhat different strategy: we shall show that there exists some map g_0 which is merely conjugate to g and for which the procedure from above does converge. More precisely, we shall choose g_0 so that if we set $f_{\mu(0)} = g_0 \circ h_0$ as before then h_0 fixes all iterates of turning points of g_0. It follows that $g_1 = h_0^{-1} \circ g_0 \circ h_0$ has the same extremal values as g_0 and that $\mu(1) = \mu(0)$. Thus we get that $\mu(n)$ is equal to $\mu(0)$ for all n. Because all relevant information about the map g is contained in the forward orbits of the turning points, we shall emphasize the role of these iterates in the proof below.

So let us be more specific. By assumption, the forward orbits of the turning points of g consist of a finite number, say k, distinct points $0 < z_1 < z_2 < \cdots < z_k < 1$ and l of these points are equal to $c_1 < c_2 < \cdots < c_l$. Let $\pi\colon \{1, \dots, k\} \to \{1, \dots, k\}$ be so that $g(z_i) = z_{\pi(i)}$ and let $z_{t(1)} < z_{t(2)} < \cdots < z_{t(l)}$ be the turning points of g. In other words, the graph of g goes through the points $(z_i, z_{\pi(i)}) \in [0,1] \times [0,1]$, $i = 1, \dots, k$. (The relevant information is really contained in the points $(z_i, \pi(z_i))$ which can be thought of as the graph of π. The information we shall use about g is just that it is l-modal and passes through these points. So if we connect the consecutive points by line segments we get a l-modal map as in Figure 4.1 which contains the same combinatorial information as g.) By definition

$$\left(z_{\pi(t(1))}, \dots, z_{\pi(t(l))}\right) \in V.$$

The points z_j correspond to iterates of the turning points of g and this map is assumed to have no inessential periodic attractors. From this one gets the following.

Claim. *For each $m \in \{1, \dots, k\}$ there exists s and $i = 1, \dots, l$ so that either $\pi^s(m+1) \le t(i) \le \pi^s(m)$ or $\pi^s(m) \le t(i) \le \pi^s(m+1)$ (and one of the inequalities is strict).*

Proof of Claim: Since the turning points of g are eventually periodic, the corresponding interval $[z_m, z_{m+1}]$ would otherwise eventually be mapped into a periodic homterval. Since g has no inessential periodic attractors, this is impossible. □

Let us show that there exists a parameter μ for which f_μ is essentially combinatorially equivalent to g. For this let W be the space of points $(x_1, \dots, x_k) \in \mathbb{R}^k$ with $0 < x_1 < x_2 < \cdots < x_k < 1$. Clearly W is a simplex. Unfortunately, this simplex is not closed. Let ∂W be the set of points $(x_1, \dots, x_k) \in \mathbb{R}^k$ with $0 \le x_1 \le x_2 \le \cdots \le x_k \le 1$ such that at least one equality holds. This set is

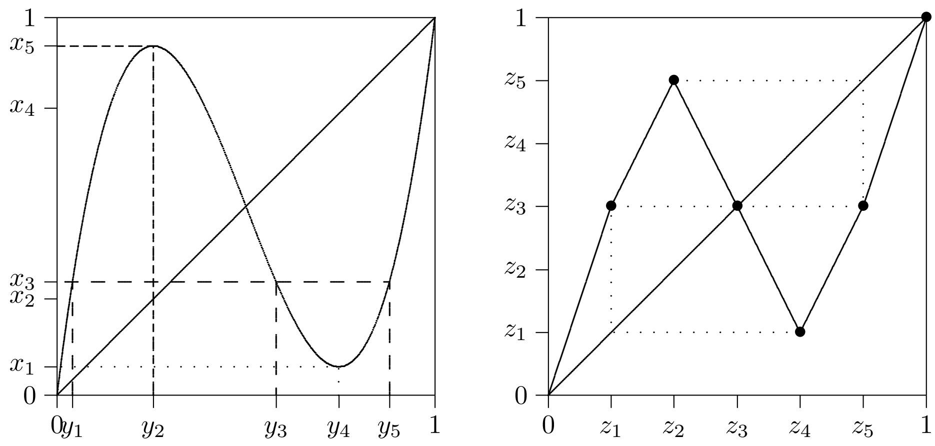

Fig. 4.1: The points $(x_1, \ldots, x_k)$ and the corresponding points $(y_1, \ldots, y_k) = T(x_1, \ldots, x_k)$. Here $\pi(1) = 3$, $\pi(2) = 5$, $\pi(3) = 3$, $\pi(4) = 1$ and $\pi(5) = 3$. Note that the map T depends on the family f_μ and is not dynamical: the second iterate of T in general uses a pullback by two different maps from the family. Only at a fixed point of T the pullback has a dynamical meaning.

the boundary of W and $W \cup \partial W$ is closed. Now take $x = (x_1, \ldots, x_k) \in W$. Since $(z_{\pi(t(1))}, \ldots, z_{\pi(t(l))})$,

$$(x_{\pi(t(1))}, \ldots, x_{\pi(t(l))}) \in V.$$

Since $F\colon \Delta \to V$ is a homeomorphism, there exists therefore a unique parameter $\mu = \mu(x_1, \ldots, x_k)$ such that

$$f_\mu(c_i(\mu)) = x_{\pi(t(i))} \text{ for } i = 1, \ldots, l. \tag{4.3}$$

Such a point $\mu(x_1, \ldots, x_k)$ exists because the map

$$F\colon \Delta \to (f_\mu(c_1(\mu)), \ldots, f_\mu(c_l(\mu))) \in V$$

is surjective and therefore has $(x_{\pi(t(1))}, \ldots, x_{\pi(t(l))})$ as a value. It is unique because we have assumed that F is a homeomorphism which is also the reason that $\mu(x_1, \ldots, x_k)$ depends continuously on $(x_1, \ldots, x_k) \in W$. Furthermore, there exists a unique point $(y_1, \ldots, y_k) \in W$ depending continuously on $(x_1, \ldots, x_k)$ such that a) the $t(j)$-th coordinate of this vector, $y_{t(j)}$, is the j-th turning point of f_μ (and therefore $f_{\mu(x_1,\ldots,x_k)}(y_{t(j)}) = x_{\pi(t(j))}$); b) if $i \notin \{t(1), \ldots, t(l)\}$ then there exists j such that $x_{t(j-1)} < x_i < x_{t(j)}$¿ Then define y_i to be the unique point such that

$$f_{\mu(x_1,\ldots,x_k)}(y_i) = x_{\pi(i)} \tag{4.4}$$

and such that y_i is in the j-th interval of monotonicity of $f_{\mu(x_1,\ldots,x_k)}$, i.e., $y_{t(j-1)} < y_i < y_{t(j)}$, see Figure 4.1. This choice is possible, because by definition of the ordering, $x_{\pi(i)}$ is contained in the interval connecting $x_{\pi(t(j-1))}$

and $x_{\pi(t(j))}$. Since these points are equal to respectively $f_\mu(y_{t(j-1)})$ and $f_\mu(y_{t(j)})$ it follows that one can choose y_i as above.

The map

$$T(x_1, \dots, x_k) = \{(y_1, \dots, y_k) \text{ as in } (4.4)\}$$

defines a continuous single-valued map T from W into itself. We call T the Thurston map associated to the family f_μ.

It suffices to show that T has a fixed point $(x_1, \dots, x_k) \in W$, because then there exists a parameter μ such that $f_\mu(x_i) = x_{\pi(i)}$ and such that $f_\mu(c_i(\mu)) = x_{\pi(t(i))}$. From the choice of the pullback above, it follows that $x_{t(i)} = c_i(\mu)$ and that the forward iterates of $c_1(\mu), \dots, c_l(\mu)$ are ordered the same way in $[0,1]$ as the forward iterates of the turning points of g. Hence f_μ and g are combinatorially equivalent. Note that in this setting we are not primarily interested in finding a good parameter μ but in finding a good realizable 'spatial structure' for the 'ordered structure' of the forward orbit of the turning points. Unfortunately, the existence of such a fixed point in W does not follow immediately from Brouwer's Fixed Point Theorem since W is not closed. Even if T extends continuously to the closed simplex $W \cup \partial W$ and therefore has fixed points in this larger simplex, these fixed points of T could, a priori, all be in ∂W (and therefore of no use). In the exercises below we see that some of the fixed points could, indeed, be in ∂W.

Exercise 4.1. Show that T is not a homeomorphism from W onto W. (Hint: it is not surjective because $y_{t(i)}$ are turning points of f_μ; by a parameter dependent coordinate change we may assume that the positions of the turning points of f_μ do not vary and therefore $y_{t(i)}$ can only take one value c_i. It is also easy to show explicitly that T need not be injective.)

Exercise 4.2. Show that T can have fixed points in ∂W and also periodic points in W. (Hint: take the quadratic family $f_\mu(x) = \mu x(1-x)$. Then $c = 1/2$ is the turning point of f_μ. Let g be some continuous unimodal map such that $g^4(c) = c$ and $g^2(c) < c < g^3(c) < g(c)$. That is, the graph of g goes through points $(x_i, x_{\pi(i)})$ where $0 < x_1 < x_2 < x_3 < x_4 < 1$ and π is simply the permutation on $\{1, \dots, 4\}$ defined by $\pi(1) = 3$, $\pi(2) = 4$, $\pi(3) = 2$ and $\pi(4) = 1$. It is easy to see that there exists a parameter μ such that $f_\mu^2(c) < c = f_\mu^4(c) < f_\mu^3(c) < f_\mu(c)$ and also μ' such that $f_{\mu'}^2(c) = c < f_{\mu'}^3(c) = f_{\mu'}^4(c)$. It follows that the map T from above has a fixed point (x_1, x_2, x_3, x_4) in W but also one of the form (x, x, y, y) in ∂W.)

Exercise 4.3. Exercise 4.3 Show that even if T can be extended to the boundary of W it does not necessarily map ∂W into ∂W. (Hint: if for example precisely two of the points $x_1 < \cdots < x_k$, say $x_{\pi(i)} < x_{\pi(i)+1}$, are extremely close together and if $\pi(i \pm 1) \neq \pi(i) + 1$, then none of the corresponding points $y_1 < \cdots < y_k$ are too close together; this is because these two nearby points are pulled back through different branches of f_μ.)

Exercise 4.4. Suppose that at least one of the turning points of g is in the basin of an inessential periodic attractor. Show that in this case all fixed points of T may be contained in the boundary of W. (Hint: let f_μ be the quadratic family and g be a unimodal map with turning point c such that, for example, $g^2(c) < g^3(c) < g^4(c) =$

$g^8(c) < g^6(c) < c < g^5(c) < g^7(c) < g(c)$. Note that the orbit c is mapped onto the orbit $g^4(c) < g^6(c) < g^5(c) < g^7(c)$. The interval $[g^4(c), g^6(c)]$ is a periodic homterval of period two. Therefore there is no quadratic map for which the orbit of the critical point is ordered in the same way.)

For simplicity we define for $x = (x_1, \ldots, x_k) \in W$,

$$d(x, \partial W) = \min_{i=2,\ldots,k} |x_i - x_{i-1}|.$$

We claim that if T has no fixed point in W then there exists an interior deformation of $id\colon W \to W$ to a non-surjective map. We shall prove this in the following two lemmas.

Lemma 4.1. *For any sequence $x(n) \in W$ converging to some $x \in \partial W$,*

$$\lim_{n\to\infty} \frac{|T(x(n)) - x(n)|}{d(x(n), \partial W)} = \infty.$$

Proof. Suppose by contradiction this is false. Then there exists a sequence $x_n \to \partial W$ and a constant $K < \infty$ such that

$$\frac{|T(x(n)) - x(n)|}{d(x(n), \partial W)} \leq K \tag{4.5}$$

for all n. Writing $x(n) = (x_1(n), \ldots, x_k(n))$ and $(y_1(n), \ldots, y_k(n)) = T(x(n))$. Using (4.5),

$$\begin{aligned} |x_{\pi(i)}(n) - x_{\pi(j)}(n)| &\geq |y_{\pi(i)}(n) - y_{\pi(j)}(n)| \\ &\qquad - |x_{\pi(j)}(n) - y_{\pi(j)}(n)| - |x_{\pi(i)}(n) - y_{\pi(i)}(n)| \\ &\geq |y_{\pi(i)}(n) - y_{\pi(j)}(n)| - 2 \cdot |x(n) - T(x(n))| \\ &\geq |y_{\pi(i)}(n) - y_{\pi(j)}(n)| - 2K \cdot d(x(n), \partial W). \end{aligned} \tag{4.6}$$

By definition, $f_{\mu(n)}(y_i(n)) = x_{\pi(i)}(n)$. Because $x(n) \to x \in \partial W$ and because of (4.5) we have that $y(n) \to x$. From the Mean Value Theorem,

$$\frac{|y_i(n) - y_j(n)|}{|x_{\pi(i)}(n) - x_{\pi(j)}(n)|} = \frac{1}{|Df_{\mu(n)}(z(n))|} \tag{4.7}$$

for some $z(n)$ which is between $y_i(n)$ and $y_j(n)$. Hence taking

$$C = \max_{\mu,x} |Df_\mu(x)|,$$

one gets

$$C \cdot |y_i(n) - y_j(n)| \geq |x_{\pi(i)}(n) - x_{\pi(j)}(n)| \tag{4.8}$$

for all n. Combining this with (4.6) one gets a constant $K_1 < \infty$ with

$$|y_{\pi(i)}(n) - y_{\pi(j)}(n)| \leq C \cdot |y_i(n) - y_j(n)| + K_1 \cdot d(x(n), \partial W).$$

Repeating this one gets for each $s \in \mathbb{N}$ constants $C_s > 0$ and $K_s < \infty$ such that

$$|y_{\pi^s(i)}(n) - y_{\pi^s(i+1)}(n)| \le C_s \cdot |y_i(n) - y_{i+1}(n)| + K_s \cdot d(x(n), \partial W) \tag{4.9}$$

for each $n \in \mathbb{N}$. Take $m \in \{1, \dots, k-1\}$ such that $|y_m(n) - y_{m+1}(n)| = d(y(n), \partial W)$ (of course m might depend on n). Because $x(n) \to x \in \partial W$ and because of (4.5) we have that $y(n) \to \partial W$. So some points collapse in the limit. Let us show that this implies that one of the turning points and one of its neighbours also collapse in the limit. So take the integer s corresponding to m as in the Claim at the beginning of this step. Note that we use here that g has no inessential attractors because otherwise such an integer s might not exist.) Then $y_{\pi^s(m)}(n)$ and $y_{\pi^s(m+1)}(n)$ lie on different sides of a turning point of $f_{\mu(n)}$ and so there exists $r \in \{1, \dots, l\}$ such that either

$$|y_{t(r)}(n) - y_{t(r)+1}(n)| \le |y_{\pi^s(m)}(n) - y_{\pi^s(m+1)}(n)|$$

or

$$|y_{t(r)}(n) - y_{t(r)-1}(n)| \le |y_{\pi^s(m)}(n) - y_{\pi^s(m+1)}(n)|.$$

Let us assume we are in the former case. Then we get from (4.9),

$$\begin{aligned} d(y(n), \partial W) &= |y_m(n) - y_{m+1}(n)| \\ &\ge C_s' \cdot |y_{\pi^s(m)}(n) - y_{\pi^s(m+1)}(n)| - K_s' \cdot d(x(n), \partial W) \\ &\ge C_s' \cdot |y_{t(r)}(n) - y_{t(r)+1}(n)| - K_s' \cdot d(x(n), \partial W). \end{aligned}$$

Because $x(n) \to x \in \partial W$ and because of (4.5) we have that $y(n) \to \partial W$. Hence, using (4.7),

$$\frac{|y_{t(r)}(n) - y_{t(r)+1}(n)|}{|x_{\pi(t(r))}(n) - x_{\pi(t(r)+1)}(n)|} \to \infty$$

because $y_{t(r)}(n)$ is a turning point of $f_{\mu(n)}$. Therefore,

$$\frac{d(y(n), \partial W)}{d(x(n), \partial W)} \ge C_s' \cdot \frac{|y_{t(r)}(n) - y_{t(r)+1}(n)|}{|x_{\pi(t(r))}(n) - x_{\pi(t(r)+1)}(n)|} - K_s' \cdot \frac{d(x(n), \partial W)}{d(x(n), \partial W)} \to \infty.$$

Of course this contradicts (4.5). □

Lemma 4.2. *If $T\colon W \to W$ has no fixed points in the open set W then there exists an interior deformation H_t of $id\colon W \to W$ such that H_1 is not surjective (in other words, the identity map $id\colon W \to W$ is not persistently surjective).*

Proof of Lemma 4.2: The idea of the proof of this lemma is to show that the direction of the vector $T(x) - x$ 'points inward' if $x \in W$ is near ∂W. We should emphasize that we shall not use that T extends continuously to ∂W. In order to construct a deformation as above, we will construct curves in W which are roughly speaking the 'geodesics in terms of a hyperbolic metric' on W; these curves are used to deform the map T. We shall construct these curves by mapping W diffeomorphically onto a ball and considering the curves in W which correspond to straight lines in the ball, see Figure 4.2.

More precisely, let ρ map a stereographic projection from the closure of W homeomorphically onto the closed ball B centred in 0 and with radius l. This map is defined as follows. Since $W \subset B$ there exists z_0 in the interior of $W \cap B$. For $z \in \partial W$, let $\rho(z) \in \partial B$ be the intersection of the infinite ray from z_0 through z with ∂B and then interpolate ρ linearly on each such ray.

We get from Lemma 4.1 that, if $v \in \mathbb{R}^k$ is a limit of

$$\lim_{x(n)\to x} \frac{\rho(T(x(n))) - \rho(x(n))}{|\rho(T(x(n))) - \rho(x(n))|} \tag{4.9}$$

for some sequence $x(n) \in W$ converging to $x \in \partial W$, then v is either tangent to ∂B or points inwards. Since T has no fixed points in W, $\rho(T(z)) \neq \rho(z)$ for all $z \in W$. Let $\psi(z) \in \partial B$ so that $\rho(z)$ lies on the straight line between $\rho(T(z))$ and $\psi(z)$. Now let $\psi_t \colon W \to B$, $t \in [0,1)$ be defined by $\psi_t(z) = \rho(z) + t(\psi(z) - \rho(z))$. Because of (4.9), the vector from $\rho(z)$ to $\rho(T(z))$ points inwards, and therefore $|\rho(z) - \psi(z)| \to 0$ when $z \to \partial W$. Hence ψ_t extends continuously to ∂W and ψ_1 is equal to $\rho(z)$ on ∂W. In particular, $H_t = \rho^{-1} \circ \psi_t \colon W \to W$ is the required deformation which moves $x \in W$ along the 'hyperbolic geodesic' connecting x and $T(x)$ towards the boundary of W. □

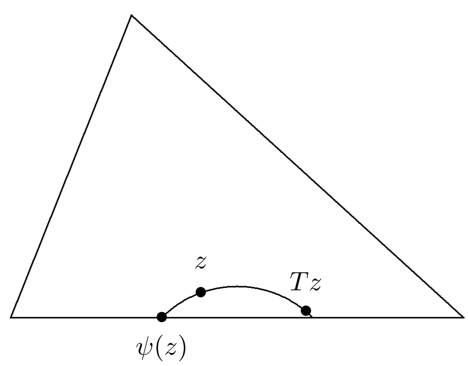

Fig. 4.2: The map ψ from W to ∂W if T has no fixed points.

From the previous lemma it follows that if T has no fixed points then one can find an interior deformation from the identity map on W to a non-surjective map. Since W is a finite-dimensional simplex this is impossible, see for example Dold (1972). This concludes the proof of the theorem in this case.

Step 2: $F \colon \Delta \to V$ is a homeomorphism and each turning point of g is either eventually periodic or belongs to the basin of an essential periodic attractor

Let us now generalize the previous situation and allow g to have periodic attractors. Let us first explain the difficulty. Of course, if one of the turning points is not eventually periodic, then there are infinitely many distinct iterates of turning points and therefore the set W from Step 1 would have to be taken infinite-dimensional. As is well known, the theorem on retracts is rather more delicate in the infinite-dimensional case. Moreover, the method we used to show that the boundary of W is repelling would not work if W is infinite-dimensional. Therefore, we will consider only a finite piece of the orbit of each turning point and use this to make sure that the dynamics in the remaining pieces of the orbits is 'correct'.

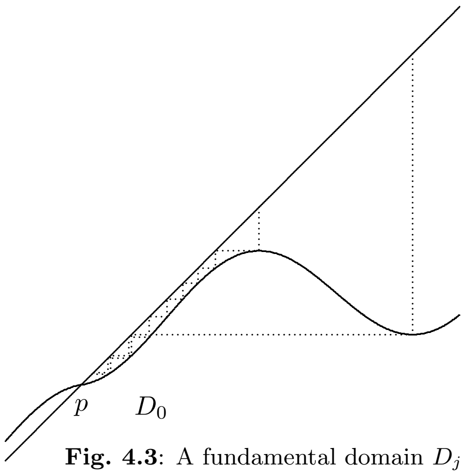

Fig. 4.3: A fundamental domain D_j.

Let $C(g) = \{c_1, \ldots, c_l\}$ and

$$C_k(g) = \bigcup_{i=0}^{k} g^i(C(g)).$$

For each essential periodic orbit choose a point p_j on this orbit such that the component B_j of the immediate basin of this orbit which contains p_j also contains a turning point $z_{a(j)}$. Choose p_j and $z_{a(j)}$ so that all iterates of g are monotone on the interval connecting p_j and $z_{a(j)}$. Moreover, choose a closed subset D_j of B_j so that

1. each boundary point of D_j is an iterate of a turning point of g;

2. the boundary of D_j contains two forward iterates of $z_{a(j)}$; D_j consists of one component if p_j is orientation reversing or a turning point; D_j can consist of one or two components if p_j is orientation preserving (it has two components precisely if p_j attracts different turning points from each side);

3. the forward orbit of each turning points which is attracted to p_j (but not mapped into this orbit) intersects the interior of D_j at most once and the closure of D_j at least once;

4. the smallest interval containing D_j, $z_{a(j)}$ and p_j is a homterval contained in the basin of p_j.

The subset D_j is called the *fundamental domain* of p_j. For convenience, let $\kappa(j)$ be the period of p_j and denote the iterates of $z_{a(j)}$ in the boundary of D_j by

$$z_{\pi^{\sigma(j)\kappa(j)}(a(j))} = z_{r(j)}$$

and

$$z_{\pi^{(\sigma(j)+1)\kappa(j)}(a(j))} = z_{\pi^{\kappa(j)}(r(j))};$$

(In other words $g^{\sigma(j)\kappa(j)}(z_{a(j)}) = z_{r(j)}$ and $g^{(\sigma(j)+1)\kappa(j)}(z_{a(j)}) = g^{\kappa(j)}(z_{r(j)})$.)

Next we take for each turning point c_i of g the smallest positive integer $n(i)$ so that either

$$g^{n(i)}(c_i) \in C_{n(i)-1}(g) \text{ or } g^{n(i)}(c_i) \in \cup(D_j \cup \{p_j\}). \tag{4.10}$$

Then $c_i, \ldots, g^{n(i)}(c_i)$, $i = 1, \ldots, l$ and $p_j, g(p_j), g^2(p_j), \ldots$ consists of a finite number, say k, distinct points $0 < z_1 < z_2 < \cdots < z_k < 1$ such that l of these points are equal to $c_1 < c_2 < \cdots < c_l$. From this choice, there are no points in $\{z_1, \ldots, z_k\}$ between D_j and the periodic attractor p_j and each of the points of the periodic orbit $O(p_j)$ corresponds to a point z_i. Let $p_j = z_{d(j)}$. This defines the space W as before. Note that the map π such that $z_{\pi(i)} = g(z_i)$ from the previous step is only defined on a subset of $\{1, \ldots, k\}$. Let $z_{l(j)}$ be so that $[z_{l(j)}, z_{r(j)}]$ is the smallest interval containing D_j and $z_{d(j)}$. If D_j has two components then $z_{l(j)}$ and $z_{r(j)}$ are the 'external points' of D_j. If D_j has one component then one of the points coincides with $z_{d(j)}$ and the other is in the 'external' boundary of D_j. Note that $\pi(r(j))$, $\pi(d(j))$ and $\pi(l(j))$ are well defined.

Let us now define the Thurston map associated to this family. Let $x \in W$. Choose as before $\mu(x) \in \Delta$ so that $f_{\mu(x)}$ has extremal values $x_{\pi(t(i))}$. Now we define $(y_1, \ldots, y_k)$ as follows. If $z_i \notin \cup_j \text{int}\,(z_{l(j)}, z_{r(j)})$ then we define y_i, as before, to be the unique point with $f_{\mu(x)}(y_i) = x_i$ that lies in the m-th interval of monotonicity of $f_{\mu(x)}$ where m is so that z_i lies in the m-th interval of monotonicity of g. This defines in particular the points $y_{l(j)}$ and $y_{r(j)}$. Next let L_j be the orientation preserving affine map which sends $z_{l(j)}$ and $z_{r(j)}$ to $y_{l(j)}$ and $y_{r(j)}$. Now define y_i when $z_i \in (z_{l(j)}, z_{r(j)})$ by $y_i = L_j(z_i)$. Thus $(y_1, \ldots, y_k)$ is well defined. Now we shall prove the analogue of Lemma 4.1 in this case.

Lemma 4.3. *For any sequence $x(n) \in W$ converging to some $x \in \partial W$,*

$$\lim_{n\to\infty} \frac{|T(x(n)) - x(n)|}{d(x(n), \partial W)} = \infty.$$

Proof. Suppose by contradiction that there exists $K < \infty$ and a sequence $x(n) \to \partial W$ for which

$$\frac{|T(x(n)) - x(n)|}{d(x(n), \partial W)} \le K. \tag{4.11}$$

As before, write $x(n) = (x_1(n), \dots, x_k(n))$ and $(y_1(n), \dots, y_k(n)) = T(x(n))$. As in Lemma 4.1, for each $s \in \mathbb{N}$ there are constants $C_s > 0$ and $K_s < \infty$ such that

$$|y_{\pi^s(i)}(n) - y_{\pi^s(i+1)}(n)| \le C_s \cdot |y_i(n) - y_{i+1}(n)| + K_s \cdot d(x(n), \partial W) \tag{4.12}$$

for each $n \in \mathbb{N}$ provided $z_{\pi^\sigma(i)}, z_{\pi^\sigma(i+1)} \notin \cup(z_{l(j)}, z_{r(j)})$ for $\sigma = 0, \dots, s-1$. Take $m \in \{1, \dots, k-1\}$ such that $|y_m(n) - y_{m+1}(n)| = d(y(n), \partial W)$ (of course m might depend on n). If it is not the case that both z_m and z_{m+1} are contained in some component of a basin of an essential periodic attractor, then (4.11) gives a contradiction exactly as in Lemma 4.1. So we may assume that (z_m, z_{m+1}) is contained in some component of the basin of an essential periodic attractor of g. Take s so that $z_{\pi^s(m)}$ or $z_{\pi^s(m+1)}$ is in $(z_{l(j)}, z_{r(j)})$ with s minimal for some j. From inequality (4.12),

$$|y_{\pi^s(m)}(n) - y_{\pi^s(m+1)}(n)| \le C_s \cdot |y_m(n) - y_{m+1}(n)| + K_s \cdot d(x(n), \partial W).$$

Hence one can find $m' \in \{1, \dots, k\}$ such that $z_{m'}, z_{m'+1} \in [z_{l(j)}, z_{r(j)}]$ with

$$|y_{m'}(n) - y_{m'+1}(n)| \le C_s \cdot |y_m(n) - y_{m+1}(n)| + K_s \cdot d(x(n), \partial W).$$

By definition, $y_{l(j)}, y_{d(j)}, y_{m'}, y_{m'+1}, y_{r(j)}$ are images under f_μ of the corresponding points $z_{l(j)}, z_{d(j)}, z_{m'}, z_{m'+1}, z_{r(j)}$ under an affine map. It follows that there is a universal constant $\tau < \infty$ such that

$$\begin{aligned} |y_i(n) - y_{i'}(n)| &\le \tau \cdot |y_{m'}(n) - y_{m'+1}(n)| \\ &\le \tau \cdot C_s \cdot |y_m(n) - y_{m+1}(n)| + \tau \cdot K_s \cdot d(x(n), \partial W) \\ &\le \tau \cdot C_s \cdot d(y(n), \partial W) + \tau \cdot K_s \cdot d(x(n), \partial W) \end{aligned} \tag{4.13}$$

for each i, i' with $z_i, z_{i'} \in [z_{l(j)}, z_{r(j)}]$. Therefore if (4.11) holds then $y(n) \to \partial W$ and (4.13) gives

$$\begin{aligned} &|y_{\pi^{\sigma(j)\kappa(j)}(a(j))}(n) - y_{\pi^{(\sigma(j)+1)\kappa(j)}(a(j))}(n)| \\ &\quad = |y_{\pi^{\kappa(j)}(r(j))}(n) - y_{r(j)}(n)| \to 0 \end{aligned} \tag{4.14}$$

as $n \to \infty$. We claim that (4.11) implies that either i) the distance of the interval $[y_{l(j)}(n), y_{r(j)}(n)]$ to the turning point $y_{a(j)}$ tends to zero, or, ii) $Df_{\mu(n)}$ tends to zero on the entire interval of monotonicity of $f_{\mu(n)}$ containing $[y_{l(j)}(n), y_{r(j)}(n)]$. Indeed, if the last possibility does not hold then we get by Assumption 4 that $Df_{\mu(n)}$ also does not go to zero on any subinterval of this interval of monotonicity. Now we have for $s = 1, \dots, \sigma(j)+1$,

$$\begin{aligned} &|f_{\mu(n)}(y_{\pi^{(s-1)\kappa(j)}(a(j))}(n)) - f_{\mu(n)}(y_{\pi^{s\kappa(j)}(a(j))}(n))| \\ &\quad = |x_{\pi^{s\kappa(j)}(a(j))}(n) - x_{\pi^{(s+1)\kappa(j)}(a(j))}(n)| \\ &\quad \le |y_{\pi^{s\kappa(j)}(a(j))}(n) - y_{\pi^{(s+1)\kappa(j)}(a(j))}(n)| \ + \ 2d(x(n), y(n)) \end{aligned} \tag{4.15}$$

and by assumption all the points $y_{\pi^{s\kappa(j)}(a(j))}(n)$, $s = 0, \dots, \sigma(j)+1$ are contained in the same interval of monotonicity. Because of (4.14) and (4.11) the terms in

(4.15) converge to zero as $n \to \infty$ when $s = \sigma(j) + 1$. Because $Df_{\mu(n)}$ does not tend to zero on any subinterval of this interval of monotonicity we get

$$\lim_{n\to\infty} |y_{\pi^{s\kappa(j)}(a(j))}(n) - y_{\pi^{(s+1)\kappa(j)}(a(j))}(n)| = 0$$

for $s = \sigma(j)$. Continuing in this way we get

$$\lim_{n\to\infty} |y_{\pi^{(s-1)\kappa(j)}(a(j))}(n) - y_{\pi^{s\kappa(j)}(a(j))}(n)| = 0$$

for all $s = 1, \dots, \sigma(j) + 1$. So all these points are close to the turning point $y_{a(j)}(n)$ and since $|y_{l(j)}(n) - y_{r(j)}(n)| \to 0$ this completes the proof of the claim. Both i) and ii) imply that $Df_{\mu(n)}$ tends to zero on the interval on $[y_{l(j)}(n), y_{r(j)}(n)]$. In particular, it follows from the Mean Value Theorem that

$$\frac{|y_{l(j)}(n) - y_{r(j)}(n)|}{|x_{\pi(l(j))}(n) - x_{\pi(r(j))}(n)|} = \frac{|y_{l(j)}(n) - y_{r(j)}(n)|}{|f_{\mu(n)}(y_{l(j)}(n)) - f_{\mu(n)}(y_{r(j)}(n))|} \to \infty$$

as $n \to 0$. Hence, using (4.13),

$$\frac{d(y(n), \partial W)}{d(x(n), \partial W)} \geq \frac{1}{\tau C_s} \cdot \frac{|y_{l(j)}(n) - y_{r(j)}(n)|}{|x_{\pi(l(j))}(n) - x_{\pi(r(j))}(n)|} - \frac{K_s}{C_s} \cdot \frac{d(x(n), \partial W)}{d(x(n), \partial W)} \to \infty$$

as $n \to \infty$. This clearly contradicts (4.11). □

Using Lemmas 4.3 and 4.2, T has a fixed point in W as before. By construction this implies that there is an order preserving map sending $c_i, \dots, g^{n(i)}(c_i)$ into $c_i(\mu), \dots, f_\mu^{n(i)}(c_i(\mu))$ for each $i = 1, \dots, l$. It remains to be shown that the same holds for the infinite orbits of the turning points of f_μ. But this follows by construction. If $T(y) = y$ then $f_{\mu(y)}(y_i) = y_{\pi(i)}$ whenever $\pi(i)$ is defined except when $z_i = z_{d(j)} \in (z_{l(j)}, z_{r(j)})$ for some j. (So in this case z_i is the attracting point $z_{d(j)}$ and D_j consists of two components. The reason we do not have equality in that case is that $y_{d(j)}$ is then defined as the image under some affine map and not as some preimage under $f_{\mu(y)}$.) So if D_j consists of one component then it follows by construction that $y_{d(j)}$ is a periodic point of $f_{\mu(y)}$ with the same period as $z_{d(j)}$. Let m be this period. It also follows that $f^m_{\mu(y)}$ is monotone on each interval $[y_{d(j)}, y_i]$. This implies that each y_i is attracted to a periodic point $y'_{d(j)}$ in this interval with the same period m or perhaps with period $m/2$ (this can only happen if $f^m_{\mu(y)}$ is orientation reversing in $y_{d(j)}$). In general $y'_{d(j)}$ need not be equal to $y_{d(j)}$ but then $y'_{d(j)} \sim y_{d(j)}$ where $\sim$ is the equivalence relation from above. It follows that f and g are essentially conjugate. If D_j consists of two components then $y_{d(j)}$ is not necessarily a periodic point of $f_{\mu(y)}$ (because, in this case, $T(y) = y$ does not imply $f_{\mu(y)}(y_i) = y_{\pi(i)}$ when $i = d(j)$). Let $n(j)$ be the period of $z_{d(j)}$. Then, since D_j consists of two components, $g^{n(j)}$ is orientation preserving near $z_{d(j)}$ and $g^{n(j)}(z_i) > z_i$ for z_i in one component of D_j and $g^{n(j)}(z_i) < z_i$ for z_i in the other component of D_j. Moreover, $g^{n(j)}$ has no turning points in the interval connecting these two components. Hence, since $T(y) = y$, the same holds for $f^{n(j)}_{\mu(y)}$: and $f^{n(j)}_{\mu(y)}(y_i) > y_i$ for z_i in one component of D_j and $f^{n(j)}_{\mu(y)}(y_i) < y_i$ for z_i in the other component of D_j and $f^{n(j)}$ is monotone on the interval connecting these points. It follows that there exists a periodic attractor $\tilde{y}_{d(j)}$ which attracts all points y_i for which z_i is in one component of D_j and (perhaps a different) periodic attractor $\hat{y}_{d(j)}$ which attracts all points y_i corresponding to z_i in the other component of D_j. Moreover, $f^{n(j)}_{\mu(y)}$ is monotone on $[\tilde{y}_{d(j)}, \hat{y}_{d(j)}]$. It follows that $f_{\mu(y)}$ and g are essentially conjugate.

Remark. We should note that the points y_i with $z_i \in D_j$ are affine images of corresponding points z_i. Moreover, because of the assumption made on the family f_μ at the beginning of this section, the slope of this affine map can be bounded from above and below in terms of $\sigma(j)\kappa(j)$.

Step 3: The proof of Theorem 4.1 under the assumption that $F\colon \Delta \to V$ is a homeomorphism

Let us first construct a sequence of l-modal maps g_n tending to g in the C^0 topology such that g_n satisfies the following two properties:

1. the first n iterates of the turning points of g and g_n coincide;
2. each turning point of g_n is eventually periodic or contained in the basin of an essential periodic attractor.

If g already satisfies Property 2, then we can simply take $g_n = g$. Otherwise let $C(g) = \{c_1, \dots, c_l\}$ and

$$C_k(g) = \bigcup_{i=0}^{k} g^i(C(g)).$$

Since 2) is not satisfied at least one of the turning points, say c_1, is not contained in the basin of an essential periodic attractor and also has the property that $g^k(c_1) \notin C_{k-1}(g)$ for all $k \geq 1$. (Otherwise each turning point which is not contained in the basin of an essential periodic attractor has a finite orbit and therefore is eventually periodic.) Choose $\delta > 0$ so small that distinct points from $C_n(g)$ are at least δ apart and let $\tilde{n}$ be the smallest integer for which the δ neighbourhood of $g^{\tilde{n}}(c_1)$ contains a point x from $C_{\tilde{n}-1}(g)$. Since the sequence $c_1, g(c_1), g^2(c_1), \dots$ has accumulation points, such an integer $\tilde{n}$ exists and by definition of δ one has $\tilde{n} > n$. Next choose a map $\tilde{g}$ which coincides with g except on the δ neighbourhood of $g^{\tilde{n}-1}(c_1)$, so that $\tilde{g}^{\tilde{n}}(c_1) = \tilde{g}(g^{\tilde{n}-1}(c_1))$ is equal to some point $x \in C_{\tilde{n}-1}$. We should emphasize that the first $\tilde{n}-1$ iterates of the turning points of g and $\tilde{g}$ coincide. Next repeat this argument successively for each turning point which is not contained in the basin of an essential periodic attractor and for which $g^k(c_1) \notin C_{k-1}(g)$ for all $k \geq 1$. In this process replace n each time by the previous integer $\tilde{n}$ and g by the previous map $\tilde{g}$. In this way we eventually end up with a map $\hat{g}$ and an integer $\hat{n}$ such that for each turning point c which is not contained in the basin of periodic attractor,

$$\hat{g}^{\hat{n}}(c) \in C_{\hat{n}-1}(\hat{g}).$$

It follows that $\hat{g}$ satisfies Properties 1 and 2. By taking n sufficiently large and δ sufficiently small, we can construct a sequence of l-modal maps g_n tending to g in the C^0 topology with the required properties. Whether or not g_n has wandering intervals will play no role. Also it is not necessary that g_n is differentiable.

From the previous step there exist parameters $\mu(n)$ such that the forward iterates of the turning points of $f_{\mu(n)}$ and g_n are essentially combinatorially equivalent. Let μ' be a limit point of the sequence of $\mu(n)$, (such a limit exists by compactness) and let $f = f_{\mu'}$. Since the first n iterates of the turning points of g and g_n coincide, the fact that g_n and $f_{\mu(n)}$ are essentially combinatorially equivalent implies that if $c_{i(1)}$ and $c_{i(2)}$ are turning points and $g_n^k(c_{i(1)}) < g_n^m(c_{i(2)})$ then $f_{\mu(n)}^k(c_{i(1)}(\mu(n))) < f_{\mu(n)}^m(c_{i(2)}(\mu(n)))$. Therefore, by taking limits we get that $g^k(c_{i(1)}) < g^m(c_{i(2)})$ implies $f^k(c_{i(1)}) \le f^m(c_{i(2)})$. So it remains to show that $f^k(c_{i(1)}) \ne f^m(c_{i(2)})$.

Let us first show that $f^k(c_{i(1)}) \ne f^m(c_{i(2)})$ when $c_{i(1)}$ or $c_{i(2)}$ is contained in the basin of a periodic attractor p_j with period $\sigma(j)$ and let $\kappa(j)$ be as in the previous step. This is easy. Indeed, g_n is equal to g on the basin of periodic attractors. If $f^k(c_{i(1)}) = f^m(c_{i(2)})$ then

$$\lim_{n\to\infty} |f_{\mu(n)}^k(c_{i(1)}(\mu(n))) - f_{\mu(n)}^m(c_{i(2)}(\mu(n)))| \to 0$$

and so the iterates of these turning points which are inside D_j also tend to each other. But because of the remark at the end of the previous step and since $\sigma(j)\kappa(j)$ does not depend on n, this is impossible.

So let us assume that $c_{i(1)}$ or $c_{i(2)}$ is not contained in the basin of a periodic attractor. By contradiction assume that $f^k(c_{i(1)}) = f^m(c_{i(2)})$. Since $g^k(c_{i(1)}) \ne g^m(c_{i(2)})$ we have $g_n^k(c_{i(1)}) \ne g_n^m(c_{i(2)})$ for n sufficiently large. Therefore, by the choice of $\mu(n)$, $f_{\mu(n)}^k(c_{i(1)}) \ne f_{\mu(n)}^m(c_{i(2)})$ while

$$\lim_{n\to\infty} f_{\mu(n)}^k(c_{i(1)}) = \lim_{n\to\infty} f_{\mu(n)}^m(c_{i(2)}).$$

Let J denote the open interval connecting $g^k(c_{i(1)})$ and $g^m(c_{i(2)})$. Since g has no wandering intervals and $c_{i(1)}$ is not contained in the basin of a periodic attractor, there exists $k' \ge 0$ so that $g^{k'}(J)$ contains a turning point. So if J_n denotes the open interval connecting $g_n^k(c_{i(1)})$ and $g_n^m(c_{i(2)})$, then $g_n^{k'}(J_n)$ contains also a turning point for n sufficiently large. Since g_n and $f_{\mu(n)}$ are essentially combinatorially equivalent, for the corresponding segment $\tilde{J}_n$ connecting $f_{\mu(n)}^k(c_{i(1)})$ and $f_{\mu(n)}^m(c_{i(2)})$ one has again that $f_{\mu(n)}^{k'}(J_n)$ contains a turning point. Since the length of this segment $\tilde{J}_n$ tends to zero as $n \to \infty$, this implies that $f^{k'+m}(c_{i(2)}) = f^{k'+k}(c_{i(1)})$ is a turning point $c_{i(3)}$ of f. Since $f_{\mu(n)}^k(c_{i(1)}) \ne f_{\mu(n)}^m(c_{i(2)})$ and $\lim_{n\to\infty} f_{\mu(n)}^k(c_{i(1)}) = \lim_{n\to\infty} f_{\mu(n)}^m(c_{i(2)})$, for all n large either $f_{\mu(n)}^{k'+k}(c_{i(1)}) \ne c_{i(3)}$ or $f_{\mu(n)}^{k'+m}(c_{i(2)}) \ne c_{i(3)}$ and

$$\lim_{n\to\infty} f_{\mu(n)}^{k'+k}(c_{i(1)}) = \lim_{n\to\infty} f_{\mu(n)}^{k'+m}(c_{i(2)}) = c_{i(3)}. \tag{4.16}$$

By the construction in Step 2a, $g^{k'+k}(c_{i(1)}) = c_{i(3)}$ (respectively, $g^{k'+m}(c_{i(2)}) = c_{i(3)}$) implies that for all n, $g_n^{k'+k}(c_{i(1)}) = c_{i(3)}$ (and similarly $g_n^{k'+m}(c_{i(2)}) =$

$c_{i(3)})$. Since either $f_{\mu(n)}^{k'+k}(c_{i(1)}) \neq c_{i(3)}$ or $f_{\mu(n)}^{k'+m}(c_{i(2)}) \neq c_{i(3)}$ this implies that either $g^{k'+k}(c_{i(1)}) \neq c_{i(3)}$ or $g^{k'+m}(c_{i(2)}) \neq c_{i(3)}$. Therefore, as before, the interval connecting either $f_{\mu(n)}^{k'+k}(c_{i(1)})$ or $f_{\mu(n)}^{k'+m}(c_{i(2)})$ to $c_{i(3)}$, is mapped after a finite number of iterates over a turning point $c_{i(4)}$. By (4.16) it follows that this iterate of f maps $c_{i(3)}$ into a turning point $c_{i(4)}$. Continuing in this way, it follows that one of the turning points of f is periodic and, since f has 'derivative' zero at turning points, therefore attracting. Moreover, $c_{i(1)}$ and $c_{i(2)}$ are both contained in the basin of this periodic attractor, which gives a contradiction.

Step 4: The proof of Theorem 4.1

Let us now drop the assumption that F is a homeomorphism. The main difference with the proof in the previous steps is that there can be many parameters μ which satisfy (4.2). More precisely, the map T becomes multi-valued when considered as a map from W into W. Even so, we are looking for a 'fixed point' of T, i.e., a point $z \in W$ for which $z \in T(z)$.

If all periodic points of g are eventually periodic then take z_i as in Step 1. To start with, let V_0 be the set of point $(v_1, \dots, v_l) \in V$ which are ordered in the same way as $z_{\pi(t(1))}, \dots, z_{\pi(t(l))}$. This means that the values at the turning points have the right ordering. Let $\Delta_0 = F^{-1}(V_0)$ and let F_0 denote the restriction of F to Δ_0. If all the integers $\pi(t(i))$ are distinct (or equivalently, if all the points $z_{\pi(t(i))}$ are distinct) then V_0 is an open subset of V and otherwise it is an open subset of a linear subspace of V. In the latter case we shall denote by ∂V_0 the boundary of V_0 as a subset of this linear subspace of V. Furthermore, define

$$\Sigma_{(v_1,\dots,v_l)} = \{x = (x_1, \dots, x_k)\,;\ 0 < x_1 < \cdots < x_k < 1 \text{ and } x_{\pi(t(i))} = v_i \ \text{ for } 1 \le i \le l\}$$

and

$$Z = \{(\mu, x)\,;\ \mu \in \Delta_0 \text{ and } x \in \Sigma_{F(\mu)}\}.$$

Note that for $\mu \in \Delta_0$ we get $\Sigma_{F(\mu)} \subset W$ and that for $\mu \notin \Delta_0$ the definition of $\Sigma_{F(\mu)}$ does not even make sense, i.e., $\Sigma_{F(\mu)}$ is the empty set. The multi-valued map $T\colon W \to W$ from above can be considered as a continuous single-valued map from Z to W. Indeed, if $(\mu, x) \in Z$ then $x \in \Sigma_{F(\mu)}$ and therefore we have, as in (4.1),

$$f_\mu(c_i(\mu)) = x_{\pi(t(i))}.$$

It follows that $T\colon Z \to W$ can be defined exactly as before by

$$(y_1, \dots, y_k) = T(\mu, x_1, \dots, x_k)$$

where

$$(y_1, \dots, y_k) \in W \text{ is such that } f_\mu(y_i) = x_{\pi(i)}$$

where y_i belongs to the j-th interval of monotonicity of f_μ if z_i belongs the j-th interval of monotonicity of g.

If some turning point of g is not eventually periodic but in the basin of some essential periodic attractor, then define z_i as in Step 2. Take Z as above and then define $T: Z \to W$ again as in Step 2.

Proof of Theorem 4.1: We are looking for $(\mu, x) \in Z$ with x in the interior of $\Sigma_{F(\mu)}$ which is a 'fixed point' in the sense that $T(\mu, x) = x$. As in Lemmas 4.1 to Lemma 4.3, one proves that if such a 'fixed point' does not exist then there exists an interior deformation $\psi_t: Z \to W$ from the projection $(\mu, x) \mapsto x$ to a map which sends Z to ∂W. Indeed, if $T(\mu, x) \neq x$ for all $(\mu, x) \in Z$ then we can define $\psi_t(\mu, x)$ to be on the 'hyperbolic geodesic' through x and $T(\mu, x)$ such that x is between $\psi_t(\mu, x)$ and $T(\mu, x)$. This is done exactly as in the proof of Lemma 4.2. Because the estimates of Lemmas 4.1 and 4.3 still hold it follows also as in Lemma 4.2 that $\psi_t: Z \to W$ extends continuously to the boundary of Z and that ψ_1 is the projection map $(\mu, x) \to x$.

Let $\hat{F}: Z \to W$ be the projection

$$\hat{F}(\mu, x) = x.$$

If T has no fixed point then

$$\hat{F}_t(\mu, x) = \psi_t(\mu, \hat{F}(\mu, x))$$

defines an interior deformation of $\hat{F}: Z \to W$ to a map sending Z to ∂W. So in order to show that T has 'fixed points', it remains to show that $\hat{F}: Z \to W$ is persistently surjective.

To prove this, let us first show that $F: \Delta_0 \to V_0$ is persistently surjective. So assume by contradiction that this is not the case. Then there exists an interior deformation F_t of $F: \Delta_0 \to V_0$ to a map sending Δ_0 to the boundary of V_0. If V_0 is an open set then $G_t: \Delta \to V$ defined by

$$G_t = \begin{cases} F_t & \text{on } \Delta_0 \\ F & \text{outside } \Delta_0 \end{cases}$$

is an interior deformation of $F: \Delta \to V$. We shall extend this deformation by taking a point p in the interior of V_0 and defining for each $x \in V \setminus \{p\}$ the arc $[0, 1] \mapsto \phi_t(x)$ which starts in x and moves x with constant speed towards ∂W along the ray starting at p and going through x. Clearly ϕ_t depends continuously on t and since $G_1(\Delta) \subset \mathrm{cl}\,(V \setminus V_0)$, it follows that

$$F_t = \begin{cases} G_{2t} & \text{for } t \in [0, 1/2] \\ \phi_{2t-1} \circ G_1 & \text{for } t \in [1/2, 1] \end{cases}$$

is an interior deformation of $F: \Delta \to V$ to a map sending Δ to the boundary of V, contradicting the assumption made on F in the statement of the theorem. If V_0 is a linear subspace of V then one can define G_t similarly by extending $F_t: \Delta_0 \to V_0$ to a small neighbourhood U of Δ_0 so that $F_t: U \setminus \Delta_0 \to V \setminus V_0$ is a well defined deformation. Hence the theorem follows from the next lemma. $\square$

Lemma 4.4. *If $F\colon \Delta_0 \to V_0$ is persistently surjective then so is $\hat{F}\colon Z \to W$.*

Proof. Let $\mathrm{pr}\colon W \to V_0$ be defined by $\mathrm{pr}(x_1,\dots,x_k) = (x_{\pi(1)},\dots,x_{\pi(l)})$. First we claim that for each $(v_1,\dots,v_l) \in \mathrm{int}\,(V_0)$, there exists a continuous map

$$H_{(v_1,\dots,v_l)}: \{w \in W\,;\ \mathrm{pr}(w) \in \mathrm{int}\,(V_0)\} \ \to\ \Sigma_{(v_1,\dots,v_l)}$$

which maps ∂W into the boundary of $\Sigma_{(v_1,\dots,v_l)}$ and which depends continuously on $(v_1,\dots,v_l) \in \mathrm{int}\,(V_0)$. Indeed, using a piecewise affine homeomorphism on $[0,1]$ we can send adjacent points from $\{x_{\pi(t(1))},\dots,x_{\pi(t(l))}\}$ to adjacent points from $\{v_1,\dots,v_l\}$. This homeomorphism induces the map from above. Now assume by contradiction that there exists an interior deformation of $\hat{F}_t\colon Z \to W$ from $\hat{F}$ to a map which sends Z into the boundary of W. If there exists $\mu \in \Delta_0$ such that $\mathrm{pr} \circ \hat{F}_1(\mu \times \Sigma_{F(\mu)}) \not\subset \partial V_0$ then, by the definition of an interior deformation, we get $\mathrm{pr} \circ \hat{F}_t(\mu \times \Sigma_{F(\mu)}) \not\subset \partial V_0$ for each $t \in [0,1)$. Hence, $\tilde{F}_t\colon \Sigma_{F(\mu)} \to \Sigma_{F(\mu)}$ defined by

$$\tilde{F}_t(x) := H_{\mathrm{pr}(\hat{F}_t(\mu,x))} \circ \hat{F}_t(\mu,x)$$

is an interior deformation and one has $\tilde{F}_0 = id$. But since $\hat{F}_1$ maps Z into ∂W one gets that $\tilde{F}_1(\Sigma_{F(\mu)}) \subset \partial\Sigma_{F(\mu)}$ and therefore $\tilde{F}_t$ is an interior deformation from the identity map on $\Sigma_{F(\mu)}$ to a map sending this space to its boundary. This is impossible because $\Sigma_{F(\mu)}$ is a finite-dimensional simplex as before. It follows that for each $\mu \in \Delta_0$, $\mathrm{pr} \circ \hat{F}_1(\mu \times \Sigma_{F(\mu)}) \in \partial V_0$. Take a continuous function $w\colon \Delta \to \Sigma_{F(\mu)}$ such that for all $j = 1,\dots,l$ all the coordinates of $w_i(\mu)$ of $w(\mu)$ with $w_i(\mu) \in (v_j, v_{j+1})$ are equally spaced in this interval; here $(v_1,\dots,v_l) = F(\mu)$, Then $\tilde{F}_t\colon \Delta_0 \to V_0$ defined by $\tilde{F}_t(\mu) = \mathrm{pr} \circ \hat{F}_t(\mu, w(\mu))$ is an interior deformation to a map sending Δ_0 to ∂V_0. Therefore $F\colon \Delta_0 \to V_0$ is not persistently surjective, a contradiction. □

Remark. In Section II.10 we shall show that the Thurston map T associated to the quadratic family $f_\mu = \mu x(1-x)$ is a contraction. This will imply that if the turning point of f_μ is eventually periodic then there exists no parameter $\mu' \neq \mu$ such that $f_{\mu'}$ is combinatorially equivalent to f_μ. This result uses ideas from Milnor, Douady, Hubbard and Sullivan, see Milnor (1983). In Section VI.4 we shall give a second proof of this last statement. This second proof is due to Sullivan and uses quasiconformal deformations. Whether this result also holds for more general families of maps is still an open question.

Exercise 4.5. Write a computer program which finds, given a l-modal map g for which the orbits of the turning points are finite, a polynomial l-modal map P which is essentially conjugate to g. (Hint: because of Remark 5 at the beginning of this section, the map $F\colon \Delta \to V$ corresponding to the canonical l-parameter family of polynomials $P_{\underline{a}}$ is a homeomorphism. Therefore the appropriate map $P_{\underline{a}}$ can be 'constructed' as in Step 1. Because of the results in Section 10 of this chapter the associated Thurston map is a contraction and therefore its fixed point can be found by Picard iteration.) Most pictures in this book were made with such a program.

5 Families of Maps and Renormalization

In this section we shall introduce the concept of renormalization. In particular, we shall show that many maps from a full family of multimodal maps are 'infinitely renormalizable'.

5.a: Restrictive intervals

A piecewise monotone map can sometimes be decomposed into 'smaller' pieces. To formalize this notion we shall introduce the notion of restrictive intervals.

Definition. Let $f\colon I \to I$ be a multimodal map. A closed proper subinterval J of I is called *restrictive* with *period* $n \geq 1$ for f if

1. the interiors of $J, \dots, f^{n-1}(J)$ are disjoint;
2. $f^n(J) \subset J,\ f^n(\partial J) \subset \partial J$;
3. at least one of the intervals $J, \dots, f^{n-1}(J)$ contains a turning point;
4. J is maximal with respect to these properties: if $J' \supset J$ is a closed interval which is strictly contained in I and such that the previous properties also hold for J' (for the same integer n) then $J' = J$.

In the unimodal case, restrictive intervals are also called *central*. We say that J is a *maximal restrictive interval* if there exists no restrictive interval J' (whose period might be different from the period of J) which strictly contains J. The

reason to introduce this notion is that it allows us to consider pieces of the dynamics on a finer scale:

Definition. The map $f^n\colon J \to J$ is called the *return map* or the *renormalization* of f to J. If $\phi\colon J \to I$ is an affine map sending J onto I then

$$f \mapsto \mathcal{R}(f;J) = \phi \circ f^{n(f)} \circ \phi^{-1}\colon I \to I$$

is called the *renormalization operator* associated to J. (Of course, there are two such maps ϕ with opposite orientations; if the original map is unimodal then one could choose the orientation of ϕ so that the unimodal map $\mathcal{R}(f;J)$ is increasing on the left lap.) If such a restrictive interval exists of period ≥ 2, then f is called *renormalizable*. The class of maps which are renormalizable is denoted by $\mathcal{D}$.

It is easy to see that if $f\colon I \to I$ is l-modal and $f^n\colon J \to J$ is the renormalization of a restrictive interval J with period n, then $f^n\colon J \to J$ has at least one and at most $2^l - 1$ turning points. In particular, if f is unimodal and J is a restrictive interval which contains the turning point then $f^n\colon J \to J$ is a unimodal map and f folds J onto $f(J)$ and maps $f^i(J)$ homeomorphically onto $f^{i+1}(J)$ for $i = 1, \ldots, n-1$.

As we shall see in Section III.4 and V.6, many maps have no restrictive intervals. On the other hand, we shall construct in this section a large class of maps which have infinitely many restrictive intervals.

Let us first give an example which shows that the return map associated to a restrictive interval of a l-modal map need not be l-modal.

Example. Let $f\colon [0,1] \to [0,1]$ be a bimodal map as in Figure 5.1. Clearly, the restrictive interval J has period 2 and $f^2\colon J \to J$ is 3-modal. Of course, this might indicate that the modality of a return map of $f^2\colon J \to J$ to a restrictive interval $J_1 \subset J$ could again be larger. However, as we indicated above this is not the case: the modality of these return maps is bounded by $2^l - 1$ where in this case $l = 2$. Note also that the 3-modal map $f^2\colon J \to J$ is of a very special form: there exists an involution $\tau\colon J \to J$ such that $f^2 \circ \tau = f^2$.

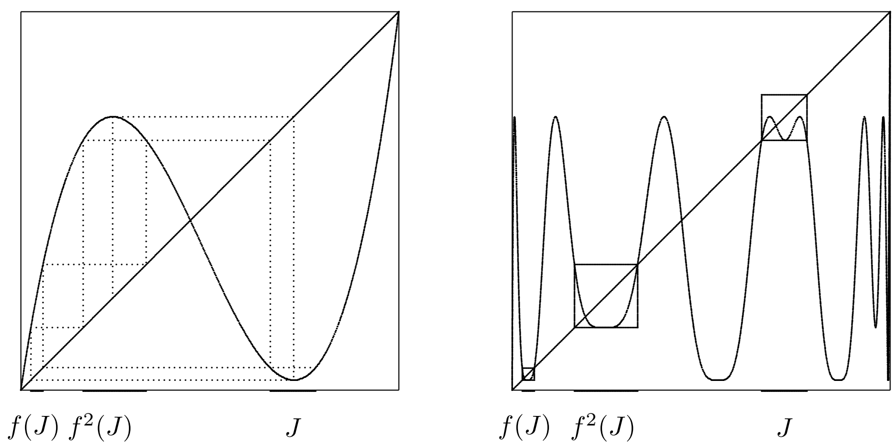

Fig. 5.1: J is a restrictive interval for the bimodal map f. The return map of J is 3-modal, whereas the return map to $f^2(J)$ is unimodal.

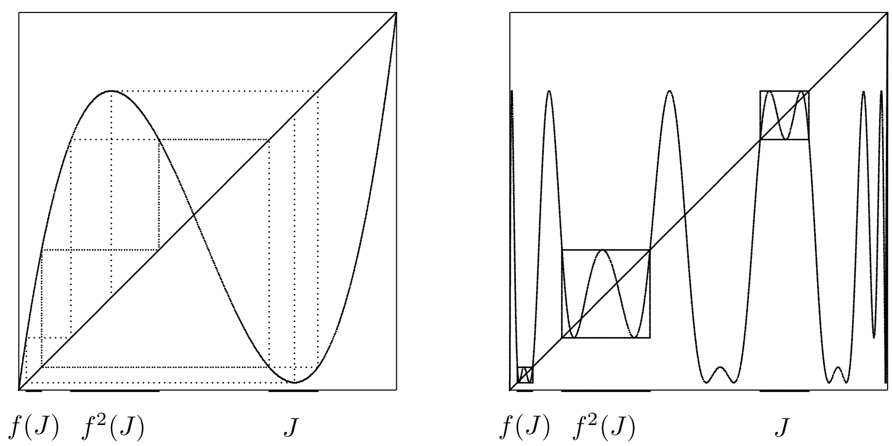

Fig. 5.2: J is a restrictive interval for the bimodal map f. The return map of J is 3-modal. Notice that the return map to J is surjective: therefore some maps arbitrarily near f have no restrictive interval of period three.

Sometimes, and this will be done in the last chapter, it is more natural to consider renormalizations of f to periodic intervals which are minimal in the following sense:

Definition. We say that J is a *unimodal interval* for some interval map g if J contains precisely one turning point of g, $g(J) = J$ and if no subinterval of J has these properties. This implies that $J = [g^2(c), g(c)]$ for some turning point c of g. We say that J is a periodic unimodal interval for f of period n if it is a unimodal interval for f^n.

Let us show 1) how to construct restrictive intervals, 2) that the first n iterates of a unimodal interval of period n are automatically disjoint and 3) that restrictive intervals define a 'filtration' of the space.

Lemma 5.1. *Let $f : I \to I$ be l-modal.*

1. *If $n \geq 2$ and J is an interval such that $f^n(J) \subset J$, such that $J, \dots, f^{n-1}(J)$ are disjoint and such that one of these intervals contains a turning point then J is contained in a restrictive interval of the same period.*
2. *If J is a unimodal interval of period n then $J, \dots, f^{n-1}(J)$ are pairwise disjoint and one of the intervals $f^k(J)$ is of the form $[f^{2n}(c), f^n(c)]$ where c is a turning point of f.*
3. *If J and J' are restrictive intervals whose interiors have a non-empty intersection and with periods ≥ 2 then one of these intervals is contained in the other. In particular, f has at most l different orbits of maximal restrictive intervals. Moreover, if J and J' are maximal restrictive intervals of period $n, n' > 1$ with precisely one common point then $n = n' = 2$ and $f(J) \subset J'$.*
4. *If f has no wandering intervals then there exists $N < \infty$ such that any restrictive interval J of period $\geq N$ contains only one turning point.*

Remark. 1. We should emphasize that a restrictive interval J of period n can be contained in a strictly larger restrictive interval J'. In that case J' has period m and n/m is an integer ≥ 2. 2. If f is a C^1 map with a turning point c of period n then Statement 1) of Lemma 5.1 implies that c is contained in a restrictive interval of period n.

Proof of Lemma 5.1: First we claim that if J is a closed interval with the property that

$$f^i(J) \text{ and } f^n(J) \text{ have no interior point in common} \tag{5.1}$$

for $i = 1, \dots, n-1$ then $f^i(J)$ and $f^j(J)$ have no interior point in common for $0 \le i < j < n$. Indeed, otherwise $f^{n-j+i}(J)$ and $f^n(J)$ would also have an interior point in common, contradicting (5.1).

To prove Statement 1, let us first assume that J is such that (5.1) and

$$f^n(J) \subset J, \tag{5.2}$$

are satisfied and such that there exists no larger closed interval J' for which (5.1) and (5.2) are also satisfied. We claim that $f^n(\partial J) \subset \partial J$. Indeed, suppose by contradiction that

$$p \in \partial J \text{ and } f^n(p) \in \operatorname{int}(J). \tag{5.3}$$

This implies that there exists a small neighbourhood V of p such that (5.2) still holds for $V \cup J$. By the maximality of J the interval $V \cup J$ cannot satisfy (5.1) for any neighbourhood V of p. This and (5.1) implies that there exists $0 \le i < n$ with $q = f^i(p) \in \partial f^n(J)$. Since $f^i(J) \cap J = \emptyset$ and $f^n(J) \subset J$ this implies $q \in \partial J$. But then, because of (5.3), $f^{n-i}(q) = f^n(p) \in \operatorname{int}(J)$ which contradicts (5.1). This completes the proof of the claim and of Statement 1.

Let us now prove Statement 2. So let J be a unimodal interval of period n. If $f^i(J) \cap f^n(J) \neq \emptyset$ then this interval is invariant. However, by looking at the graph of f^n on $J \cup f^i(J)$ and using that J and $f^i(J)$ are both n-periodic unimodal intervals one sees immediately that J and $f^i(J)$ can only have a common boundary point. (If this happens then f^n maps J surjectively onto itself.) Using the first part of the proof, the second statement follows.

So let us prove the third assertion and assume that J, J' are both restrictive intervals with periods $n \le n'$. Then each of the two collections $J, \dots, f^{n-1}(J)$ and $J', \dots, f^{n'-1}(J')$ consist of intervals with disjoint interiors and $f^n(J) \subset J$ and $f^{n'}(J') \subset J'$. Therefore, if one of the intervals $f^i(J')$ intersects precisely s of the intervals $J, \dots, f^{n-1}(J)$ then the same holds for all of the intervals $J', \dots, f^{n'-1}(J')$. So if one of the intervals $J', \dots, f^{n'-1}(J')$ contains a component of $I \setminus (J \cup \cdots \cup f^{n-1}(J))$ then each of the intervals $J', \dots, f^{n'-1}(J')$ intersects precisely $s \ge 3$ of the intervals $J, \dots, f^n(J)$. But since both the first and the last collection consists of disjoint intervals (in the sense explained above), this implies that the two intervals from the first collection $J, \dots, f^{n-1}(J)$ which are situated most to the left and most to the right intersect at most one of the intervals $J', \dots, f^{n'-1}(J')$. Therefore $n' = n - 1 < n$, a contradiction. So each of the intervals from the collection $J', \dots, f^{n'-1}(J')$ intersects at most one of the intervals from the collection $J, \dots, f^{n-1}(J)$. Hence, if T_i denotes the union of $f^i(J)$ with the intervals $J', \dots, f^{n'-1}(J')$ which intersect $f^i(J)$, then $T_0, \dots, T_{n-1}$ are all disjoint. In particular, the orbit $T_0, \dots, f^{n-1}(T_0)$ consists of disjoint intervals and $f^n(T_0) \subset T_0$. From the maximality of J it follows that $T_0 = J$ and if J' has a non-empty intersection with J then $J' \subset T_0 = J$ (unless perhaps $T_0 = I$ but then $n = 1$ and $n' = 2$). Next assume that J and J' are maximal restrictive intervals of period $1 < n \le n'$ with precisely one common point. Then either $n' = n$ or $n' = 2n$ in which case both endpoints of J have period n and $f^n(J')$ also has a common point with J. Moreover, if some iterate of J intersects the interior of J' then n is even and $f^{n/2}(J) \subset J'$. If $n = n' \geq 3$ or if $n = n' = 2$ and no iterate of J intersects the interior of J' then the first part of this theorem implies that $\hat{J} = J \cup J'$ is also contained in a restrictive interval of period n, contradicting the maximality of J. If $n' = 2n$ then it follows similarly that $\hat{J} = J \cup J' \cup f^n(J')$ is contained in a restrictive interval of period n and again we get a contradiction. This concludes the proof of the third assertion.

The last statement is obvious: from the third statement there exists otherwise a nested sequence of restrictive intervals J_n containing two turning points and with period $n(k) \to \infty$. Then $T = \cap J_n$ is a non-trivial interval, all forward iterates of T are disjoint and T is not contained in the basin of a periodic attractor. Hence T is a wandering interval, a contradiction. □

Corollary 5.1. *If f and g are two l-modal maps with turning points $c_1, \dots, c_l$ and $\tilde{c}_1, \dots, \tilde{c}_l$ respectively which are essentially conjugate then they have the same number of restrictive intervals of period ≥ 2. If J is a restrictive interval of period $n \geq 2$ for f then there exists a restrictive interval $\tilde{J}$ of period n for g such that for any $i = 1, \dots, l$ and any $k \in \mathbb{N}$*

$$f^k(c_i) \in J \text{ if and only if } g^k(c_i) \in \tilde{J}.$$

Proof. Let $h\colon I/\sim \to I/\sim$ be the map sending orbits of turning points of f to orbits of turning points of g (modulo the equivalence relation of being in the basin of an inessential attractor). Let J be a restrictive interval of f with period n and let J' be the smallest interval in J containing all forward iterates of turning points which are contained in J. It follows that $J', \dots, f^{n-1}(J')$ are disjoint and $f^n(J') \subset J'$. From the definition of a restrictive interval, we even get that the boundary points of $f^i(J'), f^j(J')$ cannot be attracted to the same periodic orbit for $0 \leq i < j < n$. Hence the equivalence relation $\sim$ does not identify points in these intervals. Therefore, taking $\tilde{J} = h(J')$ one has that $\tilde{J}', \dots, g^{n-1}(\tilde{J}')$ are disjoint and $g^n(\tilde{J}') \subset \tilde{J}'$. From Statement 1 of Lemma 5.1 it follows that J' is contained in a restrictive interval of period n. □

As we will see in Section III.4, the dynamics of points which do not enter these restrictive intervals is quite simple to describe. Hence, renormalization is a very natural and powerful tool to decompose the dynamics of a map into simpler pieces.

Exercise 5.1. Show that if J is a restrictive interval of f with period $n > 2$ then $J, \dots, f^{n-1}(J)$ are disjoint.

5.b: Renormalizations within full families

Take a family $f_\mu\colon I \to I$, $\mu \in \Delta$ of l-modal maps satisfying the assumptions made at the beginning of Section 4 and with a restrictive interval J_μ of period n for each $\mu \in \Delta' \subset \Delta$. First we want to define when $f^n\colon J_\mu \to J_\mu$, $\mu \in \Delta'$, is

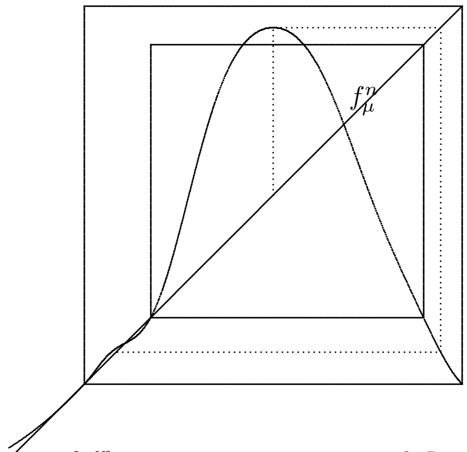

Fig. 5.3: The restriction of f_μ^n to some restrictive interval J_μ is drawn for $\mu = 0$. In this case, J_μ may not depend continuously on the parameter because the graph on f_μ^n is tangent to the diagonal at one of the boundary points of J_μ for $\mu = 0$.

a 'full family'. The first complication in the definition of this notion is that, in general, a restrictive interval does not necessarily depend continuously on the parameter:

Example. Let $f_\mu \colon [0,1] \to [0,1]$, be a one-parameter family of unimodal maps depending continuously on the parameter μ such that f_0 is as in Figure 5.3. The map f_μ has a restrictive interval J_μ of period n for each μ near 0. However, the (endpoints of the) intervals J_μ do not depend continuously on the parameter: $\lim_{\mu \downarrow 0} J_\mu$ strictly contains J_0 because for $\mu > 0$ 'one of the periodic points has disappeared'. In addition, the dependence of the fixed points of f_μ^n on the

parameter can be discontinuous if μ tends to 0 if $f_0^n(x)$ has an interval consisting of fixed points. For this reason we shall assume that

$$\text{the number of fixed points of } f_\mu^n \text{ is finite for each } \mu \text{ and each } n \in \mathbb{N}. \tag{5.4}$$

Because a restrictive interval does not depend continuously on the parameter, we will now consider only subsets $\Delta' \subset \Delta$ such that there exists n such that f_μ has a restrictive interval J_μ for each $\mu \in \Delta'$ of period n which is continuous in the sense that

$$\begin{aligned} &\#\left(f^i(J_\mu) \cap TP(f_\mu)\right) \text{ is independent of } \mu \in \Delta' \text{ for each } 0 \le i < n \\ &\text{and each point in this set varies continuously with } \mu \in \Delta'. \end{aligned} \tag{5.5}$$

Here $TP(f_\mu)$ denotes the set of turning points of f_μ. This assumption implies that the modality of $f_\mu^n \colon J_\mu \to J_\mu$ is constant as μ varies in Δ'. Next, as is shown in Example 5.1, return maps have a special structure. Hence the following

Definition. Let $f_\mu\colon I \to I$, $\mu \in \Delta$, be a family as above with a restrictive interval J_μ for each $\mu \in \Delta'$. We say that $f_\mu^n\colon J_\mu \to J_\mu$, $\mu \in \Delta'$ is a *full family of renormalized maps* if for each 'reasonable' $g\colon I \to I$ there exists $\mu \in \Delta'$ such that f_μ^n is essentially conjugate to g. Here g is *reasonable* (for this restrictive interval) if it is of the form $g_{n-1} \circ \cdots \circ g_0$ where $g_i\colon I \to I$ is a continuous map satisfying $g_i(\partial I) \subset \partial I$ and having the same orientation and modality as $f_\mu\colon f_\mu^i(J_\mu) \to f_\mu^{i+1}(J_\mu)$. Because of (5.5) this modality does not depend on $\mu \in \Delta'$. Note that a reasonable map associated to a bimodal map might be 3-modal but then two extremal values coincide.

Let us now state an analogue of Theorem 4.1 in this setting. To do this we shall generalize the notion of persistently full families. This generalization will allow some endpoints to depend discontinuously on the parameter and be also useful for families of renormalized maps. Let

$$\begin{aligned}\hat{V} = \{&(v_1, \ldots, v_k) \in I^k \text{ for which there exists} \\ &\text{a } k\text{-modal reasonable map } g\colon I \to I \text{ such that} \\ &v_i \text{ is equal to the } g\text{-value of the } i\text{-th turning point of } g\}.\end{aligned} \tag{5.6}$$

Definition. Let $f_\mu\colon I \to I$ be a l-modal family as before having a restrictive interval J_μ of period n for each $\mu \in \Delta'$ as in (5.5). We say that $f_\mu^n\colon J_\mu \to J_\mu$ is *persistently surjective* if the following conditions are met.

1. There exists an interval $I_\mu = [a(\mu), b(\mu)]$ with $a(\mu), b(\mu) \in J_\mu$ depending continuously on $\mu \in \Delta'$ such that f_μ^n is monotone on each component of $J_\mu \setminus I_\mu$ and such that each point in $J_\mu \setminus I_\mu$ is either periodic or contained in the basin of a periodic attractor; this periodic orbit has period one or two for $f_\mu^n\colon J_\mu \to J_\mu$ because $f_\mu^n(\partial J_\mu) \subset \partial J_\mu$.
2. The map $F\colon \Delta \to \hat{V}$ defined below is persistently surjective in the sense defined at the beginning of Section 4.

Here F is defined as follows. Let $A_\mu\colon \mathbb{R} \to \mathbb{R}$ be the continuous and piecewise affine map which sends $[a(\mu), b(\mu)]$ affinely onto $[0,1]$, the interval $(-\infty, a(\mu)]$ onto 0, $[b(\mu), \infty)$ onto 1 and define $F\colon \Delta \to \hat{V}$ by

$$F(\mu) = \left(A_\mu(f_\mu^n(\hat{c}_1(\mu))), \ldots, A_\mu(f_\mu^n(\hat{c}_k(\mu)))\right).$$

Here $\hat{c}_1(\mu), \ldots, \hat{c}_k(\mu)$ are the turning points of $f_\mu^n\colon J_\mu \to J_\mu$. The reason for introducing $a(\mu)$ and $b(\mu)$ in this definition is that in order to define a continuous map $F\colon \Delta \to \hat{V}$ we have to identify the interval J_μ with $[0,1]$ in a way which depends continuously on μ. Since the endpoints of J_μ may depend discontinuously on μ we insist that we can choose points $a(\mu), b(\mu) \in J_\mu$ which do depend continuously on μ and which are 'morally' the same as the endpoints of J_μ. In Theorem 5.2 in the next subsection we will show that this can be done for many families.

Theorem 5.1. *Let $f_\mu\colon I \to I$, $\mu \in \Delta$, be a family satisfying the smoothness conditions from the beginning of Section 4. Let J_μ and $\Delta' \subset \Delta$ be as above and assume that $f_\mu^n\colon J_\mu \to J_\mu$, $\mu \in \Delta'$, is a family of maps which is persistently surjective. Then $f_\mu^n\colon J_\mu \to J_\mu$, $\mu \in \Delta'$, is a full family of renormalized maps.*

Proof. Because of the Assumptions 1 and 2 in the definition above, the proof goes precisely as before. □

5.c: Renormalizations within persistently surjective families

Consider a persistently surjective family f_μ, $\mu \in \Delta$, of l-modal maps. Let us show that there is a subset $\tilde{\Delta} \subset \Delta$ such that for each of its components Δ' one has the following properties. For each $\mu \in \Delta'$, the map f_μ has a restrictive interval J_μ of some period n and 'some combinatorial type' such that J_μ depends continuously on $\mu \in \Delta'$ as in (5.5) above and such that $f_\mu^n\colon J_\mu \to J_\mu$, $\mu \in \Delta'$, is again a full family of renormalized maps. To be more specific, we associate to each l-modal map $f\colon I \to I$ a non-renormalizable interval map $\sigma(f)$ as follows.

Let $x \approx y$ if the interval $[x, y]$ is in the interior of a maximal restrictive interval of f. Furthermore we say that f and $\hat{f}$ are $\approx$*-combinatorially equivalent* if there exits an order preserving bijection $h\colon I/\!\approx \to I/\!\approx$ such that

$$\begin{array}{ccc}
\bigcup_{i=1}^{l}\bigcup_{n\geq 0} f^n(c_i)/\!\approx & \xrightarrow{f/\approx} & \bigcup_{i=1}^{l}\bigcup_{n\geq 0} f^n(c_i)/\!\approx \\
{\scriptstyle h}\downarrow & & \downarrow{\scriptstyle h} \\
\bigcup_{i=1}^{l}\bigcup_{n\geq 0} \hat{f}^n(c_i)/\!\approx & \xrightarrow{g/\approx} & \bigcup_{i=1}^{l}\bigcup_{n\geq 0} \hat{f}^n(c_i)/\!\approx
\end{array}$$

commutes.

Theorem 5.2. *Let $f_\mu\colon I \to I$, $\mu \in \Delta$, be a persistently surjective family of l-modal maps satisfying the smoothness conditions made at the beginning of Section 4 and satisfying (5.4). Let $\hat{f}\colon I \to I$ be a l-modal, non-renormalizable map with periodic turning points $c_{m(1)}, \dots, c_{m(r)}$ of period $n(1), \dots, n(r)$ with the same orientation as maps from the family f_μ. Then there exists a connected subset Δ_0 of Δ such that for each $\mu \in \Delta_0$ and each $j = 1, \dots, r$ the following properties hold.*

1. *There exists a maximal restrictive interval J_μ^j of period $n(j)$ containing $c_{m(j)}(\mu)$ (and no other turning points of f_μ) and depending continuously on μ as in (5.5) above.*
2. *The maps f_μ and g are $\approx$–combinatorially equivalent;*
3. *$f_\mu^{n(j)}\colon J_\mu^j \to J_\mu^j$, $\mu \in \Delta_0$, is again a persistently surjective family of renormalized maps. In particular, because of Theorem 4.2, $f_\mu^{n(j)}\colon J_\mu^j \to J_\mu^j$, $\mu \in \Delta_0$, is a full family of renormalized maps.*

Proof. Let $\Delta' = \{\mu \in \Delta\,;\ f_\mu$ and $\hat f$ are $\approx$–combinatorially equivalent$\}$. Since $c_{m(j)}$ is a periodic point of period $n(j)$ for $\hat f$ for $\mu \in \Delta'$ one has that f_μ has a maximal restrictive interval J^j_μ of period $n(j)$ containing $c_{m(j)}(\mu)$. Indeed, otherwise the equivalence class containing $c_{m(j)}(\mu)$ consists of just one point. Because f_μ and $\hat f$ are $\approx$-combinatorially equivalent it follows that $c_{m(j)}(\mu)$ is periodic with period $n(j)$. Since f_μ is C^1 and therefore has derivative zero at turning points, this implies that there exists an interval $\tilde J^j_\mu$ containing $c_{m(j)}(\mu)$ for each $j = 1, \dots, r$ and each $\mu \in \Delta'$ such that $\tilde J^j_\mu, \dots, f_\mu^{n(j)-1}(\tilde J^j_\mu)$ are disjoint and $f_\mu^{n(j)-1}(\tilde J^j_\mu) \subset \tilde J^j_\mu$. Because of Lemma 5.1, this interval is contained in a maximal restrictive interval J^j_μ of the same period. This contradiction shows that such restrictive intervals necessarily exist. Since $\hat f$ is not renormalizable, J^j_μ contains precisely one turning point of f_μ. Furthermore, J^j_μ is continuous in the sense that

$$f^i_\mu(J^j_\mu) \bigcap TP(f_\mu)$$

consists of at most one point for each $\mu \in \Delta'$ and each $0 \le i < n(j)$ and each of the points in these sets depends continuously on $\mu \in \Delta'$. From Theorem 4.1 it immediately follows that $f^{n(j)}_\mu \colon J^j_\mu \to J^j_\mu$, $\mu \in \Delta'$, is a full family of renormalized maps. Indeed, one can modify f_μ to a map $\tilde f$ in each of the iterates of J^j_μ so that the return map of $\tilde f$ to J^j_μ becomes equal to an arbitrary 'reasonable map g, see the definition in Section 5.b. Because of Theorem 4.1 there is a parameter value μ such that f_μ is essentially conjugate equivalent to $\tilde f$. This implies that $f^{n(j)}_\mu \colon J^j_\mu \to J^j_\mu$ is essentially conjugate to g and by construction f_μ is still $\approx$-combinatorially equivalent to $\hat f$.

So it remains to show that there exists a connected component Δ_0 of Δ' such that each of the families $f^{n(j)}_\mu \colon J^j_\mu \to J^j_\mu$, $\mu \in \Delta_0$, is again persistently surjective. Let us first show that the intervals I^j_μ from above exist. If J^j_μ depends continuously on μ then we simply take $I^j_\mu = J^j_\mu$. If it is not continuous at $\mu = \mu_0$, then (5.4) implies that $f^{n(j)}_{\mu_0} \colon J^j_{\mu_0} \to J^j_{\mu_0}$ has some inessential periodic attractors of period $n(j)$ exactly as was depicted in Figure 5.3. Therefore, in the discontinuous case, we can choose I_μ so that its endpoints are in the basins of these inessential periodic attractors. More precisely, let K^j_μ be the maximal neighbourhood of ∂J^j_μ such that each component of K^j_μ is a homterval, $f^{n(j)}_\mu(K^j_\mu) \subset K^j_\mu$ and $f^{n(j)}_\mu(\partial K^j_\mu) \subset \partial K^j_\mu$. From the continuity assumptions on f_μ,

$$K^j = \{(\mu, x)\,;\ \mu \in \Delta' \text{ and } x \in K^j_\mu\}$$

is a closed set. Because for each $\mu \in \Delta'$ there exists at least one $x \in K^j_\mu$ and because of (5.4), there exist continuous functions $\Delta' \ni \mu \mapsto a(\mu), b(\mu) \in K^j_\mu$ such that $I^j_\mu = [a(\mu), b(\mu)]$ has the required properties. Now suppose by contradiction that there exists no component Δ_0 of Δ' such that the family $f^{n(j)}_\mu \colon J^j_\mu \to J^j_\mu$, $\mu \in \Delta_0$, is persistently surjective. So let $\hat V$ be the set from (5.6) corresponding to J^j_μ. To reach a contradiction we shall use that each interior deformation of $F \colon \Delta \to \hat V$ does in fact come from an 'interior deformation' of f_μ. To make this more precise, take a component Δ_0 of Δ'. Since $F \colon \Delta \to \hat V$ is not persistently surjective we can continuously deform f_μ inside I^j_μ to a map $\tilde{f}_\mu$ for $\mu \in \Delta_0$ such that the image of at least one of the turning points of $\tilde{f}^{n(j)}_\mu \colon J^j_\mu \to J^j_\mu$ is mapped into $J^j_\mu \setminus I^j_\mu$. Note that the map $\tilde{F} \colon \Delta \to \hat{V}$ corresponding to $\tilde{f}^{n(j)}_\mu \colon J^j_\mu \to J^j_\mu$ is an interior deformation of F and therefore also persistently surjective. But by construction $\tilde{f}^{n(j)}_\mu \colon J^j_\mu \to J^j_\mu$, $\mu \in \Delta'$, is not a full family. But since $\tilde{F}$ is persistently surjective, $\tilde{f}^n_\mu \colon J^j_\mu \to J^j_\mu$ with $\mu \in \Delta'$ is full according to Theorem 5.1, and this contradicts the previous observation. □

Remark. 1. In the unimodal case, the combinatorial type of a unimodal map with a periodic turning point is determined by a permutation on a finite set. So in this case we can take a finite set $X = \{x_1, \ldots, x_n\}$ endowed with an order relation $\prec$. We say that a permutation $\sigma \colon X \to X$ is *unimodal* with respect to the order relation $\prec$ if it satisfies the following condition. Embed X monotonically into the real line, draw the graph of σ on $\mathbb{R}^2$ and connect the consecutive points of the graph by a line segment. If the curve so obtained is the graph of a unimodal map then we say that the permutation is unimodal. The resulting map is renormalizable if X is the disjoint union of p sets X_i each containing m points and such that

1. each X_i is mapped by σ onto some X_j;
2. for each $i \neq j$, either $X_i \prec X_j$ or $X_j \prec X_i$ (here $X_i \prec X_j$ means that $x_i \in X_i$, $x_j \in X_j$ implies $x_i \prec x_j$).

Similarly, to each renormalizable unimodal map $f \colon I \to I$ we associate a non-renormalizable unimodal permutation $\sigma(f) \colon X(f) \to X(f)$ as follows. Let $x \approx y$ if x and y are both contained in the same restrictive interval. Moreover, let

$$X(f) = \bigcup_{i=0}^{l} \bigcup_{k \geq 0} f^k(c_i)/\approx$$

with the ordering induced from the ordering on I. Since f is unimodal and renormalizable, $X(f)$ is a finite set and $\sigma(f)(x) = f(x)/\approx$ defines a permutation. So in the unimodal case the above theorem can be stated as follows. Let $f_\mu \colon I \to I$, $\mu \in \Delta$, be a persistently surjective family of unimodal maps satisfying the smoothness conditions made at the beginning of Section 4. Let $\sigma \colon X \to X$ be a non-renormalizable unimodal permutation. Then there exists a connected subset Δ_0 of Δ such that for each $\mu \in \Delta_0$, i) there exists a maximal restrictive interval J_μ of period $\#X$, ii) $\sigma(f_\mu) = \sigma$ and iii) $f^n_\mu \colon J^i_\mu \to J^i_\mu$, $\mu \in \Delta_0$, is again a persistently surjective family of renormalized maps. 2. There are only a countable number of combinatorial types of renormalizations possible in the unimodal case as we saw above. In the multimodal case, there are an uncountable number of combinatorial types of maps with one periodic turning point: the combinatorial type also depends on the orbit of the other turning points. 3. There is also a version of the above theorem where g has modality

$l' \le l$ with $l - l'$ is even. In this case, the corresponding restrictive intervals contain several turning points. 4. If in the previous theorem $f_\mu \colon I \to I$, $\mu \in \Delta$, is a full family which is not persistently surjective then such a connected subset Δ_0 as in Statement 3 of that theorem need not exist. 5. We should note that each equivalence class J of $\approx$ is an open interval. This implies that $I/\approx$ is not Hausdorff. If we would have changed the definition of $\approx$ so that all points in one restrictive interval J are equivalent, then we would get another problem: in order to make sure that $\approx$ defines an equivalence relationship (that transitivity holds) we would also have to impose that J has no common point with another maximal restrictive interval. However, this is precisely what happens if f is a unimodal map and J has period two (this is the best known case related to period doubling, see Example 5.3 below). For this reason we have introduced the notion of $\approx$-combinatorial equivalence.

Example. The restrictive interval J of the unimodal map f drawn in Figure 5.4 corresponds to the cyclic permutation $\sigma(f)$ on $\{J_0, J_1, \ldots, J_4\}$ where these intervals are ordered as $J_2 \prec J_0 \prec J_3 \prec J_4 \prec J_1$. Because this last set consists of 5 elements and 5 is a prime number, this permutation is non-renormalizable.

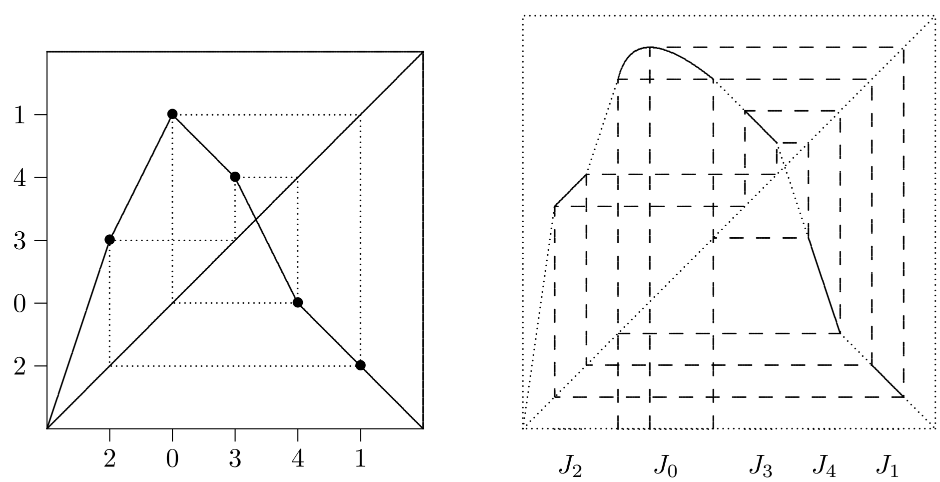

Fig. 5.4: A unimodal map with a restrictive interval corresponding to a cyclic permutation on five elements. By Theorem 5.1 there exists a quadratic map with a restrictive interval of the same type. In this case $f^5(J_0) = J_0$.

5.d: Examples of unimodal maps with solenoidal attractors

We say that a map is *infinitely often renormalizable* if it has restrictive intervals of arbitrary high period. For reasons which will become clear in Section III.4 we shall say that we have a *solenoidal attractor* in this case. Inside each full family one has renormalizable maps and, because of Statement 2 of the previous theorem, the return maps to the restrictive intervals form again a full family. Therefore there are many maps which are infinitely often renormalizable. In this subsection we shall make this more explicit in the unimodal case. Indeed, let $\mathcal{D}$ be the space of renormalizable unimodal maps. Furthermore, let σ_0 be a non-renormalizable unimodal permutation and let

$$\mathcal{D}_{\sigma_0} = \{f \in \mathcal{D}\,;\ \sigma(f) = \sigma_0\}.$$

As we saw in the previous theorem this set contains a full family of maps. So applying the previous theorem again, the set

$$\mathcal{D}_{\sigma_0,\sigma_1} = \{f \in \mathcal{R}\,;\ \sigma(f) = \sigma_0,\ \mathcal{R}(f) \in \mathcal{D} \text{ and } \sigma(\mathcal{R}(f)) = \sigma_1\}$$

contains a full family if σ_0 and σ_1 are unimodal non-renormalizable permutations. In general, letting $\mathcal{D}_n$ be the set of unimodal maps such that $f, \mathcal{R}(f), \ldots, \mathcal{R}^{n-1}(f)$ are renormalizable, the set

$$\mathcal{D}_{\sigma_0,\sigma_1,\ldots,\sigma_{n-1}} = \{f \in \mathcal{D}_n\,;\ \sigma(\mathcal{R}^i f) = \sigma_i, i = 0, \ldots, n-1\}$$

contains a full family of unimodal maps if $\sigma_0, \ldots, \sigma_{n-1}$ are unimodal, non-renormalizable permutations. Note that

$$\mathcal{R}(\mathcal{D}_{\sigma_0,\sigma_1,\ldots,\sigma_n}) = \mathcal{D}_{\sigma_1,\ldots,\sigma_n},$$

i.e., the renormalization operator acts as a shift map.

Theorem 5.3. *Let f_μ, $\mu \in \Delta$ be a full family of C^1 unimodal maps and let σ_i be a sequence of non-renormalizable, unimodal permutations. Then for each $n \in \mathbb{N}$, the set*

$$\{\mu \in \Delta\,;\ f_\mu \text{ is infinitely renormalizable } \sigma(\mathcal{R}^i(f_\mu)) = \sigma_i, i = 0, \ldots, n\}$$

is a closed, non-empty and contains an interval $\Delta_{\sigma_0,\sigma_1,\ldots,\sigma_n}$ such that $\mathcal{R}^n(f_\mu)$, $\mu \in \Delta_{\sigma_0,\sigma_1,\ldots,\sigma_n}$ is a full family. Furthermore, these intervals lie nested, i.e., $\Delta_{\sigma_0,\sigma_1,\ldots,\sigma_n} \subset \Delta_{\sigma_0,\sigma_1,\ldots,\sigma_{n-1}}$. In particular, $\Delta_{\sigma_0,\sigma_1,\ldots}$ is non-empty and $\Delta_\infty = \cup\Delta_{\sigma_0,\sigma_1,\ldots}$ contains a Cantor set.

Proof. Follows inductively from the previous theorem. □

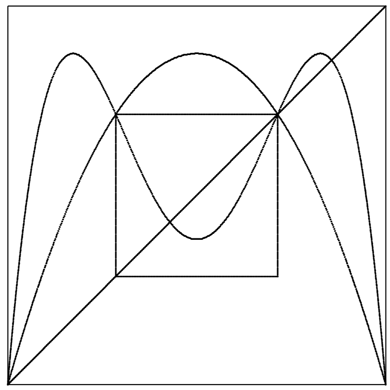

Fig. 5.5: The second iterate of a map with a restrictive interval of period two.

Example. (The *Feigenbaum maps*) Let σ be the permutation on two elements and let f_μ be a full family. Let $\Delta_\sigma = \{\mu\,;\ \sigma(f_\mu) = \sigma\}$. Then f_μ^2 maps a restrictive interval J_μ into itself for all $\mu \in \Delta_\sigma$, see Figure 5.5. For each $\mu \in \Delta_{\sigma,\sigma,\dots}$ let J_n be the corresponding restrictive interval of $\mathcal{R}^n(f_\mu)$. Then the f_μ-orbit of this interval consists of 2^n intervals with disjoint interiors and the closure of J_{n+1} is contained in the interior of J_n, see Figure 5.6. In particular, the set

$$K_n = \bigcup_{i=0}^{2^n-1} f^i(J_n)$$

consist of 2^n intervals (2^{n-1} of which with disjoint interiors) and lie nested. In particular,

$$K = \bigcap_{n\geq 0} K_n$$

contains a Cantor set. When $\mu \in \Delta_{\sigma,\sigma,\dots}$ then f_μ is called the map at *the accumulation of period doubling* or the *Feigenbaum* map. Metric properties of the set K were discovered independently by Feigenbaum (1978), (1979) and Coullet and Tresser (1978); a rigorous treatment of this will be presented in Chapter VI.

Example. Let f_μ be a full unimodal family and σ_1 and σ_2 be two distinct unimodal permutations. Then for each $\underline{\sigma} = (a_k)_{k\geq 0}$ for which $a_k \in \{\sigma_1, \sigma_2\}$ there exists a parameter μ such that $f \in \mathcal{D}_{\underline{\sigma}}$. It follows that there exists (at least) a Cantor set of parameters for which the corresponding maps are infinitely renormalizable of this type.

Example. Let σ_i be non-renormalizable, unimodal permutations on $a(i)$ elements. If $f \in \mathcal{D}_{\sigma_0}$, then f has a restrictive interval J_1 which is mapped after $a(0)$ steps into itself. If $f \in \mathcal{D}_{\sigma_0,\sigma_1}$ then J_1 contains a restrictive interval J_2 of $\mathcal{R}(f)$ and the orbit under $f^{a(0)}|J_1$ of J_2 consists of $a(1)$ intervals. Hence the orbit under f of J_2 consists of $q(1) = a(0)a(1)$ intervals. Similarly, the orbit

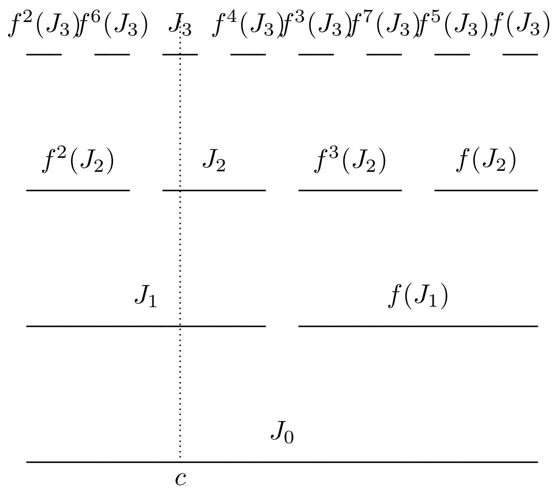

Fig. 5.6: The orbit of the intervals J_3 for the unimodal period doubling map. This interval has period eight. Notice that the unimodal map $f^2\colon J_1 \to J_1$ has a minimum (while $f\colon K_0 \to J_1$ has a maximum). Because of $f(J_1)$ is to the right of J_1 this implies that $f^2(J_2)$ is to the left of J_2. Since f maps $f(J_1)$ in an orientation reversing way onto J_1, the positions of $f(J_2)$ and $f^3(J_2)$ are as shown. Continuing in this way, one can gets that the orbit of J_3 is as above.

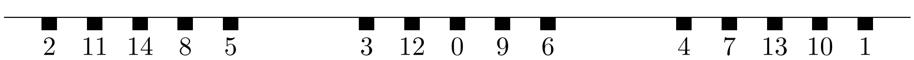

Fig. 5.7: The orbit of J_2 for a unimodal map $f \in \mathcal{D}_{\sigma_0,\sigma_1}$ when σ_0 is the cyclic permutation on $\{0,1,2\}$ with $0 \prec 1 \prec 2$ and σ_1 is the cyclic permutation on $\{0,1,2,3,4\}$ with $2 \prec 3 \prec 0 \prec 4 \prec 1$. Notice that the unimodal map $f^2\colon J_1 \to J_1$ has a minimum while the unimodal map associated to σ_1 has a maximum. Therefore the intervals $J_2, f^3(J_2), f^6(J_2), f^9(J_2), f^{12}(J_2)$ in J_0 which of course correspond to $0,1,2,3,4$ are ordered as $f^3(J_2) \prec f^12(J_2) \prec J_2 \prec f^9(J_2) \prec f^6(J_2)$. Because $f^2\colon f(J_1) \to J_1$ is orientation reversing and $f\colon f^2(J_1) \to J_1$ is orientation preserving, the orbit of J_2 is as shown.

under f of J_n consists of $q(n) = a(0)a(1)\dots a(n)$ intervals. So the action of f on J_n can be reconstructed by considering $\sigma_0, \sigma_1, \dots, \sigma_{n-1}$, see Figure 5.7. If we let

$$K_n = \bigcup_{i=0}^{q(n)-1} f^i(J_n)$$

then we call

$$K = \bigcap_{n \geq 0} K_n$$

a *solenoidal attractor*. We shall motivate this terminology in Section III.4. In Chapter VI we shall study metric properties of the set K in the case that the integers $a(i)$ are bounded.

6 Piecewise Monotone Maps can be Modelled by Polynomial Maps

We will prove in this section that any C^1 unimodal map of the interval is semi-conjugate to a quadratic map and that the semi-conjugacy is strictly monotone in the backward orbit of the turning point. This result is due to Guckenheimer (1979) and will follow from the discussion of the last section together with some properties of maps which satisfy the assumption that their Schwarzian derivative is negative.

That a quadratic map is described by a very simple mathematical formula is not very useful for the understanding of its dynamics because this property is not preserved under iteration: the n-th iterate of the map is a polynomial of degree 2^n. Singer (1978) made the following fundamental observation: if a map has negative Schwarzian derivative then all of its iterates also have this property. (Independently Allwright (1978) observed something similar.) Furthermore, quadratic maps turn out to have negative Schwarzian derivative. Therefore, rather than restrict attention to quadratic maps, we will study the dynamical properties of maps with negative Schwarzian derivatives. In the same paper, Singer proved that such maps have a finite number of attracting periodic orbits, if they have a finite number of turning points. This, because each of these orbits must attract at least one critical point or one boundary point. Later Guckenheimer (1979) showed, for unimodal maps with negative Schwarzian derivative, that any interval whose points have the same itinerary must be contained in the basin of the unique attracting periodic orbit. In particular, if the map has no attracting periodic orbit, the backward orbit of its turning point is dense. Combining these two results with Corollary 1 of Theorem 4.1 of the last section, we shall show that the quadratic family is the "universal model" for C^1 unimodal maps.

Let us recall the definition of the Schwarzian derivative (we already used this derivative in Section I.3). If $f\colon I \to I$ is a C^3 map and $Df(x) \neq 0$, the

Schwarzian derivative of f at x is defined as

$$Sf(x) = \frac{D^3 f(x)}{Df(x)} - \frac{3}{2}\left(\frac{D^2 f(x)}{Df(x)}\right)^2.$$

From the definition, the following formula for the Schwarzian derivative of the composition of two functions follows immediately by the chain rule,

$$S(g \circ f)(x) = Sg(f(x)) \cdot |Df(x)|^2 + Sf(x).$$

Hence the Schwarzian derivative of the iterates of f is given by

$$Sf^n(x) = \sum_{i=0}^{n-1} Sf(f^i(x)) \cdot |Df^i(x)|^2.$$

Therefore, if a map has negative Schwarzian derivative, so do all its iterates.

Next we shall state two analytical properties of maps with negative Schwarzian derivatives we will use in this section. The first of these is the Minimum Principle. We will give some additional background to the Schwarzian derivative and properties similar to the Minimum Principle in Chapter IV.

Lemma 6.1. (Minimum Principle) *Let T be a closed interval with endpoints a, b and $f\colon T \to \mathbb{R}$ a map with negative Schwarzian derivative. If $Df(x) \neq 0$ for all $x \in T$ then*

$$|Df(x)| > \min\{|Df(a)|, |Df(b)|\},\ \forall x \in (a, b).$$

Proof. At a critical point y of the function $x \mapsto |Df(x)|$ we have $D^2 f(y) = 0$. Hence $0 > Sf(y) = \frac{D^3 f(y)}{Df(y)}$, i.e., $D^3(f(y))$ and $Df(y)$ have different signs. Therefore, y is a local maximum of Df if $Df(y) > 0$ or a local minimum if $Df(y) < 0$. Consequently, the function $x \mapsto |Df(x)|$ cannot have a local minimum in the interior of the interval. Hence its minimum must be in the boundary. □

Before stating the second property we will derive some conclusions from the Minimum Principle. As before we say that the *basin* of a periodic point p is the set of points whose ω-limit set contains p. We say that a periodic point p of period n is *attracting* and that $O(p)$ is a *attracting periodic orbit* if its basin contains an open set. The *immediate basin* of a periodic point p is the union of the connected components of its basin which contain a point from $O(p)$. The periodic point is called a *hyperbolic attractor* if $|Df^n(p)| < 1$, a *hyperbolic repeller* if $|Df^n(p)| > 1$ and *neutral* if $|Df^n(p)| = 1$. Notice that the immediate basin of a hyperbolic attractor is the union of n open intervals, where n is the period. Finally, we say that c is a *critical point* of a C^1 map f if $Df(c) = 0$. It is called *non-degenerate* if $D^2 f(c) \neq 0$. As in Section 4 we say that a periodic attractor is *essential* if it contains a turning point in its basin. (In particular, any l-modal map can have at most l essential periodic attractors.)

Theorem 6.1. **(Singer)** *If $f\colon I \to I$ is a C^3 map with negative Schwarzian derivative then*

1. *the immediate basin of any attracting periodic orbit contains either a critical point of f or a boundary point of the interval I;*
2. *each neutral periodic point is attracting;*
3. *there exists no interval of periodic points.*

In particular, the number of non-repelling periodic orbits is bounded if the number of critical points of f is finite. Moreover, if all critical points of f are turning points then f has at most two inessential periodic attractors (containing a boundary point of ∂I in its basin).

Proof. Let p be an attracting periodic point which does have a boundary point of I in its immediate basin. Let n be the period of p and let T be the connected component of its basin containing p. Then $f^n(T) \subset T$ and, since p does not attract a boundary point of I, $f^n(\partial T) \subset \partial T$. If there exists $x \in T$ such that $Df^n(x) = 0$ then, for some $0 \le j \le n-1$, $f^j(x)$ is a critical point which belongs to $f^j(T)$ and this interval is clearly contained in the immediate basin of $f^j(p)$. Thus the theorem is verified in this case. So assume, by contradiction, that $Df^n(x) \neq 0$ for all $x \in T$. Let $m = n$ if $Df^n(x) > 0$ for all $x \in T$ and let $m = 2n$ if $Df^n(x) < 0$ for all $x \in T$. Since T is a component of the basin this implies $f^m(T) = T$, $Df^m(x) > 0$ for all $x \in T$ and $f^m(x) = x$ for $x \in \partial T$. If $x \in \partial T$ then $Df^m(x) \ge 1$ because otherwise x would be a two-sided attractor. But since $x \in \partial T$ this would contradict that T is contained in the basin of p. From the Minimum Principle, it follows that $Df^m(w) > 1$ for all $w \in \operatorname{int}(T)$ and this is impossible since $f^m(T) = T$. This proves Statement 1. If p is a neutral periodic point of period n then $Df^{2n}(p) = 1$ and $Df^{2n}(x) \ge 1$ for x near p; by the Minimum Principle this is impossible. Statement 3) follows from the claim below the statement of Theorem 3.1 and because of the Minimum Principle. □

Corollary 6.1. *Assume that $f\colon I \to I$ is a unimodal C^3 map with precisely one critical point (i.e., no inflection points). If f has negative Schwarzian derivative and the fixed point of f in ∂I is repelling then f has at most one attracting periodic orbit. In particular, any map $Q\colon [0,1] \to [0,1]$ from the quadratic family $Q_\mu = \mu x(1-x)$, $\mu \in [0,4]$ has at most one attracting periodic orbit.*

Proof. Because f is unimodal, $f(\partial I) \subset \partial I$. Since the fixed point in ∂I of f is repelling it follows from the previous result that each periodic attractor is essential. □

Another important property of maps with negative Schwarzian derivative is the Koebe Principle. We will come back to this principle and related principles extensively in Chapter IV. A version of this principle was first used and proved in Van Strien (1981) and later rediscovered by Johnson and Guckenheimer, see Guckenheimer (1987). In order to state this principle it is convenient to introduce the following terminology. Let $U \subset V$ be two intervals. We say that V contains a *δ-scaled neighbourhood of U* if each component of $V \setminus U$ has at least length $\delta|U|$.

Macroscopic Koebe Principle. (See Theorem IV.3.3). There exists a positive function $B_0 \colon \mathbb{R}^+ \to \mathbb{R}^+$ with the following property. Let f be C^3, $Sf < 0$ and suppose that for some pair of intervals $M \subset T$, and some $n \in \mathbb{N}$ $f^n|T$ is a diffeomorphism. If ϵ is so that $f^n(T)$ contains a ϵ-scaled neighbourhood of $f^n(M)$ then

$$T \text{ is a } B_0(\epsilon)\text{-scaled neighbourhood of } M.$$

In the remainder of this section we will show that unimodal maps with negative Schwarzian derivatives have no wandering intervals. Here we say as before that $J \subset I$ is a *wandering interval* for a map $f \colon I \to I$ if the intervals $\{J, f(J), \ldots, f^n(J), \ldots\}$ are pairwise disjoint and if J is not contained in the basin of an attracting periodic orbit of f. Now we come to the main topic of this section: maps satisfying some regularity conditions have no wandering intervals. So let us say that c is a *critical point* of a C^1 map f if $Df(c) = 0$. We say that f is *non-flat* at a point c if there exists a C^2 diffeomorphism $\phi \colon \mathbb{R} \to I$ with $\phi(0) = c$ such that $f \circ \phi$ is a polynomial near the origin.

Theorem 6.2. *Let $f \colon I \to I$ be a C^2 map such that f is non-flat at each critical point. Then f has no wandering intervals.*

From this theorem, which will be proved in Chapter IV, it follows that maps which are essentially conjugate are often even conjugate:

Corollary 6.2. *Let S be the class of C^3 maps $f \colon I \to I$ satisfying the following properties:*

1. *$D^2f(c) \neq 0$ at each point $c \in I$ with $Df(c) = 0$,*
2. *$Sf < 0$,*
3. *$|Df(x)| > 1$ if $x \in \partial I$ and f has no one-sided periodic attractors.*

Then $f, g \in S$ are conjugate if and only if they are combinatorially equivalent.

Proof. Because of Theorem 6.1, each periodic attractor contains a critical point in its basin. Moreover, all critical points are turning points. Because of the Minimum Principle, the immediate basin of two periodic attractors can have no boundary point in common and there exists no interval consisting of periodic points of constant period. The corollary therefore follows from Theorem 6.2 and the corollary to Theorem 3.1. □

Since the proof of Theorem 6.2 is not that easy we shall prove it in Chapter IV (in even greater generality). In this section we shall prove it when the map is unimodal and has negative Schwarzian derivative. This result is due to Guckenheimer (1979). The proof we give here serves as an introduction to the proof given in Chapter IV for the multimodal case.

Theorem 6.3. **(Guckenheimer)** *Let $f: I \to I$ be a C^3 unimodal map with negative Schwarzian derivative and such that $D^2f(c) \neq 0$ at the unique critical point c of f. Then f has no wandering intervals.*

Remark. If f is symmetric then there is a very simple proof of Theorem 6.3. (We should remark that for any unimodal map there are new coordinates in which the map becomes symmetric; however, the property that the Schwarzian derivative of the maps is negative may get lost under this coordinate change.) This proof is given in Exercise 6.1 below. Step 3 of the proof we will give here is different from the original proof of Guckenheimer. He did not make use of the Koebe Principle, but used the arguments given in Exercise 6.1 to show that even if f is not symmetric that there exists $\rho > 0$ such that the sequence of closest approach $f^{n(k)}(J)$ defined below satisfies $|f^{n(k+1)}(J)| \geq \rho \cdot |f^{n(k)}(J)|$. Combining this with the first part of Step 2 below and using the Minimum Principle, completes the argument. The main advantage of our proof is that it makes no use of periodic points. We prefer this because periodic points cannot be used in the multimodal analogue of this theorem.

Proof of Theorem 6.3 Since f is unimodal, there exists a map $\tau: I \to I$ such that $f(\tau(x)) = f(x)$ and $\tau(x) \neq x$ for $x \neq c$. Since $Df^2(c) \neq 0$, the map τ is Lipschitz and in particular there exists a number $\kappa \in (0,1)$ such that

$$|\tau(J)| \geq \kappa |J|$$

for each interval J not containing the turning point. In order to prove the theorem suppose, by contradiction, that f has a wandering interval J. By considering an iterate of J instead of J we may assume that no iterate of J contains the critical point of f. So f^n is a homeomorphism on J for all $n \geq 0$. Furthermore, we may assume that J is not contained in a larger wandering interval. This last assumption implies that no interval T which strictly contains J is a homterval. Indeed, by Corollary 1 of Lemma II.3.2, either all points of a homterval T are contained in the basin of a periodic attractor or T is also a wandering interval. Since $T \supset J$ and J is a wandering interval, the first alternative is impossible. Therefore T is also a wandering interval. But this contradicts the maximality of J.

Next we have from Corollary 2 of Theorem I.2.2 that $f^n(J)$ must accumulate at the critical point c. Therefore there exists a sequence of integers $n(k) \to \infty$ such that $f^{n(k)}(J) \to c$. Hence we can consider the sequence of 'closest approach' to the critical point defined inductively as follows:

$$n(1) = 0,$$

$$n(k+1) = \min\{j > n(k)\,;\; f^j(J) \subset (f^{n(k)}(J), \tau(f^{n(k)}(J)))\}.$$

Here $(K, \tau(K))$ denotes the smallest interval which connects K and $\tau(K)$ (and whose intersection with these intervals is empty). Furthermore, let

$$[K, \tau(K)] = K \cup (K, \tau(K)) \cup \tau(K).$$

Now the remainder of the proof goes in three steps.

Step 1: Let $T_{n(k)} \supset J$ be the largest interval on which $f^{n(k)}$ is a homeomorphism. We claim that either

$$f^{n(k)}(T_{n(k)}) \supset [c, f^{n(k-1)}(J)]$$

or that the first interval contains the mirror image of the second, i.e.,

$$f^{n(k)}(T_{n(k)}) \supset [c, \tau(f^{n(k-1)}(J))].$$

Indeed, let L and R be the components of $T_{n(k)} \setminus J$. By maximality of $T_{n(k)}$ there exists an integer $0 \leq l < n(k)$ such that $f^l(L)$ contains c (in its boundary). At the same time

$$f^l(J) \bigcap \left[f^{n(k-1)}(J), \tau(f^{n(k-1)}(J))\right] = \emptyset$$

for $l \neq n(k-1)$ since $l < n(k)$ and by definition of the sequence $n(k)$. It follows that $f^l(L \cup J)$ contains $[f^{n(k-1)}(J), c]$ or $[\tau(f^{n(k-1)}(J)), c]$. Hence $f^{n(k)}(L \cup J)$ can be contained neither in $[f^{n(k-1)}(J), c]$ nor in $[\tau(f^{n(k-1)}(J)), c]$. Indeed, otherwise one of the intervals $[f^{n(k-1)}(J), c]$ or $[\tau(f^{n(k-1)}(J)), c]$ would contain $f^{n(k)}(L \cup J)$. In particular one of the intervals $f^l(L \cup J)$ or $\tau(f^l(L \cup J))$ would contain $f^{n(k)}(L \cup J)$ and hence $f^{n(k)-l}$ would map one of these intervals $f^l(L \cup J)$ or $\tau(f^l(L \cup J))$ monotonically into itself and consequently J would be attracted by a periodic attractor (with period $n(k)-l$), a contradiction to the assumption that J is a wandering interval. Similarly, $f^{n(k)}(R \cup J)$ cannot be contained in $[f^{n(k-1)}(J), c]$ or in its mirror image $[\tau(f^{n(k-1)}(J)), c]$. Combining this it follows that $f^{n(k)}(T_{n(k)})$ contains $[c, f^{n(k-1)}(J)]$ or $[c, \tau(f^{n(k-1)}(J))]$.

Step 2: Next we show that the gap between $f^{n(k)}(J)$ and c is much smaller than the size of $f^{n(k)}(J)$ for k large. More precisely, we will show that

$$\lim_{n \to \infty} \frac{|(f^{n(k)}(J), c)|}{|f^{n(k)}(J)|} \to 0. \tag{6.1}$$

Of course it is enough to show that for each $\epsilon > 0$

$$|(f^{n(k)}(J), c)| \leq \epsilon |f^{n(k)}(J)| \tag{6.2}$$

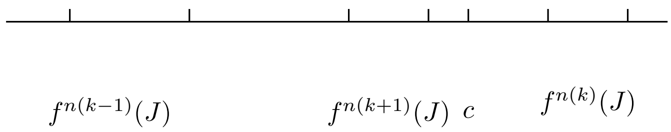

Fig. 6.1: The intervals $f^{n(k)}(J)$ tend to c, but not necessarily monotonically. In Step 2 it is shown that $f^{n(k)}(J)$ is not too large compared to the gap between this interval and c.

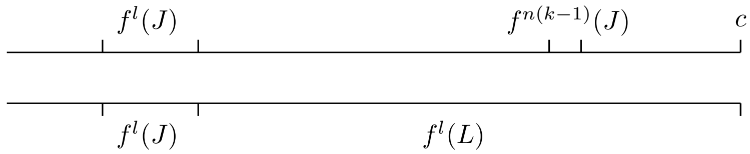

Fig. 6.2: $f^l(L \cup J)$ contains either $[f^{n(k-1)}(J), c]$ or its symmetric. So if $f^{n(k)}(L \cup J)$ is contained in $[f^{n(k-1)}(J), c]$ or in its symmetric $[\tau(f^{n(k-1)}(J)), c]$ then $f^{n(k)-l}$ would map the interval $f^l(L \cup J)$ or the interval $\tau(f^l(L \cup J))$ monotonically into itself and J would be attracted by a periodic attractor with period $n(k) - l$.

provided k is large enough (and of course we may even assume that $\epsilon \in (0, \kappa)$ where κ is as in the beginning of the proof). Because the intervals $f^n(J)$ are all disjoint, there exist infinitely many integers k with

$$|f^{n(k)}(J)| \leq |f^{n(k-1)}(J)|. \tag{6.3}$$

Take such an integer k. By Step 1, one of the components of $f^{n(k)}(T_{n(k)} \setminus J)$ contains either $f^{n(k-1)}(J)$ or its mirror image (note that this is the component which is 'further away from the turning point'), and therefore

$$\min\Big(|f^{n(k-1)}(J)|, |\tau(f^{n(k-1)}(J))|\Big) \geq \kappa \cdot |f^{n(k-1)}(J)| \geq \kappa \cdot |f^{n(k)}(J)|.$$

So on this side of $f^{n(k)}(J)$ one has 'space'. On the other side of $f^{n(k)}(J)$ (i.e., the side which is closer to the turning point), the interval $f^{n(k)}(T_{n(k)})$ contains $(f^{n(k)}(J), c)$. If (6.2) would fail then we would also get space on the other side and $f^{n(k)}(T_{n(k)})$ would at least contain an ϵ'-scaled neighbourhood of $f^{n(k)}(J)$ where $\epsilon' = \min(\kappa, \epsilon)$. But then the Koebe Principle implies that $T_{n(k)}$ is a $\rho(\epsilon')$-scaled neighbourhood of J. So we get an interval T which is strictly larger than J on which $f^{n(k)}$ is a homeomorphism for all $k \geq 0$. But then $f^n|T$ is a homeomorphism for all $n \geq 0$ and therefore T is a homterval which strictly contains J. As we have seen above, this gives a contradiction. Consequently, for k sufficiently large (6.3) implies (6.2). But since $\epsilon \in (0, \kappa)$ and either $f^{n(k+1)}(J)$ or its symmetric is contained in $(f^{n(k)}(J), c)$, inequality (6.2) gives

$$|f^{n(k+1)}(J)| \leq |f^{n(k)}(J)|.$$

Thus we get back the assumption (6.3) with $k+1$ instead of k and so (6.2) holds also for $k+1$ instead of k. Repeating this both (6.2) and (6.3) hold for all subsequent integers k. This completes the proof of this step.

Step 3: Next we claim that

$$n(k+2) - n(k+1) \leq n(k+1) - n(k) \tag{6.4}$$

for k large enough. In order to prove this we first show the following. If (6.4) would not hold then there would be an interval $T \supset f^{n(k)}(J)$ which is mapped by $f^{n(k+1)-n(k)}$ monotonically onto $[f^{n(k)}(J), \tau(f^{n(k)}(J))]$. So let $T \supset f^{n(k)}(J)$ be the maximal interval which is mapped by $f^{n(k+1)-n(k)}$ monotonically into $[f^{n(k)}(J), \tau(f^{n(k)}(J))]$. If the image of T under this map is not the entire interval then there exists $0 \leq l < n(k+1) - n(k)$ such that $f^l(T)$ contains c in its boundary. But $f^l(f^{n(k)}(J))$ is certainly not contained in $[f^{n(k)}(J), \tau(f^{n(k)}(J))]$ because $l + n(k) < n(k+1)$ and by the definition of the sequence of $n(k)$. Hence $f^l(T)$ contains either $f^{n(k)}(J)$ or its symmetric. But then $f^{n(k+1)-n(k)}(T)$ contains $f^{n(k+1)-l}(J)$. So we would get that $f^{n(k+1)-l}(J)$ is contained in $[f^{n(k)}(J), \tau(f^{n(k)}(J))]$. But since $n(k) < n(k+1) - l \leq n(k+1)$ this is impossible unless $l = 0$ again by the definition of the sequence of $n(k)$. But if $l = 0$ then T contains $[f^{n(k)}(J), c]$ and therefore $f^{n(k+1)}(J)$ or its mirror image. Thus

$$[f^{n(k)}(J), \tau(f^{n(k)}(J))] \supset f^{n(k+1)-n(k)}(T)$$

would contain $f^{n(k+1)-n(k)+n(k+1)}(J)$ and by definition of $n(k+2)$ this gives $n(k+2) \leq n(k+1) + n(k+1) - n(k)$ and (6.4) holds. So if (6.4) does not hold then we get the required interval T.

But if we let $M = f^{n(k)}(J)$ then Step 2 implies that the space $(f^{n(k)}(J), c)$ between $f^{n(k)}(J)$ and c is much smaller than the size of $f^{n(k)}(J)$ for k large. In particular, $[f^{n(k)}(J), \tau(f^{n(k)}(J))]$ is a 1-scaled neighbourhood of $f^{n(k+1)-n(k)}(M) = f^{n(k+1)}(J)$ for k sufficiently large. So $f^{n(k+1)-n(k)}(T)$ 'has space around' $f^{n(k+1)-n(k)}(M)$. But then the Koebe Principle implies that this space can be pulled back. More precisely, T contains a $\rho(1)$-scaled neighbourhood of M where $\rho(1) > 0$ is a universal number. Since T does not contain c, one of the components of $T \setminus M$ is contained in $(f^{n(k)}(J), c)$ and therefore

$$|(f^{n(k)}(J), c)| \geq \rho(1)|M| = \rho(1)|f^{n(k)}(J)|.$$

But this contradicts again (6.1) for large k. All these contradictions show that (6.4) needs to hold for k sufficiently large. Now we can easily complete the proof. Since (6.4) holds, $n(k+1) - n(k)$ is eventually equal to some integer a for all k sufficiently large. In particular, since the intervals $f^{n(k)}(J)$ tend to c, it follows that c is an attracting fixed point of f^a and that J is in its basin. This contradicts that J is a wandering interval. □

Exercise 6.1. If f is symmetric one can simplify the previous proof considerably. In that case it follows from the Minimum Principle that $|f^{n(k+1)}(J)| > |f^{n(k)}(J)|$ and so no wandering intervals can exist. Prove this in three steps. Step 1) Let

$$V_k = \{y;\ f^j(y) \notin [y, \tau(y)] \text{ for all } 0 < j < k, f^k(y) \in [y, \tau(y)]\}$$

and let T be a connected component of V_k. Show that: i) $Df^k(y) \neq 0$ for all $y \in T$ and ii) for each $y \in \partial T$, $f^k(y) \in \{y, \tau(y)\}$. (Hint: since $f^j(y) \notin [y, \tau(y)]$, $f^j(y)$ is not equal to the critical point of f for $0 \le j < k$ and hence $Df^k(y) \neq 0$. This proves i). Since y is in the boundary of T, it follows from the definition of V_k that $f^j(y) \in \{y, \tau(y)\}$ for some $j = 0, 1, \dots, k$. If $j = k$ then ii) follows. So we are left with the case that there exists $0 < l < k$ with $f^l(y) \in \{y, \tau(y)\}$ and such that $f^j(y) \notin [y, \tau(y)]$ for $j = 1, \dots, l-1$. Let $m, p \in \mathbb{N}$ be such that $k = ml + p$ with $0 \le p < l$. Then $f^k(y) = f^p f^{ml}(y)$ and since $f^l(y) \in \{y, \tau(y)\}$ and $f(y) = f(\tau(y))$ this gives $f^k(y) \in f^p(\{y, \tau(y)\})$. If $p > 0$ then $f^k(y) = f^p(y)$ and since $f^k(y) \in [y, \tau(y)]$ this gives a contradiction with the definition of l. So $p = 0$ and therefore again $f^k(y) \in \{y, \tau(y)\}$.) Step 2) Show that $|Df^k(z)| > 1$ for all $z \in T$. (Hint: as in Step 1, if we take $y \in \partial T$, we get $\hat{y} = f^k(y) \in \{y, \tau(y)\}$ and therefore $\hat{y}$ is a fixed point of f^k: $f^k(\hat{y}) = f^k(f^k(y)) \in f^k(f^k(\{y, \tau(y)\})) = f^k(y) = \hat{y}$. Furthermore, f cannot have an attracting periodic point because otherwise, by Theorem 6.1, this attracting periodic point would attract the critical point and therefore J would be contained in the basin of this periodic point and therefore not be a wandering interval. Because $f^k(\hat{y}) = \hat{y}$ this gives $|Df^k(\hat{y})| \ge 1$. On the other hand, by the symmetry hypothesis, $Df^k(y) = -Df^k(\tau(y))$. Therefore $|Df^k(y)| = |Df^k(\tau(y))| \ge 1$ for both boundary points y of T. From the Minimum Principle it follows that $|Df^k(z)| > 1$ for all $z \in T$. Step 3) By the definitions of V_k and of the sequence of closest approach $n(k)$ we have $f^{n(k)}(J) \subset V_{n(k+1)-n(k)}$. Therefore Step 2 implies that $|f^{n(k+1)}(J)| > |f^{n(k)}(J)|$.

From the previous theorem and the results from Sections 3 and 4 one gets that quadratic maps form a good model for unimodal maps. In Chapter IV we will prove Theorem 6.2 (more generally the non-existence of non-wandering intervals will be proved for a very large class of maps which includes for example all analytic interval maps). From all this we get the following

Theorem 6.4. *If $f: I \to I$ is a l-modal C^1 map then there exist a polynomial l-modal map $P: I \to I$ and a semi-conjugacy $h: I \to I$ between f and P, i.e., h is continuous, monotone and*

$$P \circ h = h \circ f.$$

Furthermore, h is strictly monotone in the backward orbits of the turning points of f. In fact, h merely collapses wandering intervals and the basins of periodic attractors which do not attract a turning point.

Proof. Let P be the l-modal polynomial map which is essentially conjugate to f and which has negative Schwarzian derivative, repelling boundary points and non-degenerate critical points. By the Corollary to Theorem 4.1 such a map P exists. It follows from Theorem 6.1 that P has only essential periodic attractors. We will show that there is a semi-conjugacy from f to P. Let c_i be the turning points of f and let B be the union of the basins of all periodic attractors of f. Similarly, let $\tilde{c}_i$ and $\tilde{B}$ be the corresponding objects for P. Since f and P are combinatorially equivalent, c_i is contained in the immediate basin of a periodic attractor if and only if the same holds for $\tilde{c}_i$. So let $B_{e,0}$ be the union of immediate basins of essential periodic attractors for f and similarly let $\tilde{B}_0$ the union of immediate basins of periodic attractors of P (all periodic attractors of

P are essential). Extend h to a conjugacy $h\colon B_{e,0} \to \tilde{B}_0$. As in Theorem 3.1 and the corollary below this theorem, h extends to a homeomorphism from

$$\cup_{i=1}^{l} \cup_{n\in\mathbb{Z}} f^n(c_i) \bigcup B$$

to the set

$$\cup_{i=1}^{l} \cup_{n\in\mathbb{Z}} P^n(c_i) \bigcup \tilde{B}.$$

From Theorem 6.2, P has no wandering intervals, the last set is dense. It follows that h extends to a semi-conjugacy from f to P. □

Remark. 1. The semi-conjugacy between f and P of Theorem 6.4 collapses i) the basin of every non-essential attracting periodic point of f, ii) wandering intervals of f and iii) intervals of periodic points. However, if f has negative Schwarzian derivative, all critical points of f are of quadratic type and the boundary points of I are not contained in the immediate basin of a periodic attractor then f has no inessential periodic attractors, no wandering intervals and no intervals of periodic points, see Theorems 6.1 and 6.2. Hence in this case the semi-conjugacy is in fact a conjugacy. 2. In the next chapter we shall show that each map with negative Schwarzian derivative whose turning points are in the basin of hyperbolic periodic orbits is structurally stable.

Exercise 6.2. Take a quadratic map Q. Let $O_1, \dots, O_k$ be periodic orbits such that the critical orbit does not accumulate on $O_1, \dots, O_k$. Show that there exists a C^∞ unimodal map f which is semi-conjugate to Q such that the inverse of each point which is eventually mapped into $O_1 \cup \dots \cup O_k$ under this semi-conjugacy consists of an interval. In particular, f has k periodic attractors, and h maps the immediate basins of these periodic attractors onto the periodic orbits $O_1, \dots, O_k$. Later on, in Chapter IV, we will show that any C^∞ map with non-flat critical points can be constructed in this way. (Hint: because the critical orbit of Q does not accumulate onto the orbits O_i one can modify Q near these orbits without changing the kneading invariant. So modify Q into a unimodal map f which coincides with Q outside a small neighbourhood of $O_1 \cup \dots \cup O_k$ and such that f has very small derivative in each point of $O_1 \cup \dots \cup O_k$. Then f has a periodic attractor near O_i.)

7 The Topological Entropy

In the next three sections we will discuss the relationship between the dynamics of an interval map and an important topological invariant, called topological entropy. We will not use this relationship in the remainder of this book. This topological invariant, which appeared for the first time in Adler et al. (1965), is defined for continuous maps of compact metric space and is a measure of the dynamical complexity of the map. It measures the growth rate as n tends to infinity of the number of different orbits of length n if we use a precision ϵ to distinguish two orbits. For continuous piecewise monotone maps of the interval we will show that the topological entropy coincides with the growth of the number of points in the backward orbits of the turning points which is equal to the "lap number" of the iterates of the map, i.e., the number of maximal intervals of monotonicity of the iterate. This result was first proved by Misiurewicz and Szlenk (1980) and Young (1981). We will follow the exposition of L.S. Young.

Definition. Let X be a compact metric space with metric d and $f\colon X \to X$ be a continuous function. A subset $E \subset X$ is (n,ϵ)-*separated* if for any $x, y \in E$ with $x \neq y$, there is an integer j such that $0 \le j < n$ and $d(f^j(x), f^j(y)) > \epsilon$. A set $F \subset X$ is said to (n,ϵ)-*span* another set K if for each $x \in K$ there exists $y \in F$ such that $d(f^j(x), f^j(y)) \le \epsilon$ for all $0 \le j < n$.

If $K \subset X$ is a compact subset, we denote by $r_n(\epsilon, K)$, or $r_n(\epsilon, K, f)$, the smallest cardinality of any set $F \subset K$ that (n,ϵ)-spans K and by $s_n(\epsilon, K)$ the largest cardinality of any set $E \subset K$ which is (n,ϵ)-separated. Define:

$$r(\epsilon, K) = \limsup_{n\to\infty} \frac{1}{n} \log r_n(\epsilon, K),$$

$$s(\epsilon, K) = \limsup_{n\to\infty} \frac{1}{n} \log s_n(\epsilon, K).$$

Notice that, for a given "precision" $\epsilon > 0$, $s_n(\epsilon, K)$ is the maximum number of different orbits of length n starting at K. Let us first show that these numbers r_n and s_n are strongly related.

Lemma 7.1. *1. If* $\epsilon_1 < \epsilon_2$ *then* $r(\epsilon_1, K) \ge r(\epsilon_2, K)$ *and* $s(\epsilon_1, K) \ge s(\epsilon_2, K)$. *2.* $r_n(\epsilon, K) \le s_n(\epsilon, K) \le r_n(\frac{1}{2}\epsilon, K) < \infty$.

Proof. Let us first prove Statement 1). For each n we have that every (n,ϵ_2)-separated set is also a (n,ϵ_1)-separated set. Hence $s_n(\epsilon_1, K) \ge s_n(\epsilon_2, K)$. Therefore $s(\epsilon_1, K) \ge s(\epsilon_2, K)$. The proof of the other inequality is the same. In order to prove Statement 2) let us first remark that, by compactness of K, the number of disjoint balls of radius ϵ and therefore the cardinality of any (n,ϵ)-separated set is bounded. Let $E \subset K$ be a (n,ϵ)-separated set of maximal cardinality. If $x \in K$ there exists $y \in E$ such that $d(f^i(x), f^i(y)) \le \epsilon$ for all $i = 0, \ldots, n-1$ because, otherwise, $E \cup \{x\}$ would be an (n,ϵ)-separated set with bigger cardinality. Therefore E also (n,ϵ)-spans K. Thus

$$\text{Card }(E) = s_n(\epsilon, K) \ge r_n(\epsilon, K).$$

Let F be a set which $(n, \frac{1}{2}\epsilon)$-span K. If $E \subset K$ is an (n,ϵ) -separated set then, for every $x \in E$, we can choose a point $T(x) \in F$ such that $d(f^i(T(x)), f^i(x)) \le \frac{1}{2}\epsilon$, $\forall i < n$. We claim that $T(x_1) \neq T(x_2)$ if $x_1 \neq x_2$. In fact, if $T(x_1) = T(x_2)$ then $d(f^i(x_1), f^i(x_2)) \le d(f^i(x_1), f^i(T(x_1))) + d(f^i(T(x_2)), f^i(x_2)) \le \epsilon$, $\forall i < n$ and this is not possible because E is (n,ϵ)-separated. From the claim we get that $\text{Card}(E) \le \text{Card}(F)$. Hence $s_n(\epsilon, K) \le \text{Card}(F)$ and therefore $s_n(\epsilon, K) \le r_n(\frac{1}{2}\epsilon, K)$. □

From Lemma 7.1, it follows that the limit, as $\epsilon \to 0$, of both $r_n(\epsilon, K)$ and $s_n(\epsilon, K)$ exist and are equal (but this number may be equal to $+\infty$).

Definition. For a continuous map $f\colon X \to X$ of a compact metric space with metric d and a (not necessarily f-invariant) subset $K \subset X$ the *topological entropy* of f *with respect to* K is the number

$$h_t(f,K) = \lim_{\epsilon \to 0} r(\epsilon, K) = \lim_{\epsilon \to 0} s(\epsilon, K).$$

The number $h_t(f) = h_t(f,X)$ is called the topological entropy of f.

The topological entropy is a measure of the dynamical complexity of the map. Indeed, for n big and ϵ small, the number of different orbits of length n, up to precision ϵ, is of the order $e^{nh_t(f)}$.

Lemma 7.2. *If $f\colon X \to X$ is a continuous map of a compact metric space X then*

$$h_t(f^m) = mh_t(f)$$

for all integers $m > 0$.

Proof. If $F \subset X$ (mn, ϵ)-spans X with respect to f then it clearly (n,ϵ)-spans X with respect to f^m. Therefore $r_n(\epsilon, X, f^m) \le r_{mn}(\epsilon, X, f)$. This implies $h_t(f^m) \le mh_t(f)$. Since X is compact, given ϵ and n, there exists $\delta > 0$ such that if $x, y \in X$ satisfy $d(x,y) < \delta$ then $d(f^i(x), f^i(y)) < \epsilon$, $\forall i < n$. Therefore any set F that (n,δ)-spans X with respect to f^m also (mn,ϵ)-spans X with respect to f. Hence $r_{mn}(\epsilon, X, f) \le r_n(\delta, X, f^m)$ and this implies $mh_t(f) \le h_t(f^m)$. □

In the remainder of this section we will show how to calculate the topological entropy of a continuous one-dimensional map $f\colon [0,1] \to [0,1]$. We will do this using the fact that f is semi-conjugate to the shift operator on the space of symbols introduced in Section 3. Therefore it will be useful to have the following result.

Amongst other things this result will also show that the topological entropy of a continuous map $f\colon X \to X$ on a metric space X does not depend on the choice of the metric on X (and is therefore really a topological invariant).

Theorem 7.1. **(Bowen)** *Let (X,d) and (Y,d') be compact metric spaces, $f\colon X \to X$, $g\colon Y \to Y$ be continuous maps. If $\pi\colon X \to Y$ is a continuous and surjective map such that $\pi \circ f = g \circ \pi$ then*

$$h_t(g) \le h_t(f) \le h_t(g) + \sup_{y \in Y} h_t(f, \pi^{-1}(y)).$$

Proof. Since X is compact, given $\epsilon > 0$ there exists $\delta > 0$ such that if $d'(\pi(x), \pi(y)) > \epsilon$ then $d(x,y) > \delta$. Hence, if $E_n \subset Y$ is an (n,ϵ)-separated set of maximal cardinality $s_n(\epsilon, Y, g)$ and $\tilde{E} \subset X$ is a set having one and only one point in each fiber over each point of E_n, we have that $\tilde{E}$ is an (n,δ)-separated set with the same cardinality as E_n. Thus $s_n(\delta,X,f) \geq s_n(\epsilon,Y,g)$ for all integers $n \in \mathbb{N}$. From this we get easily, $h_t(f) \geq h_t(g)$.

Let us now prove the second inequality. Let

$$a = \sup_{y\in Y} h_t(f,\pi^{-1}(y)).$$

If $a = \infty$ there is nothing to prove. So we can assume that $a < \infty$. Let $r_n(\epsilon,\pi^{-1}(y))$ be the minimum cardinality of a subset of $\pi^{-1}(y)$ which (n,ϵ)-spans $\pi^{-1}(y)$ with respect to f. Since $r_n(\epsilon,\pi^{-1}(y))$ is a decreasing function of $\epsilon > 0$, we have

$$h_t(f,\pi^{-1}(y)) \geq \limsup_{n\to\infty} \frac{1}{n} \log r_n(\epsilon,\pi^{-1}(y)) \text{ for every } \epsilon > 0.$$

Fix $\epsilon > 0$ and choose $\alpha > 0$. From the previous inequality it follows that for each $y \in Y$ we can choose an integer $m(y)$ such that

$$(*) \qquad a+\alpha \geq h_t(f,\pi^{-1}(y)) + \alpha \geq \frac{1}{m(y)} \log r_{m(y)}(\epsilon,\pi^{-1}(y)).$$

Note that $h_t(f)$ and $h_t(g)$ are equal to limits of respectively $s_n(\epsilon,X,f)$ and $r_n(\delta,Y,g)$. Therefore, in order to prove the second inequality of the theorem, we need to relate $s_n(\epsilon,X,f)$ to $r_{m(y)}(\epsilon,\pi^{-1}(y))$ and $r_n(\delta,Y,g)$. In order to do this we choose a finite number of points $\{y_1,\ldots,y_p\}$ as follows. For each $y \in Y$ choose a set $F_y \subset \pi^{-1}(y)$ with the smallest possible cardinality so that it $(m(y),\epsilon)$-spans $\pi^{-1}(y)$ with respect to f. From $(*)$ we have

$$(**) \qquad a+\alpha \geq \frac{1}{m(y)} \mathrm{Card}(F_y).$$

Let $D_n(z,2\epsilon,f)$ be the neighbourhood of z defined by

$$D_n(z,2\epsilon,f) = \{w \in X\,;\ d(f^i(z),f^i(w)) < 2\epsilon,\ \text{for } 0 \leq i < n\}$$

and also define

$$U_y = \cup_{z\in F_y} D_{m(y)}(z,2\epsilon,f).$$

Since F_y is an $(m(y),\epsilon)$-spanning set of $\pi^{-1}(y)$ with respect to f, it follows that U_y is a neighbourhood of $\pi^{-1}(y)$ in X. Let $\{W_{y_1},\ldots W_{y_p}\}$ be a finite cover of Y such that $\pi^{-1}(W_{y_i}) \subset U_{y_i}$, $\forall i = 1,\ldots,p$ and let δ be the Lebesgue number of this cover (this means that for $y \in Y$ the ball $B(y,\delta)$ of radius δ around y is contained in one of the sets W_{y_i}).

We want to estimate the maximal cardinality $s_n(4\epsilon,X,f)$ of a $(n,4\epsilon)$ separated set of X from above in terms of $\mathrm{Card}(F_{y_i})$, $i \in \{1,\ldots,p\}$, and $r_n(\delta,Y,g)$. For that we take an (n,δ)-spanning set E_n for Y, with respect to g, having minimal cardinality $r_n(\delta,Y,g)$. We want to shadow the orbit of a point $y \in E_n$ by pieces of orbits of points from $\{y_1,\ldots,y_p\}$ in such a way that we can apply $(**)$ to each of these pieces. This we do as follows. Take $y \in Y$ and let $c_0(y) \in \{y_1,\ldots,y_p\}$ be such that $W_{c_0(y)} \supset B(y,\delta)$ and define $t_0(y) = 0$. Next let

$t_1(y) = m(c_0(y))$ and let $c_1(y) \in \{y_1, \ldots, y_p\}$ be such $W_{c_1(y)} \supset B(g^{t_1(y)}(y), \delta)$. Similarly, assuming that $t_0(y), \ldots, t_k(y)$ and $c_0(y), \ldots, c_k(y)$ are already defined we define $t_{k+1}(y) = t_k(y) + m(c_k(y))$ and let $c_{k+1}(y) \in \{y_1, \ldots, y_p\}$ be such that $W_{c_{k+1}(y)} \supset B(g^{t_{k+1}(y)}(y), \delta)$. Finally, let $l = l(y)$ be such that

$$\sum_{s=0}^{l-1} m(c_s(y)) = t_l(y) < n \le t_l(y) + m(c_l(y)).$$

For each $y \in E_n$, $x_0 \in F_{c_0(y)}, \ldots, x_l \in F_{c_l(y)}$, consider

$$V(y; x_1, \ldots, x_l) = \Big\{ x \in X\,;\ d(f^{t+t_s(y)}(x), f^t(x_s)) < 2\epsilon \text{ for all } 0 \le t < m(c_s(y)) \text{ and all } 1 \le s \le l(y) \Big\}.$$

We claim that

1. the family $\mathcal{V} = \{V(y; x_1, \ldots, x_l)\,;\ y \in E_n, x_s \in F_{c_s(y)}, 1 \le s \le l(y)\}$ is an open cover of X;

2. any $(n, 4\epsilon)$-separated set intersects each element of $\mathcal{V}$ in at most one point.

In order to prove 1), let $x \in X$. Since E_n is a set which (n, δ)-spans Y, there exists $y \in E_n$ such that $d(g^i(y), g^i(\pi(x)) \le \delta$ for all $j < n$. Hence for each $0 \le s \le l(y)$, $\pi \circ f^{t_s(y)}(x) = g^{t_s(y)}(\pi(x)) \in W_{c_s(y)}$. Therefore there exists $x_s \in F_{c_s(y)}$ such that $d(f^{t+t_s}(x), f^t(x_s)) < 2\epsilon$ for all $0 \le t < m(c_s(y))$. Thus $x \in V(y; x_1, \ldots x_{l(y)})$. This proves 1). If $z, w \in V(y; x_1, \ldots x_{l(y)})$ then $d(f^{t+t_s}(z), f^{t+t_s}(w) \le d(f^{t+t_s}(z), f^t(x_s)) + d(f^t(x_s), f^{t+t_s}(w)) < 4\epsilon$, for all $0 \le t < m(c_s(y))$ and all $0 \le j < n$. This proves 2).

It follows from the claim that the cardinality $s_n(4\epsilon, X, f)$ of the maximal $(n, 4\epsilon)$-separated subset of X is bounded from above by the cardinality of the covering $\mathcal{V}$. But $\mathcal{V} = \cup_{y \in E_n} \mathcal{V}_y$ and the cardinality of $\mathcal{V}_y$ is

$$\mathrm{Card}(\mathcal{V}_y) = \prod_{s=0}^{l(y)} \mathrm{Card}(F_{c_s(y)}).$$

Hence, using $(**)$,

$$\begin{aligned} \log(\mathrm{Card}(\mathcal{V}_y)) \overset{(**)}{\to} &\le (a+\alpha) \sum_{s=0}^{l(y)} m(c_s(y)) \\ &= (a+\alpha) \left(\sum_{s=0}^{l(y)-1} m(c_s(y)) + m(c_l(y)) \right) \\ &\le (a+\alpha)(n+M), \end{aligned}$$

where $M = \max\{m(y_1), \ldots, m(y_p)\}$. Hence, since $s_n(4\epsilon, X, f) \le \mathrm{Card}(\mathcal{V}) =$

$\mathrm{Card}(\cup_{y \in E_n} \mathcal{V}_y)$, the previous inequality implies

$$\begin{aligned}\frac{1}{n} \log s_n(4\epsilon, X, f) &\leq \frac{1}{n} \log \mathrm{Card}(\mathcal{V}) \\ &\leq \frac{1}{n} \log \ \mathrm{Card}(E_n) + \frac{M+n}{n}(a+\alpha) \\ &= \frac{1}{n} \log r_n(\delta, Y, g) + \frac{M+n}{n}(\alpha + a).\end{aligned}$$

If we let n go to infinity we get, since M does not depend on n,

$$s(4\epsilon, X, f) \leq r(\delta, Y, g) + \alpha + a \leq h_t(g) + \alpha + a.$$

Since α is arbitrary, this implies

$$s(4\epsilon, X, f) \leq h_t(g) + a.$$

Hence $h_t(f) \leq h_t(g) + a$ and the theorem is proved. □

Corollary 7.1. *Assume $f\colon X \to X$ and $g\colon Y \to Y$ are continuous maps of compact metric spaces and $\pi\colon X \to Y$ is a conjugacy between f and g. Then f and g have the same topological entropy. In particular, the topological entropy does not depend on the choice of the metric.*

Proof. For a finite set $K \subset X$, we have that $h_t(f, K) = 0$. Hence, since $\pi^{-1}(y)$ is a unique point, the corollary follows. □

As we have seen in Section 3, if $f\colon [0,1] \to [0,1]$ is a piecewise monotone one can associate to each point x the itinerary of x. However, this itinerary is constant on intervals on which all iterates of f are monotone. Therefore in order to apply Theorem 7.1, we will need the following

Lemma 7.3. *Let $f\colon I \to I$ be a continuous map on a compact interval I and let $K \subset I$ be an interval such that the restriction of f^n to K is monotone for every n. Given $\epsilon > 0$ there exists a constant C_ϵ such that $r_n(\epsilon, K) \leq C_\epsilon \cdot n$ for every n. In particular, the topological entropy of f with respect to K is zero.*

Proof. Choose a finite set $F_i \subset f^i(K)$ so that each point in $f^i(K)$ is at most ϵ far from F_i. Clearly, F_i can be chosen so that $\mathrm{Card}(F_i) \leq C_\epsilon$ where $C_\epsilon = |I|/\epsilon + 1$. Next let

$$F_n = \bigcup_{i=0}^{n} f^{-i}(F_i) \cap K.$$

Since f^i is a homeomorphism on K for each $i \in \mathbb{N}$, we have $\mathrm{Card}(F_n) \leq C_\epsilon \cdot n$. We claim that F_n is a (n, ϵ)-spanning set of K. Indeed, take $y \in L$ and let $y' \in F_n$ be so that no other point in F_n is nearer to y. From the choice of y' and from the fact that $f^i\colon L \to f^i(L)$ is a homeomorphism it follows that $d(f^i(y), f^i(y')) \leq \epsilon$ for $i = 0, 1, \ldots, n$. This completes the proof of the lemma. □

Definition. Let $f\colon I \to I$ be a continuous piecewise monotone map. The *lap number* of f, which we denote by $l(f)$, is the number of maximal intervals on which f is monotone. In other words, $l(f) - 1$ is the number of turning points of f.

Lemma 7.4. *Let $f\colon I \to I$ be a piecewise monotone continuous map. Then the sequence $\{\sqrt[n]{l(f^n)}\}$ converges as $n \to \infty$.*

Proof. We claim that

$$l(g \circ f) \leq l(g)l(f)$$

when f and g are continuous functions. Indeed, in the f-image of each lap of f, g has at most $l(g)$ laps. Hence the total number of laps of $g \circ f$ can be estimated by the above formula. Using a general argument one gets from this that $\sqrt[n]{l(f^n)}$ converges. Indeed, let k be a fixed integer. If $n \in \mathbb{N}$, there exist integers p, q such that $n = pk + q$ with $0 \leq q < k$. From $l(g \circ f) \leq l(g)l(f)$ we get

$$l(f^n) \leq (l(f^k))^p l(f^q).$$

Hence $(l(f^n))^{\frac{1}{n}} \leq (l(f^k))^{\frac{p}{pk+q}} \times (l(f))^{\frac{q}{kp+q}}$. As n tends to infinity we have that p tends to infinity and q remains bounded. Hence $\frac{p}{pk+q}$ tends to $\frac{1}{k}$ and $\frac{q}{pk+q}$ tends to zero. Therefore

$$\limsup_{n\to\infty} (l(f^n))^{\frac{1}{n}} \leq (l(f^k))^{\frac{1}{k}} \text{ for all } k \in \mathbb{N}.$$

Therefore

$$\limsup_{n\to\infty} (l(f^n))^{\frac{1}{n}} \leq \inf_k (l(f^k))^{\frac{1}{k}} \leq \liminf_{k\to\infty} (l(f^k))^{\frac{1}{k}}. \quad \square$$

Now we come to the main result of this section.

Theorem 7.2. **(Misiurewicz and Szlenk)** *Let $f\colon I \to I$ be a continuous, piecewise monotone map. Then the topological entropy of f is equal to the logarithm of the number $s(f) = \lim_{n\to\infty} (l(f^n))^{\frac{1}{n}}$. In particular, if f is l-modal then $s(f) \leq l$.*

Proof. Let $I = [0, 1]$ and $0 < c_1 < \cdots < c_l < 1$ be the turning points of f. Consider the intervals $I_1 = [0, c_1), I_2 = (c_1, c_2), \ldots I_{l+1} = (c_l, 1)$. We are going to consider the symbolic dynamics of f. First we define

$$\Sigma_0(f) = \{\underline{x} = (x_i)_{i\geq 0}\,;\; x_i \in \{1, \ldots, l+1\} \text{ and } \cap_{i=0}^n f^{-i}(I_{x_i}) \neq \emptyset \text{ for all } n \in \mathbb{N}\}$$

with the metric d defined by $d(\underline{x}, \underline{y}) = \sum_{i=0}^\infty \frac{1}{2^i}\delta_{x_i,y_i}$ where $\delta_{x_i,y_i} = 1$ if $x_i \neq y_i$ and $\delta_{x_i,y_i} = 0$ if $x_i = y_i$. Notice that $\Sigma_0(f)$ is compact and invariant by the shift operator $\sigma\colon \Sigma_0(f) \to \Sigma_0(f)$. Let $\Sigma_I(f) = \{\,(\underline{x}, x) \in \Sigma_0(f) \times [0, 1]\,;\; f^i(x) \in$

$\mathrm{cl}\,(I_{x_i})$ for all $i \in \mathbb{N}\}$. $\Sigma_I(f)$ is compact. Indeed, if $x \in [0,1]$ then as we saw in Section 3, $\underline{i}(x^{\pm}) \in \Sigma_0(f)$ where $\underline{i}(t)$ is the itinerary of t. Hence $\Sigma_I(f) = \{(\underline{i}(x^{\pm}), x) \in \Sigma_0(f) \times [0,1]\}$ and because $\underline{i}(y_j) \to \underline{i}(y^{\pm})$ when $y_j \to y$, we get that $\Sigma_I(f)$ is compact. Let $\sigma_I\colon \Sigma_I(f) \to \Sigma_I(f)$ be defined by $\sigma_I(\underline{x}, x) = (\sigma(\underline{x}), f(x))$. Let $\pi_1\colon \Sigma_I(f) \to \Sigma_0(f)$ and $\pi_2\colon \Sigma_I(f) \to [0,1]$ be the projections defined by $\pi_1(\underline{x}, x) = \underline{x}$ and $\pi_2(\underline{x}, x) = x$. Both π_1 and π_2 are continuous, surjective and $\pi_1 \circ \sigma_I = \sigma \circ \pi_1$ and $\pi_2 \circ \sigma_I = f \circ \pi_2$. Hence we can use Theorem 7.1 and we get

$$h_t(\sigma) \le h_t(\sigma_I) \le h_t(\sigma) + \sup_{\underline{x} \in \Sigma_0(f)} h_t(\sigma_I, \pi_1^{-1}(\underline{x})).$$

Each fibre $\pi_1^{-1}(\underline{x})$ is equal to $\underline{x} \times I(\underline{x})$ where $I(\underline{x})$ is the closure of the set of points whose itinerary is equal to $\underline{x}$. Therefore the restriction to $I(\underline{x})$ of iterates of f are monotone. Hence we can use Lemma 7.3 to conclude that $h_t(\sigma_I, \pi_1^{-1}(\underline{x}))$ is equal to zero. Hence $h_t(\sigma) = h_t(\sigma_I)$. On the other hand, the cardinality of $\pi_2^{-1}(x)$ is at most two. Therefore, again using Theorem 7.1, we get $h_t(\sigma_I) = h_t(f)$. Thus $h_t(f) = h_t(\sigma)$.

For each $\underline{x} \in \Sigma_0(f)$ and each $n > 0$ consider the cylinder $C_n(\underline{x}) = \{\underline{y} \in \Sigma_0(f)\,;\; y_i = x_i, \forall i = 0, \ldots, n-1\}$. So C_n corresponds to the intervals on which f^n is monotone. Let us denote by $\mathcal{C}_n$ the family of such cylinders. It is clear that the cardinality of $\mathcal{C}_n$ is equal to the lap number $l(f^n)$. If $C_n(\underline{x}) \neq C_n(\underline{y})$ then $x_i \neq y_i$ for some $i < n$. Using the definition of the metric d, one gets $d(\sigma^i(\underline{x}), \sigma^i(\underline{y})) > \frac{1}{2}$. Hence, if we choose one point in each element of $\mathcal{C}_n$, we get a set which $(n, \frac{1}{2})$-separated. Therefore $l(f^n) = \mathrm{Card}(\mathcal{C}_n) \le s_n(\frac{1}{2}, \Sigma_0(f), \sigma)$. Thus $\frac{1}{n}\log l(f^n) \le \frac{1}{n}\log s_n(\frac{1}{2}, \Sigma_0(f), \sigma)$ or, by taking the lim sup of both members of the previous inequality, $\log s(f) \le s(\frac{1}{2}, \Sigma_0(f), \sigma)$. Therefore

$$\log s(f) \le h_t(\sigma).$$

Let us now prove the reverse inequality. Fix an integer p. If $C_{n+p}(\underline{x}) = C_{n+p}(\underline{y})$ then

$$d(\sigma^i(\underline{x}), \sigma^i(\underline{y})) \le \frac{1}{2^p} \text{ for all } i \le n.$$

Therefore, if we choose one point in each of the elements of $\mathcal{C}_{n+p}$, we get an $(n, \frac{1}{2^p})$-spanning set. Hence

$$r_n(\frac{1}{2^p}, \Sigma_0(f), \sigma) \le \mathrm{Card}(\mathcal{C}_{n+p}) = l(f^{n+p}).$$

Thus

$$\frac{1}{n}\log r_n(\frac{1}{2^p}, \Sigma_0(f), \sigma) \le \frac{n+p}{n}\frac{1}{n+p}\log l(f^{n+p}).$$

Now, letting n tend to infinity, we get $\limsup \frac{1}{n} r_n(\frac{1}{2^p}, \Sigma_0(f), \sigma) \le \log s(f)$. Therefore $h_t(\sigma) \le \log s(f)$. □

Remark. We shall show in the next section that the lap numbers $l(f^n)$ are determined by the kneading invariant of f. Of course one can deduce this from the techniques developed in Section 3, but in the next section we will give a much more precise relationship between the kneading invariants and the lap numbers of a map.

Corollary 7.2. *If $f\colon I \to I$ is a piecewise linear continuous map with slope equal to $\pm s$ then the topological entropy of f is equal to* $\max(0, \log s)$.

Proof. Let $I = [0,1]$. If $x, y \in I$ we have $d(f^i(x), f^i(y)) \le s^i d(x,y)$. Hence the set F that partitions $[0,1]$ into equal intervals of size $\frac{\epsilon}{s^n}$ is an (n,ϵ)-spanning set. The cardinality of F is at most $\frac{s^n}{\epsilon}$. Therefore $\frac{1}{n}\log r_n(\epsilon, I, f) \le \log s - \frac{1}{n}\log\epsilon$. From this we get $h_t(f) \le \log s$.

Let J be a lap of f^n. Then $|f^n(J)| = s^n|J|$. Therefore $|J| \le s^{-n}$. Since the laps are disjoint, $l(f^n) \ge s^n$. From the previous theorem we get $h_t(f) \ge \log s$. □

In Section 9 of this chapter we will show that the topological entropy depends continuously on the map if one considers C^1 maps in the C^1 topology and restricts oneself to maps with the same number of turning points.

Exercise 7.1. Show that if $f\colon I \to I$ is a continuous interval map, then $h(f) = h(f|\Omega(f))$ where $\Omega(f)$ is the non-wandering set of f. (Hint: elaborate on the proof given in Lemma 7.3. However, this result holds also for a general continuous map on a compact space.) In particular, f has topological entropy zero whenever $\Omega(f)$ consists of only a finite number of points.

Exercise 7.2. Show that the topological entropy of $f\colon [0,1] \to [0,1]$ is at least $\log 2$ if there exist two open disjoint intervals I_1 and I_2 such that $f(I_1) \supset I_1 \cup I_2$ and $f(I_2) \supset I_1 \cup I_2$. (Hint: in this case the number of laps of f^n is at least 2^n.)

Exercise 7.3. Exercise 7.3 If f has a periodic point of period s and if s is not of the form $s = 2^n$ for some $n \in \mathbb{N}$, then f has positive topological entropy. (Hint: use Lemma 7.2, Lemma 1.5 from the proof of the theorem of Sarkovskii and the previous exercise.)

Exercise 7.4. In general the topological entropy does not depend continuously on the map in the C^0 topology. (Hint: consider, for example, the non-symmetric piecewise linear map $F\colon [-1,1] \to [-1,1]$ with slope 1 and $s_2 < -2$ to the left respectively to the right of the turning point and with $F(-1) = -1$. Show that $F^n = F$ for all $n \ge 1$ and therefore $h(F) = 0$. Perturb F to a piecewise linear continuous map F^* with slopes $s > 1$ and s_2 as in Figure 7.1. Let p be the fixed point for F^* with $|DF^*(p)| = |s_2| > 2$. Take q so that $F^*(q) = p$. Show that the second iterate of F^* maps $[q,p]$ strictly over itself as shown in this figure. Using Exercise 7.2 it follows that $h(F^*) \ge \log(2)/2$.)

8 The Piecewise Linear Model

In this section we will prove, following Milnor and Thurston (1977), that a continuous, piecewise monotone map with positive topological entropy is semi-conjugate to a continuous, piecewise linear map with constant slope and with the same entropy. Essentially this result was already proved by Parry (1966).

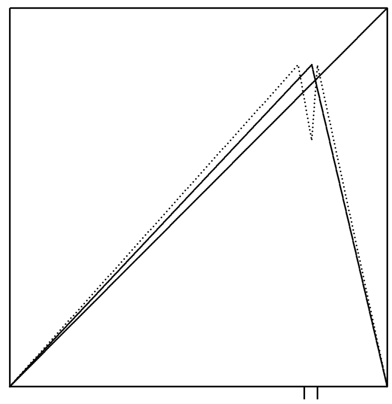

Fig. 7.1: The maps F^* and $(F^*)^2$.

Theorem 8.1. **(Parry) and (Milnor and Thurston)** *Assume that $f: I \to I$ is a continuous, piecewise (strictly) monotone map with positive topological entropy $h_t(f)$ and let $s = \exp(h_t(f))$. Then there exists a continuous, piecewise linear map $T: [0,1] \to [0,1]$ with slope $\pm s$, and a continuous, monotone increasing map $\lambda: I \to [0,1]$ which is a semi-conjugacy between f and T, i.e.,*

$$\lambda \circ f = T \circ \lambda.$$

The proof of this theorem gives also a very important relationship between the lap numbers $l(f^n)$ and the kneading invariants. In the next section we will show that often the growth rate of $l(f^n)$ depends continuously on f. Of course to show that the growth rate of numbers $l(n)$ depending on some parameter a varies continuously on a is not that easy. In order to solve this problem Milnor and Thurston associated both to the lap numbers and to the kneading invariants a power series. The product of these two power series is an extremely simple power series which is meromorphic on the Riemann sphere with a unique simple pole in $z = 1$. In this way the growth rate of $l(f^n)$ will turn out to be a special zero of the power series associated to the kneading invariants of f. To show that the zero of this last power series depends continuously on the map will be quite easy.

Let us first give a rough idea how to use power series in order to construct conjugacies. Consider the functions defined by

$$L_f(t) = \sum_{n=0}^{\infty} l(f^n)t^n$$

and

$$L_f(J;t) = \sum_{n=0}^{\infty} l(f^n|J)t^n,$$

where $l(f^n|J)$ is the lap number of f^n restricted to J. (Sometimes we shall simply write $L(t)$ and $L(J;t)$ instead of $L_f(t)$ and $L_f(J;t)$.) $L_f(t)$ is holomorphic in the disc centred at the origin of the complex plane and with radius $\frac{1}{s}$, where $s = \lim_{n\to\infty} \sqrt[n]{l(f^n)}$. From Theorem 7.2 we had that $s > 1$ if the topological entropy of f is positive. Next show that

$$\Lambda(J) = \lim_{t\to 1/s} \frac{L(J;t)}{L(I;t)}$$

exists. Once we know this, it is very easy to show that Λ defines an additive continuous probability measure on I and that $\Lambda(f(J)) = s\Lambda(J)$ whenever f is monotone on J. In this way one constructs a semi-conjugacy with a continuous piecewise linear map.

Rather than merely showing that $\Lambda(J) = \lim_{t\to 1/s} \frac{L(J;t)}{L(I;t)}$ exists, we will show that it has a meromorphic extension to $|t| < 1$. This will be done by showing that there is a rather intriguing relationship between the function $L_f(t)$ and another function $D_f(t)$ called the *kneading determinant* of f. This kneading determinant is holomorphic in the unit disc and its Taylor series is constructed from the itineraries of the turning points of f and can be simply deduced from the kneading invariants of f. A consequence of this relationship will be that $L_f(t)$ is a meromorphic function on the unit disc whose poles are contained in the set of zeros of the kneading determinant. Furthermore, the point $t = \frac{1}{s}$ is a pole of $L_f(t)$ and the semi-conjugacy will be constructed using this pole. These last results will also be crucial in the next section.

Exercise 8.1. Consider $f(x) = 4x(1-x)$. Show that $L_f(t) = \sum_{n\geq 0} 2^n t^n = \frac{1}{1-2t}$. In particular, $L_f(t)$ is a rational function with a single pole at $1/2$.

Let $f\colon I \to I$ be a l-modal map. That is, f is piecewise monotone map with l turning points $0 < c_1 < \cdots < c_l < 1$ and with $f(\partial I) \subset \partial I$. Moreover, let $c_0 = 0$ and $c_{l+1} = 1$ be the endpoints of I and let $I_0 = [0, c_1)$, $I_1 = (c_1, c_2)$, $\dots$, $I_{l+1} = (c_l, 1]$. As before, we denote by Σ the space $\{I_1, \dots, I_{l+1}, c_1, \dots, c_l\}^{\mathbb{N}}$ and Σ_0 is the shift-invariant subspace of sequences $\underline{x} = (x_0, x_1, \dots)$ such that $x_n \in \{I_1, \dots, I_{l+1}\}$ for all n. In Σ we consider again the product topology which is induced by the metric $d(\underline{x}, \underline{y}) = \sum_{n=0}^{\infty} \frac{1}{2^n} \delta_{x_n y_n}$.

Next we define for $k = 1, \dots, l+1$ and $n \geq 0$,

$$\Theta_n^k\colon \Sigma \to \{-1, -\frac{1}{2}, 0, \frac{1}{2}, 1\}$$

as follows. For $\underline{x} \in \Sigma$, let $\epsilon_0(\underline{x}) = 1$ and $\epsilon_n(\underline{x}) = \epsilon(x_0) \times \cdots \times \epsilon(x_{n-1})$ where $\epsilon(x_i)$ is the sign of x_i, i.e., $\epsilon(c_1) = \cdots = \epsilon(c_l) = 0$, $\epsilon(I_j) = 1$ if f is increasing on I_j and $\epsilon(I_j) = -1$ if f is decreasing on I_j. Then for $k = 1, \dots, l+1$ and $n \geq 0$ we define

$$\Theta_n^k(\underline{x}) = \begin{cases} +1 & \text{if } \ x_n = I_k \text{ and } \epsilon_n(\underline{x}) = 1, \\ -1 & \text{if } x_n = I_k \text{ and } \epsilon_n(\underline{x}) = -1, \\ +\frac{1}{2} & \text{if } x_n \in \partial I_k = \{c_k, c_{k-1}\} \text{ and } \epsilon_n(\underline{x}) = 1, \\ -\frac{1}{2} & \text{if } x_n \in \partial I_k = \{c_k, c_{k-1}\} \text{ and } \epsilon_n(\underline{x}) = -1, \\ 0 & \text{otherwise, i.e., if either } \epsilon_n(\underline{x}) = 0 \text{ or } x_n \notin \{c_k, I_k, c_{k+1}\}. \end{cases}$$

Note that $\Theta_0^k(\underline{x})$ is equal to $+1$ if $x_0 = I_k$ and equal to $\frac{1}{2}$ if $x_n \in \partial I_k = \{c_k, c_{k-1}\}$ and 0 otherwise.

Now let $\mathbb{D}$ be the *open* unit disc in the complex plane. For each $\underline{x} \in \Sigma$ and each $t \in \mathbb{D}$, define

$$\Theta^k(\underline{x};t) = \sum_{n=0}^{\infty} \Theta_n^k(\underline{x})t^n$$

and

$$\Theta(\underline{x};t) = (\Theta^1(\underline{x};t), \dots, \Theta^{l+1}(\underline{x};t)).$$

So the k-th component of this vector only depends on the occurrences of the terms c_k, I_k or c_{k+1} in $\underline{x}$. Since the coefficients of the Taylor series of $\Theta^k(\underline{x};t)$ are bounded, it follows that $\Theta^k(\underline{x};t)$ is indeed a holomorphic function of t on the unit disc $\mathbb{D}$. Hence $\Theta(\underline{x};t)$ is well defined for each $\underline{x} \in \Sigma$ and each $t \in \mathbb{D}$.

Now we can define a similar power series for points $\underline{x}$ which are of the form $\underline{i}_f(x)$ where $x \in I$. More precisely, define

$$\theta_n^k(x) = \Theta_n^k(\underline{i}_f(x)),$$

and

$$\theta_n(x) = (\theta_n^1(x), \dots, \theta_n^{l+1}(x)) \in \{\pm 1, \pm\frac{1}{2}, 0\}^{l+1}$$

and the *invariant coordinates* of a point $x \in I$ as

$$\theta_f(x;t) = \Theta(\underline{i}_f(x);t)) = \sum_{n=0}^{\infty} \theta_n(x)t^n.$$

Since $|\theta_n(x)| \le 1$ for each $x \in I$, each of the coordinate functions $t \mapsto \theta_f^k(x;t)$ is a holomorphic $\mathbb{C}^{l+1}$-valued function on $\mathbb{D}$. If no confusion can arise we will write θ instead of θ_f.

Of course, there is some redundancy: as we will show in Lemma 8.1 below one can reconstruct $\theta^{l+1}(x)$ from $\theta^1(x), \dots, \theta^l(x)$. For this reason, in the unimodal case one often uses a slightly simpler $\mathbb{C}$-valued holomorphic function, see for example Milnor and Thurston (1977) or Van Strien (1987). This function is defined in the exercise below.

Exercise 8.2. Let $f\colon [0,1] \to [0,1]$ be unimodal and assume for simplicity that f is monotone increasing on I_1. Let $\overline{\theta}_f(x) = \theta_f^1(x) - \theta_f^2(x)$ and write $\overline{\theta}(x) = \sum_{n\ge 0} \overline{\theta}_n(x)t^n$. Show that

$$\overline{\theta}_n(x) = \begin{cases} 1 & \text{if } f^{n+1} \text{ is increasing near } x \\ -1 & \text{if } f^{n+1} \text{ is decreasing near } x \\ 0 & \text{if } f^{n+1} \text{ has a local extremum at } x. \end{cases}$$

(Hint: $\Theta_n^1(\underline{x}) - \Theta_n^2(\underline{x}) = 1$ if $\epsilon_n(\underline{x}) = 1$ and $x_n = I_1$ or if $\epsilon_n(\underline{x}) = -1$ and $x_n = I_2$. Since $\epsilon(I_1) = 1$, it follows that $\Theta_n^1(\underline{x}) - \Theta_n^2(\underline{x}) = 1$ if and only if $\epsilon_{n+1}(\underline{x}) = 1$.)

Exercise 8.3. Consider $f(x) = 4x(1-x)$. Show that

$$\theta_f(c^-;t) = (1 - [t^2 + t^3 + t^4 + \ldots], t) = (1 - \frac{t^2}{1-t}, t)$$

$$\theta_f(c^+;t) = (t^2 + t^3 + t^4 + \ldots, 1-t) = (\frac{t^2}{1-t}, 1-t).$$

Hence the term $\overline{\theta}_f(c^-;t)$ from the previous exercise is equal to $1 - [t + t^2 + t^3 + t^4 + \ldots] = 1 - \frac{t}{1-t}$. Check that this coincides with the alternative description given in the previous exercise.

Remarks. 1. Notice that if $\underline{i}(x) \notin \Sigma_0$ then there exists $n \in \mathbb{N}$ such that $f^n(x) \in \{c_1, \ldots, c_l\}$ and therefore $\epsilon_i(\underline{i}(x)) = 0$ for $i \geq n$ and $\Theta^k(\underline{i}(x))$ is a polynomial of at most degree n for every k. Similarly, let x be in the backward orbit of the turning points, i.e., $f^n(x) = c_k$ for some $k = 1, \ldots, l+1$ and assume $x, \ldots, f^{n-1}(x)$ are not turning points. Then for each $k = 1, \ldots, l+1$, $\theta^k(x;t)$ is a polynomial in t of degree at most n. In fact, $\theta^k(x;t)$ and $\theta^{k+1}(x;t)$ are polynomials of degree equal to n and the coefficient of the t^n term is for both polynomials equal to $\frac{1}{2}$ if f^n is monotone increasing in a neighbourhood of x and $-\frac{1}{2}$ if f^n is monotone decreasing. 2. If x is not in the backward orbit of the turning points then

$$\theta_n^k(x) = \begin{cases} +1 & \text{if } f^n(x) \in I_k \text{ and } f^n \text{ is locally increasing near } x, \\ -1 & \text{if } f^n(x) \in I_k \text{ and } f^n \text{ is locally decreasing near } x, \\ 0 & \text{if } f^n(x) \notin I_k. \end{cases}$$

3. The mapping $(x,t) \to \theta(x,t)$ is not continuous as a function of the first variable. In fact, since Θ is continuous, the left and right handed limits are respectively

$$\theta(x^-;t) = \Theta(\underline{i}_f(x^-))(t),$$
$$\theta(x^+;t) = \Theta(\underline{i}_f(x^+))(t).$$

Hence θ is discontinuous precisely at the points where the itinerary map is discontinuous, i.e., at the set of backward orbits of the turning points. Notice that from the definition of $\Theta_0^k(\underline{x})$ above, for $1 \leq k \leq l+1$ and $1 \leq i \leq l$,

$$\theta_0^k(c_i^-) = \begin{cases} 1 & \text{when } k = i \\ 0 & \text{otherwise;} \end{cases}$$

and similarly

$$\theta_0^k(c_i^+) = \begin{cases} 1 & \text{when } k = i+1 \\ 0 & \text{otherwise.} \end{cases}$$

So let $\mathcal{H}(\mathbb{D})$ be the space of holomorphic $\mathbb{C}^{l+1}$-valued functions on $\mathbb{D}$ endowed with the compact open topology: a basis of neighbourhoods for a map $\phi \in \mathcal{H}(\mathbb{D})$

is $V(\epsilon, K) = \{\psi \in \mathcal{H}(\mathbb{D})\,;\; |\psi(t) - \phi(t)| < \epsilon,\, \forall t \in K\}$, where $K \subset \mathbb{D}$ is a compact set and ϵ is a positive real number. From the previous expressions it follows that the map $\theta(x) \in \mathcal{H}(\mathbb{D})$ defined by $t \mapsto \theta(x;t)$ is not continuous in $x \in I$. On the other hand, if we consider the map $\Theta(\underline{x}) \in \mathcal{H}(\mathbb{D})$ defined by $t \mapsto \Theta(\underline{x};t)$ where $\underline{x}$ is contained in the space of *all* sequences in Σ, then $\Sigma \ni \underline{x} \mapsto \Theta(\underline{x}) \in \mathcal{H}(\mathbb{D})$ is continuous. Indeed, let $\underline{x} \in \Sigma$, $K \subset \mathbb{D}$ be a compact set and $\epsilon > 0$. Choose $r < 1$ such that $|t| \leq r$ for all $t \in K$. Choose $N \in \mathbb{N}$ such that $2\frac{r^N}{1-r} < \epsilon$. Let $\mathcal{V}$ be a neighbourhood of $\underline{x} \in \Sigma$ such that if $\underline{y} \in \mathcal{V}$ then $y_i = x_i$, $\forall i < N$. Then $|\Theta^k(\underline{x})(t) - \Theta^k(\underline{y})(t)| \leq \sum_{n=N}^{\infty} |\Theta_n^k(\underline{x}) - \Theta_n^k(\underline{y})| t^n \leq \sum_{n=N}^{\infty} 2|t^n| = \frac{2|t|^N}{1-|t|} \leq 2\frac{r^N}{1-r} < \epsilon$. Therefore Θ is a continuous map.

In the next lemma it is shown that $\theta^1(x), \ldots, \theta^{l+1}(x)$ are strongly related. Of course this is not surprising: the coefficients of the power series $\theta^k(x)$ are determined by the visits of x to $\mathrm{cl}\,(I_k)$ and so are related to the coefficients of the remaining power series.

Lemma 8.1. *For every $x \in I$ we have the following identity:*

$$\sum_{k=1}^{l+1} (1 - \epsilon(I_k)t)\theta^k(x;t) = 1.$$

Proof. We claim that

$$\text{(i)} \qquad \sum_{k=1}^{l+1} \theta_0^k(x) = 1;$$

$$\text{(ii)} \qquad \sum_{k=1}^{l+1} \epsilon(I_k)\theta_n^k(x) = \sum_{k=1}^{l+1} \theta_{n+1}^k(x) \text{ for } n \geq 0.$$

Indeed, if $x \in I_j$ then $\theta_0^j(x) = 1$ and $\theta_0^k(x) = 0$ for $k \neq j$ and therefore (i) holds. On the other hand, if $x = c_k$ then $\theta_0^k(x) = \frac{1}{2}$, $\theta_0^{k+1} = \frac{1}{2}$ and $\theta_0^j(x) = 0$ for all $j \neq k, k+1$ and so again (i) holds. This completes the proof of (i). If $\epsilon_n = \epsilon(i_0(x)) \times \cdots \times \epsilon(i_{n-1}(x)) = 0$ then $\theta_n^k(x) = \theta_{n+1}^k(x) = 0$ for all $k = 1, \ldots, l+1$ and therefore both the left and the right hand side of (ii) is zero. Suppose now that $\epsilon_n \neq 0$. If $f^n(x) \in I_k$ and $f^{n+1}(x) \in I_m$ then $\theta_{n+1}^m(x) = \epsilon(I_k)\theta_n^k(x)$, $\theta_{n+1}^j(x) = 0$ for $j \neq m$ and $\theta_n^j(x) = 0$ for $j \neq k$. Hence both members in (ii) are equal to $\epsilon(I_k)\theta_n^k(x)$. If $f^n(x) = c_k$ then the second member in (ii) is equal to zero and the first member is equal to $\epsilon(I_k)\theta_n^k(x) + \epsilon(I_{k+1})\theta_n^{k+1}(x)$ which is also equal to zero, since $\epsilon(I_k) = -\epsilon(I_{k+1})$ and $\theta_n^k(x) = \epsilon_n \cdot \frac{1}{2} = \theta_n^{k+1}(x)$. Finally, if $f^n(x) \in I_k$ and $f^{n+1}(x) = c_m$ the first member of (ii) is equal to $\epsilon(I_k)\theta_n^k(x)$ and the second member is equal to $\theta_{n+1}^m(x) + \theta_{n+1}^{m+1}(x) = \theta_n^k(x)\epsilon(I_k)\frac{1}{2} + \theta_n^k(x)\epsilon(I_k)\frac{1}{2} = \epsilon(I_k)\theta_n^k(x)$. Thus the claim is proved.

From the claim we get

$$\sum_{k=1}^{l+1}\left[[1-\epsilon(I_k)\,t]\cdot\sum_{n=0}^{N}\theta_n^k(x)t^n\right]=\sum_{n=0}^{N}\sum_{k=1}^{l+1}\left(\theta_n^k(x)t^n-\epsilon(I_k)\theta_n^k(x)\,t^{n+1}\right)$$
$$=1-\sum_{k=1}^{l+1}\epsilon(I_k)\theta_N^k(x)t^{N+1}.$$

Therefore

$$\lim_{N\to\infty}\sum_{k=1}^{l+1}\left[(1-\epsilon(I_k)t)\sum_{n=0}^{N}\theta_n^k(x)t^n\right]=1$$

uniformly on any compact subset of $\mathbb{D}$. □

Definition. The $l\times(l+1)$ matrix

$$\begin{pmatrix} N_{1,1}(t) & N_{1,2}(t) & \dots & N_{1,l+1}(t)\\ N_{2,1}(t) & N_{2,2}(t) & \dots & N_{2,l+1}(t)\\ \vdots & \vdots & \ddots & \vdots\\ N_{l,1}(t) & N_{l,2}(t) & \dots & N_{l,l+1}(t)\end{pmatrix},$$

where $N_{i,j}(t)$ are the holomorphic functions defined by

$$N_{i,j}(t)=\theta^j(c_i^+;t)-\theta^j(c_i^-;t)$$

is called the *kneading matrix.* The i-th row of this matrix, i.e., the holomorphic mapping $N_i\colon\mathbb{D}\to\mathbb{C}^l$ defined by

$$N_i(t)=\theta(c_i^+;t)-\theta(c_i^-;t)$$

is the *kneading vector* associated to the i-th turning point. The kneading matrix carries all the combinatorial information on the map f because it allows us to recover the itinerary of the turning points, and because of Lemma 8.1 the columns of this matrix are related in a very neat way.

Let us write $[N_{i,j}(t)]=\sum_{n=0}^{\infty}[N_{i,j}^n]t^n$, where $[N_{i,j}^n]$ is an $l\times(l+1)$ matrix of integers. For $n>0$, each entry of this matrix is in $\{0,\pm2\}$ because $\theta_n^j(c_i^+)=-\theta_n^j(c_i^-)\in\{0,\pm1\}$ for $n>0$. Using Remark 3 above we get that the matrix $[N_{i,j}^0]$ is equal to

$$\begin{pmatrix} -1 & 1 & 0 & 0 & \dots & 0 & 0\\ 0 & -1 & 1 & 0 & \dots & 0 & 0\\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots\\ 0 & 0 & 0 & 0 & \dots & -1 & 1\end{pmatrix}.$$

In the next lemma we shall use Lemma 8.1 to show that the determinants of the matrices which we get by deleting one of the columns of the kneading matrix are all equal. So we will call this holomorphic function the *kneading determinant* D_f of f.

Lemma 8.2. *Let $D_i(t)$ be the determinant of the $l \times l$ matrix obtained from the kneading matrix by deleting the i-th column. Then*

1. *D_i is a holomorphic function on $\mathbb{D}$;*
2. *the function $D_f(t)$ defined by $D_f(t) = (-1)^{i+1}\frac{D_i(t)}{1-\epsilon(I_i)t}$ is independent of i (this function is called the kneading determinant of f;*
3. *$D_f(0) = 1$.*

Exercise 8.4. Consider $f(x) = 4x(1-x)$. Using Exercise 8.3, show that the kneading matrix $(N_{11}(t)\, N_{12}(t))$ of f is equal to

$$\begin{pmatrix} -1 + \frac{2t^2}{1-t} & 1-2t \end{pmatrix} = \begin{pmatrix} \frac{(t+1)(2t-1)}{1-t} & 1-2t \end{pmatrix}.$$

Similarly, the kneading matrix of $f(x) = 2x(1-x)$ is equal to

$$\begin{pmatrix} -1 - \frac{2t}{1-t} & 1 \end{pmatrix}.$$

Exercise 8.5. Consider $f(x) = 4x(1-x)$. Using Exercise 8.4 show that

$$D_f(t) = \frac{2t-1}{1-t} = -1 + [t + t^2 + t^3 + \dots].$$

Proof of Lemma 8.2 *Let $N^1, \dots N^{l+1}$ be the columns of the kneading matrix and let $[N^1, \dots, \hat{N}^i, \dots N^{l+1}]$ be the matrix obtained from the kneading matrix by deleting the i-th column. From Lemma 8.1, it follows that*

$$(*) \qquad \sum_{j=1}^{l+1} (1-\epsilon(I_j)t) N^j(t) = 0.$$

Using $(*)$ and the fact that the determinant is an alternating $l+1$-linear function of the columns, we get

$$\begin{aligned}
&(-1)^{l+2}(1-\epsilon(I_{l+1})t) \cdot \mathrm{Det}[N^1, \dots, \hat{N}^i, \dots, N^l, N^{l+1}] \\
&\quad = \mathrm{Det}[N^1, \dots, \hat{N}^i, \dots, N^l, (-1)^{l+2}(1-\epsilon(I_{l+1})t)N^{l+1}] \\
&\quad \overset{(*)}{\to}= \mathrm{Det}[N^1, \dots, \hat{N}^i, \dots, N^l, (-1)^{l+1} \sum_{j \neq l+1} (1-\epsilon(I_j)t)N^j] \\
&\quad = \sum_{j \neq l+1} \mathrm{Det}[N^1, \dots, \hat{N}^i, \dots, N^l, (-1)^{l+1}(1-\epsilon(I_j)t)N^j] \\
&\quad = \mathrm{Det}[N^1, \dots, \hat{N}^i, \dots, (-1)^{l+1}(1-\epsilon(I_i)t)N^i] \\
&\quad = (-1)^{l-i}\, \mathrm{Det}[N^1, \dots, (-1)^{l+1}(1-\epsilon(I_i)t)N^i, \dots, \hat{N}^{l+1}] \\
&\quad = (-1)^{2l-i+1}(1-\epsilon(I_i)t)\mathrm{Det}[N^1, \dots, N^i, \dots, \hat{N}^{l+1}] \\
&= (-1)^{i+1}(1-\epsilon(I_i)t)D_{l+1}.
\end{aligned}$$

This proves Statement 2). Statement 1) is obvious and Statement 3) is true because the matrix $[N^2(0), \dots, N^l(0)]$ is lower triangular with a 1 in each term *in the diagonal.* $\square$*Our next aim is to show how the kneading matrix and the number of laps of iterates of f are related. In order to do this we first show that the discontinuities of $x \mapsto \theta(x;t)$ are of a very special type.*

Lemma 8.3. *If $x \in I$ is such that $f^n(x)$ is a turning point c_k and $f^i(x)$ is not a turning point for $0 \le i < n$ then*

$$\theta(x^+;t) = \theta(x;t) + \frac{1}{2}t^n N_k(t);$$

$$\theta(x^-;t) = \theta(x;t) - \frac{1}{2}t^n N_k(t).$$

and $\theta(x;t)$ is a polynomial map of degree n. Here N_k is the kneading vector associated to the k-th turning points (i.e., the k-th row of the kneading matrix), as before.

Remark. In particular,

$$\theta(c_k^+;t) = \theta(c_k;t) + \frac{1}{2}N_k(t),$$

$$\theta(c_k^-;t) = \theta(c_k;t) - \frac{1}{2}N_k(t).$$

Since $\theta(c_k;t) = (0,\dots,0,1/2,1/2,0,\dots,0)$ where the terms $1/2$ are in the k-th and $k+1$-th position, this implies that each one of the terms $\theta(c_k^+;t)$, $\theta(c_k^-;t)$ and $N_k(t)$ determines the other two. So all, except the constant terms, in the power series of $\theta(c_k^+;t)$ and $-\theta(c_k^-;t)$ coincide.

Exercise 8.6. Show that if $f\colon [0,1] \to [0,1]$ is unimodal, i.e., $l = 1$, and f is increasing on I_1 that then

$$D_f(t) = \frac{-N_{1,1}(t)}{1+t} = \frac{N_{1,2}(t)}{1-t}$$

and therefore $2D_f(t) = -N_{1,1} + N_{1,2} = -\theta^1(c^+) + \theta^1(c^-) + \theta^2(c^+) - \theta^2(c^-) = -\overline{\theta}_f(c^+) + \overline{\theta}_f(c^-)$. In particular,

$$D_f(t) = \overline{\theta}_f(c^-) = -\overline{\theta}_f(c^+).$$

Compute this for $f(x) = 4x(1-x)$ using the terms $\overline{\theta}_n$ from Exercise 8.2 and compare this with the result from Exercise 8.5.

Proof of Lemma 8.3 Let $\underline{i}_f(x^-) = \underline{x}^-$, $\underline{i}_f(x^+) = \underline{x}^+$ and $\underline{i}_f(x) = \underline{x}$. Furthermore, let $\underline{i}_f(c^-) = \underline{v}$ and $\underline{i}_f(c^+) = \underline{w}$. Then, $x_i^+ = x_i^- = x_i$, $i < n$, and because $f(c^+) = f(c^-)$, $x_{n+k}^+ = x_{n+k}^- = v_k = w_k$, $k \ge 1$. So $\epsilon_k(\underline{v}) = -\epsilon_k(\underline{w})$ for $k \ge 1$. Finally, if f^n is locally increasing near x then for $k \ge 0$,

$$\epsilon_{n+k}(\underline{x}^-) = \epsilon_n(\underline{x}) \times \epsilon_k(\underline{v}) = \epsilon_k(\underline{v}) = \frac{1}{2}\left(\epsilon_k(\underline{v}) - \epsilon_k(\underline{w})\right)$$

and

$$\epsilon_{n+k}(\underline{x}^+) = \epsilon_n(\underline{x}) \times \epsilon_k(\underline{w}) = \epsilon_k(\underline{w}) = -\frac{1}{2}\left(\epsilon_k(\underline{v}) - \epsilon_k(\underline{w})\right);$$

otherwise, if f^n is locally decreasing near x then one has for $k \geq 1$, again that

$$\epsilon_{n+k}(\underline{x}^-) = \epsilon_n(\underline{x}) \times \epsilon_k(\underline{w}) = \epsilon_k(\underline{v}) = \frac{1}{2}\left(\epsilon_k(\underline{v}) - \epsilon_k(\underline{w})\right)$$

and

$$\epsilon_{n+k}(\underline{x}^+) = \epsilon_n(\underline{x}) \times \epsilon_k(\underline{v}) = -\epsilon_k(\underline{v}) = -\frac{1}{2}\left(\epsilon_k(\underline{v}) - \epsilon_k(\underline{w})\right).$$

From this and from the definition of the function Θ it follows that all the coefficients, except possibly those corresponding to the term t^n, of the power series $\theta(x^+;t)$ and $\theta(x;t) + \frac{1}{2}t^n N_k(t)$ coincide. Using Remark 1 above Lemma 8.1, the lemma follows immediately. □In

the next lemma we introduce functions $\Gamma_i(J)(t)$ which are holomorphic on the disc $|t| < 1/s$ where s is the growth rate of $l(f^n)$. These functions are closely related to the maps $t \mapsto L(J;t)$ defined in the introduction of this section.

Lemma 8.4. *Let J be an interval in I and*

$$\Gamma_{i,n}(J) = \#\{x \in \text{int}\,(J)\,;\;\; f^n(x) = c_i \text{ and } f^k(x) \text{ is not a turning point for } k < n\}.$$

Then, for each $i = 1, \ldots, l$, the function

$$\Gamma_i(J)(t) = \sum_{n=0}^{\infty} \Gamma_{i,n}(J)t^n$$

is holomorphic on the disc $\{t \in \mathbb{D}\,;\; |t| < \frac{1}{s}\}$ where, as before, s is the limit of $\sqrt[n]{l(f^n)}$ as $n \to \infty$.

Proof. Since $\Gamma_{i,n}(J) \leq l(f^n|J) \leq l(f^n)$, we get

$$\limsup_{n\to\infty} \sqrt[n]{\Gamma_{i,n}(J)} \leq \limsup_{n\to\infty} \sqrt[n]{l(f^n)} = s.$$

Therefore the radius of convergence of the Taylor series of $\Gamma_i(J)$ is at least equal to $\frac{1}{s}$. □

In the next theorem we will give a precise relationship between the power series $\Gamma_i(J)(t)$ and the kneading vectors $N_i(t)$ associated to the i-th turning point. From this relationship it will follow that the map $t \mapsto L(J;t)$, which is holomorphic on $|t| < 1/s$, has a meromorphic extension to the disc $\mathbb{D}$. We shall also use Theorem 8.2 in Section 9.

Theorem 8.2. *Let J be an open interval in I with endpoints $a < b$ and $\Gamma_i(J)$ the holomorphic function defined in Lemma 8.4. For every $t \in \{\mathbb{C}\,;\; |t| < \frac{1}{s}\}$ we have*

$$\theta(b^-;t) - \theta(a^+;t) = \sum_{i=1}^{l} \Gamma_i(J)(t) \times N_i(t).$$

Here $\Gamma_i(J)(t) \times N_i(t)$ is the product of the scalar function $\Gamma_i(J)(t)$ and the kneading vector $N_i(t) \in \mathbb{C}^{l+1}$ associated to the i-th turning point.

Proof. Let $n \in \mathbb{N}$ and let $F_n = \cup_{k \le n} \cup_{i=1}^{l} E_{i,k}(J)$ where $E_{i,k}(J)$ is the set of points $x \in \operatorname{int}(J)$ for which $f^k(x) = c_i$ and $f^j(x)$ is not a turning point for $j < n$. The previous lemma shows that the cardinality of these sets does not increase too fast as $n \to \infty$. The finite set F_n is equal to the set of turning points of $f^{n+1}|J$. Notice that if $x < y$ are two consecutive points of F_n then, for every $z \in (x, y)$, the Taylor series of $\theta(x^+;t), \theta(y^-;t)$ and of $\theta(z;t)$ coincide up to the order n. Hence $\theta(b^-;t) - \theta(a^+;t)$ has the same partial sum up to order n as the holomorphic map $\sum_{x \in F_n} (\theta(x^+;t) - \theta(x^-;t))$. By Lemma 8.3, the n-th partial sum of $\theta(b^-;t) - \theta(a^+;t)$ is therefore equal to the n-th partial sum of

$$\sum_{i=0}^{l} \sum_{k=0}^{n} \sum_{x \in E_{i,k}(J)} t^k N_i(t) = \sum_{i=0}^{l} \sum_{k=0}^{n} \Gamma_{i,k}(J)\, t^k N_i(t).$$

Therefore the required equality holds up to terms of order n. Since this is true for all $n \in \mathbb{N}$ the theorem follows. □

Corollary 8.1. *Let $J = (a, b) \subset I$ as before. Then the function $\Gamma_i(J)(t)$ has a meromorphic extension to $\mathbb{D}$, for each $i = 1, \dots, l$. The poles of these meromorphic functions can only be at the zeros of the kneading determinant $D_f(t)$.*

Proof. By Lemma 8.2, the determinant D_{l+1} of the $l \times l$ matrix $[N^1, \dots, N^l]$ (obtained from the kneading matrix by deleting the $l+1$-th column of the kneading matrix) is a holomorphic function on $\mathbb{D}$ which is not identically zero. Hence there exists an $l \times l$ matrix $[M_{i,j}]$ with meromorphic entries which is the inverse of $[N^1, \dots, N^l]$. From Theorem 8.2 we get for each $j = 1, \dots, l$,

$$\theta^j(b^-;t) - \theta^j(a^+;t) = \sum_{i=1}^{l} \Gamma_i(J)(t) \times N_{i,j}(t).$$

Thus

$$\begin{aligned}\sum_{j=1}^{l} (\theta^j(b^-;t) - \theta^j(a^+;t))\, M_{j,k}(t) &= \sum_{j=1}^{l} \sum_{i=1}^{l} \Gamma_i(J) N_{i,j}(t) M_{j,k}(t) \\ &= \sum_{i=1}^{l} \Gamma_i(J)(t) \sum_{j=1}^{l} N_{i,j}(t) M_{j,k}(t) = \sum_{i=1}^{l} \Gamma_i(J)(t) \delta_{i,k} = \Gamma_k(J)(t).\end{aligned}$$

Hence $\Gamma_k(J)(t)$ is a meromorphic function. Furthermore, multiplying both members by the kneading determinant

$$D_f(t) = \frac{D_{l+1}(t)(-1)^{l+2}}{1 - \epsilon(I_{l+1})t},$$

we get that $D_f(t)\, \Gamma_k(J)(t)$ is equal to

$$\sum_{j=1}^{l-1} \frac{\left(\theta^j(b^-;t) - \theta^j(a^+;t)\right) \cdot M_{j,k}(t) \cdot D_{l+1}(t) \cdot (-1)^{l+2}}{1 - \epsilon(I_{l+1})t}.$$

Since the $l \times l$ matrix $[M_{i,j}]$ is the inverse of $[N^1, \dots, N^l]$, and since the determinant of the matrix $[N^1, \dots, N^l]$ is equal to D_{l+1}, we get by Cramer's rule that $M_{j,k}(t) D_{l+1}(t)$ is the determinant of a $(l-1) \times (l-1)$ submatrix of $[N^1, \dots, N^l]$ for each $j, k = 1, \dots, l$. Since this last determinant is holomorphic on $\mathbb{D}$, all this implies that $D_f(t)\Gamma_k(J)(t)$ is also holomorphic on $\mathbb{D}$. □All the work done so

far in this section was aimed at proving that $t \mapsto L_f(t)$, which is holomorphic on $|t| < 1/s$, has a meromorphic extension to $\mathbb{D}$ with a pole in $1/s$. This will be shown in the next corollary.

Corollary 8.2. *The function $L_f(t) = L(I;t) = \sum_{n=0}^{\infty} l(f^n)t^n$ is meromorphic on $\mathbb{D}$ and its poles are contained in the set of zeros of the kneading determinant $D_f(t)$. Furthermore, $L_f(t) = L(I;t)$ has a pole at the point $t = \frac{1}{s}$.*

Proof. Let J be the interior of I. The number of turning points of f^n in J is $\sum_{p=0}^{n-1} \sum_{i=1}^{l} \Gamma_{i,p}(J)$. Therefore the number of laps of f^n is

$$l(f^n) = \sum_{p=0}^{n-1} \sum_{i=1}^{l} \Gamma_{i,p}(J) + 1.$$

Thus

$$L_f(t) = \sum_{n=0}^{\infty} l(f^n)t^n = \sum_{i=1}^{l} \sum_{n=0}^{\infty} \sum_{p=0}^{n-1} \Gamma_{i,p}(J)t^n + \frac{1}{1-t}.$$

From the formula for the product of two power series we get

$$\frac{t}{1-t} \sum_{n=0}^{\infty} a_n t^n = (\sum_{n=0}^{\infty} t^{n+1})(\sum_{n=0}^{\infty} a_n t^n) = \sum_{n=0}^{\infty} (\sum_{p=0}^{n-1} a_p) t^n.$$

Using this expression in the formula for $L_f(t)$ we get

$$L_f(t) = \frac{1}{1-t} + \sum_{i=1}^{l} \frac{t}{1-t} \Gamma_i(J)(t).$$

Hence, from the previous corollary, $L_f(t)$ is meromorphic on $\mathbb{D}$ and the poles of $L_f(t)$ are contained in the set of poles of $\Gamma_i(t)$ which is a subset of the set of zeros of the kneading determinant. It remains to prove that $L_f(t)$ has a pole at the point $t = \frac{1}{s}$. Since the coefficients of the Taylor series of $L_f(t)$ are positive numbers,

$$|L_f(t)| \le \sum_{n=0}^{\infty} l(f^n)|t|^n.$$

Hence $\lim_{t \to \frac{1}{s}} \sum_{n=0}^{\infty} l(f^n)|t|^n = \infty$ because, otherwise, $L_f(t)$ would be bounded in the disc of radius $\frac{1}{s}$ and therefore it could be extended holomorphically to a bigger disc. This is a contradiction because the radius of convergence of $L_f(t)$ is $\frac{1}{s}$. □

Next we are going to define a probability measure in I which will give the semiconjugacy between f and a piecewise linear map. If $J \subset I$ is a closed interval then, as we have seen in the previous corollary, the map

$$L_f(J;t) = \sum_{n=0}^{\infty} l(f^n|J)t^n$$

converges for $|t| < \frac{1}{s}$ and, in fact, extends to a meromorphic function on the unit disc with poles contained in the set of zeros of the kneading determinant. Hence $\frac{L(J;t)}{L(I;t)}$ is a meromorphic function. However, the point $t = \frac{1}{s}$ is a removable singularity of the meromorphic function $\frac{L(J;t)}{L(I;t)}$ because $0 \le L(J;t) \le L(I;t)$ for $0 < t < \frac{1}{s}$. In particular, the limit

$$\Lambda(J) = \lim_{t \to 1/s} \frac{L(J;t)}{L(I;t)}$$

exists and satisfies the inequality $0 \le \Lambda(J) \le 1$.

Lemma 8.5. *Assume that $s > 1$.*

1. *If the intervals J_1 and J_2 have only a boundary point in common then*

$$\Lambda(J_1 \cup J_2) = \Lambda(J_1) + \Lambda(J_2);$$

2. *the number $\Lambda(J)$ depends continuously on the endpoints of J;*

3. *if f is monotone on J then*

$$\Lambda(f(J)) = s\Lambda(J).$$

Proof. Since $l(f^n|J_1) + l(f^n|J_2)$ differs from $l(f^n|J_1 \cup J_2)$ by at most one, we have that the difference of the meromorphic functions,

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

is bounded by $\Sigma|t^n| = \frac{1}{1-|t|} < \infty$ for $|t| \le \frac{1}{s} < 1$. Dividing by $L(I;t)$ and passing to the limit as $t \to \frac{1}{s}$ we get Statement 1) since $\lim_{t \to \frac{1}{s}} |L(I;t)| = \infty$. To prove Statement 3), we notice that, since $f|J$ is a homeomorphism, then, $l(f^{n+1}|J) = l(f^n|f(J))$. This clearly implies that $L(J;t) = 1 + tL(f(J);t)$. Therefore

$$\Lambda(J) = \lim_{t \to \frac{1}{s}} \frac{1 + tL(f(J);t)}{L(I;t)} = \frac{1}{s}\Lambda(f(J)).$$

This proves Statement 3). In order to prove Statement 2), let $J = [a,b]$. For any $n \in \mathbb{N}$ we can choose $x > b$ so that the interval $[b,x]$ is contained in a single lap of f^n. Then, using Statements 1) and 3) we get

$$\Lambda([a,x]) = \Lambda(J) + \Lambda([b,x]) \text{ and}$$

$$\Lambda([b,x]) = s^{-n}\Lambda(f^n([b,x])) \leq s^{-n}.$$

Since $s > 1$, the proof is finished. □

Now we can prove the main result of this section.

Proof of Theorem 8.1 As before let $I = [c_0, c_{l+1}]$. Let us define $\lambda\colon I \to [0,1]$ by $\lambda(x) = \Lambda([c_0, x])$. From Lemma 8.5, it follows that λ is a continuous, surjective and monotone map. We claim that if $x, y \in I$ are such that $\lambda(x) = \lambda(y)$ then $\lambda(f(x)) = \lambda(f(y))$. So let us consider a partition $J_1, \ldots, J_n$ of the interval $[x,y]$ such that $f|J_i$ is monotone. From Lemma 8.5 we have that $0 = \Lambda([x,y]) = \sum_{i=1}^n \Lambda(J_i) = \frac{1}{s}\sum_{i=0}^n \Lambda(f(J_i))$. Since $[f(x), f(y)] \subset \cup_{i=1}^n f(J_i)$, this gives $\Lambda([f(x), f(y)]) \leq \sum_{i=1}^n \Lambda(f(J_i)) = 0$. Therefore $|\lambda(f(x)) - \lambda(f(y))| = \Lambda([f(x), f(y)]) = 0$. This proves the claim.

From the previous claim it follows that the mapping $T\colon [0,1] \to [0,1]$ defined by $T(x) = \lambda(f(\lambda^{-1}(x)))$ is well defined, continuous and satisfies $\lambda \circ f = T \circ \lambda$. It remains to show that T is piecewise linear. Take $x \in I_k = [c_{k-1}, c_k]$. We assume first that f is monotone increasing on I_k. Then,

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

But

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

Therefore

$$T(\lambda(x)) = a_k + s\lambda(x)$$

where a_k is the constant $\lambda(f(c_k)) - s\lambda(c_{k-1})$. Hence T has slope s in $\lambda(I_k)$. Similarly, if f is decreasing in I_k we get that T has slope $-s$ in $\lambda(I_k)$. □

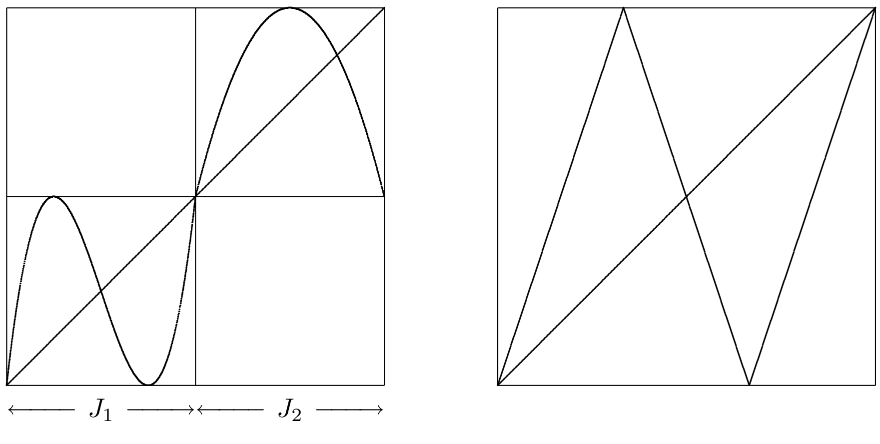

Fig. 8.1: In the map drawn on the left the growth rates of $l(f^n|J_1)$ and $l(f^n|J_2)$ are different where J_1 and J_2. This map is semi-conjugate to the piecewise linear map drawn on the right; the semi-conjugacy collapses the whole interval J_2.

Remark. 1. From the Corollary below Theorem 5.2, a continuous piecewise map with slopes equal to $\pm s$ has topological entropy equal to $\log s$. Therefore T has the same topological entropy as f. 2. The semi-conjugacy λ of the theorem may collapse one or several laps of T into a point. In fact, suppose I can be partitioned into f-invariant intervals $J_1, \ldots, J_k$, see Figure 8.1 on the left. Then, $l(f^n) = \sum_{i=1}^{n} l(f^n|J_i)$. From this it follows that if we define $s_i = \lim_{n\to\infty} \sqrt[n]{l(f^n|J_i)}$, then $s = \max\{s_1, \ldots, s_k\}$. Furthermore, from the construction of λ it is clear that any interval J_i of the partition with growth number $s_i < \max\{s_1, \ldots, s_k\}$ must be collapsed to a point. 3. As we will see in Section III.4 one can use Theorem 8.2 to give a canonical decomposition of the non-wandering set, see Jonker and Rand (1981). It will follow that there are uncountably many non-combinatorially equivalent maps which have the same topological entropy. Hence, even for unimodal maps, topological entropy is only a very rough invariant.

Exercise 8.7. Let $f\colon [-1,1] \to [-1,1]$ be as in Figure 8.1 on the left. Show that the growth rate of $l(f^n)$ is equal to 3. It follows that the semi-conjugacy from the previous theorem has the property that $\lambda[0,1] = \{1\}$ and that $\lambda[-1,0] = [0,1]$. In particular, the piecewise linear map T is as in Figure 8.1 on the right.

Exercise 8.8. Show that any unimodal $f\colon [0,1] \to [0,1]$ with $s(f) < \sqrt{2}$ has no periodic points of odd period $p \geq 3$. (Hint: use the Exercise 5.3.)

Exercise 8.9. Even if λ is a homeomorphism it is in general not a diffeomorphism. (Hint: if λ were differentiable then eigenvalues at a periodic point of period n would have to be of the form s^n. In fact, this type of argument can be used to show that in general h is not even absolutely continuous, see Exercise V.3.1.)

Exercise 8.10. Show that the continuous piecewise monotone maps from Theorem 8.1 have sensitive dependence on initial conditions. (Hint: use a similar idea as in the proof of Proposition 5.2.)

9 Continuity of the Topological Entropy

In this section we will prove, using the tools developed in the last section, that the topological entropy $h_t(f)$ depends continuously on $f\colon I \to I$ as long as we consider only maps f in the space of C^1 piecewise monotone maps with a fixed lap number. This result was first proved in Milnor and Thurston (1977). As we have seen in Theorem 8.2 the topological entropy of f is equal to the logarithm of the number $s(f) = \lim_{n\to\infty} \sqrt[n]{l(f^n)}$. First we will show that $\frac{1}{s(f)}$ is the closest zero to the origin of kneading determinant D_f. Next we will analyze how the kneading determinant varies with the mapping. Although this determinant certainly does not depend continuously on the mapping, we will show that its zeros do vary continuously. The reason for this is that the change of the kneading matrix due to a perturbation of the map is of a very special nature. Indeed, as we will see, the kneading matrix of the perturbed map can be obtained from the original kneading matrix by a small perturbation of the matrix followed by some elementary operations on the matrix (which do not change the kneading determinant) and subsequently by multiplying some rows by holomorphic functions which are non-zero in the unit disc.

We consider the space $C^r(I,I), r \geq 0$ of C^r maps of a compact interval I endowed with the C^r topology: this is defined by the norm $||f-g||_r = \sup\{|f(x)-g(x)|,\dots,|D^r f(x) - D^r g(x)|\,;\ x \in I\}$. We fix l points $c_1 < \cdots < c_{l-1} < c_l$ in the interior of I and denote by $P^r = P^r(I;c_1,\dots,c_l)$ the subspace of piecewise monotone maps whose turning points are exactly the points $c_1,\dots,c_l$ and such that the boundary of I is mapped into itself.

The main result of this section is the following

Theorem 9.1. **(Milnor and Thurston)** *The function $P^1 \to \mathbb{R}$ which associates to each mapping $g \in P^1(I;c_1,\dots,c_l)$ its topological entropy $h_t(g)$ is continuous.*

Of course one can also apply this theorem also to the space $P^{2,l}$ of C^2 maps (with the C^2 topology) which are l-modal and whose critical points are non-degenerate (i.e., the second derivative is non-zero at critical points). Indeed, if $f \in P^{2,l}$ then every map g which is C^2 close to f is also l-modal and by the Implicit Function Theorem there exists a C^2 coordinate change $h\colon I \to I$ which is C^2 close to the identity so that the turning points of f and $h \circ g \circ h^{-1}$ coincide. It follows that $P^{2,l} \ni f \mapsto h_t(f)$ is continuous.

To prove this theorem we need some lemmas. For simplicity write $I = [c_0, c_{l+1}]$.

Lemma 9.1. *Let $f \in P^1(I;c_1,\dots,c_l)$ and $D_f(t)$ be its kneading determinant, $s = \lim_{n\to\infty} \sqrt[n]{l(f^n)}$ and $t = \frac{1}{s}$. Then $t = \frac{1}{s}$ is a zero of $D_f(t)$ and $D_f(t) \neq 0$ if $|t| < \frac{1}{s}$.*

Proof. As before let $N_1,\dots,N_l$ be the rows and $N^1,\dots,N^{l+1}$ the columns of the $l \times (l+1)$ kneading matrix of f. From Theorem 8.2 we get

$$\theta(c_k^-;t) - \theta(c_0;t) = \sum_{i=1}^{l} \Gamma_i([c_0,c_k])(t) N_i(t)$$

for every $1 \leq k \leq l$. Since $\theta(c_k^-;t) = \theta(c_k;t) - \frac{1}{2}N_k(t)$, we have

$$\theta(c_k;t) - \theta(c_0;t) = \sum_{i=1}^{l} \Gamma_i([c_0,c_k])(t) N_i(t) + \frac{1}{2}N_k(t).$$

If we set $\Gamma_{i,k} = \Gamma_i([c_0,c_k])(t) + \frac{1}{2}\delta_{i,k}$, where $[\delta_{i,k}]$ denotes the identity matrix, then

$$\theta(c_k;t) - \theta(c_0;t) = \sum_{i=1}^{l} \Gamma_{i,k}(t) N_i(t)$$

or, written out in components,

$$\theta^j(c_k;t) - \theta^j(c_0;t) = \sum_{i=1}^{l} \Gamma_{i,k}(t) N_{i,j}(t) \qquad (*)$$

for every $1 \leq j \leq l+1$. Let $A(t)$ be the $l \times l$ matrix defined by $A(t) = [A_{j,k}]_{1 \leq j,k \leq l}$ where $A_{j,k} = \theta^j(c_k;t) - \theta^j(c_0;t)$. Since we only use $(*)$ for $j = 1, \ldots, l$, the matrix A is the product of the transpose of the matrix $[\Gamma_{i,j}]$ with the matrix $[N^1, \ldots, N^l]$. The last matrix is constructed from the kneading matrix by deleting the last column, and, therefore, its determinant is equal to $D_{l+1}(t)$. Since the entries of all the above matrices are holomorphic functions on the disc of radius $\frac{1}{s}$, it follows that, for $t < \frac{1}{s}$, $D_{l+1}(t) = 0$ implies that the determinant of $A(t)$ is also zero. Since the kneading determinant $D_f(t)$ has the same zeros as $D_l(t)$ in the unit disc, the determinant of $A(t)$ vanishes if $|t| < s$ and $D_f(t) = 0$. So, to prove that $D_f(t)$ has no zeros in the disc of radius $\frac{1}{s}$ it is enough to prove that the matrix $A(t)$ is non-singular for $|t| < \frac{1}{s}$.

Let us compute the matrix $A(t)$ and its determinant. Since $f(\partial I) \subset \partial I$ we have that $\theta^i(c_0;t) = 0$ if $i \neq 1, l+1$. We have also that

$$\theta^j(c_k;t) = \begin{cases} 0 & \text{if } j \notin \{k, k+1\} \\ \frac{1}{2} & \text{if } j = k, k+1. \end{cases}$$

Hence the $l \times l$ matrix $A(t)$ is equal to

$$\begin{pmatrix} \frac{1}{2} - \theta^1 & -\theta^1 & -\theta^1 & \ldots & -\theta^1 & -\theta^1 \\ \frac{1}{2} & \frac{1}{2} & 0 & \ldots & 0 & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} & \ldots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \ldots & \frac{1}{2} & 0 \\ 0 & 0 & 0 & \ldots & \frac{1}{2} & \frac{1}{2} \end{pmatrix}$$

where $\theta^1 = \theta^1(c_0;t)$. By using elementary row operations on the above matrix we can eliminate an even number of terms $-\theta^1$ from the first row and we get that the determinant as $A(t)$ is equal to the determinant of

$$\begin{pmatrix} \frac{1}{2} - \theta^1(c_0;t) & 0 & 0 & \ldots & 0 & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 & \ldots & 0 & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} & \ldots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \ldots & \frac{1}{2} & 0 \\ 0 & 0 & 0 & \ldots & \frac{1}{2} & \frac{1}{2} \end{pmatrix}$$

if l is odd and

$$\begin{pmatrix} \frac{1}{2} & 0 & 0 & \ldots & 0 & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 & \ldots & 0 & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} & \ldots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \ldots & \frac{1}{2} & 0 \\ 0 & 0 & 0 & \ldots & \frac{1}{2} & \frac{1}{2} \end{pmatrix}$$

if l is even. Therefore, $\det(A(t)) = (\frac{1}{2})^l$ if l is even and $\det(A(t)) = (\frac{1}{2})^{l-1}(\frac{1}{2} - \theta^1(c_0;t))$ if l is odd. Since

$$\theta^1(c_0;t) = \begin{cases} 1 & \text{if } f(c_0) = c_{l+1} \\ 1 + t + t^2 + \cdots = \frac{1}{1-t} & \text{if } f(c_0) = c_0. \end{cases}$$

it follows that the determinant of $A(t)$ is non-zero when t is in the unit disc. Therefore the kneading determinant is non-zero in the disc of radius $\frac{1}{s}$. Since the point $\frac{1}{s}$ is a pole of the function $L_f(t)$ (Corollary 2 of Theorem 8.2) it must be a zero of the kneading determinant. □

Lemma 9.2. *If f belongs to the space $P^0 = P^0(I; c_1, \ldots, c_l)$ of continuous piecewise monotone maps and the forward orbit of each turning point does not contain turning points then for each $i, j = 1, \ldots, l$ the mappings $P^0 \to \mathcal{H}(\mathbb{D})$, $g \mapsto N_{i,j}(t;g)$ is continuous at f. Here $\mathcal{H}(\mathbb{D})$ is the space of holomorphic maps on $\mathbb{D}$ with the supremum metric. In particular, the kneading determinant $D_g(t)$ is continuous at f.*

Proof. Let $\epsilon > 0$ and $K \subset \mathbb{D}$ be a compact set. Choose $r > 0$ and $N \in \mathbb{N}$ such that $|t| < r$ for all $t \in K$ and $4\frac{r^{N+1}}{1-r} < \epsilon$. Since there is no turning point in the forward orbit of c_k we can choose a neighbourhood $\mathcal{N}$ of f in P^0 such that if $g \in \mathcal{N}$ then the itineraries $\underline{i}_f(c_k)$ and $\underline{i}_g(c_k)$ coincide up to the order N for every $k = 1, \ldots, l$. Hence the Taylor series of $N_{k,j}(t;f)$ and $N_{k,j}(t;g)$ coincide up to the order N. Since all the coefficients of these powers series are in $\{0, \pm1, \pm2\}$, we have:

$$|N_{k,j}(t;f) - N_{k,j}(t;g)| \le \sum_{n=N+1}^{\infty} 4|t|^n \le 4\frac{r^{N+1}}{1-r} < \epsilon. \quad \square$$

If there are turning points in the forward orbit of some turning point then the kneading matrix is no longer continuous. However we have the following:

Lemma 9.3. *Let f belongs to the space P^0 of piecewise monotone continuous maps. Assume that no turning point of f is a periodic point. Then the mapping $P^0 \to \mathcal{H}(\mathbb{D})$, $g \mapsto D_g$ is continuous at f.*

Proof. To simplify the exposition let us suppose that $f^p(c_i) = c_j$ and that for $k \neq i$ there is no turning point in the forward orbit of c_k. The general situation will follow from the same ideas.

We claim that if $\epsilon > 0$ and $K \subset \mathbb{D}$ is a compact set then there exists a neighbourhood $\mathcal{N}$ of f such that for every $g \in \mathcal{N}$ the following conditions are satisfied: i) if $k \neq i$ then

$$|N_{k,m}(t;f) - N_{k,m}(t;g)| < \epsilon \text{ for all } t \in K \text{ and } m = 1, \ldots, l;$$

ii) for $k = i$ one of the following holds for all $t \in K$ and $m = 1, \dots, l$:

$$|N_{i,m}(t; f) - N_{i,m}(t; g)| < \epsilon \text{ or}$$

$$|N_{i,m}(t; f) - 2t^p N_{j,m}(t; f) - N_{i,m}(t; g)| < \epsilon.$$

The proof of (i) is exactly the same as in the previous lemma: take some large number $N \in \mathbb{N}$. Because $k \neq i$, the turning point c_k is never mapped onto a turning point. Hence, the itineraries of $\underline{i}_f(c_k^{\pm})$ and $\underline{i}_g(c_k^{\pm})$ coincide up to order N provided g is sufficiently close to f. Hence the Taylor series of $\theta_f(c_k^{\pm}; t)$ and $\theta_g(c_k^{\pm}; t)$ coincide up to order N for $k \neq i$ and (i) follows as in Lemma 9.2. Since $f^p(c_i) = c_j$ (and p is minimal with this property), one has

$$\underline{i}_f(c_i^-) = \big(i_0(c_i^-), i_1(c_i^-), \dots, i_{p-1}(c_i^-)\big) \cdot \sigma^p(\underline{i}(c_j^-))$$

or

$$\underline{i}_f(c_i^-) = \big(i_0(c_i^-), i_1(c_i^-), \dots, i_{p-1}(c_i^-)\big) \cdot \sigma^p(\underline{i}(c_j^+))$$

depending on $f^p(c_i^-) = c_j^-$ or $f^p(c_i^-) = c_j^+$. (Here $\underline{a} \cdot \underline{b}$ is the concatenation of symbols as defined below Lemma I.1.1.) Because $f^p(c_i) = c_j$ and no forward iterate of c_j meets a turning point, provided g is sufficiently close to f, the first N itineraries of c_i for f and g coincide with possibly the exception of the p-th itinerary; hence, for the first N itineraries one of the following two equalities hold:

$$\underline{i}_g(c_i^-) = \big(i_0(c_i^-), i_1(c_i^-), \dots, i_{p-1}(c_i^-)\big) \cdot \sigma^p(\underline{i}(c_j^-))$$

or

$$\underline{i}_g(c_i^-) = \big(i_0(c_i^-), i_1(c_i^-), \dots, i_{p-1}(c_i^-)\big) \cdot \sigma^p(\underline{i}(c_j^+)).$$

From this (ii) easily follows.

From the claim, it follows that for every $t \in K$, the kneading matrix of g is either near to the kneading matrix of f or near to a matrix which is obtained from the kneading matrix of f by an elementary row operation. Since elementary row operations do not change the determinant, the kneading determinant of g is uniformly near to the kneading determinant of f on the compact set K.

In the general situation, the kneading matrix has some other discontinuities, but all of them are of the same type: they correspond to elementary row operations and we get the same conclusion. □

Next we analyze the discontinuities of the kneading matrix due to the presence of periodic turning points. Here we do need the C^1 topology and the following two lemmas.

Lemma 9.4. *Suppose that $f^p(c_i) = c_i$ where f is of class C^1. If $g \in P^1$ is sufficiently close to f in the C^1 topology, then*

$$g^{np}(c_i^+) - c_i, \quad n = 1, 2, 3, \dots$$

all have the same sign (here we define the sign of this number as the sign of $g^{np}(x) - c_i$ for $x > c_i$ sufficiently close to c_i). Furthermore, given $\epsilon > 0$ there exists $\delta > 0$ such that $|g^m(c_i) - f^m(c_i)| < \epsilon$ for all $m \in \mathbb{N}$ if $||g - f||_1 < \delta$.

Proof. Since $f^p(c_i) = c_i$ and $Df^p(c_i) = 0$ one has that $|Df^p(x)| \le \frac{1}{4}$ for all x in a small neighbourhood J of c_i. If g is C^1 close to f then g^p is also C^1 close to f^p and consequently, $|Dg^p(x)| \le \frac{1}{2}$ for all $x \in J$ and there exists a unique attracting fixed point x of g^p in J. It follows that the distance between $g^{np}(c_i)$ and x goes monotonically to zero as $n \to \infty$. In particular $g^{np}(c_i) - c_i$ has the same sign for all $n \in \mathbb{N}$. □

Exercise 9.1. For $a \le 2$ the kneading matrix of $f(x) = 2x(1-x)$ is equal to

$$\begin{pmatrix} -1 - \frac{2t}{1-t} & 1 \end{pmatrix}$$

and for $a > 2$ but $|a-2|$ small, the kneading matrix of $f(x) = ax(1-x)$ is equal to

$$\begin{pmatrix} -1 & 1 - \frac{2t}{1+t} \end{pmatrix}.$$

Notice that the last matrix is equal to the first matrix times $(1-t)/(1+t)$. Similarly, the kneading matrix of $f(x) = a_0x(1-x)$ is equal to

$$\begin{pmatrix} -1 & 1 - \frac{2t}{1+t} \end{pmatrix} = \begin{pmatrix} -1 & 1 - \frac{2t-t^2}{1-t^2} \end{pmatrix}.$$

when $a_0 = 1 + \sqrt{3}$. For a close to a_0 the kneading matrix of $f(x) = ax(1-x)$ is either the same or equal to

$$\begin{pmatrix} -1 + \frac{2t^2}{1+t^2} & 1 - \frac{2t}{1+t^2} \end{pmatrix}.$$

(Hint: the turning point has period 2 for this last map.)

Lemma 9.5. *Let c_i be a periodic point of period p of $f \in P^1$ and suppose that the orbit of c_i contains no other turning points. Then:*

1. *the i-th kneading vector $N_i(t; f)$ is of the form $\frac{1}{1-t^p}P(t)$ where $P\colon \mathbb{C} \to \mathbb{C}^l$ is a polynomial map of degree p;*
2. *if $g \in P^1$ is close enough to f then the i-th kneading vector $N_i(t; g)$ of g is equal to*
$$N_i(t; f) \text{ or to } \frac{1-t^p}{1+t^p} N_i(t; f).$$
The first case occurs when $g^p(c_i^+) - c_i$ and $f^p(c_i^+) - c_i$ have the same sign, and the second case otherwise.

Proof. As we have seen before,

$$N_{i,j}(t) = \theta^j(c_i^+; t) - \theta^j(c_i^-; t) = 2\theta^j(c_i^+; t)$$

for $j \ge 1$. Moreover, since $f^p(c_i) = c_i$ we get from the previous lemma that the coefficients $\theta^j(c_i^\pm)$, $j = 1, 2, \dots$ have period p for x close to c_i. More precisely, we have that

$$N_i(t; f) = \begin{pmatrix} 0 & \cdots & 0 & -1 - \dfrac{2t^p}{1-t^p} & 1 & 0 & \cdots & 0 \end{pmatrix} + \frac{t}{1-t^p}\tilde{Q}(t)$$

when f^p has a local maximum at c_i and

$$N_i(t;f) = \begin{pmatrix} 0 & \cdots & 0 & -1 & 1+\dfrac{2t^p}{1-t^p} & 0 & \cdots & 0 \end{pmatrix} + \frac{t}{1-t^p}\tilde{Q}(t)$$

when f^p has a local minimum at c_i, where $\tilde{Q}(t)$ is a C^{l+1} valued polynomial of degree $p-2$. If f^p has a local maximum at c_i then for each g sufficiently C^1 close to f,

$$N_i(t;g) = \begin{cases} N_i(t;f) & \text{if } g^p(c_i) \le c_i \\ \begin{pmatrix} 0 & \cdots & 0 & -1 & 1-\frac{2t^p}{1+t^p} & 0 & \cdots & 0 \end{pmatrix} + \frac{t}{1+t^p}\tilde{Q}(t) & \text{if } g^p(c_i) > c_i. \end{cases}$$

In the latter case the periodic orbit is orientation reversing. A similar statement holds when f^p has a local minimum at c_i: in that case

$$N_i(t;g) = \begin{cases} N_i(t;f) & \text{if } g^p(c_i) \ge c_i \\ \begin{pmatrix} 0 & \cdots & 0 & -1+\frac{2t^p}{1+t^p} & 1 & 0 & \cdots & 0 \end{pmatrix} + \frac{t}{1+t^p}\tilde{Q}(t) & \text{if } g^p(c_i) < c_i. \end{cases}$$

Statements 1) and 2) follow immediately from this. □

Exercise 9.2. Let $f\colon [0,1] \to [0,1]$ be a C^1 bimodal map with turning points in $1/3$ and $2/3$ and such that $f(0) = 0$, $f(1) = 1$, $f(1/3) = 2/3$, $f(2/3) = 1/3$. Then the kneading matrix of f is equal to

$$\begin{pmatrix} N_1(t;f) \\ N_2(t;f) \end{pmatrix} = \begin{pmatrix} -1 & 1-\frac{2t-2t^2}{1-t^2} & 0 \\ 0 & -1+\frac{2t-2t^2}{1-t^2} & 1 \end{pmatrix}.$$

If g is also a bimodal map with turning points in $1/3$ and $2/3$, sufficiently C^1 close to f such that $g^2(1/3) = 1/3$ then the kneading matrix of g is equal to

$$\begin{pmatrix} N_1(t;g) \\ N_2(t;g) \end{pmatrix} = \begin{pmatrix} -1-\frac{2t^2}{1-t^2} & 1 & -\frac{2t}{1-t^2} \\ \frac{2t}{1-t^2} & -1 & 1+\frac{2t^2}{1-t^2} \end{pmatrix} = \begin{pmatrix} \dfrac{1+t^2}{1-t^2} & \dfrac{-2t}{1-t^2} \\ \dfrac{2t}{1-t^2} & \dfrac{1+t^2}{1-t^2} \end{pmatrix} \begin{pmatrix} N_1(t;f) \\ N_2(t;f) \end{pmatrix}.$$

if $g(2/3) < 1/3$. The reason for this is that when g is C^1 close to f then $g^2(1/3) < 1/3$. Similarly, this matrix is equal to

$$\begin{pmatrix} N_1(t;g) \\ N_2(t;g) \end{pmatrix} = \begin{pmatrix} -1 & 1-\frac{2t^2-2t^3}{1-t^2} & -2t \\ 0 & -1+\frac{2t-2t^2}{1-t^2} & 1 \end{pmatrix} = \begin{pmatrix} 1 & -2t \\ 0 & 1 \end{pmatrix} \begin{pmatrix} N_1(t;f) \\ N_2(t;f) \end{pmatrix}$$

if $g(2/3) > 1/3$.

Lemma 9.6. *Let $c_{i(1)}$ be a periodic point of period p of $f \in P^1$ and suppose that the orbit of c_i contains the turning points $c_{i(1)}, c_{i(2)}, \ldots, c_{i(k)}, c_{i(k+1)} = c_{i(1)}$ (in this order). Then for $g \in P^1$ sufficiently close enough to f one has*

$$\begin{pmatrix} N_{i(0)}(t;g) \\ \vdots \\ N_{i(k)}(t,g) \end{pmatrix} = B(t) \begin{pmatrix} N_{i(0)}(t;f) \\ \vdots \\ N_{i(k)}(t,f) \end{pmatrix},$$

and

$$\begin{pmatrix} N_{i(0)}(t;f) \\ \vdots \\ N_{i(k)}(t,f) \end{pmatrix} = C(t) \begin{pmatrix} N_{i(0)}(t;g) \\ \vdots \\ N_{i(k)}(t,g) \end{pmatrix},$$

where $B(t)$ and $C(t)$ are $k \times k$ matrices with rational coefficients.

Proof. We shall leave the elementary but tedious proof to the reader as it uses the ideas from Lemmas 9.3 and 9.5 and Exercise 9.2. It is based on analyzing what happens to the kneading vectors if one of the turning points 'falls off' the orbit of $c_{i(1)}$. From this it easily follows that

$$\begin{pmatrix} N_{i(0)}(t;g) \\ \vdots \\ N_{i(k)}(t,g) \end{pmatrix} \text{ and } \begin{pmatrix} N_{i(0)}(t;f) \\ \vdots \\ N_{i(k)}(t,f) \end{pmatrix}$$

are related by a matrix $B(t)$. The (i,i)-th coefficient of $B(t)$ is equal to 1 or $\dfrac{1+t^p}{1-t^p}$. When $i \neq j$ then the (i,j)-th coefficient of $B(t)$ is equal to 0, $\dfrac{\pm 2t^{b(i,j)}}{1-t^p}$ or to $\pm 2t^{b(i,j)}$, where $b(i,j) > 0$ is the smallest integer so that $f^{b(i,j)}(c_i) = c_j$. The reason for this is as follows. For example the effect of a perturbation of g on the j-th coefficient of the i-th kneading vector is due to a change of sign of $g^{a(j)}(c_{i(j)}) - c_{i(j+1)}$. This can be described by a row operation as in Lemma 9.3 but now perhaps with a periodic effect. This explains the (i,j)-th coefficient of $B(t)$. A possible 'change in orientation' as in Lemma 9.5 explains the (i,i)-th coefficient of $B(t)$. □

Proof of Theorem 9.1 Let $f \in P^1$ and $K \in \mathbb{D}$ be a compact disc that contains the point $\frac{1}{s}$. Since $\frac{1}{s}$ is a zero of the kneading determinant $D_f(t)$ it follows from the Cauchy integral formula that for each $\epsilon > 0$ there exists δ such that if $\phi\colon \mathbb{D} \to \mathbb{C}$ is an analytic function with $|\phi(t) - D_f(t)| < \delta$ for all $t \in K$ then ϕ has a zero in an ϵ-ball around $\frac{1}{s}$. Clearly, since $D_f(t) \neq 0$ for $|t| < \frac{1}{s}$, it also follows that $\phi(t) \neq 0$ for $|t| \leq \frac{1}{s} - \epsilon$. If f has no periodic turning point then, by lemmas 9.2 and 9.3, there exists a neighbourhood $\mathcal{N}$ of f such that $|D_g(t) - D_f(t)| < \delta$ for any $g \in \mathcal{N}$. Therefore the growth number of g, $s(g)$ satisfies $|\frac{1}{s(g)} - \frac{1}{s}| < \epsilon$.

If some turning point of f is periodic, we can use Lemmas 9.2-9.6 to get a neighbourhood $\mathcal{N}$ of f so that if $g \in \mathcal{N}$ then the matrix $[N_{i,j}(t;g)]_{1\leq i,j\leq l}$ is close to the matrix $[N_{i,j}(t;f)]_{1\leq i,j\leq l}$ multiplied (on the left) with a matrix $B(t)$. Also $[N_{i,j}(t;f)]_{1\leq i,j\leq l}$ is close to the matrix $[N_{i,j}(t;g)]_{1\leq i,j\leq l}$ multiplied (on the left) with a matrix $C(t)$. Hence

$$|\det(C(t))D_g(t) - D_f(t)| = |D_g(t) - \det(B(t))D_f(t)| < \delta$$

for all $t \in K$. It follows from the previous argument that the zeros of $D_g(t)$ and $D_f(t)$ are ϵ-close to each other on $\mathbb{D}$. Hence the growth number $s(g)$ of g also satisfies $|\frac{1}{s} - \frac{1}{s(g)}| < \epsilon$. This proves the continuity of the topological entropy since $h_t(g) = \log s(g)$. □

Exercise 9.3. Let f_a be a family of C^1 unimodal maps depending continuously on a in the C^1 topology. Assume that $\nu(f_a)$ increases at $a = a'$. Show that this implies that the turning point of $f_{a'}$ is periodic and that the attracting hyperbolic periodic orbit near the turning point of f_a for a near a' changes from orientation preserving for $a < a'$ to orientation reversing for $a > a'$.

Exercise 9.4. Suppose that f_a is a family of C^1 unimodal maps depending continuously on the parameter in the C^1 topology. Assume that Δ is an interval of parameter values such that for each $a \in \Delta$, f_a has a (possible one-sided) periodic attractor. Show that f_a has the same topological entropy for each $a \in \Delta$.

10 Monotonicity of the Kneading Invariant for the Quadratic Family

In this section we will consider one parameter families $f_a\colon [0,1] \to [0,1]$ of unimodal maps. In order to show that the kneading invariant depends monotonically on the parameter it suffices to show the following statement:

(∗) if the turning points of f_a and $f_{a'}$ are eventually periodic
and their kneading invariants are equal
then $a = a'$.

Similarly, as we will see, a similar statement holds if the turning points of f_a and $f_{a'}$ are eventually periodic.

Theorem 10.1. **(Sullivan, Milnor, Douady and Hubbard)**
Let $f_a\colon [0,1] \to [0,1]$ be the quadratic family $f_a(x) = ax(1-x)$. Then f_a satisfies condition (∗).

A similar result holds for polynomials of higher degree (but then some conditions on the order of the critical points have to be made). In this section we will give a proof of this result using Teichmüller theory and the Thurston pullback map from Section II.4. The background for the proof of this theorem will only be developed in Section III.1 and in Chapter VI. Since Theorem 10.1 is so important we shall give another proof, due to Sullivan, of this fact in Section VI.4. This other proof uses deformations of quasiconformal maps.

Proof of Theorem 10.1. Let us show that the map T from the proof of Theorem II.4.1 can be used to prove that the quadratic family $f_\mu = \mu x(1-x)$ has the following property: if μ is so that the critical point of f_μ is eventually periodic then there exists no parameter $\mu' \neq \mu$ such that $f_{\mu'}$ is combinatorially equivalent to f_μ. This result uses ideas from Sullivan, Milnor, Douady and Hubbard, see Milnor (1983). So in order to prove Theorem 10.1 it is enough to show that there exists a unique fixed point of the Thurston map $T\colon W \to W$ from Section 4.

In order to prove the fixed point of T is unique, we will show that T is a contraction on the space if we endow W with a 'convenient' metric. For this we need some background in Teichmüller theory (for some background in this, see the Appendix and Chapter VI). Let $\bar{\mathbb{C}} = \mathbb{C} \cup \{\infty\}$ be the Riemann sphere. Consider two punctured Riemann spheres

$$\bar{\mathbb{C}} \setminus \{0, x_1, x_2, \ldots, x_k, 1, \infty\} \text{ and } \bar{\mathbb{C}} \setminus \{0, x_1', x_2', \ldots, x_k', 1, \infty\}.$$

Next define the distance of these punctured spheres to be $\log K$ where K is the smallest number for which there exists a K-quasiconformal homeomorphism $\phi\colon \bar{\mathbb{C}} \to \bar{\mathbb{C}}$ sending x_i to x_i' with $\phi(0) = 0$, $\phi(1) = 1$, $\phi(\infty) = \infty$ which is isotopic (in these punctured spheres) to a homeomorphism preserving the real axis. (We will discuss quasi-conformal maps in the first section of Chapter III much more thoroughly.) According to Teichmüller theory this 'best' homeomorphism exists and is unique. It is called the *Teichmüller map*. In fact, this homeomorphism is quasiconformal and its dilatation is constant. (These terms are explained in Section III.1.) If its dilatation K is equal to 1 then this homeomorphism is conformal. Otherwise there is a unique pair of foliations on $\bar{\mathbb{C}} \setminus \{0, x_1, \ldots, x_k, 1, \infty\}$ and on $\bar{\mathbb{C}} \setminus \{0, x_1', \ldots, x_k', 1, \infty\}$ both of which are smooth and mutually transverse except in a finite number of points. In these special points the foliations have a 'prong' singularity: the foliations are locally mapped by $z \mapsto z^a$ on the pair of foliations consisting of horizontal and vertical lines for some $a \geq 1$. The number of leaves (of one of the foliations) emanating from the singularity is called the number of prongs. These foliations can be characterized as follows. The directions where ϕ stretches most respectively least together define a pair of foliations on $\bar{C} \setminus \{0, x_1, \ldots, x_k, 1, \infty\}$. The corresponding directions for ϕ^{-1} define a similar foliation on $\bar{\mathbb{C}} \setminus \{0, x_1', \ldots, x_k', 1, \infty\}$. The total number of singularities of each of these singularities counted with their index (i.e., the number of prongs) depends only on k. For more on this, see Casson and Bleiler (1988), Gardiner (1987) and Fathi et al. (1979).

Now note that the preimage of such a pair of foliations under any quadratic map on $\bar{\mathbb{C}}$ has either an additional singularity or a singularity with a larger number of prongs. Indeed, if the critical value is for example a regular point (in that case both foliations are transverse at the critical value of the quadratic map) then the preimage has a 2-prong singularity at its preimage. For each other singular point of the foliations there are precisely two preimages which also have singularities. From this argument it follows that the total number of singularities counted with their 'index' increases by taking preimages under a quadratic map.

Now we can prove the result announced: of course we can view f_μ as a quadratic map on the Riemann sphere. Assume that the critical point of f_μ is eventually periodic and denote the forward orbit of the critical point by $\{x_1, \ldots, x_k\}$ as in the proof of Theorem II.4.1. Now put a special metric on the space W from the proof of Theorem II.4.1: we define the distance between $\{x_1, x_2, \ldots, x_k\}$ $\{x_1', x_2', \ldots, x_k'\} \in W$ to be the distance between the corresponding punctured

Riemann spheres

$$\bar{\mathbb{C}} \setminus \{0, x_1, x_2, \dots, x_k, 1, \infty\} \text{ and } \bar{\mathbb{C}} \setminus \{0, x'_1, x'_2, \dots, x'_k, 1, \infty\}$$

i.e., the Teichmüller metric as described above. This metric is equivalent to the usual Euclidean metric on W. So take two points $\{x_1, x_2, \dots, x_k\}$ and $\{x'_1, x'_2, \dots, x'_k\}$ in W and the corresponding 'optimal' homeomorphism $\phi \colon \bar{\mathbb{C}} \to \bar{\mathbb{C}}$. Now one can take the pullback of the homeomorphism ϕ: the map

$$\tilde{\phi} = f^{-1}_{\mu(x'_1,\dots,x'_k)} \circ \phi \circ f_{\mu(x_1,\dots,x_k)}$$

is well defined because the branch point of the first map is mapped by ϕ onto the branch point of the second map. Moreover, $\tilde{\phi}$ sends x_i to x'_i (and $0, 1, \infty$ to $0, 1, \infty$). Moreover, $\tilde{\phi}$ also is isotopic to a map which preserves the real axis and therefore isotopic to the ϕ. Since f_μ is a conformal map, this implies that the quasiconformal constant of $\tilde{\phi}$ is also precisely K. Certainly the best $\hat{K}$ for which there exists a quasiconformal homeomorphism $\hat{\phi}$ sending $(x_1, \dots, x_k)$ onto $(x'_1, \dots, x'_k)$ is not larger than K. So T is non-expanding in terms of the Teichmüller metric. Using the results from Teichmüller theory mentioned above one can even show that this quasiconformal constant is strictly smaller than K and so that T is a contraction. Indeed, associated to the best $\hat{K}$ as above there exists a unique quasiconformal homeomorphism $\hat{\phi}$ with quasiconformal constant $\hat{K}$. Now if $K = \hat{K}$ then, by the uniqueness of the Teichmüller map, $\hat{\phi} = \tilde{\phi}$. This would imply that the preimages of the foliations mentioned above under the maps $z \mapsto \lambda z(1-z)$ must be these unique foliations again. However, the preimage of such a pair of foliations has a larger multiplicity and from the uniqueness of these foliations it follows that this is impossible. In this way it turns out that T becomes (not necessarily uniformly) contracting in this metric. In particular, T can have at most one fixed point. □

It is easy to show that condition $(*)$ implies that the kneading invariant of the family is monotone:

Corollary 10.1. *Let $f_a \colon [0,1] \to [0,1]$ be a unimodal family consisting of C^1 maps depending continuously on a. If condition $(*)$ is satisfied then*

$$a \mapsto \nu(f_a)$$

is monotone.

Proof. Let us suppose by contradiction that the kneading sequence of f_a is not monotone in a. Then there exists a local minimum of $a \mapsto \nu(f_a)$. Note that $a \mapsto \nu(f_a)$ is constant on each interval in the parameter space such that the turning point of f_a is not periodic for each parameter value in this interval. Combining this gives that there exist two parameters $a \neq a'$ such that f_a and $f_{a'}$ both have periodic turning points and the same periodic kneading sequences. This contradicts $(*)$. □

Corollary 10.2. *Let $f_a\colon [0,1] \to [0,1]$ be a unimodal family consisting of C^1 maps depending continuously on a. If condition $(*)$ is satisfied then*

$$a \mapsto h_{top}(f_a)$$

is monotone.

Proof. This follows simply from the fact that the kneading invariant uniquely determines the topological entropy. Indeed, as we have seen in Exercise 9.1, $\nu(f_a)$ is only discontinuous at $a = a'$ if the turning point of $f_{a'}$ is periodic of some period n and if it is not periodic of period n for $a > a'$ sufficiently close to a'. By the previous corollary $\nu(f_a)$ increases at such a parameter value a' and therefore the hyperbolically attracting periodic orbit near the turning point of f_a for a near a' changes from orientation preserving for $a < a'$ to orientation reversing if $a > a'$. Similarly, the lap numbers of $l(f_a^n)$ only change at $a = a'$ if the turning point of f'_a is periodic. Moreover, the lap number increases if the orientation of the periodic orbit changes from orientation preserving to reversing as the parameter increases. Combining this, proves the corollary. □

Remark. 1. Of course the kneading invariant is not strictly monotone: if $f_{a'}$ has an attracting hyperbolic periodic point then the kneading invariants of f_a and $f_{a'}$ are the same for a near a'. 2. From the previous conditions it does not quite follow that the number of periodic orbits of f_a of a given period increases as a increases. If f_a undergoes a periodic doubling (or pitch fork) bifurcation then the number of periodic orbits changes even though the kneading invariant stays the same. However, between any two consecutive periodic doubling bifurcations of different periods (say n and $2n$) the turning point becomes periodic and then the kneading invariant does change. For the quadratic family one can also show that there exists at most one parameter a' for which $f_{a'}$ has a periodic weakly attracting orbit (with eigenvalue -1) with a given kneading type. So for the quadratic family the number of periodic orbits of f_a of a given period increases as a increases. 3. For general families of unimodal maps f_a one certainly cannot hope to get monotonicity for the kneading sequence. Even for a family $f_a\colon [0,1] \to [0,1]$ of the form $f_a(x) = a \cdot f(x)$ this need not be the case. In fact, even when f has negative Schwarzian derivative one can give counter-examples, see Zdunik (1984), Nusse and Yorke (1988a) and Koljada (1989).

11 Some Historical Comments and Further Remarks

There are many refinements of Sarkovskii's theorem. Some of these results not only deal with the periods of periodic orbits, but also with their combinatorial type. Exercise 1.3 gives an example of such a result. We should also mention a proof of Sarkovskii's theorem (in the unimodal case) in which a circle map with a non-trivial rotation interval is associated to a unimodal interval map; the result then follows from Exercise I.1.5, see Gambaudo and Tresser (1991). There are also Sarkovskii type results for continuous maps of degree one on a circle for example in Block et al. (1980). More recently, results were obtained for maps defined on more general one-dimensional spaces. An example of this is the space 'Y' (where three intervals are glued together at one of their endpoints). On this space an analogue of the theorem of Sarkovskii can be proved (but with a more delicate ordering). We shall not go into details of these results in this book but refer to the book of Alsedá et al. (1990), and also to Alsedá (1991), Baldwin (1987), (1988), Bernhardt (1982), (1987), Blokh (1991) and Misiurewicz and Nitecki (1989). Sometimes, using Thurston's results on pseudo-Anosov diffeomorphisms and traintracks, these Sarkovskii type of results can be applied to diffeomorphisms on surfaces, see Gambaudo et al. (1980).

The kneading theory of Section 3 can also be developed for maps of the circle. There are difficulties however because of the circular ordering, see Alsedà and Mañosas (1990) and also Barkmeijer (1988). The tower construction of Hofbauer, described in Section 3b, also makes it pretty clear that some kneading sequences corresponding to non-renormalizable maps should be more 'irrational' than others. It is not impossible that provided a map has a certain growth rate of the sequence S_n defined in Section 3b, results similar to those in Section I.3 for circle diffeomorphisms with Diophantine rotation numbers can be obtained. We will come back to this in Section V.7.

The result of Section 4, that many multimodal families are full, answers a question of Milnor (the unimodal case was dealt with in Milnor and Thurston (1977)). This result can be used to find a map f_μ within a full l-modal family which is essentially conjugate to a given l-modal map g. The parameter μ we are looking for can be found as the fixed point of the Thurston map associated to g. However, although we have shown in Section 4 that this Thurston map does have a fixed point if the family f_μ satisfies some mild conditions, it need not be attracting. Indeed, the examples mentioned at the end of Section 10 show that the Thurston map associated to a map g can have several fixed points. On the other hand, if the critical orbits of g are finite and the family is polynomial then the Thurston map is a contraction, as we have seen in Section 10 of this chapter. Therefore, in this case the polynomial f_μ can be found by Picard iteration. More generally we would like to state:

Conjecture 1. For any l-modal map g and any analytic l-modal full family f_μ, the corresponding Thurston map does not have attracting periodic orbits of period > 1 or strange attractors: almost all points $x \in W$ tend to a fixed point.

The algorithm suggested by the proof in Section 4 was used to draw the pictures in this chapter.

The renormalization results of Section 5 are usually only stated in the unimodal case. However, there are several papers dealing with renormalization in the bimodal case; for example, in the papers of MacKay and Tresser, bimodal families of maps of the circle and the set of parameters for which the corresponding maps have zero topological entropy are studied.

There are several results showing that the topological entropy of a continuous interval map is determined by its periodic orbits, see for example Jonker (1981), Baldwin (1987) and Block and Coppel (1989). There are several algorithms to determine numerically the topological entropy of a map of the interval, see Collet et al. (1983), Góra and Boyarsky (1991) and Block and Keesling (1991).

The kneading determinant from Section 8 is related to a zeta function, see Milnor and Thurston (1977). The properties of another zeta function are related to the spectrum of the Perron-Frobenius operator (this operator acts on densities of measures and is defined in Sections V.2 and V.3), see Keller (1984) and also Baladi and Keller (1990).

Misiurewicz (1989) has some results on the continuous dependence of the topological entropy for maps which are not C^1 or piecewise monotone, see also Misiurewicz and Shlyachkov (1989).

As we have seen in Section 10 of this chapter each bifurcation of the map $x \mapsto ax(1-x)$ creates periodic orbits as a increases. However, for general families of the form $x \mapsto a \cdot f(x)$ this is not the case. In this direction we would like to state the following conjectures.

Conjecture 2. If $f\colon [0,1] \to [0,1]$ is unimodal, $Sf < 0$ and f is symmetric then the kneading invariant of the $x \mapsto a \cdot f(x)$ is monotone in a.

So other instances of monotone families can be found in Douady and Hubbard (1984).

Conjecture 3. There exists an open set U of one-parameter families of unimodal smooth maps such that any two families $\{f_a\}_{a\in[0,1]}, \{\tilde{f}_a\}_{a\in[0,1]} \in U$ are equivalent. More precisely, for each pair of these families there exists a one-parameter family of homeomorphisms $h_a\colon [0,1] \to [0,1]$ and a reparametrization ρ such that

$$h_a \circ f_a = \tilde{f}_{\rho(a)} \circ h_a.$$

It is also not clear whether these homeomorphisms could be chosen in such a way that they would depend continuously on the parameter, because at the parameters where saddle-node bifurcations occur one could get moduli of stability as in Newhouse et al. (1983).

We expect that kneading invariants of generic families vary not too erratically on the parameter.

Conjecture 4. Generic one-parameter families of unimodal maps have the property that their kneading invariants vary piecewise monotonically.

The corresponding statement has been proved for families of circle diffeomorphisms in the C^1 topology, see de Melo and Pugh (1991).

For families of bimodal maps, for example the family $f_{a,b}(x) = ax^3 + bx^2 + (1 - a - b)x$, one can also ask whether the entropy depends 'monotonically' on the parameter. Milnor (1990) made this question precise in the following way:

Conjecture 5. The set of parameters (a, b) for which the topological entropy of the map $f_{a,b}\colon [0,1] \to [0,1]$ is equal to s is connected for each $s \geq 0$.

The numerical studies from Milnor (1990) support this conjecture.

Chapter III.

Structural Stability and Hyperbolicity

In this chapter we want to analyze which one-dimensional systems are structurally stable. In Chapter I this question was quite easy to answer: a circle diffeomorphism is structurally stable if and only if all periodic points of f are hyperbolic. Moreover structurally stable diffeomorphisms form an open and dense set. (These statements were shown in Exercise I.4.1.) For non-invertible maps the situation is much more complicated and partly unknown. The concept of hyperbolicity of some infinite compact set will play an essential role in this discussion. As we will see in this chapter non-invertible one-dimensional dynamical systems have many infinite hyperbolic sets whereas circle diffeomorphisms have none.

Hyperbolicity plays such a key role because it can be used to establish the existence of non-invertible one-dimensional endomorphisms which are structurally stable and at the same time have a very complicated dynamics. In fact, one of the main open problems in the theory is to prove that the set of one-dimensional dynamical systems with this hyperbolicity property is dense. We will discuss this problem in more detail later.

Section 1 is devoted to the study of holomorphic dynamics in the Riemann sphere. The notion of hyperbolicity was already present in the pioneering work of Julia and Fatou. Many results in the dynamics of rational maps on the Riemann sphere have an analogue in real one-dimensional dynamics. Moreover, although the techniques are necessarily different, one can find similarities in many proofs. This is the main reason for proving some statements in this section instead of merely referring to more complete sources. The techniques in this first section do not, however, play an immediate role in the remainder of this chapter, so the reader could skip this section. However, in Chapter VI we will make extensive use of the ideas explained in Section 1.

In Section 2 we discuss the relation between hyperbolicity and structural stability for smooth endomorphisms of the circle and of the interval. We also prove that a hyperbolic invariant set for a $C^{1,\alpha}$ map has either zero Lebesgue measure or full Lebesgue measure (in which case the map is an expanding covering map of the circle).

In Section 3 we prove that maps with negative Schwarzian derivatives behave very much like rational maps and have many invariant hyperbolic sets. In Section 4 we give a fairly complete description of the structure of maps with negative Schwarzian derivative in the same spirit as in Jonker and Rand (1981) and Van Strien (1981).

Section 5 is an exposition of the results of Mañé (1985) which show that hyperbolicity is very common even in general smooth interval maps. This result deals with the dynamics of the points which stay away from the critical points (these are points where the derivative is zero). More specifically, the main result, which is an extension of the results in Section 3, states that an invariant set of a C^2 interval map is hyperbolic if it does not contain critical points and non-hyperbolic periodic points. The proof we will give is much easier than the original one in Mañé (1985) and uses the ideas from Van Strien (1990). There points whose orbits come close to critical points are also analyzed provided the critical points satisfy the so-called Misiurewicz condition. In Section 6 we give a proof of some of these results for Misiurewicz maps under the additional assumption that they satisfy the negative Schwarzian condition. These results will turn out to be important for the proof of Jakobson's Theorem in Section V.6.

1 The Dynamics of Rational Mappings

In this section we will describe some of the aspects of the dynamics of holomorphic maps $f: \bar{\mathbb{C}} \to \bar{\mathbb{C}}$. It will turn out that many of the results and concepts from this theory play a part in the real case. However, none of the results in this section is actually needed in this chapter. This section serves also as an introduction to the tools used in Chapter VI. The development of this theory started at the end of the last century and made much progress in the beginning of this century with the long papers of Julia (1918) and Fatou (1919)-(1920). More recently, this theory gained new impetus with the work of several mathematicians, in particular Sullivan, Thurston, Douady and Hubbard, and with the beautiful pictures that resulted from numerical experiments. We do not intend to make a detailed exposition of the subject. The reader can find this in the survey articles of Blanchard (1984) and Lyubich (1986). Instead we present a few basic results in order to make explicit some of the analogies between this theory and the dynamics of smooth maps defined on an interval.

The main reason for the success of the study of iterations of rational functions comes from the strong analytical tools provided by Complex Analysis. The main tool used by Julia and Fatou is the Koebe-Riemann Uniformization Theorem and one of its consequences: Montel's Theorem on normal families of holomorphic maps. Another important result is Koebe's Distortion Theorem for univalent holomorphic functions. More recently, Sullivan started the use of the Measurable Riemann Mapping Theorem to perform deformations of rational maps via quasiconformal maps, and Thurston introduced Teichmüller theory in the study of rational maps, see Sullivan (1985) and Thurston.

The Uniformization Theorem of Koebe and Riemann states that any simply connected Riemann surface is conformally equivalent, i.e., holomorphically equivalent, to either the Riemann sphere $\bar{\mathbb{C}} = \mathbb{C} \cup \{\infty\}$, the complex plane $\mathbb{C}$ or the unit disc $\mathbb{D} = \{z \in \mathbb{C};\ |z| < 1\}$. As a consequence, we have that given any Riemann surface S, there exists a holomorphic covering map $\pi\colon \tilde{S} \to S$ where $\tilde{S}$ is one the three Riemann surfaces above. If S is either the complex plane, the punctured plane $\mathbb{C} \setminus \{0\}$ or is homeomorphic to the torus, then it is called a parabolic Riemann surface and its holomorphic universal covering surface $\tilde{S}$ is the complex plane. If S is the Riemann sphere then it is an elliptic surface. All other Riemann surfaces are called *hyperbolic* and they have the unit disc as the holomorphic universal covering space (later on we shall put a special metric on such a surface S using its covering space $\mathbb{D}$). In particular, if S is the Riemann sphere minus three points then there exists a holomorphic covering map $\pi\colon \mathbb{D} \to S$. If $f_n\colon D \to \bar{\mathbb{C}}$ is a sequence of holomorphic maps whose images omit three points in the Riemann sphere, then they define a sequence of holomorphic maps $f_n\colon D \to S$, where S is the sphere minus these three points. If D is simply connected we can lift all these maps and we get a sequence of holomorphic maps $\hat{f}_n\colon D \to \mathbb{D}$ such that $\pi \circ \hat{f}_n = f_n$. Of course a sequence of bounded holomorphic maps $\hat{f}_n$ is equicontinuous and is therefore a *normal family*, i.e., its closure is a compact set in the space of holomorphic maps in D endowed with the compact open topology. From this we get that $\{f_n\}$ is a normal family of holomorphic maps. Therefore we get the following important statement.

Theorem 1.1. **(Montel)** *Let S be any Riemann surface and $f_n\colon S \to \bar{\mathbb{C}}$ be a family of holomorphic maps. If there exist at least three points in the Riemann sphere $\mathbb{C}$ which are not covered by the union of the images of all these maps, then the family $\{f_n\}$ is a normal family.*

Let $f\colon \bar{\mathbb{C}} \to \bar{\mathbb{C}}$ be a holomorphic map. Then f is a rational map, that is $f(z) = \frac{P(z)}{Q(z)}$ where P and Q are polynomials. The *critical set* of f, is the finite set $C_f = \{z \in \bar{\mathbb{C}};\ Df(z) = 0\}$. The f-images of points in $C(f)$ are called *critical values*. Clearly for each non-critical value y of f the set $f^{-1}(y)$ consists of d distinct point. Hence the restriction of f to $\bar{\mathbb{C}} \setminus f^{-1}(f(C_f))$ is a covering map of degree $d = \max\{\text{degree } P, \text{degree } Q\}$ over $\bar{\mathbb{C}} \setminus f(C_f)$. In the remainder of this section we will only study iterates of rational maps of degree $d \geq 2$.

Definition. A point x in the Riemann sphere belongs to the *Fatou set* of f if there exists an open neighbourhood V of x such that the restriction of the iterates of f to V is a normal family. Let $F(f)$ denote the Fatou set of f. The *Julia set* $J(f)$ of f, , is the complement of the Fatou set.

Let us say that a set E is completely invariant if $f(E) = E = f^{-1}(E)$. The following statement is an important consequence of Montel's Theorem.

Lemma 1.1. *Let f be a rational map of degree d. Then there exists a completely invariant set $E = E(f)$ satisfying the following properties:*

1. *the set $E(f)$ contains at most two points;*
2. *if D is an open set with a non-empty intersection with the Julia set of f, then $\cup_{n=0}^{\infty} f^n(D) \supset \bar{\mathbb{C}} \setminus E(f)$;*
3. *if the cardinality of $E(f)$ is one then f is holomorphically conjugate to a polynomial and if the cardinality of $E(f)$ is two, then f is holomorphically conjugate to either $z \mapsto \dfrac{1}{z^d}$ or $z \mapsto z^d$.*

Proof. From Montel's Theorem it follows that if D is an open set intersecting the Julia set then the iterates of D omit at most two points in the Riemann sphere. Suppose first that there is an open set D containing a point $x \in J(f)$ such that $E = \bar{\mathbb{C}} \setminus \cup_{n=0}^{\infty} f^n(D) = \{\alpha, \beta\}$. We have that $f^{-1}(E) \subset E$ because E is the complement of a forward invariant set. Since E has at most two points, it follows that it is also forward invariant. Let us show that $E(f) = E$. There are two cases to consider:

i) $E = \{\alpha, \beta\}$ with $f(\alpha) = \alpha$. Hence, $\alpha \notin f^{-1}(\beta)$ and, since $f^{-1}(\beta) \subset f^{-1}(E)$, therefore $f^{-1}(\beta) = \{\beta\}$. It follows that $f(\beta) = \beta = f^{-1}\beta)$ and $f(\alpha) = \alpha = f^{-1}(\alpha)$. So the points α and β are totally invariant. If $\phi\colon \bar{\mathbb{C}} \to \bar{\mathbb{C}}$ is a holomorphic isomorphism (i.e. a Möbius transformation: $\phi(z) = \frac{az+b}{cz+d}$) with $\phi(\alpha) = 0$ and $\phi(\beta) = \infty$ we have that $g = \phi \circ f \circ \phi^{-1}$ is a rational map satisfying

$$\begin{aligned} \infty &= g(\infty) = g^{-1}(\infty) \text{ and} \\ 0 &= g(0) = g^{-1}(0). \end{aligned}$$

The first equation implies that g is a polynomial which, together with the second equation, implies that $g(z) = cz^d$ where c is a complex constant. The Julia set of g is clearly equal to the unit circle and if W is any small disc intersecting the Julia set then $\cup_{n=0}^{\infty} g^n(W) = \bar{\mathbb{C}} \setminus \{0, \infty\}$ since the iterates of W omit the two points $0, \infty$ and cannot omit a third point. Hence, $E(f) = \{\alpha, \beta\}$ in this case.

ii) $E = \{\alpha, \beta\}$ with $f(\alpha) = \beta$. Hence, $f^{-1}(\alpha) = \beta$ and, therefore, $f(\beta) = \alpha$. As before, we take a Möbius transformation ϕ with $\phi(\alpha) = 0$, $\phi(\beta) = \infty$ and let $g = \phi \circ f \circ \phi^{-1}$. Hence, g is a rational map of degree d such that $0 = g^{-1}(\infty) = g(\infty)$. It follows that $g(z) = cz^{-d}$ where c is a constant and this is clearly holomorphically conjugate to the mapping $z \mapsto z^{-d}$. The Julia set of g is again the unit circle and the iterates of any small disc intersecting the unit circle omit precisely the points $0, \infty$. Therefore, again $E(f) = \{\alpha, \beta\}$ in this case.

Suppose now that there is no open set D intersecting the Julia set whose iterates omit two points. Then, as we have mentioned above, there is an open set D containing a point in the Julia set and whose iterates omit one point which we call α. If ϕ is a Möbius transformation which maps α to ∞, then $g = \phi \circ f \circ \phi^{-1}$ is a polynomial. The Julia set of any polynomial is contained in the finite plane (∞ is an attracting fixed point) and the iterates of any open subset of the finite plane omit the point ∞. Hence, the iterates of any disc in the finite plane intersecting the Julia set of g omits the point ∞ and (by assumption) no other point. Therefore $E(f) = \{\alpha\}$ in this case and the proof is finished. □

Remark. The completely invariant set $E(f)$ of the above lemma is called the *exceptional set* of f. One has $E(f) \cap J(f) = \emptyset$ and from Property 2 from the previous lemma one has for each $x \notin E$ that $\alpha(x) \supset J(f)$. Here $\alpha(x)$ is the set of points z for which there exist $z_k \to z$ and $n(k) \to \infty$ such that $f^{n(k)}(z_k) = x$. Moreover, if f is neither a polynomial and nor holomorphically conjugate to the mapping $z \mapsto z^{-d}$ then $E(f)$ is the empty set.

Proposition 1.1.

1. *The Fatou set and the Julia set are both completely invariant sets, i.e., $f(F(f)) = f^{-1}(F(f)) = F(f)$ and $f(J(f)) = f^{-1}(J(f)) = J(f)$.*
2. *Each orbit in the Fatou set is Lyapounov stable with respect to the spherical metric ρ: given $x \in F(f)$ and $\epsilon > 0$ there exists $\delta > 0$ such that if $\rho(x, y) < \delta$ then $\rho(f^n(x), f^n(y)) < \epsilon$ for all $n \in \mathbb{N}$.*
3. *The Julia set is a non-empty perfect set (this means that it is closed and contains no isolated points).*
4. *If the Julia set has non-empty interior then it is the whole Riemann sphere.*

Proof. It is clear from the definition that the Fatou set is open and completely invariant. Hence, the Julia set is closed and also completely invariant.

To prove Statement 2, let $x \in F(f)$ and V be a neighbourhood of x such that the family of iterates of f restricted to V is a normal family. Hence, if $W \subset V$ is a compact neighbourhood of x then $\{f^n|W\}$ is an equicontinuous family. Therefore, given $\epsilon > 0$ there exists $\delta > 0$ such that if $\rho(x, y) < \delta$ then $\rho(f^n(x), f^n(y)) < \epsilon$ for all $n \in \mathbb{N}$. Hence the orbit of x is Lyapounov stable.

Let us prove Statement 3. Suppose the whole sphere is the Fatou set. Then the iterates of f would be a normal family at each point of the whole sphere, and by compactness it would be a normal family on the whole sphere. But this is impossible since f^n is a rational map of degree nd tending to infinity with n (here we use that $d \geq 2$ because the iterates of a Möbius transformation are again Möbius transformations). Hence the Julia set is non-empty. In order to show that the Julia set is perfect let a be a point in the Julia set. Suppose first that a is not a periodic point of f. Let V be a neighbourhood of a. We want to prove the existence of a point of the Julia set in $V \setminus \{a\}$. Since a is not in the exceptional set of f, there exists an integer n such that $f^n(V)$ contains a, i.e., there exists $b \in V$ such that $f^n(b) = a$. Since the Julia set is backward invariant, we get that $b \in J(f)$. On the other hand, $b \neq a$ because a is not periodic. Hence every point in the Julia set which is not a periodic point is an accumulation point of the Julia set. Suppose now that a is a periodic point of period k in the Julia set. Then a is not a super-attracting periodic point (a is super-attracting if $Df^k(a) = 0$ or in other words if there is a critical point in the orbit of f) because in such points $\{f^n\}$ is equicontinuous. It follows that there exists $b \in f^{-1}(a)$, $b \neq f^{k-1}(a)$. Since b is not a periodic point and it is in the Julia set, it follows from the previous argument that there exists a sequence $z_n \to b$ with $z_n \in J(f)$ and $z_n \neq b$. Hence $f(z_n) \to a$, $f(z_n) \in J(f)$ and $f(z_n) \neq a$. Thus a is also an accumulation point of the Julia set in this case.

Finally, we prove Statement 4. Suppose the interior of the Julia set is non-empty and let D be an open disc contained in the Julia set. Since the family of the restriction of the iterates of f to D is not a normal family, it follows from Montel's Theorem that $\cup_{n\in\mathbb{N}} f^n(D)$ omits at most two points in the Riemann sphere. By the invariance of the Julia set, we have that $\cup_{n\in\mathbb{N}} f^n(D)$ is contained in the Julia set. Hence $J(f)$ is equal to the sphere because it is compact. □

To get further properties of the dynamics of rational maps it is convenient to use the Poincaré metric, see the Appendix.

Definition. We call a Riemann surface S *hyperbolic* if $\mathbb{D}$ is the universal covering of S. So in this case there exists a holomorphic universal covering map $\pi\colon \mathbb{D} \to S$. The *hyperbolic metric* on S is the unique Riemannian metric on S such that π is a local isometry between the hyperbolic metric of $\mathbb{D}$ and the hyperbolic metric of S.

Using this hyperbolic metric it is often extremely easy to show that f is expanding on some sets. For this one uses the following version of the Lemma of Schwarz which states that a holomorphic map from $\mathbb{D}$ into itself which fixes 0 is either a rotation or is everywhere contracting.

Lemma 1.2 (Lemma of Schwarz for Riemann surfaces)**.** *Let $f\colon S_1 \to S_2$ be a holomorphic map between hyperbolic Riemann surfaces. Then either f is a holomorphic covering map and a local isometry (of the hyperbolic metric) or it strictly contracts the hyperbolic metric.*

Proof. See the Appendix. □

Definition. Let $f\colon \bar{\mathbb{C}} \to \bar{\mathbb{C}}$ be a rational map and let p be a periodic point of period n of f. We say that p is an *attractor* if $0 < |Df^n(p)| < 1$, it is a *super-attractor* if $Df^n(p) = 0$, it is a *repeller* if $|Df^n(p)| > 1$ and if $|Df^n(p)| = 1$ it is called an *indifferent* periodic point. The *basin* of p is the set of points whose ω-limit set is the orbit of p and the *immediate basin* is the union of the connected components of the basin that contain points of the orbit of p.

Theorem 1.2. **(Julia)** *If p is an attracting periodic point of a rational map $f\colon \bar{\mathbb{C}} \to \bar{\mathbb{C}}$ then the immediate basin of p contains a critical point of f. In particular, the number of periodic attracting orbits is bounded by the number of critical points. (Note that ∞ is a critical point and also a super-attracting fixed point if f is a polynomial.)*

Proof. Let B_p be the connected component of the basin of p that contains p. Then $f^i(B_p) = B_{f^i(p)}$ is the connected component of the basin containing $f^i(p)$ and the immediate basin of p is $\cup_{i=0}^{n-1} f^i(B_p)$. Hence there is a critical point of f in the immediate basin of p if and only if there exists a critical point of f^n in B_p. Suppose, by contradiction, that the map $g = f^n$ has no critical point in B_p. We claim that this implies that g has an analytic inverse on B_p. Indeed, take neighbourhoods $U \subset V$ of p with $g(V) = U$ and $V \setminus U$ an annulus. By writing V as the union of a nested collection of simple closed curves surrounding p one can easily see that g must be injective on U if g has no critical points. So one can define g^{-1} on U. Similarly, one can define g^{-1} on $g^{-n}(U)$ for each U. In this way we get that g^{-1} is a well defined analytic map on B_p. It follows that g is a covering map and, therefore, it is a local isometry with respect to the hyperbolic metric of B_p. But this is not possible because $p \in B_p$ is an attracting fixed point of g. □

Remark. One can even prove that the immediate basin of an attracting periodic point contains a critical point whose forward orbit consists of an infinite number of distinct points. The above proof does not guarantee the last condition.

Theorem 1.3. **(Julia)** *If f is a rational map of degree $d \geq 2$, then the number of non-repelling periodic orbits of f is bounded by $3(2d-2)$.*

Proof (sketch). By the previous theorem, it follows that the number of attracting periodic orbits is bounded by the number of critical point which is bounded by $2d - 2$. Using a similar, but more complicated, argument one can prove that if p is a periodic point of period n and $Df^n(p) = 1$ then there must exist a critical point in the immediate basin of p. Hence the number of attracting periodic orbits plus the number of periodic orbits of this type is bounded by the number of critical points. It remains to estimate the cardinality of the set $\mathcal{I}$ of periodic points p such that if n is the period of p then $|Df^n(p)| = 1$ but $Df^n(p) \neq 1$. Given such a periodic point, it follows from the Implicit Function Theorem that there exist a neighbourhood V of f in the space of rational maps and a holomorphic function $\phi_p\colon V \to \bar{\mathbb{C}}$ such that $\phi_p(f) = p$ and $\phi_p(g)$ is a periodic point of period n of g. Then the function $\lambda_p\colon V \to \mathbb{C}$, $\lambda_p(g) = Dg^n(\phi_p(g))$ is holomorphic and it is not constant. Hence it maps any neighbourhood of f onto a neighbourhood of $Df^n(p)$. In particular, there is an open set of maps in this neighbourhood consisting of maps with an attracting periodic orbit of period n. We claim that if the cardinality of $\mathcal{I}$ is N, we can perturb the map in order to get a map g which has $\frac{1}{2}N$ attracting periodic orbits. Indeed, let $w \mapsto f_w$ be a holomorphic one parameter family of rational maps of degree d with $f_0 = f$ and, for each such periodic point p of f let us denote by $\lambda_p(w)$ the restriction of $\lambda_p \colon V \to \mathbb{C}$ to this one parameter family. Hence for each p, λ_p is a non-constant holomorphic function in a neighbourhood of 0. In this neighbourhood let us consider the function $\sigma(w) = \sum_{p \in \mathcal{I}} \operatorname{sign}(\log|\lambda_p(w)|)$, where $\operatorname{sign}(t)$ denotes the sign of the real number t (with $\operatorname{sign}(0) = 0$). Thus σ is an integer valued function. For each p, the set $\{w\,;\ |\lambda_p(w)| = 1\}$ is a real analytic curve passing through zero and $|\lambda_p(w)| - 1$ changes sign along this curve. So take a line L through 0 which is transversal to each of these curves in 0. Then $L \ni w \mapsto |\lambda_p(w)| - 1$ changes sign at 0 for each $p \in \mathcal{I}$. In particular, $\sigma(w) = -\sigma(-w)$ for $w \neq 0$ and $w \in L$ sufficiently close to 0. So there exists $w \in L$ arbitrarily close to 0 such that $\sigma(w) \leq 0$ and $|\lambda_p(w)| \neq 1$ for all p. This clearly implies that at least half of the N periodic points of f_w are attractors. This proves the claim. Hence the total number of periodic orbits of this type is bounded by twice the number of critical points. This proves the theorem. □

Theorem 1.4. *The Julia set is the closure of the set of repelling periodic orbits of f.*

Proof. Since the number of non-repelling periodic orbits is finite, it is enough to prove that the Julia set is the closure of the repelling periodic points of f. Let p be a non-periodic point in the Julia set. We may also assume that p is not a critical value. Given a neighbourhood V of p, we want to prove the existence of a periodic point in V. Let p_1 and p_2 be such that $f(p_1) = f(p_2) = p$ (note that we have assumed all along that degree$(f) \geq 2$). Since p is not a critical value of f, there exists a neighbourhood $U \subset V$ of p and local inverses of f, $\phi_i \colon U \to U_i$ such that $\phi_i(p) = p_i$. Taking U small, we can assume that the neighbourhoods U, U_1, U_2 are pairwise disjoint. The family $f^n|U$ is not a normal family because p is in the Julia set. Hence $g_n(x) = \frac{f^n(x)-\phi_1(x)}{f^n(x)-\phi_2(x)} \frac{x-\phi_2(x)}{x-\phi_1(x)}$ is not a normal family. Therefore, by Montel's Theorem $\cup_{n \geq 0} g_n(U)$ cannot omit the points $\{0, 1, \infty\}$. Thus, because $\phi_1(x) \neq \phi_2(x)$ for all $x \in U$, there exist $n \in \mathbb{N}$ and $x \in U$ such that either $f^n(x) = x$ or $f^n(x) = \phi_1(x)$ or $f^n(x) = \phi_2(x)$. In all of these cases, x is a periodic point of f. □

Definition. Let $K \subset \bar{\mathbb{C}}$ be an invariant set for the rational map f. We say that K is a *hyperbolic* set if there exist constants $C > 0$ and $\lambda > 1$ such that

$$||Df^n(x)|| > C\lambda^n,$$

for all $n \in \mathbb{N}$. Here $||\cdot||$ is any Riemannian metric on the Riemann sphere.

The above definition does not depend on the choice of the metric of the sphere, because any two Riemannian metrics are equivalent. If we change the metric, the only thing that changes is the constant C. Because of the Lemma of

Schwarz for Riemann surfaces (Lemma 1.2) it is natural to consider the Poincaré metric on an appropriate neighbourhood of K:

Theorem 1.5. *Let $f\colon \bar{\mathbb{C}} \to \bar{\mathbb{C}}$ be a rational map and $S = \bar{\mathbb{C}} \setminus P$ where P is the closure of the forward orbit of the critical points of f. If S is connected and $K \subset S$ is a compact invariant set, then K is a hyperbolic set.*

Proof. Let $||v||_x^S$ be the norm of the tangent vector v at the point x with respect to the hyperbolic metric of S. We claim that for each $x \in S$ with $f(x) \in S$ one has $||Df(x)(v)||_{f(x)}^S > ||v||_x^S$, i.e., f expands the hyperbolic metric of S. Indeed, let R be the connected component of $f^{-1}(S)$ that contains x. Then $R \subset S$ and since the restriction of f to R is a proper holomorphic map without critical points and it maps R onto S, it is a covering map. Therefore, it is a local isometry of the hyperbolic metrics, namely,

$$||Df(x)(v)||_{f(x)}^S = ||v||_x^R.$$

On the other hand, the inclusion map $i\colon R \to S$ is also holomorphic but it is not a covering map because it is not surjective. Hence, since S is connected, one gets from Lemma 1.2 that

$$||v||_x^S = ||Di(x)(v)||_x^S < ||v||_x^R.$$

This proves the claim. Since $K \subset S$ is compact and invariant we have, from the claim, that there exists $\lambda > 1$ such that $||Df(x)(v)||_{f(x)}^S > \lambda ||v||_x^S$. If $||\cdot||$ is some Riemannian metric in the Riemann sphere, there exists a constant $C > 0$ such that $\frac{1}{C}||v||_x < ||v||_x^S < C||v||_x$ for every $x \in K$ and every tangent vector v at x. Hence, $||Df^n(x)(v)||_{f^n(x)} \geq \frac{1}{C}||Df^n(x)(v)||_{f^n(x)}^S > \frac{1}{C}\lambda^n||v||_x^S > (\frac{1}{C})^2\lambda^n||v||_x$. □

Corollary 1.1. *If f is a rational map such that each critical point of f is in the basin of an attracting periodic point or is a super-attractor, then the Julia set of f is a hyperbolic set.*

Corollary 1.2. *If f is a rational map such that each critical point is eventually mapped onto a repelling periodic point then $J(f) = \mathbb{C}$ and any closed forward invariant set which does not contain a critical point is hyperbolic. (Note that this corollary is not applicable if f is polynomial, because then ∞ is a critical point and a super-attractor.)*

Proof of Corollary 1.2. In this case, the forward orbits of the critical points form a finite set P and $S = \bar{\mathbb{C}} \setminus P$ is connected. Clearly P is contained in the Julia set because every point of P falls into an expanding periodic point. Hence the backward orbit of P is also in the Julia set. If x is not in the backward orbit of P then $f^n(x) \in S$ for every $n \geq 0$. Let $y \in S$ be a point in the ω-limit of x (it is clear that the ω-limit set of x cannot be a subset of P since the periodic points in P are repelling). Since $||Df(z)||^S_{f(z)} > 1$ for $z, f(z) \in S$, there exists $\lambda > 1$ such that $||Df(z)||^S_{f(z)} > \lambda$ for each z near y. From this one gets for $f^{n_i}(x) \to y$ that $||Df^{n_{i+1}}(x)||^S \geq 1 \cdot \lambda \cdot ||Df^{n_i}(x)||^S$ and therefore that $||Df^{n_i}(x)||^S$ tends to infinity. The hyperbolic metric and the spherical metric are equivalent in a compact neighbourhood of $\{x, y\}$. Therefore, $||Df^{n_i}(x)||$ also tends to infinity. This implies that x belongs to the Julia set because f^{n_i} cannot be a normal family in a neighbourhood of x. Hence the Julia set is the whole Riemann sphere. Since $||Df(x)(v)||^S_{f(x)} > ||v||^S_x$ for $x, f(x) \in S$ the last statement follows from compactness and the fact that each point in P is eventually periodic and repelling. □

Theorem 1.6. **(Structural Stability of $J(f)$)**
Let f be a rational map of degree d whose Julia set is a hyperbolic set. Then there exists a neighbourhood $\mathcal{N}$ of f in the space of rational maps of degree d such that for any $g \in \mathcal{N}$, the Julia set of g is also hyperbolic and there exists a homeomorphism $h_g\colon J(f) \to J(g)$ which is a conjugacy between $f|J(f)$ and $g|J(g)$.

Before proving Theorem 1.6, let us prove the existence of an *adapted metric* to a hyperbolic invariant set.

Lemma 1.3. *Let K be a hyperbolic invariant set for a rational map f. Then there exists a Riemannian metric $||\cdot||$ on $\bar{\mathbb{C}}$ and $\lambda > 1$ such that*

$$||Df(x) \cdot v||_{f(x)} > \lambda ||v||_x$$

for every $x \in K$ and $v \in T_x\bar{\mathbb{C}}$.

Proof. Let $|\cdot|_x$ be any Riemannian metric on the Riemann sphere. Since K is a hyperbolic invariant set, there exist $C > 0$ and $\alpha > 1$ such that

$$|Df^n(x)|_{f(x)} > C\alpha^n |v|_x$$

for every $v \in T\bar{\mathbb{C}}_x$ and for every $x \in K$. Let $N > 0$ be such that $C\alpha^N > 1$. Define

$$||v||_x = \frac{1}{N} \sum_{n=0}^{N-1} |Df^n(x) \cdot v|_{f^n(x)}.$$

Clearly $||\cdot||_x$ is a Riemannian metric on the Riemann sphere. Since any two Riemannian metrics on the Riemann sphere are equivalent, it follows that there exists a constant $C_1 > 0$ such that $\frac{1}{C_1}|v|_x < ||v||_x < C_1|v|_x$ for all $x \in \bar{\mathbb{C}}$ and all $v \in T_x\bar{\mathbb{C}}$. Thus,

$$||Df(x) \cdot v||_{f(x)} = \frac{1}{N} \sum_{n=0}^{N-1} |Df^n(f(x)) \cdot Df(x) \cdot v|_{f^{n+1}(x)} =$$

$$= ||v||_x + \frac{1}{N}(|Df^N(x)\cdot v|_{f^N(x)} - |v|_x) \geq ||v||_x + \frac{1}{N}(C\alpha^N - 1)|v|_x \geq$$

$$\geq \frac{(1 + (C\alpha^N - 1))}{NC_1}||v||_x.$$

Taking $1 < \lambda < 1 + \frac{(C\alpha^N - 1)}{NC_1}$ the lemma follows. □

Proof of Theorem 1.6. By taking an adapted metric, we can assume that $|Df(x)| > 1$ for all $x \in J(f)$. From the compactness of $J(f)$, there exist $\lambda > 1$, a neighbourhood V of $J(f)$ and a neighbourhood $\mathcal{N}$ of f such that $|Dg(x)| > \lambda$ for all $x \in V$ and for all $g \in \mathcal{N}$. Let $\epsilon > 0$ be such that for all $x \in J(f)$, the ball $B(x, \epsilon)$ is contained in V. By shrinking $\mathcal{N}$, we can assume that for all $x \in J(f)$ and all $g \in \mathcal{N}$, $g(B(x, \epsilon)) \supset B(f(x), \epsilon)$. Therefore, for each integer n, and each $x \in J(f)$ the set $W_{g,n}(x) = \{y\,;\ g^i(y) \in B(f^i(x), \epsilon),\ \text{for all } 0 \leq i \leq n\}$ is a non-empty set with diameter at most equal to $2\epsilon\lambda^{-n}$. Hence there exists a unique point $h_g(x)$ such that $g^n(h_g(x)) \in B(f^n(x), \epsilon)$ for all integer n. Clearly $h_g(f(x)) = g(h_g(x))$. It is easy to verify that h_g is continuous and one-to-one. Thus h_g is a homeomorphism from $J(f)$ onto the compact set $h_g(J(f)) \subset V$ conjugating f and g. Since h_g maps the periodic points of f in $J(f)$ into the periodic points of g in $h_g(J(f))$, we get that $h_g(J(f))$ contains all periodic points of g except for, possibly, a finite number of them (corresponding to the periodic points of f which are not in the Julia set). Hence, $h_g(J(f))$ is the Julia set of g. □The lemma below plays an important role in extending

conjugacies defined on a set to its closure, see Mañé et al. (1983). We will skip the proof because it is not needed in this book.

Lemma 1.4. *Let $X \subset \bar{\mathbb{C}}$ be a set and $\phi_\lambda\colon X \to \bar{\mathbb{C}}$ be a parametrized family of maps, where the parameter λ belongs to an open set $W \subset (\bar{\mathbb{C}})^k$, satisfying the following conditions:*

1. *$\phi_0(z) = z$ for all $z \in X$;*
2. *ϕ_λ is one-to-one for every $\lambda \in W$;*
3. *for each $z \in X$, the map $W \to \bar{\mathbb{C}}$, $\lambda \mapsto \phi_\lambda(z)$ is holomorphic.*

Then, for each $\lambda \in W$, there exists a unique extension $\phi_\lambda\colon \bar{X} \to \bar{\mathbb{C}}$ to the closure $\bar{X}$ of X which is continuous and satisfies Properties 1, 2 and 3 above.

Theorem 1.7. **(Mañé et al.)** *Let $f\colon \bar{\mathbb{C}} \to \bar{\mathbb{C}}$ be a rational map satisfying the following properties:*

1. *each critical point of f is contained in the basin of a hyperbolic attracting periodic point;*

2. *the critical points are non-degenerate, i.e., the second derivative of the map does not vanish at the critical points;*

3. *there is no critical point in the forward orbit of a critical point.*

Then f is structurally stable.

Proof (Sketch). If W is a small neighbourhood of f then Properties 1, 2 and 3 above are clearly satisfied for all $g \in W$. Next we take a neighbourhood V of the attracting periodic points of f and define a holomorphic family of continuous maps $h_g \colon V \to h_g(V) \subset \bar{\mathbb{C}}$ (close to the identity) which conjugate f with g and map the forward orbit of the critical points of f into the forward orbit of the corresponding critical points of g. Then we extend h_g to a conjugacy in the backward orbit of V. This is the union of the basins of the periodic attractors of f which is dense in $\bar{\mathbb{C}}$. Indeed, because the Poincaré metric of the complement of the closure of the forward orbits of the critical point is expanded and it is equivalent to the spherical metric in any compact subset, $\bar{\mathbb{C}} \setminus V$ cannot contain an open set. Therefore we can use Lemma 1.4. to extend h_g continuously to $\bar{\mathbb{C}}$. □

Remark. One should compare the proof of this theorem with the proof of Theorem II.3.1. Instead of Lemma 1.4, we use there the simple fact that a monotone map of a dense subset of the interval whose image is a dense subset, extends continuously to a homeomorphism of the whole interval.

It is not known if the set of rational maps satisfying the above properties is dense in the space of rational maps. This is one of the main open problems in this theory. By the above theorem, a positive answer to this question would give a characterization of the structurally stable rational maps. From Mañé et al. (1983) and Sullivan and Thurston (1986), it follows that the set of structurally stable rational maps is dense.

The dynamics of a rational map on its Fatou set is completely described by an important theorem due to Sullivan. Since the Fatou set is completely invariant, the image of one of its connected components is another connected component. A connected component U of the Fatou set, also called a domain of f, is *periodic* if there is an integer n such that $f^n(U) = U$. It is called *wandering* if it is not eventually periodic. Julia and Fatou had already a complete understanding of the dynamics of f^n on U. There are five possibilities:

1. there is a hyperbolic attracting fixed point of f^n in U and U is contained in the basin of this fixed point;

2. there is a super-attracting fixed point of f^n in U and U is in the basin of this fixed point;

3. there exists a unique fixed point of f^n in the boundary of U, this fixed point is rationally indifferent, and U is in the basin of this fixed point;

4. the restriction of f^n to U is holomorphically conjugate to an irrational rotation $R_\lambda\colon \mathbb{D} \to \mathbb{D}$, $R_\lambda(z) = \lambda z$;

5. the restriction of f^n to U is holomorphically conjugate to an irrational rotation of an annulus: $R_\lambda\colon A_r \to A_r$, $R_\lambda(z) = \lambda z$, where $A_r = \{z\,;\; r < |z| < 1\}$.

That the possibilities 4) and 5) really do occur for rational maps was proved by Siegel and Herman, respectively. That the number of domains of type 1), 2), 3) and 4) is finite follows from Theorem 3.3. The non-existence of wandering domains as well as the finiteness of domains of type 5) was proved in Sullivan (1985) using the Measurable Riemann Mapping Theorem to perform quasiconformal deformations. More precisely:

Theorem 1.8. **(Sullivan)** *If $f\colon \bar{\mathbb{C}} \to \bar{\mathbb{C}}$ is a rational map then:*

1. *every domain is eventually periodic;*

2. *the number of periodic domains is finite.*

To prove the above theorem Sullivan introduced in this theory the idea of quasiconformal deformations that had been developed earlier in the theory of Kleinian groups. The main technical tool here is the Measurable Riemann Mapping Theorem. This also plays a fundamental role in the study of many dynamical properties of polynomials, see Douady (1984), Douady and Hubbard (1984), (1985) and Shishikura (1987).

In Chapter IV we will prove the analogue of this theorem in the real case. Although the proof of this theorem cannot be generalized to real dynamics we want to give a sketch of the main ideas involved because some of these tools are also important in the theory of real one-dimensional dynamics as we will see in the last chapter. The most important tool in this proof is the Measurable Riemann Mapping Theorem which allows one to construct deformations of conformal maps. This deformation theory is accomplished through changing the conformal structure so that it is still close to being conformal.

Before stating this generalized Riemann Mapping Theorem we need some definitions. For more detailed information on these definition, see the Appendix and for example Ahlfors (1987) and Lehto (1987).

Definition. Let $\phi\colon \Omega \to \Omega'$ be so that its partial derivatives $\bar{\partial}\phi(z)$ and $\partial\phi(z)$ are almost everywhere defined. Here

$$\partial\phi(z) = 1/2(\phi_x - i\phi_y) = 1/2(u_x + v_y) + i/2(v_x - u_y)$$

and

$$\bar{\partial}\phi(z) = 1/2(\phi_x + i\phi_y) = 1/2(u_x - v_y) + i/2(v_x + u_y)$$

where $w = \phi(z)$, $z = x+iy$ and $w = u+iv$. At each point where these derivatives are defined we define the *Beltrami coefficient* $\mu\colon \Omega \to \mathbb{C}$ of ϕ by

$$\mu_\phi(z) = \frac{\bar{\partial}\phi(z)}{\partial\phi(z)}.$$

Next we say that ϕ is *K-quasiconformal* if

a) $\phi\colon \Omega \to \Omega'$ is an orientation preserving homeomorphism between the open sets Ω and Ω';

b) the real $Re(\phi)$ and imaginary part $Im(\phi)$ of ϕ are absolutely continuous on almost all verticals and on almost all horizontals in the sense of Lebesgue;

c) there exists $k < 1$ such that for

$$\mu_\phi(z) = \frac{\bar{\partial}\phi(z)}{\partial\phi(z)}$$

one has

$$|\mu_\phi(z)| \le k \text{ for almost all } z \in \Omega$$

where

$$K \le \frac{1+k}{1-k}.$$

Remark. If ϕ is orientation preserving then $|\mu_\phi(z)| < 1$ almost everywhere. Furthermore, it is not hard to see that if ϕ is differentiable at z then $D\phi(z)$ sends the ellipses $r \cdot [e^{i\theta} + \mu_\phi(z)e^{-i\theta}]$, $r \ge 0$, to circles and the eccentricity (i.e., the ratio of the major to the minor axis of this ellipse) is equal to

$$D_\phi = \frac{1+|\mu_\phi(z)|}{1-|\mu_\phi(z)|}.$$

This will play an important role in Chapter VI.

Theorem 1.9. *If ϕ is 1-quasiconformal then it is conformal.*

Proof. See for example Ahlfors (1987) and Lehto (1987). □

The main analytic tool which is used in the proof of Sullivan's result mentioned above is the following

Theorem 1.10. **(Ahlfors-Bers)**

1. *For each measurable $\mu\colon \bar{\mathbb{C}} \to \{z\,;\ |z| < 1\}$ such that $||\mu|| \le k < 1$ there exists a unique $K = \frac{1+k}{1-k}$-quasiconformal map ϕ such that*

$$\mu_\phi = \mu \text{ a.e.}$$

such that $\phi(0) = 0$, $\phi(1) = 1$, $\phi(\infty) = \infty$.

2. *If for each $z \in \bar{\mathbb{C}}$, $\mu_t(z)$ depends analytically on a parameter $t \in \bar{\mathbb{C}}^l$ then $t \mapsto \phi_t(z)$ is also analytic for all $z \in \bar{\mathbb{C}}$.*

A *Beltrami coefficient* or a *complex structure* is a measurable mapping $z \to \mu(z)$ such that $|\mu(z)| < 1$ in almost every point. We call this a Beltrami coefficient or a complex structure because at every z for which $\mu(z)$ is defined, one gets the ellipses $r \cdot [e^{i\theta} + \mu(z)e^{-i\theta}]$, $r \geq 0$, which are based at z with eccentricity $\frac{1+|\mu(z)|}{1-|\mu(z)|}$. The previous theorem states that for any Beltrami coefficient μ with $||\mu|| < 1$, there exists a quasiconformal map ϕ which sends the Beltrami coefficient μ to the Beltrami coefficient 0. This means that in every point z where ϕ is differentiable, $D\phi(z)$ sends the ellipses $r \cdot [e^{i\theta} + \mu(z)e^{-i\theta}]$, $r \geq 0$ centred at z to circles.

Let us explain in a few words how this theorem is used to construct deformations of conformal maps which are conjugate to the original map. So let h be a quasiconformal conjugacy between two conformal maps f and g. Because f and g are conformal, any Beltrami coefficient is preserved by these maps. In particular, the Beltrami coefficient corresponding to the Beltrami coefficient of h is preserved by f. Indeed, consider the ellipses corresponding to $\mu(z)$ based at z. $Dh(z)$ sends these ellipses to circles and since g is conformal, $Dg \circ h(z)$ sends these ellipses to circles based in $g(h(z))$. By definition Dh^{-1} sends these circles to the ellipses corresponding to $\mu(h^{-1} \circ g \circ h(z)) = \mu(f(z))$. So the fact that f and g are conjugate implies that Df sends the ellipses corresponding to $h(z)$ to those corresponding to $h(f(z))$. More formally, because f and g are analytic,

$$\mu(f(z))\frac{\overline{f'(z)}}{f'(z)} = \frac{\bar{\partial}h\ (f(z))}{\partial h\ (f(z))}\frac{\overline{f'(z)}}{f'(z)} = \frac{\bar{\partial}h\ (z)}{\partial h\ (z)} = \mu(z).$$

So if we take $\mu_t = t \cdot \mu$ with $t \in \mathbb{C}$ and $|t| < 1$ we get another Beltrami coefficient which is also kept invariant by f, i.e.,

$$(*) \qquad \mu_t(f(z))\frac{\overline{f'(z)}}{f'(z)} = \mu_t(z).$$

We can use the Ahlfors-Bers Theorem to get a family of quasiconformal homeomorphisms h_t with Beltrami coefficient μ_t. Because of $(*)$, $h_t \circ f \circ h_t^{-1}$ is 1-quasiconformal (see also Lemma VI.4.1) and therefore conformal. Thus one gets an arc of conformal maps connecting f to g; each map in this arc is quasiconformally conjugate to f.

The idea behind Sullivan's Theorem is the following: if f has a wandering domain then one can use this to construct an arbitrarily high-dimensional space of conformal deformations of f. But since the space of rational maps of a certain degree is bounded, it follows that wandering domains cannot exist.

A similar result will be shown in the real case in Chapter IV. However, in that case no analogue of the Measurable Riemann Mapping Theorem is known and therefore different ideas will have to be used.

2 Structural Stability and Hyperbolicity

In this section we will introduce the important concept of hyperbolicity in real one-dimensional dynamics and prove that an endomorphism with 'enough' hyperbolicity in the relevant part of the dynamics is structurally stable. This result was first proved by Shub (1969) for maps of the circle without attracting periodic points (see Section II.2), by Nitecki (1970) for maps of the circle without critical points and finally by Jakobson (1971) for endomorphisms with critical points.

Here N will denote either a compact interval of the real line or of the unit circle. Let $C^r(N,N)$, $r \geq 1$, be the space of C^r endomorphisms of N such that, if the boundary ∂N of N is not empty, then it is invariant by the maps in $C^r(N,N)$. That is, if $f \in C^r(N,N)$ then $f(\partial N) \subset \partial N$ and therefore every $x \in \partial N$ is either x a fixed point, a periodic point of period 2 or $f(x)$ is a fixed point of f.

In $C^r(N,N)$ we consider the C^r-topology which is defined by the metric

$$d_r(f,g) = \max_{1 \leq i \leq r} \{d(f(x), g(x)), |D^i f(x) - D^i g(x)|\,;\ x \in N\}.$$

Here, d is the usual metric in N. A sequence of maps f_n converges to a map g in the C^r topology if and only if f_n converges uniformly to g together with all its derivatives up to the order r. It is easy to see that with the C^r metric, $C^r(N,N)$ is a complete metric space. Hence, by Baire's Theorem, a countable intersection of open and dense subsets of $C^r(N,N)$ is also a dense subset of $C^r(N,N)$.

Definition. An endomorphism $f \in C^r(N,N)$ is C^r *structurally stable* if there exists a neighbourhood $\mathcal{N}$ of f in $C^r(N,N)$ such that for each $g \in \mathcal{N}$, there exists a homeomorphism $h_g\colon N \to N$ which is a conjugacy between f and g, i.e., $g \circ h_g = h_g \circ f$. (As we shall see below, a C^2 endomorphism can be C^2-structurally stable without being C^1-structurally stable.)

One of the main open questions in one-dimensional dynamics is the characterization of the structurally stable endomorphisms in $C^r(N,N)$ for $r \geq 2$. We will define in this section a class of endomorphisms which are structurally stable. It is conjectured that this class is dense in the space of all endomorphisms and, therefore, coincides with the space of structurally stable endomorphisms. Before going into this, we will establish some properties which are necessary for the stability of endomorphisms.

2.a: Necessary conditions for structural stability

Note that a conjugacy h between two endomorphisms f and g, must map a turning point of f into a turning point of g and a periodic point of period n of f into a periodic point of period n of g. From this we get that a structurally stable map f must satisfy some conditions at each periodic point and at each critical point.

Definition. Let f be a periodic point of period n of $f \in C^r(N,N)$. If $0 \neq |Df^n(p)| \neq 1$ we say that p is a *hyperbolic* periodic point (the reason we do not allow $Df^n(p)$ to be zero in this definition is that the structural stability of f implies $Df^n(p) \neq 0$, see Corollary 2 below). It is an *attracting* periodic point if $0 < |Df^n(x)| < 1$ and a *repelling* periodic point if $|Df^n(p)| > 1$. If $Df^n(p) = 0$ then p is called a *super-attracting* periodic point. The basin of p is the set $B(p) = \{x\,;\ \omega(x) \in \{p, \dots, f^{n-1}(p)\}\}$. If p is a hyperbolic periodic point and f is not constant on any interval then $B(p)$ is an open set.

Definition. If c is a critical point of f, i.e., $Df(c) = 0$ then we say that c is *non-degenerate* if f is C^2 and $D^2f(c) \neq 0$.

Note that a non-degenerate critical point of a map is always a turning point, i.e., f is not monotone on any neighbourhood of the critical point. Let us prove that for 'most' maps all periodic points are hyperbolic and all critical points are non-degenerate.

Proposition 2.1. *Take $f \in C^r(N,N)$ and let $r \geq 1$. Then there exists a map $g \in C^r(N,N)$ arbitrarily close to f in the C^r topology, such that all periodic orbits of g are hyperbolic. If, furthermore, $r \geq 2$, then all critical points of g are non-degenerate.*

Proof. The space $C^r(N,N)$ is a Baire space. Therefore, the proposition follows from the following claims:

1) Let $\mathcal{U}_n \subset C^r(N,N)$, $r \geq 1$, be the space of maps whose periodic points of period smaller or equal to n are all hyperbolic. Then $\mathcal{U}_n$ is an open and dense subset of $C^r(N,N)$. Furthermore, given $f \in \mathcal{U}_n$, there exists a neighbourhood $\mathcal{N} \subset \mathcal{U}_n$ of f and continuous functions $p_i \colon \mathcal{N} \to N, i = 1, \dots, k_n$, such that $\{p_i(g)\,;\ i = 1, \dots, k_n\}$ is the set of periodic points of g of period smaller or equal to n.

2) Let $\mathcal{U} \subset C^r(N,N)$, $r \geq 2$, be the space of maps whose critical points are all non-degenerate. Then $\mathcal{U}$ is an open and dense subset of $C^r(N,N)$. Furthermore, given $f \in \mathcal{U}$, there exists a neighbourhood $\mathcal{N} \subset \mathcal{U}$ of f and continuous functions $c_i \colon \mathcal{N} \to N$, $i = 1, \dots, d$, such that $\{c_i(g)\,;\ i = 1, \dots, d\}$ is the set of critical points of g.

In order to prove these claims notice that if the graph of f^i is transversal to the diagonal Δ of $N \times N$ for all $i \leq 2n$ then all the periodic points of f of period $\leq n$ are hyperbolic and, by the Implicit Function Theorem they depend continuously on the map. So the openness of $\mathcal{U}_n$ follows from the Implicit Function Theorem. From transversality techniques the density of $\mathcal{U}_n$ follows, see the Appendix. In fact, this statement is also true for endomorphisms of compact manifold of any dimension. The second claim also follows from transversality techniques, because the critical points of f are nondegenerate if and only if the map $x \mapsto (x, Df(x))$, from N to $N \times \mathbb{R}$ is transversal to $N \times \{0\}$. The details are left to the reader. □

In the next two corollaries we will show that the two generic properties from the previous proposition are satisfied for every structurally stable map.

Corollary 2.1. *Assume that $f \in C^r(N, N)$, $r \geq 1$, is structurally stable then each periodic orbit of f is hyperbolic (and in particular f has only a finite number of periodic points of each period).*

Proof. Let $\mathcal{N}$ be a neighbourhood of f such that any map in $\mathcal{N}$ is conjugate to f. By Proposition 2.1, there exists a map $g \in \mathcal{N}$ such that all periodic points of g are hyperbolic and g has a finite number of turning points which are non-degenerate. Hence, the number of periodic points of g of any given period is finite. Since turning points and periodic points are preserved by conjugacies, all maps in $\mathcal{N}$ have the same number of turning points and the same number of periodic points of period n for each integer n. In particular, the number of periodic points of a given period and the number of turning points of f is finite.

Suppose, by contradiction that f has a periodic point p of period n such that $|Df^n(p)| = 1$. Let V be a neighbourhood of p such that f has no periodic point of period $2n$ in V and, furthermore, has no periodic point of period n in $V \setminus \{p\}$. We will show that $|Df^n(p)| = 1$ implies that there exists $g \in \mathcal{N}$ such that the number of periodic points of period $\leq 2n$ of f and g is different. Let Φ be a positive C^∞ function which is equal to 1 near p and equal to 0 outside V.

First suppose that $Df^n(p) = 1$ and that p is attracting (resp. repelling) from both sides. Take $g(x) = f(x) + \epsilon\Phi(x) \cdot (x - p)$. Then $g^n(p) = p$ and $Dg^n(p) > 1$ or $Dg^n(p) < 1$ depending on the sign of ϵ. Therefore, one can find $g \in \mathcal{N}$ so that g^n has at least three fixed points in V, see Figure 2.1. It follows that f and g have a different number of periodic points of period n, a contradiction. Next suppose that $Df^n(p) = 1$ and that p is attracting from one side and take $g(x) = f(x) + \epsilon\Phi(x)$. Then $Dg^n(p) = 1$ and $g^n(p) > p$ or $g^n(p) < p$ depending on the sign of ϵ. Therefore one can find $g \in \mathcal{N}$ so that g^n has at least two fixed points in V, see Figure 2.1. Again this gives a contradiction. Similarly, if $Df^n(p) = -1$ then $Df^{2n}(p) = 1$ and p is either attracting or repelling from both sides (this holds because f maps one component of $V \setminus \{p\}$ to the other). So perturb f to $g(x) = f(x) + \epsilon\Phi(x) \cdot (x - p)$ as in the previous case. Again the number of periodic points of g and f of period $\leq 2n$ is different for an appropriate choice of ϵ.

We should remark that if $r \geq 3$ then we can choose perturbations of a special form, see the exercises at the end of the next section. In this case these perturbations give rise to generic bifurcations. □

Remark. The above corollary is also true for endomorphisms of higher-dimensional manifolds, see Shub (1969).

Fig. 2.1: The number of periodic points of f and g from the proof of Corollary 1 differs.

The next corollary motivates us to consider structurally stable maps in $C^r(N,N)$ only for $r \geq 2$.

Corollary 2.2. *If $f \in C^1(N,N)$ is structurally stable then f has no critical points. If $f \in C^r(N,N)$, $r \geq 2$, is structurally stable then all critical points of f are non-degenerate.*

Proof. Let $\mathcal{N}$ be a neighbourhood of f such that each map in $\mathcal{N}$ is conjugate to f. By Proposition 2.1, there exists a map $g_1 \in \mathcal{N}$ such that g_1 has a finite number of turning points. Since turning points are preserved by conjugacies, all maps in $\mathcal{N}$ have a finite number of turning points. On the other hand, if f has a critical point, it is easy to see that we can approximate f in the C^1-topology by a map g_2 with an infinite number of turning points (here it is important that we use the C^1 topology).

Furthermore, if f has a degenerate critical point c then arbitrarily C^r-close to f one can find a map g which has a larger number of turning points than f. Again we get a contradiction. □

2.b: Hyperbolicity and the Axiom A condition

The previous corollary shows that hyperbolicity of periodic orbits and non-degeneracy of critical points is necessary for structural stability. Now we shall introduce the same notion of hyperbolicity for general forward invariant sets K. This notion formalizes when f is uniformly expanding along K. It will turn out in Theorem 2.3 below that hyperbolicity of the complement of the basins of periodic attractors and the necessary conditions from above are also sufficient for structural stability.

Definition. Let $f\colon N \to N$ be a C^r map, $r \geq 1$, where N is either the circle or a compact interval of the real line. A subset $K \subset N$ is a *hyperbolic repelling* (or for short, *hyperbolic*) set of f if K is forward invariant and there exist constants $C > 0$ and $\lambda > 1$ such that

$$|Df^n(x)| > C\lambda^n$$

for all $x \in K$ and all $n \in \mathbb{N}$. Notice that a periodic orbit p of period n was called hyperbolic if $0 < |Df^n(p)| \neq 1$; in particular, such periodic orbits are allowed to be attracting. However, it is unnecessary to define this notion for more general sets because we will see in Section 5 and Chapter IV that if $|Df^n(x)| < \frac{1}{C}\lambda^{-n}$ for all $n \in \mathbb{N}$, $x \in K$, then all points in K are attracted to a finite union of attracting hyperbolic periodic orbits.

Lemma 2.1. *Let $K \subset N$ be a compact invariant set of a C^1 map $f\colon N \to N$. Then K is a hyperbolic set if and only if for each $x \in K$ there exists an integer $n = n(x)$ such that $|Df^n(x)| > 1$.*

Proof. If K is hyperbolic we take n so that $C\lambda^n > 1$. So let us prove the reverse implication. So suppose that $|Df^{n(x)}(x)| > 1$ for every $x \in K$. By compactness of K and continuity of the derivative of f, there exists a finite cover $V_1, \dots, V_k$ of K by open sets, integers $n_1, \dots, n_k$ and numbers $\lambda_1, \dots, \lambda_k > 1$, such that $|Df^{n_i}(x)| > \lambda_i$ for every $x \in V_i$ and every $i = 1, \dots, k$. Let $n_0 = \max\{n_i\,;\ 0 < i \leq k\}$, $\lambda_0 = \min\{\lambda_i\,;\ 0 < i \leq k\}$ and $a = \min\{|Df(x)|\,;\ x \in K\}$. Choose an integer m so big that $\lambda_0^m \cdot a^{n_0} > 1$. Let $\tilde{n} = (m+1) \cdot n_0$.

We claim that $|Df^{\tilde{n}}(x)| > 1$ for every $x \in K$. Indeed, choose $i_i \in \{1, \dots, k\}$ so that $x \in V_{i_1}$, i_2 so that $f^{n_{i_1}}(x) \in V_{i_2}$, i_3 so that $f^{n_{i_1}+n_{i_2}}(x) \in V_{i_3}$ and so on. Since $\tilde{n} > m \cdot n_0$, there exist integers $s = s(x) \geq m$ and $m_0 < n_0$ such that $\tilde{n} = n_{i_1} + \cdots + n_{i_s} + m_0$. Therefore,

$$|Df^{\tilde{n}}(x)| \geq \lambda_{i_1} \times \cdots \times \lambda_{i_s} \times |Df^{m_0}(f^{n_{i_1}+\cdots+n_{i_s}}(x))| \geq \lambda_0^m \cdot a^{m_0} > 1$$

which proves the claim.

From the claim and the compactness of K we get the existence of $\lambda > 1$ such that $|Df^{\tilde{n}}(x)| > \lambda^{\tilde{n}}$ for every $x \in K$. Now, taking $C = \min\{\dfrac{a^i}{\lambda^i}\,;\ 1 \leq i \leq \tilde{n}\}$ we get $|Df^i(x)| \geq C\lambda^i$ if $i \leq \tilde{n}$. If n is any integer we can write $n = s\tilde{n} + t$ with $t < \tilde{n}$ and we get

$$|Df^n(x)| = |Df^{s\tilde{n}}(x)| \cdot |Df^t(f^{s\tilde{n}}(x))| \geq \lambda^{s\tilde{n}} \cdot C \cdot \lambda^t = C\lambda^n. \quad \square$$

Now we define when a map satisfies the so-called Axiom A conditions. This notion will play an essential role throughout the remainder of this chapter.

Definition. Definition We say that a map $f \in C^r(N,N)$, $r \geq 1$, satisfies the *Axiom A* conditions if:

1. f has a finite number of hyperbolic periodic attractors;
2. $\Sigma(f) = N \setminus B(f)$ is a hyperbolic set, where $B(f)$ is the union of the basins of the hyperbolic attractors of f.

We should point out that this definition is slightly different from the usual definition for diffeomorphisms on manifolds. The reason one considers the set $\Sigma(f)$ rather than the non-wandering set $\Omega(f)$ is the fact that $\Omega(f)$ is not necessarily backward invariant for non-invertible maps f, see Exercise 2.1 below. If f satisfies the Axiom A, then all critical points of f are in the basins of hyperbolic periodic attractors and all periodic points in $\Sigma(f)$ are hyperbolic repellors. Furthermore, if $J \subset \Sigma(f)$ is an interval then, since $\Sigma(f)$ is completely invariant, all the iterates $f^n(J)$ also belong to $\Sigma(f)$. Since $|Df^n(x)| > C\lambda^n$ for all $x \in J$ and $\lambda > 1$ it follows that $\Sigma(f) = N = S^1$ and $f^n(J) = S^1$ for n big enough. Therefore, either $\Sigma(f)$ is totally disconnected or f is an expanding map of the circle. Note that this argument also shows that Axiom A maps do not have wandering intervals.

Exercise 2.1. Assume that $f\colon [-1,1] \to [-1,1]$ is C^∞ and satisfies the following properties. $\Omega(f)$ is hyperbolic, all the critical points of f are non-degenerate, f has a finite number of critical points and the orbits of the critical points of f are disjoint. Show that these conditions do not imply that f is structurally stable. (Hint: take $f\colon [-1,1] \to [-1,1]$ such that $f[-1,0] \subset [0,1]$, $f(0) = 1$ and $f[0,1] \subset [0,1]$ with a turning point in $(-1,0)$ which is mapped onto a repelling periodic orbit of $f\colon [0,1] \to [0,1]$ and note that $\Omega(f) = \Omega(f|[0,1])$.) It will be shown in Theorem 2.3 that f is structurally stable if in addition $\Sigma(f)$ is hyperbolic.

Exercise 2.2. Show that if $f\colon N \to N$ satisfies the Axiom A conditions, then the periodic points of f are dense in the non-wandering set. (Hint: as remarked above we may assume that $\Sigma(f)$ is totally disconnected because otherwise $N = S^1$ and f is an expanding map of the circle. So take a non-wandering point x which is not periodic. Then there exists a sequence $n(k) \to \infty$ and a sequence of points $x_k \to x$ with $f^{n(k)}(x_k) \to x$. Now let $J_{n(k)}$ be the component containing x_k of the set $\{y\,;\ f^i(y) \notin B_0(f) \text{ for all } 0 \leq i < n(k)\}$ (since x is non-wandering this component is non-empty). Here $B_0(f)$ is the union of the immediate basins of periodic attractors. By the forward invariance of $B_0(f)$,

$f^{n(k)}(J_{n(k)})$ is a component of $N \setminus B_0(f)$. Since $B_0(f)$ only has a finite number of components this implies that $f^{n(k)}(J_{n(k)}) \supset J_{n(k)}$ for sufficiently large k. Since by assumption Σ contains no intervals, $|J_{n(k)}| \to 0$ and it follows that arbitrarily close to x there are periodic points.)

Exercise 2.3. Show that the periodic points of a C^∞ map $f\colon [0,1] \to [0,1]$ need not be dense in $\Omega(f)$. (Hint: choose $0 < a < b < c < 1$ and let f be so that $f(a) = f(b) = f(c) = c$, $f[0,a] \subset [a,c]$, $f[a,b] \subset [c,1]$, $f[b,c] = [a,c]$ and $f[c,1] \subset [c,1]$, see Figure 2.2. Then no point in $[0,b]$ is periodic and yet a is non-wandering.) One can also show that the Axiom A conditions defined above exclude the existence of cycles. (We should note that L.S. Young (1979) has shown that for each $r \geq 0$, any C^r endomorphism can be C^r approximated by an endomorphism for which the periodic orbits are dense. Even for $r = 1$ the corresponding result is much more difficult in higher dimensions. In fact, for diffeomorphisms in higher dimensions this property has only been proved for $r = 1$: it follows from Pugh's closing lemma.)

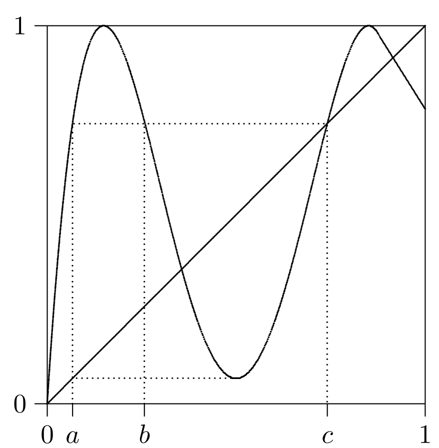

Fig. 2.2: The point a is a non-wandering point of the map $f\colon [0,1] \to [0,1]$ from Exercise 2.3, but there are no periodic points in $[0,b]$. This shows that the periodic points are in general not dense in the non-wandering set.

2.c: The density of Axiom A

In contrast to the situation in higher dimensions, Axiom A maps occur commonly in one-dimensional dynamical systems. For example, in the next section, we will prove that any map with negative Schwarzian derivative and whose critical points are in the basin of hyperbolic periodic attractors, satisfies the Axiom A. This result of Misiurewicz was generalized by Mañé (1985) for general C^2 maps. He proved that any C^2 map whose periodic points are hyperbolic and whose critical points are in the basins of periodic attractors satisfies the Axiom A. More precisely, Mañé proved the following remarkable theorem. Here we let

$C(f)$ be the set of critical points of a map f, i.e.,

$$C(f) = \{x\,;\ Df(x) = 0\}.$$

Moreover, let $B_0(f)$ be the union of the immediate basins of the periodic attractors of f.

Theorem 2.1. **(Mañé)** *Let N be a compact interval of the real line or the circle and let $f\colon N \to N$ be a C^2 map. Let U be a neighbourhood of the set $C(f)$ of critical points of f. Then*

1. *All periodic orbits of f contained in $N \setminus U$ of sufficiently large period are hyperbolic and repelling.*
2. *If all periodic orbits of f which are contained in $N \setminus U$ are hyperbolic, then there exists $C > 0$ and $\lambda > 1$ such that*

$$|Df^n(x)| \geq C\lambda^n$$

whenever $f^i(x) \in N \setminus (U \cup B_0(f))$ for all $0 \leq i \leq n-1$.

In particular, if all critical points of f are contained in the basin of periodic attractors and all periodic points of f are hyperbolic then f is Axiom A.

Proof. We will give a proof of this result in Section 5 of this chapter.

Furthermore, Jakobson (1981) has shown the following:

Theorem 2.2. **(Jakobson)** *The set of maps satisfying the Axiom A is dense in $C^1(N, N)$.*

Proof. For maps having no critical points this result follows immediately from Proposition 2.1 and Theorem 2.1 (each C^1 map can be approximated by a C^2 map in the C^1 topology). The proof of this result for maps with critical points is sketched in Exercises 2.4-2.7 below. □The density of Axiom A

in the space $C^r(N, N)$, $r \geq 2$, is one of the most important open questions in one-dimensional dynamics. By the result of Mañé this density property would follow from

Conjecture 1. The set of maps in $C^r(N, N)$, $r \geq 2$, whose critical points are in the basin of hyperbolic attractors forms a dense set.

This is still an open question even in the space of maps with negative Schwarzian derivative. Recently the proof of the following conjecture was announced by Świątek (1992b).

Conjecture 2. The set of parameters a for which the critical point of $f_a(x) = ax(1-x)$, $0 < a \leq 4$ is in the basin of hyperbolic attractors, is dense.

In the C^1 topology the situation is much easier, as we will see in the next exercises. Below, in Section 2.D, we shall show that the sets of maps (or parameters) from the previous conjectures are certainly open.

Exercise 2.4 (A C^1 closing lemma for interval maps)**.** Let $f \in C^r(N,N)$, $r \geq 1$, have a finite number of critical points and suppose that c is a critical point which is recurrent, i.e., $c \in \omega(c)$. Show that C^1 close to f there exists a C^r map g such that g coincides with f outside a neighbourhood of c, has a unique critical point c' near c and c' is in the basin of an attracting periodic point of g. Whether one can choose g to be C^2 close to f is unknown. (Hint: Let n be an integer so that $f^n(c)$ is very close to c and so that $d(f^k(c),c) > d(f^n(c),c)$ for $0 < k < n$. Let I be the interval connecting $f^n(c)$ and c. Since c is recurrent one can choose n such that I is arbitrarily small and such that there exists an integer $m > n$ with $f^m(c) \in I$. Choose m minimal with this property. Now distinguish two cases. i) $f^m(c)$ is closer to c, i.e., $d(f^m(c),c) \leq d(f^m(c),f^n(c))$. Then let z between $f^m(c)$ and $f^n(c)$ be such that $d(f^m(c),z) = d(z,f^n(c))$ and let U be a small neighbourhood of the segment connecting c and z such that $f^k(c) \notin U$ for all $0 < k < m$. Modify f on U to a C^r map g as follows. Let $\hat{c} = f^m(c)$. Choose g such that $g(\hat{c})$ is equal to $f(c)$ and such that $g|U$ has a unique critical point in $\hat{c}$. It is not hard to see that g can be chosen arbitrarily C^1 close to f provided we choose $f^n(c)$ sufficiently close to c and therefore $|Df|$ is sufficiently close to 0 on U. Moreover, since g and f coincide outside U, $g^m(\hat{c}) = g^{m-1}(f(c)) = f^m(c) = \hat{c}$. So $\hat{c}$ is the unique critical point of g in I and it is periodic. ii) $f^m(c)$ is closer to $f^n(c)$, i.e., $d(f^m(c),f^n(c)) \leq d(f^m(c),c)$. Then let z between c and $f^m(c)$ be such that $d(c,z) = d(z,f^m(c))$ and let U be a small neighbourhood of the segment connecting z and $f^n(c)$ such that $f^k(c) \notin U$ for all $0 < k < m$ with $k \neq n$. Modify f on U to a C^r map g as follows. Let $\hat{c} = f^m(c)$ and choose g such that Dg is almost zero on the segment between $f^n(c)$ and $f^m(c)$ and such that $g(f^n(c)) = f(f^n(c))$. Then $g^{m-n}(U)$ is almost equal to $g^{m-n-1}(f^{n+1}(c)) = f^m(c) = \hat{c}$. It follows that if we take $|Dg|$ sufficiently small on U then g has an attracting fixed point p of period $m-n$ on U near $\hat{c}$. Moreover g has the same critical points as f and c is in the basin of p. It is not hard to see that g can be chosen arbitrarily C^1 close to f provided we choose $f^n(c)$ sufficiently close to c because then $|Df|$ is sufficiently close to 0 on U.)

Exercise 2.5. Let $f \in C^r(N,N)$, $r \geq 1$ with a finite number of critical points and suppose $\omega(c)$ contains a critical point $d \neq c$. Show that C^1 close to f there exists a C^r map g such that g coincides with f outside a neighbourhood of d, has a unique critical point d' near d and either $f^n(c) = d$ for some n or $c \in B(g)$. (Hint: use the same ideas as in Exercise 2.4.)

Exercise 2.6. a) Let $f \in C^r(N,N)$ with $r \geq 2$ and suppose that f has no critical points and no neutral periodic points. (So $|Df^n(p)| \neq 1$ when p is a periodic point with period n.) Show that f is Axiom A. (Hint: if N is an interval this is trivial. If $N = S^1$ it follows immediately from Mañé's result that $N \setminus B(f)$ is hyperbolic.) b) Let $f \in C^r(N,N)$ with $r \geq 2$ and suppose that f has no critical points. Show that there exists C^r close to f a C^r Axiom A map g. (Hint: by Proposition 2.1 there exists C^r close to f a C^r map g whose periodic points are all hyperbolic. So the claim follows from part a) of this exercise.)

Exercise 2.7. In this exercise we will see that if $f \in C^r(N,N)$ with $r \geq 2$ then C^1 close to f there exists a C^r map g such that all its critical points are contained in the basin of periodic attractors. The proof proceeds in a three steps. a) Let $f \in C^r(N,N)$ with $r \geq 2$. Assume that f has a critical point c, all periodic points of f are hyperbolic and that $\omega(c) \cap C(f) = \emptyset$. Show that in this case, there exists C^r near to f a C^r map g which coincides with f except near c such that for some integer $n > 0$ one has $g^n(c) \in C(g) \cup B(g)$. (Hint: if this were not the case then there would exist a neighbourhood U of $f(c)$ such that no point in U is eventually mapped into $C(g) \cup B(g)$. It follows in particular that U would be a homterval (i.e., all iterates of f are monotone on U) and that no point in U would be in the basin of a periodic attractor. As we have proved in Lemma II.3.1 this would imply that U is a wandering interval. In particular the length of $f^n(U)$ would tend to zero. However, $\omega(c) \cap C(f) = \emptyset$ and since $f(c) \in U$ this implies that $f^n(U)$ does not accumulate on $C(f)$. But this would contradict the Theorem of Schwartz from Chapter I, more precisely it would contradict Corollary 2 to Theorem I.2.2.) b) Let $f \in C^r(N,N)$ with $r \geq 2$. Assume that f has a critical point c and that all periodic points of f are hyperbolic. Show that there exists map g such that all critical points of g are contained in the basins of periodic attractors. (Hint: from a) we may assume that for each critical point c, either $\omega(c) \subset C(f)$ or $f^n(c) \in C(f) \cup B(f)$ for some $n > 0$. Then, using Exercises 2.4 and 2.5, we can find C^1 close to f a C^r map g such that for each critical point of g, $g^n(c) \in C(g) \cup B(g)$ for some $n > 0$. Since this holds for each critical point there are two possibilities. Firstly, there exists a sequence $k(i) \to \infty$ with $g^{k(i)}(c) \in C(g)$. Because $C(g)$ consists of only a finite number of points, c is eventually mapped onto a periodic (super attracting) point of $C(g)$ in this case. Secondly, there exists $n \geq 0$ with $g^n(c) \in C(g)$ and $g^k(c) \notin C(g)$ for $k > n$ and therefore $g^n(c) \in B(g)$.) c) Prove that C^1 close to f there exists a map g such that all critical points of g are contained in the basins of periodic attractors. (Hint: simply use b) and Proposition 2.1.)

2.d: Axiom A implies stability

The main reason for defining the notion of hyperbolic sets is that they are persistent under perturbations.

Theorem 2.3 (Stability of hyperbolic sets). *Suppose that K is a compact hyperbolic invariant set for some map $f \in C^r(N,N)$, $r \geq 1$. Then there exists a neighbourhood $\mathcal{N}$ of f such that for each $g \in \mathcal{N}$ there exists a compact hyperbolic invariant set $K(g)$ for g and a homeomorphism $h_g \colon K \to K(g)$ conjugating f and g.*

Proof. By changing the metric in N we can assume that the constant C of the definition of hyperbolicity of K is equal to one, see Lemma III.1.3. Using this and the compactness of K, we get an open neighbourhood V of K, a neighbourhood $\mathcal{N}$ of f and a constant $\lambda > 1$ such that

$$(*) \qquad |Dg(x)| > \lambda \text{ for all } x \in V \text{ and all } g \in \mathcal{N}.$$

Suppose that $\epsilon > 0$ is so small that V contains a 2ϵ neighbourhood of K. From $(*)$ we certainly have that the set of points $\tilde{K}(g)$ of points y such that $g^n(y) \in V$ for all $n \geq 0$ is hyperbolic. Let us now show that there exists a subset $K(g)$

of $\tilde{K}(g)$ with the required properties. Again we conclude from $(*)$, if $g \in \mathcal{N}$ then $g(B(x, \frac{t}{\lambda})) \supset B(f(x), t)$ for each $t \in (0, \epsilon]$ and each $x \in K$. Since K is f-invariant, for each $x \in K$ the set

$$I_n(x) = \{y \in B(x, \epsilon)\,;\; g^i(y) \in B(f^i(x), \epsilon),\ \text{for all } 0 \le i \le n\}$$

is a non-empty interval with length smaller than $(\frac{1}{\lambda})^n$. From this we get the existence of a unique point $h(x)$ such that $g^n(h(x)) \in B(f^n(x), \epsilon)$ for every positive integer n. From the definition it follows immediately that $h(f(x)) = g(h(x))$. Also $h(K) \subset \tilde{K}(g)$ because the forward orbit of g through the point $h(x)$ is contained in an ϵ neighbourhood of K and therefore in V. We claim that h is one-to-one. Indeed, if $h(x) = h(y)$ then $f^n(y) \in B(f^n(x), 2\epsilon)$ for every positive integer n. By the choice of ϵ and since $f^n(x) \in K$ for all $n \in \mathbb{N}$, it would follow that the f-iterates of the interval bounded by x and y remain in V. This is not possible because f is expanding in V. So simply take $K(g) = h(K)$. Then h is a conjugacy between f and g on K. □

The next theorem shows that the Axiom A condition defined above plays the same role as the Axiom A and no-cycle condition for diffeomorphisms on higher-dimensional manifolds.

Theorem 2.4 (Σ-stability of Axiom A maps). *For each $f \in C^r(N, N)$, $r \ge 1$, satisfying the Axiom A there exists a neighbourhood $\mathcal{N}$ of f such that each $g \in \mathcal{N}$ satisfies the Axiom A and there exists a homeomorphism $h_g\colon \Sigma(g) \to \Sigma(f)$ conjugating f and g. Furthermore, h_g depends continuously on g in the C^0-topology.*

Proof. If $\Sigma(f)$ is not totally disconnected then f is an expanding map of the circle and the result was already proven in Section II.2. So we may assume that f has some attracting periodic point. By changing the metric in N we can assume that the constant C of the definition of hyperbolicity of $\Sigma(f)$ is equal to one, see Lemma III.1.3. Using this and the compactness of $\Sigma(f)$, we get an open neighbourhood W of $\Sigma(f)$, a neighbourhood $\mathcal{N}_1$ of f and a constant $\lambda > 1$ such that $|Dg(x)| > \lambda$ for all $x \in W$ and all $g \in \mathcal{N}_1$.

Let $\{p_i\,;\; i = 1, \dots, k\}$ be the set of hyperbolic attracting periodic points of f and let n_i be the period of p_i. Since $|Df^{n_i}(p_i)| < 1$, there exist an open interval U_i containing p_i such that its closure $\bar{U}_i$ is contained in the immediate basin of p_i, and $f^{n_i}(\bar{U}_i) \subset U_i$. By the Implicit Function Theorem there exist a neighbourhood $\mathcal{N}_2 \subset \mathcal{N}_1$ of f and continuous functions $p_i\colon \mathcal{N}_2 \to U_i$ such that $p_i(f) = p_i$ and $p_i(g)$ is a hyperbolic periodic attractor of g of period n_i if $g \in \mathcal{N}_2$. Shrinking $\mathcal{N}_2$ if necessary we can assume that $\bar{U}_i$ is in the immediate basin of the periodic attractor $p_i(g)$ and $g^{n_i}(\bar{U}_i) \subset U_i$. Let $U = \cup U_i$. From the construction it follows that if $g \in \mathcal{N}_2$ and $g^n(x) \in U$ then x is in the basin of one of the periodic attractors $p_i(g)$.

Let V be a neighbourhood of $\Sigma(f)$ such that $\bar{V} \subset W$. Let $\epsilon > 0$ such that for each $x \in \bar{V}$, the ball of centre x and radius 2ϵ in N is contained in W. For each x in the compact set $N \setminus V$, there exists a neighbourhood U_x of x, a neighbourhood $\mathcal{N}_x \subset \mathcal{N}_2$ of f and an integer n_x such that $g^{n_x}(U_x) \subset U$ if $g \in \mathcal{N}_x$. By taking a finite cover of the compact set $N \setminus V$ and intersecting the corresponding neighbourhoods of f we get a neighbourhood $\mathcal{N}_3 \subset \mathcal{N}_2$ of f and an integer n_0 such that for every $x \in N \setminus V$ and $g \in \mathcal{N}_3$, $g^{n_0}(x) \in U$. Hence the complement of the union $B(g)$ of the basins of the hyperbolic attractors $p_i(g)$ is contained in V and since $|Dg(x)| > \lambda > 1$ for $x \in W$, $\Sigma(g) = N \setminus B(g) \subset W$ is hyperbolic. This proves the first part of the theorem. Notice that $\Sigma(g)$ is equal to the set of points of V whose positive orbits remain in V.

Exactly as in the proof of the previous theorem there exist $\epsilon > 0$ and a neighbourhood $\mathcal{N} \subset \mathcal{N}_3$ of f such that for $g_1, g_2 \in \mathcal{N}$ there exists a unique point $h_{g_1}(x)$ such that $g_2^n(h_{g_1}(x)) \in B(g_1^n(x), \epsilon)$ for every positive integer n. Moreover, h_{g_1} is one-to-one, $h_{g_1}(g_1(x)) = g_2(h_{g_1}(x))$ and $h_{g_1}(\Sigma(g_1)) \subset \Sigma(g_2)$. Finally, h_{g_1} is a homeomorphism from $\Sigma(g_1)$ to $\Sigma(g_2)$ since $h_{g_2}(h_{g_1}(x)) = x$ for every $x \in \Sigma(g_1)$. □

Thus we have shown that the set Σ corresponding to an Axiom A map is stable under small perturbations. In the next result we will show that the Axiom A condition and an additional condition is sufficient for structural stability. This additional condition was shown to be necessary for structural stability in Corollary 2 of Proposition 2.1. These additional conditions can be regarded as the analogue of the condition on the transversality of invariant manifolds condition for diffeomorphisms on higher-dimensional manifolds.

Theorem 2.5 (Structural stability of Axiom A maps without cycles). *Let $f \in C^r(N, N)$, $r \geq 2$, satisfy the Axiom A together with the following conditions:*

i) *the critical points of f are non-degenerate;*

ii) *if c_1, c_2 are critical points and $f^n(c_1) = f^m(c_2)$ then $n = m$ and $c_1 = c_2$.*

Then f is structurally stable.

Proof. If f has no critical points then it is an expanding map of the circle and structurally stable, by the Corollary of Theorem II.2.1.

Suppose now that f has a critical point. Because $\Sigma(f) = N \setminus B(f)$ is hyperbolic this implies that each critical point must be contained in $B(f)$; in particular, the set of attracting periodic orbits of f is non-empty. Choose a point p_i, $i = 1, \ldots, k$ in each hyperbolic attracting periodic orbit of f and let n_i be the period of p_i. Let $c_1, \ldots, c_d$ be the critical points of f. Choose compact intervals V_i in the basin of p_i such that $f^{n_i}(V_i) \subset \operatorname{int}(V_i)$, f^{n_i} is monotone in V_i and such that the boundary of V_i does not contain points in the orbit of a critical point of f. The compact set $D_i = V_i \setminus \operatorname{int}(f^{n_i}(V_i))$ is a union of two intervals and it is a fundamental domain for the basin of p_i in the sense that each orbit in the basin of p_i has at least one and at most two points in D_i (the last situation occurs if and only if the orbit hits the boundary of D_i). In particular, for every critical point c_j, there exist a positive integer m_j and $i(j) \in \{1, \dots, k\}$ such that $f^{m_j}(c_j) \in D_{i(j)}$.

Let $\mathcal{N}$ be a neighbourhood of f such that there exist continuous functions $p_i : \mathcal{N} \to N, i = 1, \dots, k$ and $c_j : \mathcal{N} \to N, j = 1, \dots, d$ such that for each $g \in \mathcal{N}$:

1. $c_j(g)$ is a critical point of g and $c_j(f) = c_j$, for $j = 1, \dots, d$;
2. $p_i(g)$ is the only hyperbolic attracting periodic point of period n_i for g in V_i and $p_i(f) = p_i$;
3. $g^{n_i}(V_i) \subset \mathrm{int}\,(V_i)$, g^{n_i} is monotone in V_i and $g^{m_j}(c_j)$ belongs to the interior of $D_{i(j)}(g) = V_{i(j)} \setminus \mathrm{int}\,(g^{n_i}(V_{i(j)}))$;
4. the points $c_j(g), \dots, g^{m_j}(c_j(g))$, $j = 1, \dots, d$ are all distinct and
5. g satisfies the Axiom A and each forward invariant set in $N \setminus \cup_i V_i$ is hyperbolic (and repelling).

From Properties 4 and 5 it follows that the itinerary of each critical point of $g \in \mathcal{N}$ is the same as the itinerary of the corresponding critical point of f. From Property 2, 3 and 5 it follows that f and g have the same number of periodic attractors. By Properties 3 and 4, there exists an orientation preserving homeomorphism

$$h : \left(\bigcup_{j=1}^{l} \bigcup_{i \geq 0} f^i(c_j) \right) \bigcup B_0(f) \quad \to \quad \left(\bigcup_{j=1}^{l} \bigcup_{i \geq 0} g^i(c_j(g)) \right) \bigcup B_0(g)$$

such that $h(c_i) = c_i(g)$ and $g \circ h = h \circ f$. Moreover, since f and g are Axiom A, these maps do not have wandering intervals. It follows from all this and Theorem II.3.1 that f and g are topologically conjugate. $\square$

2.e: The measure of hyperbolic invariant sets

Now we will show that a hyperbolic compact forward invariant set has either Lebesgue measure zero or is equal to N. In the first section of Chapter V we will come back to this issue in much greater generality. In order to prove this we need the following concept.

We say that a map f is of class C^s where $s = k + \alpha$, $k \in \mathbb{N}$ and $\alpha \in [0, 1)$ when f is C^k and its k-th derivative satisfies a Hölder condition of order α, namely

$$|D^k f(x) - D^k f(y)| \leq C|x - y|^\alpha,$$

where C is a positive constant.

Theorem 2.6 (Hyperbolic sets have zero or full Lebesgue measure). *Let $f\colon N \to N$ be a $C^{1+\alpha}$ map with $\alpha > 0$. If $\Gamma \subset N$ is a compact, forward-invariant, hyperbolic set for f then either $\Gamma = N = S^1$ (and f is an immersion of the circle) or Γ has Lebesgue measure equal to zero.*

Proof. Since Γ is also a hyperbolic set for iterates of f, we consider f^n instead of f. So we can assume that $|Df(x)| > \lambda > 1$ for all x in a neighbourhood V of Γ. If Γ contains an interval then $\Gamma = S^1$. Indeed, if $J \subset \Gamma$ is an interval then, because Λ is invariant, $f^n(J) \subset V$ for all n and f^n has no critical point in J, because Γ is hyperbolic. Hence if $f^n|J$ is injective then $f^n(J)$ is an interval of length at least equal to $\lambda^n \mu(J)$, where $\mu(J)$ is the Lebesgue measure of J. This cannot hold for all $n \in \mathbb{N}$ because $\lambda > 1$. It follows that for some $n \in \mathbb{N}$, $f^n|J$ is not injective and since f^n has no critical points, this implies that $N = S^1 = f^n(J)$. So in this case N cannot be an interval and f is an immersion of the circle.

Therefore, we assume by contradiction that Γ has positive Lebesgue measure and contains no intervals. By Lebesgue's Density Theorem, see the Appendix, there exists a density point $a \in \Gamma$. This means that

$$(*) \qquad \lim_{\delta \to 0} \frac{\mu(B(a,\delta) \cap \Gamma)}{\mu(B(a,\delta))} = 1,$$

where $B(a,\delta)$ is the ball of radius δ and centre at a in N. Let $\epsilon > 0$ be such that $B(x,\epsilon) \subset V$ whenever $x \in \Gamma$. Since $|Df(x)| > \lambda > 1$ for every $x \in V$, for every $\delta > 0$, there exists an integer n such that $\mu(f^n(B(a,\delta))) \geq \epsilon$. Taking the smallest such n one has $f^i(B(a,\delta)) \subset V$ for all $0 \leq i < n$.

We claim that f^n has bounded distortion on $B(a,\delta)$, more precisely, there exists a constant C_1, independent of δ, such that

$$(**) \qquad \frac{|Df^n(x)|}{|Df^n(y)|} < C_1$$

for all $x, y \in B(a,\delta)$. Indeed, since f is $C^{1+\alpha}$ and the derivative of f is not zero in the closure of V, there exists $\beta > 0$ such that the map $x \mapsto \log|Df(x)|$ is C^β on V. Therefore,

$$\begin{aligned} \log \frac{Df^n(x)}{Df^n(y)} &= \sum_{i=0}^{n-1} \left(\log|Df(f^i(x))| - \log|Df(f^i(y))|\right) \\ &\leq \sum_{i=0}^{n-1} C|f^i(x) - f^i(y)|^\beta \\ &\leq \sum_{i=0}^{n-1} C\lambda^{(i-n)\beta}|f^n(x) - f^n(y)| \leq C\frac{\lambda^\beta}{\lambda^\beta - 1}. \end{aligned}$$

This proves the claim.

Hence there exists $n \geq 0$ such that f^n maps $B(a,\delta)$ diffeomorphically and with bounded distortion onto an interval J_δ of at least length ϵ. From the forward-invariance of Γ we get that $f^n(\Gamma \cap B(a,\delta)) \subset \Gamma \cap J_\delta$ and therefore, using $(**)$ and $(*)$, we conclude

$$\frac{\mu(J_\delta \cap \Gamma)}{\mu(J_\delta)} \geq \frac{\mu(f^n(\Gamma \cap B(a,\delta)))}{\mu(J_\delta)} = 1 - \frac{\mu(f^n(B(a,\delta) \setminus \Gamma))}{\mu(f^n(B(a,\delta)))}$$
$$\geq 1 - C_1 \cdot \frac{\mu(B(a,\delta) \setminus \Gamma)}{\mu(B(a,\delta))} \to 1,$$

as $\delta \to 0$. Since each of the intervals J_δ has length at least ϵ, there exists a sequence $\delta_n \to 0$ so that J_{δ_n} converges to an interval J. Therefore, $\mu(J \cap \Gamma) = \mu(J)$. Since Γ is a closed set, we get that $\Gamma \supset J$ and this contradicts the assumption that J does not contain intervals. □

Exercise 2.8. Give an example of a C^1 map $f\colon S^1 \to S^1$ of degree two with $Df(x) > 1$ for all $x \in S^1$, with a compact, forward-invariant, hyperbolic set $\Gamma \neq S^1$ which has positive Lebesgue measure. (Hint: take an interval $J \subset S^1$ and let f be as above. Taking $\Gamma_n = \{x\,;\ f^i(x) \notin J \text{ for all } i = 1, \dots, n\}$, $\Gamma = \cap_{n\geq 0}\Gamma_n$ is forward invariant and compact. In order to show that one can choose Γ to have positive measure while f is still C^1, use a construction similar to the Denjoy counter-example in Chapter I. More details can be found in Bowen (1975).)

3 Hyperbolicity in Maps with Negative Schwarzian Derivative

As we saw in the previous section hyperbolicity plays a fundamental role when one studies the stability of maps. In general it is extremely difficult to check whether a map is hyperbolic. However, as we will see in the remainder of this chapter hyperbolicity is rather common in one-dimensional systems. In this section we will prove a result due to Misiurewicz (1981) which states that maps with negative Schwarzian derivative have many hyperbolic invariant sets. In particular we will get that a map with negative Schwarzian derivative satisfies the Axiom A (and therefore is structurally stable) if its critical points are in the basin of hyperbolic attractors. In the next section we shall apply all this to get a decomposition of the non-wandering set for unimodal maps similar to the one which is known for Axiom A diffeomorphisms without cycles. In Sections 5 and 6 of this chapter we shall see that the assumption on the negative Schwarzian derivative can be dispensed with. However, since the proofs under this assumption are much easier we shall first deal with this simpler case.

Let I be a compact interval and $f\colon I \to I$ be a C^3 map with negative Schwarzian derivative. In Section II.4 we have studied some basic properties of such maps. We saw for example that the iterates of f also have negative Schwarzian derivative and that the Minimum Principle holds. This principle says that, if x is strictly between two points a, b from an interval T and $f^n|T$ is a diffeomorphism, then $|Df^n(x)| > \min\{|Df^n(a)|, |Df^n(b)|\}$. As we saw in Section II.6, Singer (1978) realized that this property could be used to show

Theorem 3.1. **(Singer)** *If $f\colon I \to I$ is a C^3 map with negative Schwarzian derivative then the number of hyperbolic attracting and of non-hyperbolic periodic points of f is bounded.*

Notice that the above theorem is the exact analogue of Theorem III.1.3 concerning the finiteness of the number of non-repelling periodic points of rational maps. For rational maps, we have seen in Theorem III.1.5, that a compact invariant set which is not accumulated by the forward orbit of a critical point is a hyperbolic set. We will show below that the same is true for maps with negative Schwarzian derivative. In fact, the statement of Theorem 3.2 below is even stronger and it does not hold for rational maps. Indeed, f is not hyperbolic in a Herman ring, see Section 1. A more general version of this theorem will be proved in Section 5.

Theorem 3.2. **(Misiurewicz)** *Let $f\colon I \to I$ be a C^3 mapping with negative Schwarzian derivative. A compact forward-invariant set K is hyperbolic if it does not contain critical points, non-hyperbolic periodic points or hyperbolic attracting periodic points.*

By Theorem 2.5 this set K has zero Lebesgue measure. In particular, if such a map $f\colon I \to I$ has no periodic attractors and non-hyperbolic periodic orbits then almost every point accumulates on critical points of f. Theorem 3.2 follows immediately from the following useful result.

Theorem 3.3. *Let $f\colon I \to I$ be a C^3 map with negative Schwarzian derivative. Let V be an open set which contains all critical points of f and contains at least one point from each non-repelling periodic orbit. Then there exist $C > 0$, $\lambda > 1$ and $K < \infty$ such that if $x \in I$ satisfies $f^i(x) \notin V$ for every $i = 1, \ldots, n-1$ then*

$$|Df^n(x)| > C\lambda^n.$$

If J is an interval so that $J, \ldots, f^{n-1}(J)$ are all outside V then

$$\frac{Df^n(x)}{Df^n(y)} \leq K \text{ for all } x, y \in J.$$

Proof. We claim that there exists an integer m such that if $f^i(x) \notin V$ for all $0 \leq i \leq m$ then $|Df^m(x)| > 1$. Indeed suppose, by contradiction, that there exist arbitrarily large integers n for which there is a point $x_n \in I$ such that $|Df^n(x_n)| \leq 1$ and $f^i(x_n) \notin V$ for every $0 \leq i \leq n$. Since the set of critical points of f and of non-repelling periodic points is finite, there exists an open set U containing each non-repelling periodic point in V and whose closure is contained in V. Let J_n be the maximal interval containing x_n such that

$$f^j(J_n) \subset I \setminus \mathrm{cl}\,(U) \text{ for all } j = 0, \ldots, n.$$

Since f^n is a diffeomorphism on J_n (because the critical points of f are in U) and $Sf^n < 0$, it follows from the Minimum Principle that there exists a component L_n of $J_n \setminus \{x_n\}$ such that

$$|Df^n(x)| \leq 1 \text{ for all } x \in L_n.$$

Let $y_n \neq x_n$ be in the boundary of L_n. By the maximality of J_n, there exists an integer $0 \leq k(n) < n$ such that $f^{k(n)}(y_n) \in \mathrm{cl}\,(U)$. Let δ be so that each component of $V \setminus U$ has length $\geq \delta$. Since $f^{k(n)}(x_n) \notin V$ and $f^{k(n)}(y_n) \in \mathrm{cl}\,(U)$ we get

$$|f^{k(n)}(L_n)| > \delta. \tag{3.1}$$

On the other hand, since $0 < |Df^n(x)| \leq 1$ for all $x \in L_n$, we get

$$|f^n(L_n)| \leq |L_n|. \tag{3.2}$$

We claim that

$$|L_n| \to 0 \text{ as } n \to \infty. \tag{3.3}$$

Indeed, if (3.3) does not hold, we can take a subsequence L_{n_i} converging to an interval L of positive length. If J is an interval whose closure is contained in the interior of L then there exists an integer j such that L_{n_i} contains J for all $n_i > j$. Since $L_n \subset J_n$, and $f^j(J_n) \cap \mathrm{cl}\,(U) = \emptyset$ for all $0 \leq j \leq n$, it follows that $f^n(J) \subset I \setminus \mathrm{cl}\,(U)$ for all integers n. From Theorem I.2.2, we get that there exists a point in J that is asymptotic to periodic attractors (Denjoy-Schwartz). This is not possible because U contains a point from each attracting periodic orbit of f. This proves (3.3).

From (3.2) and (3.3) we get $|f^n(L_n)| \to 0$. Since $|f^{k(n)}(L_n)| \geq \delta$, we can take a subsequence $n_i \to \infty$ such that the intervals $f^{k(n_i)}(L_{n_i})$ converge to an interval S of length at least equal to δ. As $|f^n(L_n)| = |f^{n-k(n)}(f^{k(n)}(L_n))| \to 0$ and $|f^{k(n)}(L_n)| \geq \delta$ we get $n - k(n) \to \infty$. If J is an interval whose closure is contained in the interior of S, then $f^{k(n_i)}(L_{n_i})$ contains J for i big enough. As before, it follows that $f^n(J)$ is contained in $I \setminus \mathrm{cl}\,(U)$ for all n. Using again Theorem I.2.2 we get that there exist points in J which are asymptotic to a periodic attractor, a contradiction because $f^n(J) \cap \mathrm{cl}\,(U) = \emptyset$ for all $n \geq 0$. This proves the claim.

From the claim, and the compactness of $I \setminus V$ it follows that there exist an integer k and $\lambda > 1$ such that $|Df^k(x)| > \lambda^k$ whenever $f^i(x) \notin V$ for all $i \leq k$. Let $\rho = \min\{|Df(x)|;\ x \in I \setminus V\}$ and $C > 0$ be such that $\rho^i > C\lambda^i$ for all $0 \leq i < k$. If n is an integer such that $f^l(x) \notin I \setminus V$ for all $l < n$, then we can write $n = jk + i$, with $0 \leq i < k$ and we have

$$|Df^n(x)| = \left(\prod_{l=0}^{j-1} |Df^k(f^{lk}(x))| \right) \times |Df^i(f^{jk}(x))| \geq$$

$$\geq (\lambda^k)^j \rho^i \geq \lambda^{jk} C \lambda^i = C\lambda^n.$$

This proves the first inequality from the statement of this theorem. The last inequality holds because there exists a constant C' (which is equal to the Lipschitz constant of $\log Df$ on the complement of V) such that

$$\left|\frac{Df(f^i(x))}{Df(f^i(y))} - 1\right| \leq C' \cdot |f^i(x) - f^i(y)|$$

for $x, y \in J$ and $i = 0, 1, \ldots, n-1$. Combining this with the first inequality gives the second inequality. □

Corollary 3.1. *Let $f\colon I \to I$ be a map with negative Schwarzian derivative, a finite number of critical points and such that all critical points of f are in the basin of the hyperbolic attractors of f. Then f satisfies the Axiom A.*

Proof. By Singer's Theorem II.6.1, f has a finite number of hyperbolic attractors since each one of them must contain a critical point or a boundary point of I in its basin. From the previous theorem it follows that the complement $\Sigma(f)$ of the basins of the hyperbolic attractors is a hyperbolic set. □

Let us conclude this section by analyzing the dynamics near non-hyperbolic periodic points and their bifurcations.

Remark. 1. The periodic points of a map f with negative Schwarzian derivative have some special properties. Because $Sf^n(p) < 0$, one has that $D^2f^n(p) = 0$ implies $D^3f^n(p) \cdot Df^n(p) < 0$. Furthermore, if p is a periodic point of period n of f and $Df^n(p) = -1$ then a simple calculation shows that $Df^{2n}(p) = 1$, $D^2f^{2n}(p) = 0$ and $D^3f^{2n}(p) < 0$. Therefore, if f is a map with negative Schwarzian derivative and p is a non-hyperbolic periodic point of period n then p is (possibly one-sided) attracting. So only one of the following situations can occur:

1. p is a super-attractor, i.e., $Df^n(p) = 0$;
2. p is semi-stable, i.e., $Df^n(p) = 1$ and $D^2f^n(p) \neq 0$;
3. p is an orientation reversing weak-attractor of codimension 1, i.e., $Df^n(p) = -1$, $Df^{2n}(p) = 1$, $D^2f^{2n}(p) = 0$ and $D^3f^{2n}(p) < 0$;
4. p is a weak-attractor of codimension 2, i.e., $Df^n(p) = 1$, $D^2f^n(p) = 0$ and $D^3f^n(p) < 0$.

As we have seen in the previous section if f has such a periodic point, some nearby maps will have a different number of periodic points. Therefore, such periodic points are said to *bifurcate*. Precise descriptions how these periodic orbits can bifurcate is given in the exercises below. 2. The terminology 'codimension k' periodic orbit refers to the minimum number of parameters that are necessary to observe such orbits in generic (i.e., typical) families depending on these parameters. For example for periodic orbits of codimension 1 one has the following. Each C^r map f with such a periodic orbit can be approximated in the C^r topology by a map without such periodic orbits. Moreover there exist one-parameter families of maps such that for each nearby family there are parameter values for which these codimension 1 periodic orbits will occur. Similarly one can motivate the terminology codimension 2. (For more on this see the exercises at the end of this section.) 3. All non-hyperbolic periodic points of a map with negative Schwarzian derivative are attracting periodic points because its basin contains an open set.

In the next exercises we analyze the bifurcations which non-hyperbolic periodic orbits can undergo. Because of the above remark only the bifurcations discussed in the exercises below can occur in families of maps with negative Schwarzian derivative.

Exercise 3.1. If f is unimodal and has negative Schwarzian derivative then f cannot have a periodic attractor p of codimension two. (Hint: since f has negative Schwarzian, p is accumulated from both sides by a turning point. But this is impossible because f^n is orientation preserving and n is the minimal period of p. Alternatively, show that one can find a unimodal map g which coincides with f outside a neighbourhood of p, is arbitrarily C^3 near f and has two periodic attractors near p. This is impossible according to Theorem II.6.1.)

Exercise 3.2. Assume that $f_\mu\colon \mathbb{R} \to \mathbb{R}$ depends continuously on a parameter $\mu \in \mathbb{R}$ in the C^2 topology. a) Show that if $f_0(0) = 0$, $Df_0(0) = 1$ and $Df_0^2(0) \neq 0$, then there exists a neighbourhood V of 0 such that f_0 is semi-stable at 0 in V (i.e., $f^n(x) \to 0$ for $n \to \infty$ in one component of $V \setminus \{0\}$ and $f^n(x) \to 0$ for $n \to -\infty$ in the other component of $V \setminus \{0\}$). Moreover, for each parameter μ sufficiently close to 0, $f_\mu|V$ has either 0, 1 or two fixed points. For each x, either some iterate of x is outside V or the orbit of x tends to a fixed point of f_μ. b) Let in addition $\frac{d}{d\mu} f_\mu(0) \neq 0$ at $\mu = 0$, and let us assume, for simplicity, that $\frac{d}{d\mu} f_\mu(0) Df_0^s(0) > 0$. Then f_μ has no fixed points in V for $\mu > 0$, one fixed point in V for $\mu = 0$ and two fixed points in V for $\mu < 0$ (if $\frac{d}{d\mu} f_\mu(0) Df_0^s(0) < 0$ the same statements hold for $\mu < 0$, $\mu = 0$ and $\mu > 0$ respectively.) More precisely, show that locally near $(0,0)$ the set $\{(x,\mu)\,;\ f_\mu(x) = x\}$ is of the form $\mu = g(x)$ where $g(0) = 0$, $Dg(0) = 0$ and $D^2 g(0) \neq 0$. This bifurcation is called the *saddle-node bifurcation* or *fold bifurcation.*) (Hint: consider $f_\mu(x) - x = 0$. Since $\frac{d}{d\mu} f_\mu(0) \neq 0$ at $\mu = 0$ the statement follows from the Implicit Function Theorem.) c) Show that any two nearby families f_μ and $\tilde{f}_\mu$ as above are conjugate near 0 for μ near 0: show that there exists a family of conjugacies h_μ such that $h_\mu \circ f_\mu = \tilde{f}_\mu \circ h_\mu$ restricted to V. (Hint: for $\mu > 0$ choose $x \in V$ and an arbitrary order preserving homeomorphism $h_\mu\colon [x, f_\mu(x)] \to [x, \tilde{f}_\mu(x)]$. There exists a unique extension of h_μ to V such that $h_\mu \circ f_\mu = \tilde{f}_\mu \circ h_\mu$ restricted to V. For $\mu = 0$ one chooses two corresponding intervals and for $\mu < 0$ three such intervals. These intervals are called *fundamental domains.* In order to show that one can construct h_μ so that it depends continuously on the parameter one has to do more work, see Newhouse et al. (1983).

Exercise 3.3. Assume that $f_\mu\colon \mathbb{R} \to \mathbb{R}$ depends continuously on a parameter $\mu \in \mathbb{R}$ in the C^3 topology. a) Show that if $f_0(0) = 0$, $Df_0(0) = -1$ then $Df^2(0) = 1$ and $D^2 f^2(0) = 0$. b) If $f_0(0) = 0$, $Df_0(0) = -1$ and $D^3 f^2(0) < 0$ then there exists a

neighbourhood V of 0 such that f_0 is stable at 0 in V (i.e., $f^n(x) \to 0$ for $n \to \infty$ for $x \in V$). Moreover, for each parameter μ sufficiently close to 0, the map f_μ has precisely one fixed point in V, and f_μ has either zero, or two periodic points of period 2. Each point in V (except the fixed point) tends to these periodic points. c) If in addition $\frac{d}{d\mu} Df_\mu(0) > 0$ at $\mu = 0$, then f_μ has two periodic points of period 2 in V for $\mu > 0$ and no periodic points of period 2 in V for $\mu < 0$. More precisely, show that locally near $(0,0)$ the set $\{(x,\mu)\,;\ f_\mu(x) = x\}$ is of the form $x = g(\mu)$ and that the set $\{(x,\mu)\,;\ f_\mu^2(x) = x \neq f_\mu(x)\}$ is of the form $\mu = \hat{g}(x)$ where $\hat{g}(0) = 0$, $D\hat{g}(0) = 0$ and $D^2\hat{g}(0) \neq 0$. This bifurcation is called the *flip* or *period doubling* bifurcation.
(Hint: consider $f_\mu(x) - x = 0$. Since $\frac{d}{dx} f_0(0) \neq 0$ the first statement follows from the Implicit Function Theorem. The second statement is a little bit harder to prove: consider $f_\mu^2(x) - x = 0$. Since $\frac{d}{d\mu} f_0^2(0) = 0$ one can not apply the Implicit Function Theorem immediately. Show that $\frac{f_\mu^2(x)-x}{f_\mu(x)-x}$ is C^2 and that one can apply the Implicit Function Theorem to this new function by showing that the assumptions imply that $\frac{d}{d\mu}\frac{f_\mu^2(x) - x}{f_\mu(x) - x}$ is non-zero at $\mu = 0$ and $x = 0$.)

Exercise 3.4. Assume that $f_\mu \colon \mathbb{R} \to \mathbb{R}$ depends continuously on a parameter $\mu \in \mathbb{R}$ in the C^3 topology. Assume that $f_0(0) = 0$, $Df_0(0) = 1$, $D^2 f_0(0) = 0$ and $D^3 f_0(0) < 0$. So 0 is a weak-attractor of codimension 2. Show that for μ close to zero f_μ has either one fixed point near 0 (which is attracting), or three fixed points near 0 (with the middle one repelling and the other two attracting). This is sometimes called the *pitch-fork* bifurcation.

4 The Structure of the Non-Wandering Set

In this section we will describe the non-wandering set of multimodal maps $f \colon I \to I$. As usual, a point x is called *non-wandering* if for each neighbourhood U of x there exists $n \geq 1$ such that $f^n(U) \cap U \neq \emptyset$. The non-wandering set $\Omega(f)$ is the collection of all points which are non-wandering. We shall decompose the non-wandering set into simpler parts. More precisely, we would like to get a structure similar to the structure of Axiom A diffeomorphisms satisfying the no-cycle condition. For such diffeomorphisms one knows that the non-wandering set decomposes into a finite number of transitive sets and moreover one has a filtration. Using this filtration one proves the stability of these diffeomorphisms, see Smale (1967). Surprisingly, for arbitrary multimodal interval maps one has a similar (but possibly countably infinite) decomposition of the non-wandering set. This decomposition is related to possible renormalizations of the system and was first constructed in Jonker and Rand (1981) for continuous unimodal maps. In Van Strien (1981) it was shown that for unimodal maps with negative Schwarzian derivative all, except possibly one, of these parts of the non-wandering sets are hyperbolic. Here we shall deal with the multimodal case.

As a byproduct of the decomposition of the non-wandering set we shall describe the attracting sets of a multimodal map. Here we say that a forward invariant set A is a *topological attractor* if its basin $B(A) = \{x\,;\ \omega(x) \subset A\}$ satisfies

1. the closure of $B(A)$ contains intervals;
2. each closed forward invariant subset A' which is strictly contained in A has a smaller basin of attraction: $\mathrm{cl}\,(B(A)) \setminus \mathrm{cl}\,(B(A'))$ contains intervals.

(A measure theoretic analogue of this definition will be given in Section V.1.) Let us say that f has *sensitive dependence on some set* K if there exists $\delta > 0$ such that for each $x \in K$ and each interval J containing x there exists $n \in \mathbb{N}$ such that $f^n(J)$ has length $\geq \delta$. Guckenheimer (1979) defines f to be sensitive dependent on initial conditions if this last statement holds for a set K of positive Lebesgue measure. A forward invariant set A is called *transitive* if it contains a dense orbit. In the first theorem we describe attractors of interval maps.

Theorem 4.1. *If $f \colon I \to I$ be a l-modal map without wandering intervals then each attractor C of f is transitive and is of one of the following types:*

1. *periodic;*
2. *a finite union of intervals on which f acts transitively and such that (at least) one of the intervals contains a turning point; the restriction of f to these intervals has sensitive dependence on initial conditions and the restriction of f to these intervals is conjugate to a piecewise linear map;*
3. *a solenoidal attractor: this means that f acts on C as an adding machine (as explained in Section II.5); in this case f is infinitely renormalizable and*
$$C = \bigcap_{n=0}^{\infty} K_n.$$
Here for each ≥ 0,
$$K_n = \bigcup_{k=0}^{q(n)-1} f^k(J_n),$$
where J_n is a restrictive interval of period $q(n)$, $J_{n+1} \subset J_n$ and $q(n+1) = a(n) \cdot q(n)$ with $a(n) \in \{2, 3, 4, \dots\}$. Moreover, C contains a turning point whose forward orbit is dense in C.

Moreover, we have the following properties:

a. *If an attractor of f is as in 1) or 2) then its basin has non-empty interior; if it is as in case 3) then the interior of the basin is empty.*

b. *If f is C^3 and has negative Schwarzian derivative then each attractor has a critical point or a boundary point of I in its basin.*

The proof of this theorem will occupy the remainder of this section. Before starting with the proof we will state a related theorem which asserts that one can decompose the non-wandering set of a one-dimensional map by some ‘filtration’:

there exists an at most countable nested sequence of forward invariant sets K_n. Such a set K_n consists of a finite number of periodic, restrictive, intervals (for the definition of a restrictive interval, see Section II.5). In particular, each point either remains forever in $K_n \backslash K_{n+1}$ or it is eventually mapped into K_{n+1}. Axiom A diffeomorphisms have a similar, but finite, filtration, see Smale (1967). For simplicity, we say that two maps are Ω-conjugate, if there exists a conjugacy between the restriction of the two maps to their non-wandering sets.

Theorem 4.2. *Let $f\colon I \to I$ be a l-modal map without wandering intervals and let K_n, J_n and $q(n)$ be as in the previous theorem. Then there exists $N \in \mathbb{N} \cup \{\infty\}$ such that*

1. *$\Omega(f)$ can be decomposed into closed forward invariant subsets Ω_n:*

$$\Omega(f) = \bigcup_{0 \le n \le N} \Omega_n.$$

These sets Ω_n are defined as follows. Define $K_0 = I$ and K_{n+1} as the union of all maximal restrictive intervals of $f\colon K_n \to K_n$. Then K_n is a nested decreasing sequence of sets each consisting of a finite union of intervals for each finite $n \le N$. If we define $K_\infty = \cap_{n \ge 0} K_n$ when $N = \infty$ then

$$\Omega_n := \Omega(f) \cap \mathrm{cl}\,(K_n \setminus K_{n+1})$$

for $n < N$ and $\Omega_N := \Omega(f) \cap K_N$.

2. *For each finite $n \le N$, the set Ω_n is a union of transitive sets and, if $N = \infty$, $\Omega_\infty = K_\infty$ consists of solenoidal attractors as above.*

3. *The topological entropy of f is zero if and only if Ω_n consists of periodic orbits of period 2^n for every finite $n \le N$.*

4. *If Ω_n contains non-periodic points then $f^{q(n)}\colon \Omega_n \to \Omega_n$ is Ω-semi-conjugate to a piecewise linear map with constant slopes (and the semi-conjugacy is almost injective, see the proof below).*

Finally, most of these sets Ω_n are hyperbolic if f satisfies some additional conditions. Indeed, we define a *stratum on level* $n \le N$ of the filtration $(K_i)_{i \le N}$ to be a set W which is the forward orbit of a component of K_n. Such a stratum is called *critical* if $W \setminus K_{n+1}$ contains a turning point or if $W \subset K_\infty$ (here we take $K_{n+1} = \emptyset$ if $n = N < \infty$ and $K_\infty = \emptyset$ if $N < \infty$). Clearly the filtration has at most l critical strata if f is l-modal. The next theorem says that the part of the non-wandering set which is contained in a non-critical stratum of the filtration is hyperbolic.

Theorem 4.3. *Let $f\colon I \to I$ be a l-modal C^3 map, with negative Schwarzian derivative, non-flat critical points and whose periodic point(s) in the boundary of I are repelling. Then each stratum of the filtration consists of a finite number of transitive sets and, similarly, each transitive set is contained in a stratum of the filtration. A transitive set which is contained in a non-critical stratum is hyperbolic and the restriction of f to this transitive set is conjugate to a subshift of finite type and so has positive topological entropy. Each attractor of f is contained in a critical stratum.*

Remark. 1. Guckenheimer (1979) proved Theorem 4.1 in the unimodal case. If f is unimodal then f has at most one such attractor and Lebesgue almost all points tend to this attractor. In cases 1) and 3), f is only finitely often renormalizable. In the second case f is infinitely often renormalizable. In the unimodal case, f has no sensitive dependence on initial conditions in cases 1) and 2). Theorem 4.2 was proved previously by Jonker and Rand (1981) in the unimodal case. For the multimodal case, see Blokh (1983). A more abstract decomposition for general piecewise monotonic maps can be found in Hofbauer (1979), Nitecki (1982) and Hofbauer and Raith (1989).

2. Note that K_n consists of several intervals and that $f\colon K_n \to K_n$ permutes these components. One can define the notion of a restrictive interval for such a map defined on a union of intervals exactly as before.

3. The condition $Sf < 0$ can be replaced by the assumption that f is C^2 and that all attracting periodic points of f are essential, see Section 5.

In the remainder of this section we shall prove these theorems. First we shall deal with maps having zero topological entropy.

Proposition 4.1. *If $f\colon I \to I$ is multimodal with zero topological entropy and which is not renormalizable then the ω-limit of every point in I is a fixed point.*

Proof. If f is a homeomorphism then the result is obviously true. So assume that f is l-modal with $l \geq 1$. If f has an orientation reversing fixed point then because $f(\partial I) \subset \partial I$ there are two possibilities: i) there exists $p' \neq p$ such that $f(p') = f(p) = p$ or ii) f exchanges the two components of $J \setminus \{p\}$. However, ii) cannot occur because then f is renormalizable. So assume that i) holds and let $J = [p', p]$. Since p is an orientation reversing fixed point, and since f is not renormalizable, $f^2(J)$ strictly contains J and $f^2(\partial J) \subset \partial J$. Therefore, for each point x in the interior of J there exist at least two points x_1 and x_2 in J such that $f^2(x_1) = f^2(x_2) = x$. It follows that for each x in J and each integer n there exist at least 2^n points in J which are mapped to x by f^{2n}. In particular, the number of laps of f^{2n} is at least 2^n and hence, by Lemma II.7.4, we get that the topological entropy of f is at least equal to $(1/2)\log 2$. This contradicts the assumption that f has zero entropy. Therefore f has no orientation reversing fixed points. It is easy to see that this implies that the ω-limit of every point in I is a fixed point. □

Proposition 4.2. *Assume that $f\colon I \to I$ is a multimodal map with zero topological entropy. Then any restrictive interval of period ≥ 2 is contained in a restrictive interval of period 2. Any point which is not eventually mapped into a restrictive interval of period 2 is asymptotic to a fixed point.*

Proof. Suppose that J is a restrictive interval with period $n \geq 2$. We shall prove that this restrictive interval is contained in a restrictive interval of period 2. By definition $J, \dots, f^{n-1}(J)$ are disjoint and f permutes these intervals. By Lemma II.1.2, there exists a component I_1 of $I \setminus (J \cup \dots \cup f^{n-1}(J))$ (which does not contain ∂I) such that $f(I_1) \supset I_1$ and for which $f(a) \geq b$ and $f(a) \leq b$ where $a < b$ are the boundary points of I_1. Therefore the set $\{t \in I_1\,;\ f^2(t) \leq t\}$ is non-empty and we can define p to be the maximum of this set. Then p is either a periodic point of period 2 or a fixed point. If $p \in \partial I_1$ then p is contained in one of the intervals $f^i(J)$ and therefore J has period 2 and so J is as claimed. Otherwise let q be equal to the smallest $t > p$ such that $f^2(t) = p$ and if such a t does not exist then we take q to be the right endpoint of I. From the choice of p one has $q \notin I_1$. We claim that $V = [p, q]$ is mapped into itself by f^2. This is easy to see: since $f^2(p) = f^2(q) = p$ and $f^2(t) > p$ for $t \in (p, q)$ one would otherwise have that $f^2(V)$ strictly contains V and $f^2(\partial V) \subset \partial V$. As in the proof of the previous proposition this implies that $h(f) \geq (1/2)\log(2)$, a contradiction. Therefore $f^2(V) \subset V$. Moreover, $f(V) \cap V \subset \{p\}$ because otherwise there exists $x \in V$ for which $f(x) = p$ and therefore $V \ni f^2(x) = f(p) \notin V \setminus \{p\}$, a contradiction. So V is a restrictive interval of period 2. Since $q \notin I_1$, V has a non-empty intersection with at least one of the restrictive intervals $J, \dots, f^{n-1}(J)$. From Lemma II.5.1 it follows that $V \supset J$ which proves the claim. The last statement follows immediately: by collapsing all maximal restrictive intervals (and their preimages) one obtains a map which is not renormalizable and also has zero topological entropy; so this proposition follows from the previous one. □

Remark. From these two propositions it follows that f has only periodic orbits whose periods are powers of 2 if the topological entropy of f is zero. This result was proved before by Misiurewicz (1979). Next we consider maps with positive

entropy:

Proposition 4.3. *Let $f\colon I \to I$ be a C^1 multimodal map with positive topological entropy and without wandering intervals. Then f is semi-conjugate to a piecewise linear map T with slope $\pm s$ with $s > 1$ where $\log s$ is the topological entropy of f. Moreover, the following properties hold. Let X be the set of points which are eventually mapped into a restrictive interval or into the basin of a periodic attractor.*

1. *This semi-conjugacy can collapse an interval only if it is contained into X. Basins of periodic attractors are certainly collapsed.*

2. *The union of X and the set of points which are mapped eventually into an interval $V \neq I$ with $f(V) \subset V$ is dense in I. Moreover, f is conjugate to T on the complement of X.*

3. *If T has a periodic turning point c then $h^{-1}(c)$ is a restrictive interval where h is the semi-conjugacy between f and T.*

Proof. By Theorem II.8.1, there exist a monotone semi-conjugacy $h\colon I \to I$ between f and a piecewise linear map T with constant slope and with the same topological entropy as f. In other words there exists a monotone continuous surjective map with $h \circ f = T \circ h$. Let us prove Statement 1). So take a non-trivial interval J of the form $h^{-1}(x)$. Let us suppose by contradiction that the forward orbit of x does not contain a periodic point. Then the forward orbit of J consists of disjoint intervals not tending to a periodic point and so J is a wandering interval, a contradiction. Therefore some forward iterate $T^k(x)$ of x is a periodic point with period n. Hence $K = f^k(J) = h^{-1}(y)$ is also a non-trivial interval, $f^n(K) \subset K$, $f^n(\partial K) \subset \partial K$ and $K, \dots, f^{n-1}(K)$ are disjoint. So if the orbit of K contains no turning point then it is contained in the basin of a periodic attractor and otherwise it is contained in a restrictive interval. Since T has no periodic attractors h certainly collapses components of basins of periodic attractors. This completes the proof of 1).

Let us prove 2). If U is an interval in the complement of X then it is not contained in the basin of a periodic attractor. Since f has no wandering intervals this implies that $f^n(U) \cap f^m(U) \neq \emptyset$ for some $n > m \geq 0$ (where n and m minimal). Since $f^{-1}(X) \subset X$, this implies that $f^{n-m}(U') \subset U'$ where U' is the maximal interval in the complement of X which contains $f^m(U)$. So there exists $k \leq n-m$ such that $U', \dots, f^{k-1}(U')$ have disjoint interiors and $f^k(U') \subset U'$. By definition U' is not contained in a restrictive interval and so it follows from Lemma II.5.1 that $k = 1$ and that the only closed interval V containing U' for which $f(V) \subset V$, $f(\partial V) \subset \partial V$ and V contains a turning point is equal to I. It follows that U is contained in a forward invariant interval. This proves Statement 2)

Finally, if one of the turning points c_T of T is periodic with period n then f^n maps $h^{-1}(c_T)$ into itself. Now assume by contradiction that $h^{-1}(c_T)$ consists of a single point. Then, since h maps turning points of f to turning point of T, this singleton is a turning point c of f with period n. Since f is C^1 each periodic turning point is attracting and therefore h maps the immediate basin of c to c_T, so $h^{-1}(c_T)$ is not a singleton, a contradiction. □

Since the previous proposition shows that a smooth multimodal map with positive entropy is semi-conjugate to a piecewise linear map T with slopes $\pm s$, we shall now describe the non-wandering set of these piecewise linear maps.

Proposition 4.4. *Let $T\colon I \to I$ be a piecewise linear l-modal map with slope $\pm s$, $s > 1$. Then one has the following properties.*

1. *T has sensitive dependence on initial conditions.*
2. *T is only finitely often renormalizable.*
3. *The non-wandering set of T contains a finite number of intervals which are permuted by T and on each of these intervals T is transitive. (The permutation of the intervals may split into disjoint cycles.) On the complement of these intervals, the non-wandering set of T consists of a finite number of periodic points and the remainder is a subshift of finite type.*
4. *The only attractors of T are intervals.*

Proof. Take k so large that $(s^k/2^l) > 1$ and so that each periodic turning point has at most period k. Let $X = \{T^i(c_j)\,;\ j = 1, \dots, l \text{ and } i = 0, 1, \dots, k\}$. We claim that for each interval J and each integer $j \geq 0$, one of the following possibilities holds:

1. $|T^{j+k}(J)| \geq (s^k/2^l)|T^j(J)|$;
2. at least $l+1$ of the intervals $T^j(J), \dots, T^{j+k-1}(J)$ contain a turning point or one of these intervals contains at least two turning points.

Here $|J|$ denotes the length of the interval J. So let us prove this claim. Of course, $|T(J)| = s|J|$ if J contains no turning point and $|T(J)| \geq (s/2)|J|$ if J contains precisely one turning point. So if 2) does not hold then 1) holds. Since $(s^k/2^l) > 1$ one cannot be in case 1) for all $j \in \mathbb{N}$, because otherwise $|T^n(J)| \to \infty$. So there exists j so that 2) holds. Therefore either one of the intervals $T^j(J), \dots, T^{j+k-1}(J)$ contains two turning points or there exists a turning point c and $j \leq r < s < j+k$ such that $c \in T^r(J) \cap T^s(J)$ and therefore $c, T^{s-r}(c) \in T^s(J)$. If c is non-periodic then this implies that $T^s(J)$ contains two points from X; if c is periodic then $T^{s-r}(c) = c$ and $T^{i(s-r)}(T^s(J))$ contains another turning point for i large because otherwise $|T^{i(s-r)}(T^s(J))| \to \infty$ as $i \to \infty$. So in each case we get that there exists $n \in \mathbb{N}$ so that $T^n(J)$ contains two points from X. This implies the Statements 1 and 2 because, as we have shown, there are no arbitrarily small restrictive intervals. In order to prove Statement 3), take a minimal periodic interval V. By this we mean that $T^n(V) \subset V$ for some $n \in \mathbb{N}$ and that no subset of V has this property. From the first part of the proof, such an interval exists (but it might be equal to I). We claim that V is a subset of the non-wandering set of T. So assume by contradiction that there exists a closed interval $J \subset V$ such that $T^i(J) \cap J = \emptyset$ for all $i \in \mathbb{N}$. Then consider $U = \cup_{i \geq 0} T^i(J)$. Clearly T maps components of U into components of U and since T has no wandering intervals each component of U is eventually periodic. Since J is not contained in an eventually periodic component and since $T^{kn}(J) \subset V$ for each $k \geq 0$ one gets that J is eventually mapped into a periodic component U' of U which is a proper subset of V. This contradicts the minimality of V. We leave the (easy) proof of the remaining statements to the reader. □

Finally, let us describe the dynamics in the infinitely renormalizable case.

Proposition 4.5. *Let f be infinitely renormalizable and without wandering intervals. Then there exists a nested sequence of intervals I_n containing a turning point and integers $q(n)$ such that the intervals $I_n, \dots, f^{q(n)-1}(I_n)$ have pairwise disjoint interiors and such that $f^{q(n)}$ maps I_n into itself and $f^{q(n)}(\partial I_n) \subset \partial I_n$. The intersection of the nested sets*

$$F_n = \bigcup_{j=0}^{q(n)-1} f^j(I_n)$$

for $n \geq 0$ forms a Cantor set denoted by Ω_∞. The map $f\colon \Omega_\infty \to \Omega_\infty$ is minimal, has zero entropy and is conjugate to some adding machine (this notion is defined in the proof of this proposition).

Proof. Since f has no wandering interval and since the first $q(n)$ iterates of I_n are disjoint, the maximal length of each connected component of F_n tends to zero as n tends to infinity. Therefore Ω_∞ contains no intervals and so it is a Cantor set. Let us abstractly describe the dynamics of $f\colon \Omega_\infty \to \Omega_\infty$. Let $\mathbb{Z}_k$ the group of integers modulo k and $\tau_k\colon \mathbb{Z}_k \to \mathbb{Z}_k$ the translation by one. Since no point in Ω_∞ is contained in the interior of an interval of the form $f^k(I_n)$, $k = 0, 1, \dots, q(n) - 1$ one can code each point x in Ω_∞ uniquely by integers $0 \leq k_n(x) < q(n)$ by defining

$$x \in f^{k_n(x)}(I_n) \text{ for all } n \geq 0.$$

Let $\Gamma = \prod_{i=1}^{\infty} \mathbb{Z}_{q(i)}$. Endow $\mathbb{Z}_k$ with discrete topology and Γ with the corresponding product topology. If we let $\Sigma_{(q(0),q(1),\dots)}$ be the subset of elements $(m_1, m_2, \dots) \in \Gamma$ such that

$$m_{i+1} = m_i \bmod q(i)$$

then, because $f^j(I_{n+1}) \subset f^k(I_n)$ if and only if $j = k \bmod q(n)$, one has that the map

$$\Omega_\infty \ni x \mapsto (k_1(x), k_2(x), \dots) \in \Sigma_{(q(0),q(1),\dots)}$$

is well-defined for $x \in \Omega_\infty$ and surjective. Since Ω_∞ contains no intervals, this maps is injective. For simplicity write $\Sigma = \Sigma_{(q(0),q(1),\dots)}$. The action of f on Σ is given by the translation $\tau\colon \Sigma \to \Sigma$, defined by

$$\tau(k_1, k_2, \dots) = (\tau_{q(1)}(k_1), \tau_{q(2)}(k_2), \dots).$$

This translation map is a homeomorphism and it is in fact an isometry if we consider the following metric on Σ: for $(k_1, k_2, \dots), (\tilde{k}_1, \tilde{k}_2, \dots) \in \Gamma$, let n be the smallest integer such that $k_n \neq \tilde{k}_n$ and take then $d((k_1, k_2, \dots), (\tilde{k}_1, \tilde{k}_2, \dots)) =$

$\frac{1}{n}$. It follows from this that the topological entropy of τ is zero. Moreover, τ not only acts on this Cantor set but it also acts on each finite level of Σ: it simply permutes the cylinder sets of the form

$$\{(k_1, k_2, \dots, k_n, *, *, \dots)\,;\ k_i = m_i \text{ for } i = 1, \dots, n\}$$

(which have arbitrary elements at the coordinates larger than $n+1$). In particular, for each point $x \in \Sigma$ and each cylinder set there exists $n \geq 0$ such that $\tau^n(x)$ is contained in this cylinder set. Hence, τ is a minimal homeomorphism on Σ, i.e., every orbit is dense in Σ with zero topological entropy. This action is called an *adding machine*. □

Proof of Theorems 4.1-4.3. Let us first assume that $h(f) = 0$. In this case Propositions 4.1 and 4.2 imply that a point is either asymptotic to a periodic attractor or eventually mapped into a restrictive interval of period 2. For each maximal restrictive interval J one can repeat the same argument by considering $f^2 \colon J \to J$. So if f is finitely renormalizable then there exists $N < \infty$ such that the non-wandering set of f consists of periodic orbits of period 2^n with $n \leq N$. If f is infinitely renormalizable then the non-wandering set of f consists of periodic orbits of period 2^n for each $n \in \mathbb{N}$ (some of which may be attracting) and a subset of $\cap_{n\geq 0} K_n$ where K_n is the union of all restrictive intervals of period $\geq n$. In Proposition 4.5 it was proved that the dynamics of f on the closure of each orbit of this intersection is as stated. Clearly the first return time $q(n+1)$ of f to each component I_{n+1} of K_{n+1} is a multiple of q_n. This and Theorem II.6.1 imply Theorems 4.1-4.3 in this case.

So assume that $h(f) > 0$. If f has no restrictive interval then from Proposition 4.3 we get that f is semi-conjugate to a piecewise linear map T. This semi-conjugacy only collapses basins of periodic attractors. Moreover, the only attractors of T are intervals. Since f has no restrictive intervals it has no essential periodic attractors. This proves Theorems 4.1 and 4.2. Moreover, if f is as in Theorem 4.3 then, because of Theorem II.6.1, f has also no neutral or inessential periodic attractors. By Theorem II.6.2 the map f has no wandering intervals. Hence Proposition 4.3 implies that f and T are conjugate. Theorem 4.3 follows.

On the other hand, if f does have a restrictive interval then take K_1 to be the union of all maximal restrictive intervals. $\Omega_0 = \Omega(f) \cap \mathrm{cl}\,(I \setminus K_1)$ and $\Omega' = \Omega(f) \cap \mathrm{cl}\,(I \setminus K_1)$ are both forward invariant. Note that Ω_0 may contain a forward invariant interval if the set X from Proposition 4.3 is not dense in I. Since $f \colon \Omega_0 \to \Omega_0$ is semi-conjugate to a piecewise linear map; it follows by Proposition 4.4 that f is transitive on such an interval. Next consider $f \colon K_1 \to K_1$. Of course $K_1 = \cup_j J_1^j$ where J_i^j are maximal restrictive intervals. By applying the previous construction to each of these intervals J_1^j and considering the first return map to this interval we can repeat the procedure. It may terminate after a finite number, say N, of steps or this procedure may be continued infinitely often. In the latter case we get a solenoidal attractor. This proves Theorems 4.1 and 4.2. If f is as in Theorem 4.3, then f has no neutral or inessential periodic attractors. Moreover, it has no wandering intervals because of Theorem II.6.2. By definition the forward orbit $\tilde{K}$ of a non-critical component K' of $K_n \setminus K_{n+1}$ contains no turning points. Since K_{n+1} is forward invariant, all this implies that $\tilde{K}$ contains only hyperbolic repelling periodic orbits. From Theorem 3.3 it follows that each transitive set in $\tilde{K}$ is hyperbolic. By construction $f^{q(n)}: K' \to K'$ is semi-conjugate to a tent map T (or has zero topological entropy but then we are in the previous situation). Proposition 4.3 implies that the semi-conjugacy h precisely collapses points which enter K_{n+1}. It follows that h restricted to $\Omega(f) \cap \mathrm{cl}\,(K' \setminus K_{n+1})$ is almost injective: if $x, y \in \Omega(f) \cap \mathrm{cl}\,(K' \setminus K_{n+1})$ and $x \neq y$ then $h(x) = h(y)$ implies that the interval $[x, y]$ is eventually mapped into K_{n+1} and therefore (from the choice of x and y) into two endpoint of a component of K_{n+1}. It follows that h is at most two-to-one on this set. Since each component of K_{n+1} is periodic and since $f^{q(n)}: K' \to K'$ has no turning points in $K' \setminus K_{n+1}$, every turning point of T is periodic. From this one easily gets that $f^{q(n)}: \Omega(f) \cap K' \to \Omega(f) \cap K'$ is a subshift of finite type, see Van Strien (1981). It then follows immediately, that all attractors are contained in critical strata of $K_n \setminus K_{n+1}$. □

Example. In this example we construct a unimodal map $f: I \to I$ as above with positive entropy, semi-conjugate to the tent map with slope ± 2 and with an attracting fixed point in the boundary of I. Hence the complement of the basin of this attracting fixed point is a Cantor set. Let

$$f_{a,b}(x) = -\frac{1}{3}x^3 - \frac{1}{2}(a-b)x^2 + abx$$

where the parameters a and b are positive real numbers. Since $Df_{a,b}(x) = -(x+a)(x-b)$, it is easy to prove that $f_{a,b}$ has negative Schwarzian derivative, see Exercise 1.7 in Section IV.1. Moreover, it has a minimum at the point $-a$ and a maximum at the point b. If $ab < 1$, say $a = \frac{1}{2b}$, then 0 is an attracting fixed point. So let us consider the one parameter family of maps $g_b = f_{a(b),b}$, with $a(b) = \frac{1}{2b}$. Let $d = d(b)$ be the positive zero of g_b. Note that $g_b(b) = \frac{1}{6}b^3 + \frac{1}{4}b$ and therefore $g_b(b) \leq b$ if $b^2 \leq \frac{9}{2}$ and $g_b(b) > b$ if $b^2 > \frac{9}{2}$. Thus, $g_b(b) < d(b)$ if $b^2 \leq \frac{9}{2}$. On the other hand, $g_b(b) > d(b)$ if b is big enough. Therefore there exists b' such that $g_{b'}(b') = c$ and $g_b(b) < d(b)$ for $b < b'$. Thus, for $b \in (0, b']$, g_b maps the interval $[0, d(b)]$ into itself. If $\phi_b: [0, d(b)] \to I$ is the orientation preserving affine map, then $f_b = \phi_b \circ g_b \circ \phi_b^{-1}$ is a full family of unimodal maps with negative Schwarzian derivative. Therefore, for $b < b'$ but close to b', f_b satisfies the required conditions. f_b is semi-conjugate to the tent map with slopes ± 2.

Remark. 1. The numbers $\frac{q_{i+1}}{q_i} \in \mathbb{N}$ can be considered as the analogue of the continued fraction expansion for rotations of the circle. In the last chapter we shall consider unimodal infinitely renormalizable maps for which these fractions are bounded. 2. If f is unimodal and as in Theorem 3.4 then $K_n \setminus K_{n+1}$

never contains a turning point for $n < N$. So all, except of possibly one, transitive set is hyperbolic. If f is multimodal then K_{n+1} may not contain all turning points, and therefore Ω_n may contain a critical point. So in this case Ω_n is not hyperbolic. 3. In Jonker and Rand (1981) and also Collet and

Eckmann (1980), a so-called $*$-product is used to construct the decomposition in the unimodal case. This $*$-product is a way of expressing that the kneading invariant of renormalizable maps can be determined by the piecewise linear map T (which has at least one periodic turning point) and by the renormalized map $f^{q(2)}: I_2 \to I_2$. If f is unimodal then this $*$-product is quite simple to define. In the multimodal case the notation needed for such a $*$-product becomes somewhat more complicated, and therefore, just as in Van Strien (1981), we have not used this $*$-product here. These renormalizations can also be described by giving symbolic substitution rules. We shall not discuss this matter here either.

Exercise 4.1. Let $T: I \to I$ be the family of symmetric unimodal tent maps with slope $s \in (1, 2]$ and turning point c. Then the following statements hold: 1. If $\sqrt{2} < s \leq 2$ then for every open interval J there exists an integer n such that $T^n(J) \supset [T^2(c), T(c)]$. The non-wandering set of T is equal to $\{p\} \cup [T^2(c), T(c)]$ where p is the fixed point of T on the boundary of I. Moreover, T has no restrictive intervals in this case. 2. If k is such that $\sqrt{2} < s^{2^k} \leq 2$ then T has exactly k restrictive intervals $I_k \subset I_{k-1} \subset I_0 = I$ containing the turning point. I_i has period 2^i for $i = 0, 1, \ldots, k$. The non-wandering set of T consists of k periodic orbits which remain outside the forward orbit of I_k. Periodic orbits are dense in the non-wandering set of T and for each $x \in I$ there exists an integer n such that either $T^n(x) \in I_k$ or such that $T^n(x)$ is in one of these k periodic orbits. (Hint: For simplicity let $I = [0, 1]$ and assume that $T(0) = T(1) = 0$. First assume $\sqrt{2} < s \leq 2$. Exactly as in the proof of Proposition 4.4 one proves that if J is an interval then $T^n(J) \supset [T^2(c), T(c)]$ for n sufficiently large. Next show that $T: I \to I$ can be renormalized if $s^2 \leq 2$: T has two fixed points, the fixed point p in the boundary point of I and another, orientation reversing, fixed point p_1 in the interior of I. So let I_1 be the closed interval connecting p_1 to its symmetric p_1' (this is the point $p_1' \neq p_1$ with $T(p_1') = p_1$). Then T folds I_1 onto the interval $[p_1, T(c)]$. Since the length of this last interval is precisely equal to the length of I_1 times $(s/2) \times s$, we get from $s^2 \leq 2$ and since p_1 is orientation reversing, $T^2(I_1) = [T^2(c), p_1] \subset I_1$. Therefore, the first return map to I_1 is T^2. Moreover, $T^2: I_1 \to I_1$ is again a symmetric unimodal tent map with slope s^2. So $T_s: I \to I$ can be renormalized. Furthermore, any point $x \in I$ is mapped into $I_1 \cup T(I_1)$ after some iterate. Indeed, if $x \in (T(c), 1)$ then $T_s(x) \in (0, T^2(c))$. If $x, T(x), \ldots, T^j(x) \in (0, T^2(c))$ then the distance of $T^j(x)$ to 0 is s^j times the distance of x to 0. Hence, since $s > 1$, some iterate of x must be mapped into I_1. If $s^{2^k} \leq 2$ then this argument can be repeated k times.)

Exercise 4.2. Let T_s be a tent map with slope $\pm s$. If T_s has periodic points of odd period $p \geq 3$ then $s > \sqrt{2}$. (Hint: use the previous proposition.)

Exercise 4.3. Exercise 4.3 Let T be a piecewise linear l-modal map with slopes $\pm s$ where $s > 1$. Show that the union of the backward orbits of the turning points is dense. (Hint: clearly such a map cannot have periodic attractors. Moreover, it cannot have wandering intervals because then T would not have sensitive dependence. It follows from Section 3 that T has no homtervals and therefore the union of the backward orbits of the turning points is dense.)

Exercise 4.4. Show that any unimodal map with negative Schwarzian derivative which has no restrictive intervals (and no periodic attractors) has sensitive dependence on initial conditions. (Hint: in that case the map is conjugate to a tent map and so the statement follows from Proposition 4.4.)

Exercise 4.5. Suppose that f_a is a unimodal family of maps with negative Schwarzian derivative and suppose that $f_a(0) = 0$ and $Df_a(0) > 1$. Suppose that f_a has a periodic (possibly one-sided) attractor for each a in some interval Δ. Show that there exists an integer such that each periodic attractor of f_a has periodic $k \cdot 2^n$ for some $n \in \mathbb{N}$. More precisely, each component of the set $\{(a,x) \in \Delta \times N\,;\ \text{ such that } f_a^n(x) = x\}$ is a curve. If it is a closed curve, then the periodic orbit is created and later disappears again. (Hint: use the fact that the only bifurcations these maps can have are saddle-node and flip bifurcations.)

Exercise 4.6. Let f_a be as in the previous exercise and assume that $a \mapsto f_a$ is a full family. Let Δ be a maximal interval such that f_a has a periodic (possibly one-sided) attractor for each $a \in \Delta$. Let a^* be a boundary point of Δ. Show that one of the two following possibilities is satisfied. 1. f_{a^*} has a semi-stable periodic orbit of period k as in Case 2 of Remark 1 of the previous section and for $a \in \Delta$ close to a^*, this semi-stable orbit splits up in a hyperbolic stable orbit for f_a and a repelling orbit, both of period k. Furthermore for each $a \in \Delta$ each periodic orbit of f_a has period $k \cdot 2^n$ for some $n \in \mathbb{N}$. 2. f_{a^*} is infinitely renormalizable, and as $a \in \Delta$, $a \to a^*$, f_a has periodic orbits of period $k \cdot 2^n$ where $n \to \infty$. (Hint: the only way a (possibly one-sided) periodic attractor can disappear is when it is of type 2, or if its period tends to infinity. All other changes just imply the doubling or halving of the period of this periodic attractor. This follows from the exercises at the end of the previous section.)

Exercise 4.7. Consider $f_a(x) = ax(1-x)$. Show that no periodic halving of periodic attractors can occur as a increases. (Hint: use the results from Section II.10.)

5 Hyperbolicity in Smooth Maps

In this section we will prove a remarkable theorem due to Mañé (1985) which extends the hyperbolicity result we proved for maps with negative Schwarzian derivative to general C^2 maps. It states that a compact invariant set is hyperbolic if all the periodic points in the set are hyperbolic and if it contains no critical points. This result is remarkable because, for higher-dimensional systems, the assumption that all periodic points are hyperbolic usually gives no uniform estimates. And even when the periodic points carry some sort of hyperbolic structure it is in general not clear how to extend this hyperbolic structure to the closure of the set of periodic points. However, for one-dimensional maps one has good distortion results. These distortion results, and the inherent topological expansion, are the main ingredients for proving the strong results for one-dimensional maps.

The proof we give below uses many arguments from Van Strien (1990) and is much simpler than the proof in Mañé (1985). Mañé uses the lemma of Zorn to show that some maximal non-hyperbolic set cannot exist. Here a more direct argument is used. In Van Strien (1990), see also the next section, these ideas are used to show that Misiurewicz maps are almost hyperbolic. The special case where the map has negative Schwarzian derivative, is considered in the next section.

The main ingredient here is to control the distortion of each iterate f^n restricted to some interval T. As we have already seen in Section I.2, in order to get a bound which is independent of n for the distortion of $f^n|T$ we need two requirements:

1. the intervals $f^i(T), i = 0, \dots, n-1$ stay away from a neighbourhood of the critical points;
2. the sum of the lengths of these intervals, $\sum_{i=0}^{n-1} |f^i(T)|$ is bounded by a constant which does not depend on n.

In order to get a bound for the total length of the iterates of the interval T up to n we have two methods. The first one is to require disjointness. More generally, we get a bounded total length if the family of intervals have bounded intersection multiplicity. Here the *intersection multiplicity* of a collection of sets $X_1, \dots X_n$ is the maximal cardinality of a subcollection with non-empty intersection, i.e., it is equal to the maximum of distinct integers $1 \leq i_1, i_2, \dots, i_k \leq n$ such that $X_{i_1} \cap \dots \cap X_{i_k} \neq \emptyset$. Therefore, if the intersection multiplicity of the above collection of intervals is d, then the total length is bounded by d because the total length of N is equal to one.

The second method does not require bounded intersection multiplicity but does require a weak form of hyperbolicity. If there exists a constant $\lambda > 1$ independent of n, such that $|f^{i+1}(T)| > \lambda |f^i(T)|$ for all $i \leq n$ then the total length is bounded by $|f^n(T)| \cdot \frac{\lambda}{\lambda - 1} \leq \frac{\lambda}{\lambda - 1}$.

We have already used the first method to prove Denjoy's Theorem in Section I.2 and the second method was used in Section III.2. Here we will need a combination of these methods.

Theorem 5.1. **(Mañé)** *Let N be a compact interval of the real line and $f\colon N \to N$ be a C^2 map. Let U be a neighbourhood of the critical points of f. Then 1. All periodic orbits of f contained in $N \setminus U$ of sufficiently large period are hyperbolic repelling. 2. If all the periodic orbits of f contained in $N \setminus U$ are hyperbolic, then there exist $C > 0$ and $\lambda > 1$ such that*

$$|Df^n(x)| \geq C\lambda^n$$

whenever $f^i(x) \in N \setminus (U \cup B_0)$ for all $0 \leq i \leq n-1$, where B_0 is the union of the immediate basins of the periodic attractors of f contained in $N \setminus U$.

Before we prove this theorem we will derive some corollaries from it.

Corollary 5.1. *Let $f\colon N \to N$ be a C^2 map and $K \subset N$ be a compact forward invariant set. If K does not contain critical points, periodic attractors and non-hyperbolic periodic points of f, then it is a hyperbolic set.*

Proof of Corollary 1. Let U be a neighbourhood of the critical points of f which does not intersect K. By the first part of Theorem 5.1, there exists n_0 such that all periodic orbits of f contained in $N \setminus U$ of period $\geq n_0$ are hyperbolic repelling. Since the set of periodic points of period $\leq n_0$ is a compact set, there exists an open neighbourhood V of K which does not contain critical points and non-hyperbolic periodic orbits of f. Let W be a neighbourhood of K whose closure is contained in V. Using the transversality techniques of section III.2, we can approximate f by a C^2 map g which coincides with f in W and has the property that all the periodic points of g are hyperbolic. (Notice that the space $\mathcal{W}$ of C^2 maps that coincide with f in W is a closed subspace of $C^2(N,N)$ and therefore is a Baire space. Exactly as in Proposition 2.1, one can prove by induction that the set $\mathcal{W}_n \subset \mathcal{W}$ of maps whose periodic points of period less or equal n are hyperbolic is open and dense in $\mathcal{W}$. Therefore $\cap_{n=1}^{\infty} \mathcal{W}_n$ is residual, and in particular dense in $\mathcal{W}$.)

Let U be a neighbourhood of the critical points of g such that $U \cap K = \emptyset$. From Theorem 5.1 it follows that K is a hyperbolic set for g. Since $g = f$ on W, it follows that K is a hyperbolic set for f as well. □

From the above result we immediately get the following corollaries:

Corollary 5.2. *Let $f\colon N \to N$ be a C^2 map such that all critical points of f belong to the basin of a hyperbolic attracting periodic orbit. If the periodic points of f are hyperbolic then f satisfies the Axiom A.*

In Chapter V we will need that similar results hold for open sets of maps:

Corollary 5.3. *Let $f\colon N \to N$ be a C^2 map such that each of its periodic orbits is repelling and let U be a neighbourhood of the set of critical points of f. Then there exists a neighbourhood $\mathcal{N}$ of f in the C^1 topology and constants $C > 0$, $\lambda > 1$ such that if $g \in \mathcal{N}$ and $x, g(x), \dots, g^{n-1}(x) \notin U$ then*

$$|Dg^n(x)| \geq C\lambda^n.$$

Proof. From Theorem 5.1 there exist constants $C' > 0$ and $\lambda' > 1$ such that $x, f(x), \dots, f^{n-1}(x) \notin U$ implies $|Dg^n(x)| \geq C'(\lambda')^n$. It follows that there exists $\hat{n}$ such that if $x, \dots, f^{\hat{n}-1}(x)$ are outside U then $|Df^{\hat{n}}(x)| \geq 3$. Therefore, for g sufficiently close to f in the C^1 topology, one also has that $|Dg^{\hat{n}}(x)| \geq 2$ provided $x, \dots, g^{\hat{n}-1}(x) \notin U$. If $x, \dots, g^{n-1}(x) \notin U$ we can cut this orbit up into $k = [n/\hat{n}]$ pieces of length $\hat{n}$ and possibly a last piece of length $\leq \hat{n}$. On the last piece the map g does not contract more than $C_1 = \left(\inf |Dg_{N\setminus U}|\right)^{\hat{n}}$. From this it follows that there exist $\lambda > 1$ and $C > 0$ such that for each g which is C^1 close to f and for each x with $x, g(x), \dots, g^{n-1}(x) \notin U$ one has

$$|Dg^n(x)| \geq C_1 \cdot 2^k \geq C\lambda^n. \quad \square$$

From Corollary 2 and the transversality techniques of Section III.2, it follows that any C^r map, $r \geq 2$, whose critical points belong to the basin of hyperbolic attracting periodic points can be approximated closely, in the C^r topology, by a map satisfying the Axiom A. Therefore, the question of density of Axiom A is reduced to the following

Conjecture 1. The set of C^r maps, $r \geq 2$, whose critical points are all in the basins of the hyperbolic attracting periodic points is dense in $C^r(N, N)$.

Before proving Theorem 5.1 let us explain that in the space of C^r, $r \geq 2$, unimodal maps the above conjecture is reduced to a closing-lemma type of Conjecture.

Conjecture 2. Let $r \geq 2$. Any unimodal C^r map f whose critical point c is recurrent (i.e., $c \in \omega(c)$) can be approximated in the C^r topology by a unimodal map whose critical point is periodic.

Proof. that Conjecture 1 and 2 are equivalent in the case of unimodal maps Let us first show that if the critical point of a unimodal map f is not recurrent, hen f it can be C^r approximated by a unimodal map g whose critical point is in the basin of a hyperbolic attractor. Indeed, otherwise there exists a neighbourhood $\mathcal{N}$ of f in the C^r topology such that the turning point of any map in this neighbourhood is non-periodic; it follows that each $g \in \mathcal{N}$ is combinatorially equivalent to f and so its turning point is also not recurrent. Moreover, we may assume that the turning point of $g \in \mathcal{N}$ is not contained in the basin of a periodic attractor of g. Now choose $g \in \mathcal{N}$ so that all its periodic orbits are hyperbolic. Take an open neighbourhood V of c such that its closure contains no forward iterates of c, let P be the first return map to V and let D be the domain of this first return map. Clearly D is open; by the choice of V one has that c is not contained in the closure of any component of D. We claim c is in the closure of D. Before proving this claim let us show that the result follows from this claim. Indeed, because of the claim, there exist a sequence of components of D accumulating to c. For each component I of D one has $P|I = f^n$ for some n and f^n maps I monotonically onto V. So we can choose $x \in D$ arbitrarily close to c so that $P(x) = g^n(x) = c$. If we modify g on V to a map $\tilde{g}$ so that $\tilde{g}(c) = g(x)$ then $\tilde{g}^n(c) = c$ and so $\tilde{g}$ has the required properties. The C^r distance between g and $\tilde{g}$ can be made arbitrarily small by choosing x sufficiently close to c_1, compare this also to Exercise III.2.4. So it remains to prove the claim. In other words, we need to show that $\cup_{n \geq 1} g^{-n}(c)$ contains c in its closure, because then c is the closure of D. So assume by contradiction that there exists an interval neighbourhood I of c such that I contains no points of $\cup_{n\geq 1} g^{-n}(c)$. Since c is not in the basin of a periodic attractor it follows from Lemma 5.2 below that I is a wandering interval. Since the forward iterates of I do not accumulate to c and all iterates of I are disjoint, there exists a neighbourhood W of c and an integer n_0 such that $f^n(I) \cap W = \emptyset$ for all $n \geq n_0$. From Theorem 5.1 it follows that $|f^n(I)| \geq C\lambda^n$, a contradiction.

In the remainder of this section we shall prove Theorem 5.1. For the proof we need several lemmas. The first lemma gives the topological disjointness which is needed to get bounds on the non-linearity.

Lemma 5.1. *Let p be a periodic point of period n. If T is an interval that contains p and such that the intersection of $f^n(T)$ with the orbit of p is $\{p\}$, then the intersection multiplicity of the set of intervals $\{T, f(T), \ldots, f^{n-1}(T)\}$ is at most 2.*

Proof. Let J be the maximal interval which contains p and does not contain any other point of the orbit of p. Then $f^n(T) \subset J$. Let $O(p)$ denote the orbit of p. Notice that if $0 \leq i \leq n$ then $f^i(T) \cap O(p) = f^i(p)$. In fact, if $f^k(p) \in f^i(T)$ for some $0 \leq k \leq n$, $k \neq i$, then $f^{n-i+k}(p) = f^{n-i}(f^k(p)) \in f^n(T)$ and this is not possible because $f^{n-i+k}(p) \neq p$. From this the result follows immediately. □

The next lemma shows that forward iterates of an interval must either meet turning points or be contained in the basin of a periodic attractor.

Lemma 5.2. *Let $f\colon N \to N$ be a piecewise monotone continuous map. If $T \subset N$ is an interval such that $f^n|T\colon T \to f^n(T)$ is a homeomorphism for every $n \in \mathbb{N}$, then T is either a wandering interval or there exist an interval L and integers j, k such that $f^j(T) \subset L$, $f^k(L) \subset L$ and $f^k|L$ is monotone. In particular, if T is not a wandering interval then the ω-limit set of any point in T is a periodic orbit of period $\leq 2k$.*

Proof. Suppose T is not a wandering interval. If T is contained in the basin of a periodic attractor then the statement is obvious. So assume that it is not in the basin of a periodic attractor. Then the sequence of intervals $f^n(T)$ cannot be disjoint. Thus, there exist integers j, k such that $f^j(T) \cap f^{j+k}(T) \neq \emptyset$. Therefore $f^{j+k}(T) \cap f^{j+2k}(T) \neq \emptyset$ and, by induction, $f^{j+nk}(T) \cap f^{j+(n+1)k}(T) \neq \emptyset$. Hence, $L = \cup_{n=0}^{\infty} f^{j+nk}(T)$ is an interval. Clearly, $f^k(L) \subset L$ and $f^k|L$ is monotone. Therefore, all periodic points in L are fixed points of f^{2k} and every point in L is asymptotic to a periodic point. □Next it is necessary

to show that the periodic orbits carry some hyperbolic structure. First we shall show that they become 'very' expanding when the period is large. This and the disjointness from Lemma 5.1 will enable us to show that some intervals which are contained in each other decrease in length with a definite factor. This is an important step towards proving that the sum of the lengths of the iterates of some interval is universally bounded.

Lemma 5.3. *Let $f\colon N \to N$ be a C^2 map and U be a neighbourhood of the critical points of f. Then there exists a sequence K_n with $K_n \to \infty$ as $n \to \infty$ such that if p is a periodic point of period n whose orbit is entirely contained in $N \setminus U$ then $|Df^n(p)| > K_n$.*

Proof. Let M be an interval which contains N in its interior. It is easy to see that there exists a C^2 map $g\colon M \to M$ which coincides with f in $N \setminus U$ and satisfies the following conditions: i) the boundary of M is an attracting periodic point of g whose immediate basin is contained in $M \setminus N$; ii) every critical point c of g has interval neighbourhoods $W_c \subset V_c \subset U$ such that if J is a connected component of $V_c \setminus W_c$ then either $g(J)$ or $g^2(J)$ contains a repelling periodic point. Let $V = \cup V_c$ and $W = \cup W_c$. From the second property above, it follows that there exists a positive number ρ such that for every component J of $V \setminus W$ and any positive integer i, we have that $|g^i(J)| > \rho$. Since f coincides with g in $N \setminus U$ it is enough to prove the lemma for g on $M \setminus V$, instead of for f on $N \setminus U$.

Let us consider a periodic orbit of period m entirely contained in $M \setminus V$ and let p be a point in this orbit which is the closest to some critical point c, in the sense that there is no other point of the orbit of p between c and p. Let T be the maximal interval containing p such that $g^i(T) \cap W = \emptyset$ for $0 \le i < m$ and $g^m(T) \cap O(p) = \{p\}$. We claim that given ϵ, there exists m_0 such that for $m \ge m_0$ any interval T corresponding to a periodic orbit of period m has length $< \epsilon$. Indeed, if this were not the case, there would exist a sequence of intervals T_n of length at least equal to ϵ around a periodic point p_n of period $m_n \to \infty$ such that $g^i(T) \cap W = \emptyset$ for $0 \le i < m_n$ and $g^{m_n}(T) \cap O(p) = \{p\}$. By taking a subsequence, if necessary, we can assume that the intervals T_n converge to an interval T. As $m_n \to \infty$ as $n \to \infty$, $g^i(T_n) \cap W = \emptyset$ for $0 \le i \le m_n$, it follows that $g^i(T) \cap W = \emptyset$ for all $i \in \mathbb{N}$. By Corollary 2 of Theorem I.2.2, T is not a wandering interval since it does not accumulate at critical points. Hence, by Lemma 5.2, there exist integers j, k and an interval L such that $\lim_{n\to\infty} g^j(T_n) = g^j(T)$ and g^{2k} maps L homeomorphically and orientation preservingly into itself. If $g^j(T_n)$ contains no fixed point of $g^{2k}\colon L \to L$ then because $g^j(T_n) \to L$ as $n \to \infty$ one has for n sufficiently large that g^{2k} maps $g^j(T_n)$ homeomorphically into itself. But this contradicts that T_n contains a periodic point of $m_n > 2k$. So we may assume that for sufficiently large n, the intervals $g^j(T_n)$ contain a fixed point x of $g^{2k}\colon L \to L$. In this case let I_n be the interval in $g^j(T_n)$ connecting $f^j(p_n)$ and x. Since x is an orientation preserving fixed point of g^{2k} one has for $n > 2k + j$ that $g^{2k}|I_n$ is an orientation preserving homeomorphism with fixed point $x \in \partial I_n$ and either $g^{2k}(I_n) \subset I_n$ or $g^{2k}(I_n) \supset I_n$. Since $g^j(p_n) \in I_n$, in the first case $g^{j+2k}(p_n) \in g^{2k}(I_n) \subset I_n \subset g^j(T_n)$ and in the second case $g^j(p_n) \in I_n \subset g^{2k}(I_n) \subset g^{j+2k}(T_n)$. Because $n > j + 2k$ in both cases we get a contradiction with $g^{m_n}(T_n) \cap O(p_n) = \{p_n\}$. This finishes the proof of the claim.

Let T_2 be the component of $T \setminus \{p\}$ which is mapped by g^m outside $[c,p]$ and let T_1 be the other component. Since there is no other point of the orbit of p between c and p, we get, from the maximality of T either that $g^m(T_1)$ contains the component of V containing c or that there exists an integer $0 \le i < m$ such that $g^i(T_1)$ contains one component J of $V \setminus W$. Hence, from the choice of V, W and ρ above, we have that $|g^m(T_1)| \ge \rho$. Since $|T|$ tends to 0 as the period m goes to ∞, we get that $\frac{|g^m(T_1)|}{|T_1|} \to \infty$ as the period m goes to ∞. By Lemma 5.1, $\sum_{0 \le i \le m} |g^i(T)| \le 2|M|$. Therefore, since $g^i(T) \cap W = \emptyset$ for $0 \le i < m$, we get from Denjoy's theorem (or more precisely from Corollary 1 of Lemma I.2.1) that the distortion of g^m on T is bounded by a constant D_0 which is independent of m. Hence, $|Dg^m(p)| \ge \frac{1}{D_0} \cdot \frac{|g^m(T_1)|}{|T_1|}$ tends to infinity as $m \to \infty$. This proves the lemma. □

Let $f\colon N \to N$ be a C^2 map and $V \subset \mathrm{cl}\,(V) \subset U$ be neighbourhoods of the critical points of f.

Assume from now on, that all the periodic orbits entirely contained in $N \setminus V$ are hyperbolic.

Let B_0 be the immediate basin of the attracting periodic orbits which are entirely contained in $N \setminus V$. By Lemma 5.3, B_0 is a finite union of intervals. Note also that the boundary of B_0 is forward invariant. Moreover, let

$$\Gamma_n(X,Y) = \{x \in N\,;\ f^i(x) \in N \setminus \mathrm{cl}\,(X \cup Y),\ \text{for all } 0 \le i \le n\}.$$

Of course, $\Gamma_n(X,Y)$ is open.

Now we show that some intervals which are contained in each other decrease in length with a definite factor. From this we will get that the sum of the lengths of the iterates of some interval is universally bounded.

Lemma 5.4. *Let $f\colon N \to N$ be as above. Then there exist $\delta > 0$ and $\lambda > 1$ with the following property. If J is an interval such that $|J| < \delta$, $J, f(J), \ldots, f^{n-1}(J)$ are disjoint intervals contained in $N \setminus V$ and $f^n(J) \supset J$ then*

$$|Df^n(x)| > \lambda$$

for every $x \in J$.

Proof. Since the intervals $\{J, f(J), \ldots, f^{n-1}(J)\}$ are disjoint and contained in $N \setminus V$, there exists a constant D_0, independent of n, such that the distortion of f^n on J is bounded by D_0. Since $f^n(J) \supset J$, the interval J contains a periodic point of period n. Hence, by Lemma 5.3,

$$|Df^n(x)| \ge \frac{1}{D_0} \cdot K_n$$

for $x \in J$. Since $K_n \to \infty$ there exists n_0 such that

$$|Df^n(x)| \geq 2 \text{ if } n \geq n_0 \text{ and } x \in J.$$

Because all periodic points in $N \setminus V$ are hyperbolic, there is only a finite number of periodic points of period $\leq n_0$. Hence, there exist $\delta > 0$ and $\lambda > 1$ such that if x is a repelling periodic point of period $n \leq n_0$ then $|Df^n(y)| > \lambda$ for every y such that $|y - x| \leq 2\delta$. So, if $n \leq n_0$, $J, f(J), \ldots, f^{n-1}(J)$ are disjoint intervals contained in $N \setminus V$, and $f^n(J) \supset J$, then J contains a repelling periodic point of period n. Since $|J| < \delta$ we get that $|Df^n(x)| > \lambda$ for every $x \in J$. Thus we get again $|f^n(J)| > \lambda |J|$. □

The next two lemmas give some topological expansion.

Lemma 5.5. *Given $\epsilon > 0$ there exists a positive integer $n(\epsilon)$ such that if I is an interval with $|I| \geq \epsilon$ and $I \subset \Gamma_n(V, B_0)$ then $n < n(\epsilon)$.*

Proof. If the lemma is not true, there exist an $\epsilon > 0$ and a sequence I_n of intervals of length at least ϵ such that $I_n \subset \Gamma_n(V, B_0)$. By taking a subsequence we may assume that I_n converges to an interval I. It follows that $I \subset \Gamma_n$ for all n. Since the forward iterates of I do not accumulate at critical points, I cannot be a wandering interval. Hence, by Lemma 5.2. there exist integers j, k and an interval L such that $f^j(I) \subset L$ and f^k maps L monotonically into L. Hence every non-periodic point in L is in the basin of a periodic point. Consequently $f^j(I)$ must intersect B_0 and this implies that, for n big enough, $f^j(I_n)$ intersects B_0 which is a contradiction. □

Lemma 5.6. *Let* $\gamma_n = \max\{|J|;\ J \text{ is a connected component of } \Gamma_n(V, B_0)\}$. *Then $\gamma_n \to 0$ as $n \to \infty$.*

Proof. If this were not the case, then there would exist $\epsilon > 0$ and a sequence $I_{n(i)}$ of connected components of $\Gamma_n(V, B_0)$ such that $|I_{n(i)}| \geq \epsilon$. This contradicts the previous lemma. □

Now we come to the main concept which is needed to get the bound on the sum of the lengths of some orbit of intervals.

Definition. An interval J is *m-compatible* if for every $i < j \leq m$ such that $f^i(J) \cap f^j(J) \neq \emptyset$ we have $f^j(J) \supset f^i(J)$. We say that J is *(λ, m)-compatible* if it is m-compatible and if $|f^j(J)| > \lambda |f^i(J)|$ whenever $f^i(J)$ and $f^j(J)$ are both contained in an interval $f^k(J)$ for some $i < j < k \leq m$.

The use of this notion becomes immediately apparent from the next lemma.

Lemma 5.7. *Let $\lambda > 1$. If J is a (λ, m)-compatible interval then*

$$\sum_{i=0}^{m} |f^i(J)| < \frac{\lambda}{\lambda - 1} \cdot |N|.$$

Proof. Let k be the biggest integer such that there exist $i_1, \ldots, i_k$ such that none of the intervals $f^{i_1}(J), f^{i_2}(J), \ldots, f^{i_k}(J)$ are strictly contained in some interval $f^i(J)$, for some $i = 0, \ldots, m$. By the maximality of k and the fact that J is m-compatible we have that for any $0 \le i \le m$, the interval $f^i(J)$ is contained in one of the intervals $f^{i_l}(J)$ and the intervals $f^{i_1}(J), f^{i_2}(J), \ldots, f^{i_k}(J)$ are pairwise disjoint. Let us fix some l between 1 and k. Let $j_1 < j_2 < \cdots < j_s$ be the set of integers $i \le m$ such that $f^i(J) \subset f^{i_l}(J)$. Since J is (λ, m)-compatible, we have that $j_s = i_l$ and $|f^{j_{i+1}}(J)| > \lambda \cdot |f^{j_i}(J)|$ for all $i \le s - 1$. Thus, $\sum_{i=1}^{s} |f^{j_i}(J)| \le \frac{\lambda}{\lambda-1} \cdot |f^{i_l}(J)|$. Therefore,

$$\sum_{i=0}^{m} |f^i(J)| = \sum_{l=1}^{k} \sum_{f^i(J) \subset f^{i_l}(J)} |f^i(J)| \le \sum_{l=1}^{k} \frac{\lambda}{\lambda - 1} |f^{i_l}(J)| \le \frac{\lambda}{\lambda - 1} |N|,$$

because the intervals $f^{i_1}(J), \ldots, f^{i_k}(J)$ are disjoint. □

Let $\epsilon = \min\{\delta, |E|\,;\ E \text{ is a connected component of } U \setminus V\}$ where δ is as in Lemma 5.4. Moreover, let $n_0 = n(\epsilon)$ be as in Lemma 5.5. Let us show that connected components of $\Gamma_n(V, B_0)$ are (λ, m)-compatible.

Lemma 5.8. *Let J be a connected component of $\Gamma_n(V, B_0)$ which contains a point of $\Gamma_n(U, B_0)$. Then J is a $(\lambda, n - n_0)$-compatible interval.*

Proof. First, let us prove that J is an $(n - n_0)$-compatible interval. Let $0 \le i < j \le n - n_0$ be such that $f^i(J) \cap f^j(J) \ne \emptyset$. Suppose, by contradiction, that $f^j(J)$ does not contain $f^i(J)$. Then the open interval $f^i(J)$ contains a boundary point $f^j(a)$ of $f^j(J)$ where a is one of the boundary points of J. We claim that if $0 \le l \le j$ then $f^l(a)$ is not in the boundary of B_0. Indeed, otherwise $f^{j-l}(f^l(a))$ is also in the boundary of B_0 but this is impossible because $f^j(a)$ is contained in the open interval $f^i(J)$ and $J \subset \Gamma_n(V, B_0)$. Furthermore, we claim that for $l \le n - n_0$ the endpoint $f^l(a)$ does not belong to the boundary of V. Indeed, if this is not the case, then $f^l(J)$ would contain a component of $U \setminus V$ since it contains a point outside U. Hence, by Lemma 5.5 and the definition of ϵ, $f^l(J)$ cannot be in $\Gamma_{n-l}(V, B_0)$ since $n - l \ge n_0$ and therefore this contradicts $J \subset \Gamma_n(V, B_0)$ and proves our second claim. From these two claims it follows that there exists a neighbourhood W of a such that $f^j(W) \subset f^i(J)$ and $f^l(W) \cap (B_0 \cup V) = \emptyset$ for $l \le j$. Since $J \subset \Gamma_n(V, B_0)$ and since $n - i > n - j$ and $f^j(W) \subset f^i(J) \in \Gamma_{n-i}(V, B_0)$ we get that $J \cup W \subset \Gamma_n(V, B_0)$. This contradicts the fact that J is a connected component of $\Gamma_n(V, B_0)$. Therefore J is an $(n - n_0)$-compatible interval.

Let us now prove that J is $(\lambda, n - n_0)$-compatible. Let $0 \le i < j < k \le n - n_0$ be such that $f^i(J) \cup f^j(J) \subset f^k(J)$ and assume that j is the smallest integer satisfying this condition. Let $E \subset \Gamma_{j-i}(V, B_0)$ be the maximal interval containing $f^i(J)$ such that $f^{j-i}(E) \subset f^k(J)$. We claim that $f^{j-i}(E) = f^k(J)$, $E \subset f^k(E)$ and the intervals $E, f(E) \dots, f^{j-i-1}(E)$ are pairwise disjoint. Let us finish the proof, assuming the claim. From Lemma 5.4 we get that $|Df^{j-i}(x)| > \lambda$ for each $x \in E$ and therefore $|f^j(J)| > \lambda |f^i(J)|$ if j is the smallest integer between i and k such that $f^i(J) \cup f^j(J) \subset f^k(J)$. If it is not the smallest integer with the above property we use the above estimate for the smallest integer and repeat the argument taking this integer instead of j. After a finite number of, say s, steps we get $|f^j(J)| > \lambda^s \cdot |f^i(J)| > \lambda |f^i(J)|$. So it remains to prove the last claim. We note that $f^l(E) \cap \partial V = \emptyset$ if $0 \leq l < j - i$. Indeed, if this is not true then $|f^l(E)| > \epsilon$ because $f^l(f^i(J)) \subset f^l(E)$ contains a point not in U. Since $f^{j-i}(E) \subset f^k(J)$ we must have that $f^l(E) \subset \Gamma_{((j-i)-l)+(n-k)}$. This is impossible because of Lemma 5.5 and because $((j - i) - l) + (n - k) \geq n_0 = n(\epsilon)$. Therefore, $f^l(E) \cap \partial V = \emptyset$. Using this and the same argument as in the first part of this proof, we get that E is $(j - i)$-compatible. Indeed suppose, by contradiction, that $f^k(J)$ contains a boundary point of $f^{j-i}(E)$, say $f^{j-i}(x)$ where x is a boundary point of E. Since B_0 is forward invariant, we get as before that $f^l(x) \notin \partial B_0$ for every $l < j - i$. Hence, we get as before a neighbourhood X of x such that $f^{j-i}(X) \subset f^k(J)$ and $f^l(X) \cap (V \cup B_0) = \emptyset$ for $l \leq j - i$. This contradicts the maximality of E and proves that $f^{j-i}(E) = f^k(J)$. Let us now prove the statement about disjointness. If, for some $0 \leq l < j - i$ we have that $f^l(E) \cap f^k(J) \neq \emptyset$ then $f^{i+l}(J) \subset f^l(E) \subset f^{j-i}(E) = f^k(J)$. This is a contradiction because $i < i + l < j$. Hence the intervals $E, f(E), \dots, f^{j-i-1}(E)$ are pairwise disjoint. This proves the claim and finishes the proof of the lemma. □

Combining all this gives the following bound on the sum of the lengths of iterates of some interval. From this we will immediately get the required distortion results.

Proposition 5.1. *Let $f \colon N \to N$ be a C^2 map, $V \subset \mathrm{int}\,(U) \subset U$ be neighbourhoods of the set of critical points of f such that all periodic orbits which are entirely contained in $N \setminus V$ are hyperbolic. Let B_0 denote the union of the immediate basins of the periodic attractors of f whose orbits are entirely contained in $N \setminus V$. Then there exists a positive constant C_0 such that if $J \subset \Gamma_n(V, B_0)$ is an interval, then*

$$\sum_{i=0}^{n} |f^i(J)| < C_0.$$

Proof. Let J be a connected component of $\Gamma_n(V, B_0)$, with $n > n_0$, where n_0 is as in the previous lemma. Then, by Lemma 5.8 and 5.7 we have

$$\sum_{i=0}^{n} |f^i(J)| = \sum_{i=0}^{n-n_0-1} |f^i(J)| + \sum_{i=n-n_0}^{n} |f^i(J)| < \frac{\lambda}{\lambda - 1}|N| + n_0 \cdot |N| = C_0.$$

If $n \leq n_0$ the above sum is bounded by $n_0|N|$. □

Proof of Theorem 5.1. Because of Proposition 5.1 one can prove this theorem exactly in the same way as Theorem 3.2. As in this theorem it is sufficient to show that there exists k such that for each $x \in \Gamma_n(U, B_0)$ and each $n > k$, we have $|Df^n(x)| > 1$. So let us show this by contradiction. Assume that there exist sequences x_i and $n(i) \to \infty$ such that $\limsup_i |Df^{n(i)}(x_i)| \leq 1$ and $x_i \in \Gamma_{n(i)}(U, B_0)$. Since all periodic points of f which stay outside a neighbourhood of the set of critical points and which have sufficiently large period are hyperbolic repelling and since $N \setminus U$ does not contain non-hyperbolic periodic points, we can take a neighbourhood V of $C(f)$ such that the closure of V is contained in U and such that f does not have a non-hyperbolic orbit contained in $N \setminus V$. Let $\delta > 0$ be smaller than any connected component of $U \setminus V$ and also smaller than the distance between any two points in the boundary of B_0. Let J_i be the connected component of $\Gamma_{n(i)}(V, B_0)$ which contains x_i.

We claim that there exists $m(i) < n(i)$ such that $|f^{m(i)}(J_i)| > \delta$. Indeed, by the maximality of J_i, there exists either $l < n(i)$ such that $f^l(J_i)$ contains a point of the boundary of V or there exist integers $l_1 < l_2 < n(i)$ such that $f^{l_1}(x) \in \partial B_0$ and $f^{l_2}(y) \in \partial B_0$, where x, y are the endpoints of J_i. In the first case $f^l(J_i)$ contains a component of $U \setminus V$ because $f^l(J_i)$ contains $f^l(x_i)$ which is not in U. In the second case $f^{l_2}(J_i)$ contains two points in the boundary of B_0. In both cases the claim is verified.

By Proposition 5.1, the distortion of $f^{n(i)}$ on J_i is bounded. Since $\limsup_i |Df^{n(i)}(x_i)| \leq 1$ we get that $|f^{n(i)}(J_i)|/|J_i|$ is universally bounded. On the other hand, by Lemma 5.6, $|J_i| \to 0$ as $i \to \infty$. Hence, $|f^{n(i)}(J_i)| \to 0$ as $i \to \infty$. Since $|f^{m(i)}(J_i)| > \delta$, we get that $(n(i) - m(i)) \to \infty$. By taking a subsequence we may assume that $f^{m(i)}(J_i)$ converges to an interval E. As $n(i) - m(i) \to \infty$ we get that $E \subset \Gamma_k(V, B_0)$ for all k. This contradicts Lemma 5.6 and proves the theorem. □

Remark. 1. Theorem 5.1 and its corollaries also hold for C^2 circle maps. If $f: S^1 \to S^1$ is a C^2 circle map which has either a critical point or an attracting periodic point we can deduce the result for f directly from Theorem 5.1. In fact, as we have done several times before we can construct an interval map which has the same dynamics as f outside a neighbourhood of a pre-image of the critical point (or of the attracting periodic point) and we can modify the new map in this neighbourhood so that it satisfies the hypothesis of Theorem 5.1. If there is no critical point and no attracting periodic point then the map f is conjugate to a covering map of the circle and we need to modify the proof in this case. We refer to Mañé (1985) or Van Strien (1990) for the details. 2. In

the next section we will state a stronger result from which Theorem 5.1 easily follows. This result will also deal with points which come close to critical points. The main problem will be to control the non-linearity of high iterates. Because of the critical points, we cannot hope to bound the non-linearity with the tools of Section I.1. In the next chapter we will therefore develop tools which can be used to analyze this non-linearity. In particular, in the next section we will give a proof of this generalization under the additional condition that the Schwarzian derivative of f is negative.

6 Misiurewicz Maps are Almost Hyperbolic

In Theorem 5.1 only compact sets not containing critical points were considered. Because of this, the bounded non-linearity results of Section I.1 suffice for its proof. In Van Strien (1990) a more general version of Theorem 5.1 is proved which also deals with points whose orbit comes very close to critical points. This theorem can be stated as follows. Let f be C^2 and let $C(f)$ be the set of critical points of f, i.e., $C(f) = \{c;\ Df(c) = 0\}$. We say that f satisfies the *Misiurewicz* condition if there exists a neighbourhood W of $C(f)$ such that

$$\bigcup_{n\geq 1} f^n(C(f)) \cap W = \emptyset.$$

This means that the forward orbit of each critical point does not accumulate onto any critical point. Often we shall also assume that f has no periodic attractors. In this section we shall show that these maps are almost hyperbolic. So let us call a maximal interval on which f^n is a diffeomorphism a *branch* of f^n. The first result is a distortion result: f^n is almost 'polynomial' on branches of f^n. In spite of the unbounded non-linearity, the branches of iterates are well controlled: they are polynomial-like. To be more precise, there exists a finite collection $\mathcal{P}_n$ of C^∞ maps $g\colon [0,1] \to [0,1]$ with $Dg(x) \neq 0$ for $x \in (0,1)$ which have critical points of order $\leq n$ for $x = 0$ or $x = 1$. This collection has the following property. For each $n \in \mathbb{N}$ the restriction of the map f^n to any branch is equal to one of the maps g up to a map which has bounded distortion.

Theorem 6.1 (Iterates are quasi-polynomial on branches). *Suppose that $f\colon N \to N$ is a C^2 Misiurewicz map whose periodic points are all hyperbolic and repelling. Furthermore, assume that all critical points of f are non-flat. Then there exists a finite collection of smooth maps $\mathcal{P}$ and $K < \infty$ with the following property. For any $n \in \mathbb{N}$ and any branch I of f^n there exists $g \in \mathcal{P}$ such that, up to scaling, $f^n|I$ is equal to $h \circ g$. Here $h\colon [0,1] \to [0,1]$ is a diffeomorphism, which does depend on n and I, but such that*

$$\frac{|Dh(x)|}{|Dh(y)|} \leq K \text{ for all } x, y \in [0,1].$$

Clearly the set $\mathcal{P}_n$ can be chosen in such a way that it consists only of polynomial maps.

Corollary 6.1. *For each Misiurewicz map as above there exist $K < \infty$ and $\tilde{n} \in \mathbb{N}$ such that for each $n \geq \tilde{n}$ and each interval I on which f^n is a diffeomorphism, $f^n|I$ is a composition of a map with distortion bounded by K and a simple 'polynomial map'.*

For the proof Theorem 6.1 in the C^2 case we refer to Van Strien (1990). So let us assume that f is C^3 and has negative Schwarzian derivative. In this case the proof is quite easy using the Koebe Principle. We shall discuss this principle in detail in the next chapter but let us state it in the version we need here.

Let $U \subset V$ be two intervals. We say that V contains a *δ-scaled neighbourhood of U* if both components of $V \setminus U$ have at least length $\delta \cdot |U|$.

Koebe Principle. Let f have negative Schwarzian derivative. Then for each $\delta > 0$ there exists $K < \infty$ such that if $I \subset J$ are intervals, $f^n|J$ is a diffeomorphism and $f^n(J)$ contains a δ-scaled neighbourhood of $f^n(I)$ then the distortion of $f^n|I$ is bounded by K.

Proof. See the next chapter. □

For the proof of Theorem 6.1 we need the following lemma.

Lemma 6.1. *If $f\colon [0,1] \to [0,1]$ has no periodic attractors and no wandering intervals then there exists for each $\tau > 0$ some $\tilde{n} < \infty$ such that if T is an interval of length $\geq \tau$ then $f^{\tilde{n}}|T$ is not a diffeomorphism. Moreover, if f is a Misiurewicz map there exists $\rho > 0$ such that if T is a maximal interval on which f^n is monotone then $|f^n(T)| \geq \rho$.*

Proof. If the first statement did not hold one could find a sequence $k_n \to \infty$ and intervals T_n with $|T_n| \geq \tau$ and $f^{k_n}|T_n$ monotone. Taking a limit of these intervals we would get an interval T of length $\geq \tau$ with $f^n|T$ monotone for all $n \geq 0$. So T would be a homterval, but since f has no wandering intervals and no periodic attractors this is not possible, see Corollary 1 of Section II.3.

If $T = [a, b]$ is a maximal interval on which f^n is monotone then there exist $k(a), k(b) < n$ such that $f^{k(a)}(a), f^{k(b)}(b) \in C(f)$. If $k(a) \leq k(b)$ then $f^{k(b}(a) \notin W$ because f is Misiurewicz and therefore $f^k(T)$ contains a component of the set $W \setminus C(f)$. From the first part of this lemma, $n - k \leq \tilde{n}$ and so there exists a constant $\rho > 0$ (which does not depend on n) such that $|f^n(T)| \geq \rho$. □

Exercise 6.1. Show that Theorem 6.1 and Lemma 6.1 imply that each Misiurewicz map f whose periodic points are hyperbolic and repelling has an absolutely continuous invariant probability measure. (Hint: as we will see in Section V.4 the map f has an absolutely continuous invariant probability measure if for each $\epsilon > 0$ there exists $\delta > 0$ such that for all $n \in \mathbb{N}$ one has $|f^{-n}(A)| \leq \epsilon$ whenever A is a measurable set with $|A| \leq \delta$. But this follows immediately from the fact that f^n is quasi-polynomial on each branch I_n and from the fact that $|f^n(I_n)| \geq \rho$.)

Proof of Theorem 6.1. Let V be a neighbourhood of $C(f)$ such that the neighbourhood W from the definition of Misiurewicz maps contains a 3-scaled neighbourhood of V and let $\tau = |V|$. Let $\tilde{n}$ be the number corresponding to τ from the previous lemma. Now take $n \in \mathbb{N}$, a branch I of f^n and let $n_0 \in \mathbb{N}$ be the largest integer $< n_1 = n - \tilde{n}$ such that $f^{n_0}(I) \cap V \neq \emptyset$. If no such integer exists then take $n_0 = 0$. Because f has no wandering intervals we get from the previous lemma and since $f^{n-n_0}|f^{n_0}(I)$ is a diffeomorphism that

$$|f^{n_0}(I)| \leq \tau.$$

Now by assumption the critical points of f^n are outside W. We claim that

$$\frac{Df^{n_0}(x)}{Df^{n_0}(y)} \leq K \text{ for all } x, y \in I.$$

If $n_0 = 0$ then this is obvious. But otherwise $f^{n_0}(I) \cap V \neq \emptyset$ and $|f^{n_0}(I)| \leq \tau$ and so by the choice of V and τ we get $f^{n_0}(I) \subset W$. It follows that there exists an interval $J \supset I$ on which f^{n_0} is a diffeomorphism for which $f^{n_0}(J) \supset W$. Hence $f^{n_0}(J)$ contains a τ-scaled neighbourhood of $f^{n_0}(I)$. Using the Koebe Principle it follows that $f^{n_0}|I$ has a distortion which is bounded by some universal constant K and the claim follows.

Next let $I' = f^{n_0}(I)$. Since $f^i(I') \cap V = \emptyset$ for $i = 1, \ldots, n_1 - n_0 - 1$ it follows similarly that

$$f^{n_1-n_0-1}|f(I') \mapsto I'' = f^{n_1-n_0}(I)$$

is also a diffeomorphism with a universally bounded distortion. So $f^n|I$ is the composition of the following four maps

$$f^{n_0}\colon I \mapsto I', \ , \ f\colon I' \mapsto f(I') \ ,$$

$$f^{n_1-n_0-1}\colon f(I') \to f^{n_1-n_0}(I) \text{ and } f^{\tilde{n}}\colon f^{n_1-n_0}(I) \to f^n(I).$$

The first and third of these maps have bounded distortion and the second and last of these maps are simply f and $f^{\tilde{n}}$. From this one can easily deduce the theorem. □

Theorem 6.2 (Misiurewicz maps are globally expanding). *Suppose that $f\colon N \to N$ is a C^2 Misiurewicz map with non-flat critical points and with all its periodic points hyperbolic and repelling. Then there exist $K > 0$ and $\rho \in (0,1)$ such that each interval on which f^n is a diffeomorphism has at most size $K \cdot \rho^n$. Moreover, let I_0 and I_1 be two maximal intervals on which f^n is a diffeomorphism with a common boundary point then $\frac{1}{K} \leq |I_0|/|I_1| \leq K$.*

Proof of Theorem 6.2. Again we will only prove this here in the case that f has negative Schwarzian derivative. For the general case, see Van Strien (1990). Let I_n be a maximal interval on which f^n is a diffeomorphism such that $f^{n+1}|I_n$ is not a diffeomorphism. Since I_n is maximal, the endpoints of $f^n(I_n)$ are critical values of f^n and therefore outside W and since $f^{n+1}|I_n$ is not a diffeomorphism $f^n(I_n)$ contains a critical point. So $f^n(I_n)$ contains a component of W. So let I^i_{n+1} be the maximal intervals in I_n on which f^{n+1} is a diffeomorphism. Then $f^n(I^i_{n+1})$ contains a component of $W \setminus C(f)$. So the intervals $f^n(I_n)$, $f^n(I^i_{n+1})$ all have length between τ and 1. From the bounded distortion statement in the previous theorem it follows that there exists a universal constants $\tilde{\tau}_1, \tilde{\tau}_2 \in (0,1)$ such that

$$\tilde{\tau}_1 \leq \frac{|I^i_{n+1}|}{|I_n|} \leq \tilde{\tau}_2. \tag{6.1}$$

Furthermore, $f^{n+k}|I^i_{n+1}$ is *not* a diffeomorphism for some $k \leq \tilde{n}$ where $\tilde{n}$ is the number from Lemma 6.1. From the previous argument for each maximal interval I_{n+k} inside $I^i_{n+1} \subset I_n$ on which f^{n+k} is a diffeomorphism one has

$$|I_{n+k}| \leq \tilde{\tau}|I^i_{n+1}| \leq \tilde{\tau}|I_n|. \tag{6.2}$$

Clearly (6.1) and (6.2) imply the theorem. □

Exercise 6.2. Show that the previous theorem implies that any conjugacy between two Misiurewicz maps as above is quasisymmetric. We should note that later on, in Chapter VI, we shall see that two conjugate infinitely renormalizable unimodal maps of 'bounded type' are also quasi-symmetrically conjugate. Whether all smooth conjugate maps are quasi-symmetrically conjugate is an open question. We shall see in Theorem VI.4.2b that this exercise implies that there are no two parameters a for which the maps $1 - ax^2$ are conjugate and Misiurewicz (without a periodic attractor). (Hint: Let f and g be the diffeomorphisms and h the conjugacy. For any interval of the form $(x-a, x+a)$ simply choose n minimal so that $(x-a, x)$ and $(x, x+a)$ both completely contain at least one maximal interval on which f^n is a diffeomorphism. Now let $\mathcal{I}_n(f)$ be the collection of interval I which are maximal intervals on which f^n is a diffeomorphism. From the previous theorem it follows that there exists a universal number R such that at most R intervals from $\mathcal{I}_n(f)$ intersect $(x-a, x+a)$. Since this statement holds also for the other diffeomorphism g it follows that $h(x-a, x+a)$ is contained in the union (of the closure) of at most R elements of $\mathcal{I}_n(g)$ and that $h(x-a, x)$ and $h(x, x+a)$ both contain an element from this collection. It follows that the length of these intervals is comparable.)

Theorem 6.3 (Quasi-hyperbolicity for Misiurewicz maps). *Suppose that $f\colon N \to N$ is a C^2 Misiurewicz map with non-flat critical points and with all its periodic points hyperbolic and repelling. Then for each sufficiently small neighbourhood W of $C(f)$ there exist constants $\lambda > 1$, $C > 0$ such that for each $x \in N$ one has the following:*

i) if $f^j(x) \notin W$ for $0 \leq j \leq k-1$ then

$$|Df^k(x)| \geq C\lambda^k;$$

ii) if $f^k(x) \in W$ then

$$|Df^k(x)| \geq C\lambda^k;$$

iii) without any conditions one gets,

$$|Df^k(x)| \geq C\lambda^k \inf_{j=0,\dots,k-1} |Df(f^j(x))|.$$

Remark. Theorem 6.3 implies that the periodic points of f are uniformly expanding: if n is the period of a periodic point p then $|Df^n(p)| \geq C\lambda^n$. The content of statement iii) of this theorem is that the only non-hyperbolic feature of a Misiurewicz map is that one 'picks up' one small derivative even if one enters W many times.

Proof of Theorem 6.3. Although this result holds for general C^2 maps here we will only again prove the result for C^3 maps with negative Schwarzian derivative. For the proof in the general case, see Van Strien (1990). Let W_0 be a neighbourhood of $C(f)$ such that $f^n(C(f)) \cap W_0 = \emptyset$ for all $n \geq 1$. Of course we may assume that the closure of W is contained in the interior of W_0. Let $\delta > 0$ be such that the components of $W_0 \setminus W$ all have at least length δ. Statement i) follows immediately from Theorem 5.1 (for this it was not even necessary to assume that the critical points of f are non-flat). So let us prove ii): let x be such that $f^k(x) \in W$. Let I be the largest interval containing x on which $f^k|I$ is monotone. Since f is Misiurewicz all critical values of f^k are outside W_0. Therefore and because $f^k(x) \in W$, $f^k(I)$ contains one of the components of W_0. Therefore $f^k(I)$ contains a δ neighbourhood of $f^k(x)$ and by the Koebe Principle

$$|Df^n(x)| \geq \frac{1}{K}\frac{\delta}{|I|}.$$

From the previous theorem, $|I| \leq K\rho^n$ and therefore ii) follows. So let us prove iii). If there exists no $0 \leq l \leq k$ for which $f^l(x) \in W$ then the result follows immediately from i). Otherwise let $l \leq k$ be the largest number such that $f^l(x) \in W$ and we get from ii)

$$|Df^l(x)| \geq C\lambda^l$$

and from i) one has

$$|Df^{k-l-1}(f^{l+1}(x))| \geq C\lambda^{k-l-1}.$$

Combining this gives iii). □

The next theorem shows that the estimates from the previous theorem also hold for an open set of maps.

Theorem 6.4. *Suppose that $f\colon N \to N$ is a C^2 Misiurewicz map with non-flat critical points and with all its periodic orbits hyperbolic and repelling. Then there exist $C > 0$ and $\lambda > 1$ and a neighbourhood W of $C(f)$ such that for any neighbourhood $U \subset W$ of $C(f)$ there exists a neighbourhood $\mathcal{N}$ of f in the C^1 topology such that for each $g \in \mathcal{N}$ one has*

i) if $x, \ldots, g^n(x) \notin W$ then

$$|Dg^n(x)| \geq C\lambda^n;$$

ii) if $x, g(x), \ldots, g^{n-1}(x) \notin U$ and $g^n(x) \in W$ then

$$|Dg^n(x)| \geq C\lambda^n;$$

iii) if $g^j(x) \notin U$ for $0 \leq j \leq n-1$ but not necessarily $g^n(x) \in W$, then

$$|Dg^n(x)| \geq C\lambda^n \inf_{j=0,\ldots,n-1} |Dg(g^j(x))|.$$

Proof. By the previous theorem there exist a neighbourhood W of $C(f)$, $\lambda_1 > 1$ and $C_1 > 0$ such that

$$|Df^n(x)| \geq C_1(\lambda_1)^n$$

whenever $f^n(x) \in W$. Of course this statement holds for the same values of λ_1 and C_1 if we shrink W. So choose $\lambda_2 \in (1, \lambda_1)$. Since f is Misiurewicz we may assume that W is so small that $f^n(C(f)) \cap W = \emptyset$ for all $n \geq 1$ and moreover so that for any x with $x, f^n(x) \in W$ one has $C_1(\lambda_1)^n > 2(\lambda_2)^n$. So if $x, f^n(x) \in W$ then

$$|Df^n(x)| \geq 2(\lambda_2)^n. \tag{6.3}$$

Furthermore, by Theorem 5.1 and Theorem 2.2, there exist $\lambda_3 \in (1, \lambda_2)$ and $C_3 > 0$ such that for each g which is C^1 close to f and for each x with $x, g(x), \ldots, g^{n-1}(x) \notin W$ one has

$$|Dg^n(x)| \geq C_3(\lambda_3)^n. \tag{6.4}$$

This proves the first statement. Now let $\lambda \in (1, \lambda_3)$ and choose $\hat{n}$ so large that for each $n \geq \hat{n}$ and each $g \in \mathcal{N}$

$$C_3(\lambda_3)^{n-1} \inf_{x \notin U} |Dg(x)| \geq \lambda^n. \tag{6.5}$$

Clearly $\hat{n}$ strongly depends on the size of U. By (6.3) there exists a neighbourhood of f such that for each g in this neighbourhood and each x such that $x, g^n(x) \in W$ for some $n \leq \hat{n}$, one has

$$|Dg^n(x)| \geq \lambda^n. \tag{6.6}$$

(Again the neighbourhood of f depends strongly on $\hat{n}$.) Now we claim that

$$x \in W \setminus U,\ g^k(x) \in W \text{ implies } |Dg^k(x)| \geq \lambda^k. \tag{6.7}$$

Indeed, take any point $x \in W \setminus U$ and let $k \leq n$ be the smallest integer such that $g^k(x) \in W$. If $k \geq \hat{n}$ then because $g(x), \ldots, g^{k-1}(x) \notin W$ we get from (6.4) and (6.5) that

$$|Dg^k(x)| = |Dg^{k-1}(g(x))| \cdot |Dg(x)| \geq C_3(\lambda_3)^{k-1} \cdot |Dg(x)| \geq \lambda^k.$$

On the other hand if $k < \hat{n}$ then (6.7) follows from (6.6).

Let x be such that $x, g(x), \ldots, g^{n-1}(x) \notin U$ and let $0 \leq k_1 < k_2 < \cdots < k_l \leq n$ be the integers so that $g^{k_i}(x) \in W$. From (6.7),

$$|Dg^{k_{i+1}-k_i}(g^{k_i}(x))| \geq \lambda^{k_{i+1}-k_i} \tag{6.8}$$

for $i = 1, \dots, l-1$. Furthermore, from (6.4) one has

$$|Dg^{k_1}(x)| \geq C_3 \lambda^{k_1}. \tag{6.9}$$

If $g^n(x) \in W$ then $k_l = n$ and combining (6.8) and (6.9) gives the second statement of the theorem. If $g^n(x) \notin W$ then $k_l < n$ and $g^k(x) \notin W$ for $k = k_l + 1, \dots, n$. Hence, by (6.4),

$$|Dg^{n-k_l-1}(g^{k_l+1}(x))| \geq C_3 \lambda^{n-k_l-1}$$

and, using (6.8) and (6.9),

$$|Dg^n(x)| \geq C_3 \lambda^{n-1} \cdot |Dg(g^{k_l}(x))|.$$

This concludes the proof of the third statement of the theorem. □

7 Some Further Remarks and Open Questions

As we mentioned, the main open problem in this chapter is the following:

Conjecture 1. The set of Axiom A maps is dense in the C^r topology for $r \geq 2$.

The case where $r = 1$ was proved by Jakobson (1981), see Theorem III.2.2. Of course the example due to Gutierrez (1987) shows that in general a map cannot be perturbed to an Axiom A map by a C^2 small perturbation which is localized in an arbitrarily small neighbourhood of one point. In other words one will have to study global perturbations in order to prove Conjecture 1. A partial result to the C^2 closing lemma is given by Contreras (1991) and Tsujii (1992a) and (1992d). Roughly speaking, these results show that C^2 generically, maps are either Axiom A or such that they have an ergodic invariant measure with zero Liapounov exponent (note that these measures could have their support on a rather small set). A conjecture related to the previous one is of course:

Conjecture 2. The set of parameters for which $f_a(x) = ax(1-x)$ is Axiom A forms a dense set in $[0, 4]$.

Recently, Świątek has proved this old and famous conjecture which dates back to the 1930's, see Świątek (1992b). His proof uses complex extensions as in Chapter VI.

Let us be somewhat more specific. Take $f_a(x) = ax(1-x)$ and define

$$\begin{aligned} A &= \{a\,;\ f_a \text{ has a hyperbolic periodic attractor} \\ &\qquad \text{which is not a super-attractor}\}, \\ \hat{A} &= \{a\,;\ f_a \text{ has a non-hyperbolic periodic attractor}\}, \\ \tilde{A} &= \{a\,;\ f_a \text{ has a periodic super-attractor}\}, \\ NA_1 &= \{a\,;\ f_a \text{ has no periodic attractors and is only} \\ &\qquad \text{finitely often renormalizable}\}, \\ NA_2 &= \{a\,;\ f_a \text{ has no periodic attractors and is} \\ &\qquad \text{infinitely often renormalizable}\}. \end{aligned}$$

Of course, A is the set of parameters for which f_a is Axiom A. So the previous conjecture states that $\hat{A} \cup \tilde{A} \cup NA_1 \cup NA_2$ has no interior points. It follows from Section II.10 (or Section VI.4) that $\hat{A}$ and $\tilde{A}$ have no interior points. Yoccoz (1990) proved that NA_1 contains no interior points. The proof of this result goes beyond the scope of this book. From the results of Sullivan (1991), see Chapter VI, it follows that the subset of parameters $a \in NA_2$ for which f_a is of 'bounded type', also has no interior points. The proof of this last statement is based on the notion of quasisymmetry. It was Sullivan who first used this concept in this theory. In particular he showed, see Section VI.4, that Conjecture 2 follows from the following

Conjecture 3. If two C^∞ maps $f, g : N \to N$ with quadratic critical points are conjugate then they are quasi-symmetrically conjugate.

In fact, Świątek proves Conjecture 3 for quadratic maps f and g and in this way he solved Conjecture 2!

Because of the results of Yoccoz (1990) and Sullivan (1991), Conjecture 3 only remains to be proved for maps which are infinitely renormalizable and of 'unbounded type'. However, Świątek does not explicitly use these last results. His result is based on the 'divergence' of the moduli of some annuli. This last property does not hold anymore if the critical point is of higher order. Some special cases of Conjecture 3 were solved previously. For example, if f and g are Misiurewicz maps the proof of Conjecture 3 is very easy, see Exercise III.6.1. More interestingly, Jakobson and Świątek (1991b) proved Conjecture 3 for non-renormalizable maps which satisfy some 'starting' conditions. For infinitely renormalizable maps of 'bounded type' the conjecture is proved in Section VI.2 and VI.3. We should emphasize that in general two conjugate analytic unimodal maps with negative Schwarzian derivative are not quasi-symmetrically conjugate: the order of the critical point plays an important role, see Section VI.10.

Another result in the direction of Conjecture 3 was given in Nowicki and Przytycki (1990). This result states that two conjugate maps with positive Liapounov exponent, i.e., for which

$$\liminf_{n\to\infty} \frac{1}{n} \log |Df^n(f(c))| > 0$$

are Hölder conjugate (however, a quasisymmetry is Hölder continuous but the opposite is not true).

We should add that a theorem of Jakobson, which we will prove in Section V.6, tells us that the set of parameters for which the maps $f_a(x) = ax(1-x)$ is not Axiom A has positive Lebesgue measure. In fact the set of parameters for which the Liapounov exponent is positive, has positive Lebesgue measure. However, we would like to state

Conjecture 4. For almost every parameter $a \in [0,4]$ in the sense of Lebesgue one has that $f_a(x) = ax(1-x)$ has either

1. a periodic attractor, or
2. $\liminf_{n\to\infty} \frac{1}{n} \log |Df^n(f(1/2))| > 0$.

If this conjecture holds then in particular the set of parameters for which the corresponding maps are infinitely renormalizable has zero Lebesgue measure. Condition 1) implies that almost all points tend to the periodic attractor and Condition 2) implies that $\omega(x)$ is a finite union of intervals for almost every x (and the way the orbit is distributed is determined by some absolutely continuous invariant measure, see Chapter V). In Section V.5 we will see that there are certainly uncountably parameters which do not satisfy 1) or 2) or even correspond to infinitely renormalizable maps.

Chapter IV.

The Structure of Smooth Maps

In the last chapter we described the structure of the unimodal maps with negative Schwarzian derivative. The ingredients we used were: i) the combinatorial theory of Section II.3; ii) the finiteness of attractors; iii) the non-existence of wandering intervals. For ii) and iii) we used the assumption on the Schwarzian derivative.

Here we will study the structure of general smooth mappings of the interval and of the circle. Since the combinatorial theory also applies to all maps with a finite number of turning points, we only need to extend the last two ingredients mentioned above to this more general class of maps. The aim of this chapter is to show that ii) and iii) also hold under very mild assumptions. The example of Denjoy, see Theorem I.2.3, and the example from Ivanov (1992), show that it is essential to assume that f has some smoothness.

Moreover, it is necessary to assume that the map is non-flat at each of its critical points. This assumption is necessary because there exist C^∞ maps with flat critical points which have wandering intervals and for which, moreover, the periods of the attracting periodic orbits are unbounded, see Sarkovskii and Ivanov (1983), de Melo (1987). That wandering intervals can exist for C^∞ circle homeomorphisms was already shown previously by Hall (1981). Moreover, it is not enough that f is piecewise continuous; as we have seen in Exercise I.2.2 there are piecewise continuous, affine interval exchange transformations with wandering intervals.

So let us state the results from this chapter more precisely. Let N be either the interval $[-1, 1]$ or the circle S^1. We say that a critical point c of a C^2 map $f\colon N \to N$ is *non-flat* if there exists a C^2 local diffeomorphism ϕ with $\phi(c) = 0$ such that $f(x) = \pm|\phi(x)|^\alpha + f(c)$ for some $\alpha \geq 2$. If $\alpha = 2$ we say that c is a *non-degenerate* critical point. For example, if f is C^∞ and some derivative is non-zero at c then c is a non-flat critical point. The main results in this chapter

are the following two theorems:

Theorem A. *If $f: N \to N$ is a C^2 map with non-flat critical points then f has no wandering interval.*

Theorem B. *If $f: N \to N$ is a C^2 map with non-flat critical points, then there exist $\lambda > 1$ and $n_0 \in \mathbb{N}$ such that*

$$|Df^n(p)| > \lambda$$

for every periodic point p of f of period $n \geq n_0$.

Moreover, the proof of Theorem B will give

Theorem B'. *If $\mathcal{K}$ is a compact family of C^2 maps with non-flat critical points then there exist $\lambda > 1$ and $n_0 \in \mathbb{N}$ such that if $f \in \mathcal{K}$ and p is a periodic point of period $n \geq n_0$ of f then one of the following possibilities hold.*

1. *p is an attracting periodic point whose immediate basin of attraction contains a critical point;*
2. *p is in the boundary of the immediate basin of a periodic attractor which attracts a critical point;*
3. *$|Df^n(p)| \geq \lambda$.*

In particular, the number of periodic orbits of maps in $\mathcal{K}$ of type 1) and 2) and of period $\geq n_0$ is bounded by the number of critical points.

We will prove Theorem A under a more general hypothesis. It holds for maps in the collections $\mathcal{N}F^{1+Z}$ and $\mathcal{N}F^{1+bv}$ which will be defined precisely in Section 2. This class of maps includes besides the C^2 maps without flat critical points, all continuous maps that are piecewise C^2 (or C^1 and satisfy a Zygmund condition which will be discussed in Section 2.a) and are non-flat at the critical points. In particular, one may have discontinuities of the derivative. Even for maps in our class without critical points the techniques of Schwarz from Theorem I.2.2 cannot be used: the distortion of the map in a neighbourhood of a point of discontinuity of the derivative is not bounded by a constant times the length of this neighbourhood. On the other hand, our class of maps includes those for which the proof of Denjoy's Theorem I.2.1 on the non-existence of a wandering interval for circle diffeomorphisms works. So Theorem A is a natural extension of Denjoy's original ideas to maps with critical points. Because our class is so general, Theorem A is new even for non-invertible maps without critical points (e.g., covering maps of the circle) or for piecewise linear maps. Theorem B holds for the slightly smaller class of maps $\mathcal{N}F^{1+z}$ which we will also introduce in Section IV.2 below.

We want to observe that these theorems give, for real one-dimensional dynamics, a result that is the analogue of Sullivan's Structure Theorem of rational maps discussed in Section III.1. To explain this we give the following

Definition. Let N be as before and let $f\colon N \to N$ be a continuous map. We define its *Fatou set* $F(f)$ as the set of points x for which there exists a neighbourhood U such that the family $f^n|U$ is normal. This means that each sequence in this family has a subsequence which converges in the C^0 topology (to a continuous function). Next the *Julia set* $J(f)$ is defined to be the complement of $F(f)$. The Fatou set is clearly open and backward invariant. In general it is not forward invariant but it is almost forward invariant: if U is a connected component of $F(f)$ not containing a turning point then $F(U)$ is also a connected component of the Fatou set. If U is one of the finitely many components of $F(f)$ that contains some turning point then $f(U)$ is contained in the closure of another component of $F(f)$.

This definition is the same as the one given for rational maps of the Riemann sphere, see Section III.1. As we have shown in that section, for a polynomial map on the Riemann sphere

1. the Julia set is equal to the set of points which are not in the basin of a periodic attractor;
2. the Julia set is the closure of the set of repelling periodic points.

For real one-dimensional maps the first statement still holds but the second statement is false:

Lemma. *For a real one-dimensional map $f\colon N \to N$*

1. *the Julia set is equal to the set of points which are not in the basin of a periodic attractor and not contained in a wandering interval. It is also equal to the α-limit of the set of turning points of f.*
2. *however, in general, the Julia set is not the closure of the set of repelling periodic points and can, in general, be much larger than the non-wandering set of f.*
3. *if f is the restriction of a polynomial map F then $J(f) = J(F)|N$.*

Proof. We claim that if $x \in F(f)$ then there exists a neighbourhood of x which contains at most one preimage of the set of turning points. Indeed, otherwise for each neighbourhood U of x the modality of $f^n|U$ goes to infinity as $n \to \infty$ and so because x is in the Fatou set one has $\inf_{n\geq 0} |f^n(U)| = 0$. From the Contraction Principle which we shall prove in Section 5 of this chapter it follows that U is either a wandering interval or contained in the basin of a periodic attractor. On the other hand, if some point is contained in a wandering interval or in the basin of a periodic attractor then it is certainly contained in $F(f)$. This proves the claim and the first statement of the lemma. To prove the second statement consider the quadratic Feigenbaum interval map. Because of Theorem A, it has no wandering intervals and as we have seen before it has no periodic attractors. So by the first part of this lemma, the Julia set of this map is the whole interval. On the other hand, for that map the non-wandering points and in particular, the periodic points are certainly not dense in the interval. This completes the proof of the second statement. The last property follows easily from Property 2 above and Statement 1 of this lemma. □

We can now reformulate Theorems A and B as

Theorem AB. *Let $f: N \to N$ be a C^2 non-invertible map having a finite number of critical points which are non-flat. Then*

1. *all the connected components of the Fatou set of f are eventually periodic (i.e., eventually mapped into a periodic component);*
2. *the number of periodic components of the Fatou set is finite.*

Our exposition of the above theorems follows very closely Martens et al. (1991). This work is based on and extends work of Lyubich (1989) and Blokh and Lyubich (1989). Let us describe some previous versions of these theorems.

The first result in this direction was obtained by Denjoy in 1932. He proved, as we have already seen in Section I.2, that a C^1 diffeomorphism of the circle such that the logarithm of its derivative is a function of bounded variation, does not have wandering intervals. His proof relies on i) a detailed understanding of the dynamics of rotations and ii) a uniform bound on the distortion of the n-th iterate of the diffeomorphism restricted to an interval whose first n iterates are disjoint.

Guckenheimer (1979) proved the non-existence of wandering intervals for unimodal maps of the interval with negative Schwarzian derivative and no inflection points, see also Misiurewicz (1981). In Section II.4, we gave a proof of his result in a special case.

Yoccoz (1984b) proved the non-existence of wandering intervals for C^∞ homeomorphisms of the circle having only non-flat critical points. He combines techniques of Denjoy away from the critical points with some analytical estimates near the critical points which are related to the Schwarzian derivative.

In de Melo and Van Strien (1988) and (1989) the same result was shown for smooth unimodal maps (not necessarily having negative Schwarzian derivative) with a non-flat critical point and also for maps satisfying the so-called Misiurewicz condition. The main tool in this proof is the control of the distortion of the cross-ratio under iterations. This control implies that under some disjointness assumptions the diffeomorphic inverse branches of iterates of a smooth map behave very much like univalent holomorphic maps. This similarity is clear from the Minimum Principle and the Koebe Principle for real maps, see Sections 1 to 3 below.

In Lyubich (1989) and Blokh and Lyubich (1989) the non-existence of wandering intervals is proved for smooth maps for which all critical points are turning points. They introduced some very nice and powerful new tools generalizing those of Guckenheimer (1979) and used the analytical tools developed in de Melo and Van Strien (1988) and (1989).

The proof we present, coming from Martens et al. (1991), combines the analytical tools developed in de Melo and Van Strien (1988) with the topological ingredients of Lyubich and Blokh. Inflection points are also allowed in this proof. Moreover, the smoothness required in this proof is precisely the same as needed in the non-wandering results due to Denjoy in Chapter I.

Theorem B implies that the period of the attracting periodic orbits is bounded. In particular, if f is an analytic function or if all the periodic points of f are hyperbolic, the number of attracting periodic orbits is finite. So Theorem B is related to Julia's Theorem III.1.3 and to Singer's Theorem II.4.1. In these theorems, the number of attracting periodic orbits is bounded because each must contain a critical point in its basin. Of course this is no longer true for general smooth maps.

Another contribution in this direction is due to Mañé (1985). Using estimates related to Denjoy and Schwartz, he proved that Theorem B holds for maps of the circle without critical points (see Section III.5). Van Strien (1990) proved Theorem B for maps satisfying the Misiurewicz condition. In this case one even has a 'hyperbolic structure' on the set of periodic points, see Section III.6.

Let us finish this introduction by stating a corollary of Theorem B.

Corollary. In the space $\mathcal{U}^r$ of C^r unimodal maps of the interval $[-1,1]$ endowed with the C^r topology for $r \geq 3$, the set of structurally stable maps is open and dense.

Proof. See de Melo (1987). □ This corollary is the

analogue of the Mañé-Sad-Sullivan Theorem on the density of the structurally stable rational maps. Again we do not know if the structurally stable unimodal maps satisfy the Axiom A property.

Exercise 1. Show that the previous theorems imply that for any C^2 unimodal map $f\colon I \to I$ which is infinitely renormalizable and whose turning point is non-flat, f has an attracting Cantor set and its basin has positive Lebesgue measure.

1 The Cross-Ratio: the Minimum and Koebe Principle

In Denjoy's theory for circle diffeomorphisms, described in Section I.1, the main technical tool is the control of the distortion of iterates of a map on some interval under some disjointness assumptions on the iterates of this interval. The distortion of a differentiable map f on an interval T was defined as the maximal ratio of the absolute values of the derivative in two different points. This number measures the non-linearity of the map. Another way to present the same concept is to consider pairs of adjacent intervals L, R, intersecting at a common boundary point, and to measure the distortion by the map of the ratio $\mathcal{R}(L,R) = \frac{|L|}{|R|}$, i.e., to evaluate the number $\mathcal{R}(f,L,R) = \frac{\mathcal{R}(f(L),f(R))}{\mathcal{R}(L,R)}$. It is easy to see that the distortion of a differentiable map f in the interval T is bounded if and only if there is a bound for $\mathcal{R}(f, L, R)$ for any pair of adjacent intervals $L, R \subset T$.

Cross-ratios

If a map f has critical points we cannot hope to get a bound for its non-linearity: we cannot consider the ratio of the length of a pair of adjacent intervals. So, instead of considering three consecutive points, we consider four points and we measure their position by their cross-ratio. Then we determine how a map distorts the cross-ratio of such a configuration of four points. Cross-ratios were already discussed in Section III.1. They are connected with hyperbolic geometry: the hyperbolic distance between two points in the hyperbolic space is given by the logarithm of the cross-ratio of four points. Using the same expression as in Section III.1, on each bounded interval of the real line we can also define a metric which is invariant by the group of real Möbius transformations. We will prove below that a map with positive Schwarzian derivative contracts this metric in the same way as a holomorphic map of the unit disc contracts the hyperbolic metric. Hence, if the iterate of a map of negative Schwarzian derivative is diffeomorphic on a given interval it expands the hyperbolic metric of this interval. This is the main reason for many analogies between the theory on the dynamics of rational maps and the theory of the dynamics of interval maps with negative Schwarzian derivative.

In this section we will show that diffeomorphic inverse branches of interval maps with negative Schwarzian derivative (or which satisfy some bounds on the distortion of the cross-ratio) are very much like univalent holomorphic maps.

Definition. Let $J \subset T$ be open and bounded intervals in N such that $T \setminus J$ consists of intervals L and R. Define the cross-ratio of these intervals as

$$D(T, J) = \frac{|J||T|}{|L||R|},$$

(where $|I|$ denotes the length of the interval I). This cross-ratio is related to the hyperbolic metric. Indeed, let T be an open and bounded interval on N. For $x, y \in T$ let

$$\rho_T(x, y) = \log \frac{|L \cup J||J \cup R|}{|L||R|} = \log(1 + D(T, J))$$

where J is the interval bounded by the points x, y. We will return to this in Property 2 below.

Definition. If $g\colon T \to N$ is continuous and monotone and $J \subset T$ as above then we define the cross-ratio distortion of g as

$$B(g, T, J) = \frac{D(g(T), g(J))}{D(T, J)}$$

(if J has a common boundary point with T then we take $\limsup B(g,T,J_n)$ where J_n is a sequence of intervals with $J_n \uparrow J$). If $f\colon I \to I$ is continuous and $f^n|T$ is monotone then

$$B(f^n, T, J) = \prod_{i=0}^{n-1} B(f, f^i(T), f^i(J)).$$

We recall that the Schwarzian derivative of a C^3 map $g\colon T \subset \mathbb{R} \to \mathbb{R}$ is the number,

$$Sg(x) = \frac{D^3g(x)}{Dg(x)} - \frac{3}{2}\left(\frac{D^2g(x)}{Dg(x)}\right)^2, \text{ if } Dg(x) \neq 0.$$

As we have seen before in Section I.3, one has the following composition formula: $S(g\circ f) = (Sg\circ f)\cdot(Df)^2 + Sf$. Furthermore, let $\mathcal{M}$ denote the group of Möbius transformations of the real line, namely, $\phi \in \mathcal{M}$ if $\phi(x) = \frac{ax+b}{cx+d}$ where a, b, c, d are real numbers with $ad - bc \neq 0$. Let us discuss some elementary properties of the operator B defined above, the Schwarzian derivative and the group of real Möbius transformations.

Property 1. The Schwarzian derivative of $g\colon T \subset \mathbb{R} \to \mathbb{R}$ is identically zero if and only if g is the restriction of a Möbius transformation to T.

Property 1'. Let $g\colon T \to \mathbb{R}$ be a monotone map. Then $B(g, T^*, J^*) = 1$ for all pairs of intervals $J^* \subset T^*$ if and only if g is the restriction of a Möbius transformation.

Property 2. Let T be a bounded open interval of the real line. For $x, y \in T$ let

$$\rho_T(x,y) = \log\frac{|L\cup J||J\cup R|}{|L||R|} = \log(1 + D(T,J))$$

where J is the interval bounded by the points x, y. Then ρ_T is a metric in T and the group of isometries of this metric is exactly the group $\mathcal{M}_T$ of all Möbius transformations which map T onto T. Furthermore, the group $\mathcal{M}_T$ acts transitively on T, namely, given $x, y \in T$, there exists an isometry ϕ such that $\phi(x) = y$. Notice that the formula defining ρ_T is exactly the same as the one that gives the distance between two points in the hyperbolic space. Therefore, we shall call this the *hyperbolic metric* of the interval T.

Property 3. If $g\colon T \to \mathbb{R}$ has negative (resp. positive) Schwarzian derivative at all points, ϕ, ψ are Möbius transformations then $g\circ\phi$ and $\psi\circ g$ also have negative (resp. positive) Schwarzian derivative.

Property 4. If $g\colon T \to \mathbb{R}$ is a C^3 map with negative Schwarzian derivative then

$$B(g, T^*, J^*) > 1, \text{ for all pairs of intervals } J^* \subset T^* \subset T.$$

Hence, a diffeomorphism $g\colon T \to T'$ having negative (resp. positive) Schwarzian derivative expands (resp. contracts) the hyperbolic metric:

$$\rho_{T'}(g(x), g(y)) > \rho_T(x,y)$$

(respectively, $\rho_{T'}(g(x), g(y)) < \rho_T(x, y))$) for all $x, y \in T$ and $x \neq y$.

Proof of Properties 1-4. To prove Property 1, note that we may write $Sg = -2|Dg|^{\frac{1}{2}} D^2 \sqrt{\dfrac{1}{|Dg|}}$. If $Sg \equiv 0$ then $D^2 \frac{1}{|Dg|^{\frac{1}{2}}} \equiv 0$. Hence, $\frac{1}{|Dg(x)|^{\frac{1}{2}}} = cx+d$ and, therefore, $g(x) = \frac{ax+b}{cx+d}$. The converse is also easy. Property 1') is proved as follows: we have already seen in Section III.1 that any Möbius transformation preserves cross-ratios. Suppose now that g is a monotone map which preserves the cross-ratio D. Let $T = [x_0, x_1]$ and choose a point $x_2 \in (x_0, x_1)$. As we have seen in Section III.1, there exists a unique Möbius transformation ϕ such that $\phi(x_i) = g(x_i)$ for $i = 0, 1, 2$. We claim that $g(x) = \phi(x)$ for all $x \in T$. Indeed, if J is the interval bounded by x_2 and x, then $B(g, T, J) = B(\phi, T, J) = 1$. Therefore, $D(\phi(T), \phi(J)) = D(g(T), g(J))$ and, since $\phi(T) = g(T)$ and $\phi(x_2) = g(x_2)$, we get that $\phi(x) = g(x)$.

Let us prove Property 2. That ρ_T is a metric follows immediately from the formula. From Property 1' above, we get that the group of isometries coincide with the group of Möbius transformations $\mathcal{M}_T$. Let $T = (x_0, x_1)$. As there is a unique Möbius transformation ϕ satisfying, $\phi(x_i) = x_i$, $\phi(x) = y$, the group of isometries acts transitively on T. This proves the statement. If $T = (0, 1)$ then the group of orientation preserving isometries is the family of maps

$$\phi_\lambda(x) = \frac{x}{\lambda x + 1 - \lambda}, -\infty < \lambda < 1.$$

To see Property 3 we note that from Property 1 one has $S\psi = 0$ and, therefore, the composition formula from above gives that $S(\psi \circ g) = (S\psi \circ g) \cdot |Dg|^2 + Sg = Sg$ and $S(g \circ \phi) = (Sg \circ \phi) \cdot |D\phi|^2 + S\phi = (Sg \circ \phi) \cdot |D\phi|^2$.

Finally, in order to prove Property 4, we should first note that an elementary calculation shows if $Sg < 0$ then $x \mapsto Dg(x)$ has no positive local minima. This property we refer to as the *Minimum Principle* and will be generalized below. Now let $T^* = [x_0, x_1]$, $J^* = [y_0, y_1]$ and let ϕ be a Möbius transformation such that $\phi \circ g$ fixes the endpoints of T^* and the point y_0. We claim that $\phi(g(y_1)) > y_1$. In fact, suppose, by contradiction, that $\phi(g(y_1)) \leq y_1$. By the Mean Value Theorem, there exist $z_0 \in [x_0, y_0]$, $z_1 \in [y_0, y_1]$, $z_2 \in [y_1, x_1]$ such that $D(\phi \circ g)(z_0) = \dfrac{\phi(g(y_0)) - \phi(g(x_0))}{y_0 - x_0} = 1$, $D(\phi \circ g)(z_1) = \dfrac{\phi(g(y_1)) - \phi(g(y_0))}{y_1 - y_0} \leq 1$ and $D(\phi \circ g)(z_2) = \dfrac{\phi(g(x_1)) - \phi(g(y_1))}{x_1 - y_1} \geq 1$. Since $S(\phi \circ g) = Sg < 0$, this contradicts the Minimum Principle. Therefore, $\phi(g(y_1)) > y_1$. Thus

$$B(\phi \circ g, T^*, J*) = \frac{\dfrac{|\phi(g(T^*))|}{|T^*|} \dfrac{|\phi(g(J^*))|}{|J^*|}}{\dfrac{|\phi(g(L^*))|}{|L^*|} \dfrac{|\phi(g(R^*)|}{|R^*|}} = \frac{\dfrac{|\phi(g(J^*))|}{|J^*|}}{\dfrac{|\phi(g(R^*)|}{|R^*|}} > 1.$$

Since $B(\phi \circ g, T^*, J^*) = B(\phi, g(T^*), g(J^*)).B(g, T^*, J^*)$ and, by Property 1', the Möbius transformation ϕ preserves the cross-ratio, we get that $B(g, T^*, J^*) > 1$. □

The Minimum Principle for maps with a given cross-ratio distortion

In the remainder of this section we will prove that the inverse of a C^1 diffeomorphism of an interval behaves like a univalent holomorphic map if its cross-ratio distortion is bounded from below. In the next section we will show that this assumption is satisfied even for high iterates of a map provided some disjointness assumptions are met.

The main properties of univalent holomorphic maps are the Maximum and Koebe Principles. Since our maps will be like the inverse of holomorphic maps, we will now show that one has the Minimum and the Koebe Principles for C^1 maps $g\colon T \to \mathbb{R}$ for which there exists a constant $C_0 > 0$ such that

$$B(g, T^*, J^*) > C_0 \text{ for all intervals } J^* \subset T^* \subset T$$

when g is a diffeomorphism on T. Clearly we cannot take $C_0 > 1$ and also we can take $C_0 = 1$ when $Sg < 0$.

Theorem 1.1. **("Minimum Principle")** *Let $T = [a, b] \subset N$ and $g\colon T \to g(T) \subset N$ be a C^1 diffeomorphism. Let $x \in (a, b)$. If for any $J^* \subset T^* \subset T$.*

$$B(g, T^*, J^*) \ge C_0 > 0$$

then

$$|Dg(x)| \ge C_0^3 \min\{\, |Dg(a)|\, ,\, |Dg(b)|\, \}.$$

Proof. Take an arbitrary interval $T^* = [a, b]$ in T and consider

$$B_0(g, T^*) = \frac{|g(T^*)|^2}{|T^*|^2} \frac{1}{|Dg(a^*)| \cdot |Dg(b^*)|}.$$

Moreover, define

$$B_1(g, T, x) = \frac{|Dg(x)| \frac{|g(T)|}{|T|}}{\frac{|g(L)|}{|L|} \frac{|g(R)|}{|R|}},$$

where L and R are the components of $T \setminus \{x\}$. Observe that

$$B_0(g, T^*) = \lim_{J^* \to T^*} B(g, T^*, J^*), B_1(g, T, x) = \lim_{J \to x} B(g, T, J)$$

(the last limit means that both endpoints tend to x). Hence, $B_0(g, L)$, $B_0(g, R)$, $B_1(g, T, x) \ge C_0 > 0$. The first two of these three inequalities imply

$$\left(\frac{|g(L)|}{|L|}\right)^2 \ge C_0 |Dg(a)| \cdot |Dg(x)|, \left(\frac{|g(R)|}{|R|}\right)^2 \ge C_0 |Dg(x)| \cdot |Dg(b)|$$

and the last one gives

$$|Dg(x)| \frac{|g(T)|}{|T|} \ge C_0 \frac{|g(L)|}{|L|} \frac{|g(R)|}{|R|}.$$

Since $g|T$ is a diffeomorphism,

$$\min\left\{\frac{|g(L)|}{|L|}, \frac{|g(R)|}{|R|}\right\} \le \frac{|g(T)|}{|T|} \le \max\left\{\frac{|g(L)|}{|L|}, \frac{|g(R)|}{|R|}\right\}.$$

Then

$$|Dg(x)|^2 \ge C_0^2 \left(\frac{\frac{|g(L)|}{|L|}\cdot\frac{|g(R)|}{|R|}}{\frac{|g(T)|}{|T|}}\right)^2 \ge C_0^2 \min\left\{\left(\frac{|g(L)|}{|L|}\right)^2, \left(\frac{|g(R)|}{|R|}\right)^2\right\}$$
$$\ge C_0^3 \min\{|Dg(a)|\cdot|Dg(x)|, |Dg(b)|\cdot|Dg(x)|\}.$$

Hence, $|Dg(x)| \ge C_0^3 \min\{|Dg(a)|, |Dg(b)|\}$. □

The Koebe Principle for maps with a given cross-ratio distortion

Now we shall discuss one of the most powerfull tools in one-dimensional dynamics: the Koebe Principle. For maps with negative Schwarzian derivative one version of this principle was already used and proved in Van Strien (1981) and later rediscovered by Johnson and Guckenheimer, see Guckenheimer (1987). It was extended in Van Strien (1987), (1990) for maps which do not satisfy the negative Schwarzian derivative condition. This principle states that a map which satisfies bounds on the distortion of cross-ratios has bounded non-linearity 'away from its critical values'. If the map satisfies the negative Schwarzian derivative condition then also iterates satisfy this condition. So in that case we can apply the next result to each iterate. If this condition is not satisfied then we have to estimate the cross-ratio distortion. This will be done in Section 2 and we shall apply all this to iterates of maps in Section 3.

Let $U \subset V$ be two intervals. We say that V *is an ϵ-scaled* neighbourhood of U if both components of $V \setminus U$ have length $\epsilon \cdot |U|$.

Theorem 1.2. **("Koebe Principle")** *Let $C_0 \in (0,1]$, $J \subset T$ be intervals and $g\colon T \to g(T)$ a C^1 diffeomorphism. Assume that for any intervals J^* and T^* with $J^* \subset T^* \subset T$ one has*

$$B(g, T^*, J^*) \ge C_0 > 0.$$

If $g(T)$ contains a τ-scaled neighbourhood of $g(J)$, then

$$\frac{1}{K(C_0,\tau)} \le \frac{Dg(x)}{Dg(y)} \le K(C_0,\tau), \quad \forall x, y \in J,$$

where $K(C_0,\tau) = \dfrac{(1+\tau)^2}{C_0^6\tau^2}$. Moreover, if $Sg < 0$ then one can take $C_0 = 1$ and there exists a constant $\hat{K}$ which only depends on τ so that

$$\left|\frac{Dg(x)}{Dg(y)} - 1\right| \le \hat{K}\cdot\frac{|x-y|}{|J|}.$$

Remark. 1. $K(C_0, \tau)$ tends to 1 as $C_0 \to 1$ and $\tau \to \infty$. Therefore, the distortion of g is extremely small if C_0 is close to 1 and if we consider points x in an interval $J \subset T$ such that $g(J)$ is extremely small and extremely far away from the boundary of $g(T)$.

2. The last part of the principle implies that there exists $\tilde{K}$ only depending on τ such that $\frac{|Dg(x) - Dg(y)|}{Dg(y)} \leq \tilde{K} \cdot \frac{|g(x) - g(y)|}{|g(J)|}$ for all $x, y \in J$. Hence if g is C^2 then there exists a universal upper bound $C(\delta, \tau)$ with

$$\frac{|D^2 g(x)|}{|Dg(x)|^2} \leq \frac{\tilde{K}}{|g(J)|}$$

for each $x \in J$.

3. If $U \subset V$ are two intervals then the length of U in terms of the hyperbolic metric on V is equal to

$$\ell_V(U) := \log \frac{|L \cup U||U \cup R|}{|L||R|}$$

where L and R are the components of $V \setminus U$. If V contains an ϵ-*scaled* neighbourhood of U then

$$(*) \qquad \exp\left[\ell_V(U)/2\right] \leq \frac{1+\epsilon}{\epsilon} \leq \exp\left[2\ell_V(U)\right].$$

Using the expression for $K(C_0, \tau)$ we get

$$C_0^6 \cdot \exp\left[-2\ell_{g(T)}(g(J))\right] \leq \frac{|Dg(x)|}{|Dg(y)|} \leq \frac{1}{C_0^6} \cdot \exp\left[2\ell_{g(T)}(g(J))\right]$$

for all $x, y \in J$. So this ratio is bounded on an interval J by a constant which depends on C_0 and on the projective length of $g(J)$ in $g(T)$.

Proof. By rescaling, we may assume that $J = g(J) = [0,1]$ and that g is increasing. Let $a, b \in \partial T$ be such that $a < 0 < 1 < b$ and $L = [a, 0]$, $J = [0,1]$ and $R = [1, b]$. Using the operator B_0 from the proof of the Minimum Principle, we get

$$(1.1) \qquad |Dg(0)| \cdot |Dg(1)| \leq \frac{1}{C_0} \left(\frac{|g(J)|}{|J|}\right)^2 = \frac{1}{C_0}.$$

Similarly, using the operator B_1,

$$|Dg(0| \geq C_0 \cdot \frac{\frac{|g(L)|}{|L|} \frac{|g(J)|}{|J|}}{\frac{|g(L \cup J)|}{|L \cup J|}}.$$

Because $|g(J)| = |J| = 1$ and

$$\frac{|g(L)|}{|L|} \frac{|L \cup J|}{|g(L \cup J)|} \geq \frac{|g(L)|}{|g(L \cup J)|} \geq \frac{\tau}{1+\tau}$$

this gives

$$|Dg(0)| \geq \frac{C_0\tau}{1+\tau}. \tag{1.2}$$

Similarly,

$$|Dg(1)| \geq \frac{C_0\tau}{1+\tau}. \tag{1.3}$$

Combining (1.1), (1.2) and (1.3) gives

$$\frac{C_0\tau}{1+\tau} \leq |Dg(0)|, |Dg(1)| \leq \frac{1+\tau}{C_0^2\tau}.$$

From the Minimum Principle we get

$$|Dg(x)| \geq \frac{C_0^4\tau}{1+\tau}\,, \quad \text{for all } x \in [0,1]. \tag{1.4}$$

Let $U = [0,x]$ and $V = [x,1]$. Since g is a diffeomorphism, either $\frac{|g(U)|}{|U|} \leq \frac{|g(J)|}{|J|} = 1$ or $\frac{|g(V)|}{|V|} \leq \frac{|g(J)|}{|J|} = 1$. If the former holds then

$$\frac{\left[\frac{|g(U)|}{|U|}\right]^2}{|Dg(0)| \cdot |Dg(x)|} \geq C_0$$

gives

$$|Dg(x)| \cdot |Dg(0)| \leq \frac{1}{C_0} \cdot 1$$

and otherwise we get a similar estimate for $|Dg(x)| \cdot |Dg(1)|$. Using (1.2) or (1.3) this gives

$$|Dg(x)| \leq \frac{1}{C_0}\frac{1+\tau}{C_0\tau}.$$

From this and (1.4) one has

$$\frac{C_0^4\tau}{1+\tau} \leq |Dg(x)| \leq \frac{1+\tau}{C_0^2\tau}. \quad \square$$

From the Koebe Principle we get that g has bounded distortion on a certain interval. Now we shall some other versions of the Koebe Principle. We shall not need these versions until much later in the book. First we state a version which gives $C^{1+1/2}$ control on the size of the distortion. In Section 3 we will also encounter similar versions of the Koebe Principle. However, there we do not consider some abstract map g whose cross-ratio is not distorted too much, but instead consider the restriction of an iterate of some map satisfying some additional conditions.

Corollary 1.1. ('$C^{1+1/2}$ Koebe Principle') *For each $D, \tau > 0$ there exists $L(D,\tau) < \infty$ with the following property. Let $J \subset T$ be intervals and $g\colon T \to g(T)$ a homeomorphism such that $g(T)$ is a τ-scaled neighbourhood of $g(J)$. Assume that for any intervals J^* and T^* with $J^* \subset T^* \subset T$ one has*

$$|B(g,T^*,J^*) - 1| \leq D \cdot \frac{|T^*|}{|T|}.$$

Then g is $C^{1+1/2}$ in the sense that

$$\left|\frac{Dg(x)}{Dg(y)} - 1\right| \leq L(D,\tau) \cdot \left[\frac{|x-y|}{|T|}\right]^{1/2}, \quad \forall x, y \in J.$$

In particular, $\log Dg$ is Hölder with exponent $1/2$ on the interior of T.

Proof. See Exercise 1.5. □

Corollary 1.2. ('One-sided Koebe Principle') *For each $C > 0, \rho > 0$ there exists $K < \infty$ with the following property. Let $T = [a,b]$ and $g\colon T \to g(T)$ be a C^1 diffeomorphism. Assume that for any intervals J^* and T^* with $J^* \subset T^* \subset T$ one has*

$$B(g,T^*,J^*) \geq C > 0.$$

Then

$$|Dg(x)| \geq \frac{1}{K}|Dg(b)|.$$

for each $x \in [a,b]$ with

$$\frac{|g(x) - g(a)|}{|g(T)|} \geq \rho.$$

Proof. See Exercise 1.6. □

The First Expansion Principle

The next principle will play an important role in proving that the periodic points of high period are repelling and is concerned with the situation that $B(g,T,M) - 1 \geq \delta > 0$. It states that whenever an interval is mapped monotonically over itself with an expansion of the cross-ratios then the map is really 'bending' and, therefore, at some point expanding. In the next section we shall encounter a second expansion principle.

Theorem 1.3. **("First Expansion Principle")** *For every $\epsilon > 0$ and $\delta > 0$ there exists $\rho > 0$ such that the following holds. Let $J \subset T$ be intervals in N for which $g\colon T \to g(T) \subset N$ is a C^1 diffeomorphism and assume both components of $T \setminus J$ have at least length $\delta \cdot |T|$ (note that this is not the same as saying that T contains a δ-scaled neighbourhood of J). If $B(g,T,J) \geq 1 + \epsilon$ and $g(T) \supset T$ then there exists $\theta \in T$ with $Dg(\theta) \geq 1 + \rho$.*

Proof. Let L and R be the components of $T \setminus J$ and let $\xi > 0$ be so small that $(1+\epsilon)(1-\xi)^2 \geq 1 + \frac{1}{2}\epsilon$.

Case 1: First suppose that $\frac{|g(L)|}{|L|} \geq 1-\xi$ and $\frac{|g(R)|}{|R|} \geq 1-\xi$. Then, using $B(g,T,J) \geq 1+\epsilon$, we get $\frac{|g(T)|}{|T|} \cdot \frac{|g(J)|}{|J|} \geq (1+\epsilon)(1-\xi)^2 \geq 1 + \frac{1}{2}\epsilon$. So at least one of the terms $\frac{|g(T)|}{|T|}, \frac{|g(J)|}{|J|}$ is greater or equal than $\sqrt{1+\frac{1}{2}\epsilon}$. Using the Mean Value Theorem we obtain in either case $\theta \in T$ so that $|Dg(\theta)| \geq \sqrt{1+\frac{1}{2}\epsilon}$.

Case 2: Suppose that $\frac{|g(L)|}{|L|} < 1-\xi$. The case that $\frac{|g(R)|}{|R|} < 1-\xi$ goes similarly. Using $|L|, |R| \geq \delta|T|$ we get

$$\begin{aligned} \frac{|g(J\cup R)|}{|J\cup R|} &= \frac{|g(T)|-|g(L)|}{|J|+|R|} \geq \frac{|T|-|g(L)|}{|J|+|R|} > \\ &> \frac{|T|-(1-\xi)|L|}{|J|+|R|} > 1+\delta\xi. \end{aligned}$$

The Mean Value Theorem then gives $\theta \in J \cup R$ such that $|Dg(\theta)| \geq 1+\delta\xi$. Combining Case 1 and Case 2 proves the result. □

Exercise 1.1. Show that the diffeomorphism $g\colon (0,1) \to (0,1)$ defined by $g(x) = x^2$ satisfies $B(g,T,J) > 1$ for each $J \subset T \subset (0,1)$. Similarly, show that there exists no lower bound $C > 0$ such that $B(g^{-1},T,J) \geq C$ for each $J \subset T \subset (0,1)$.

Exercise 1.2. Give an example of $g\colon (0,1) \to (0,1)$ which is convex or concave but so that the property $B(g,T,J) \geq 1$ does not hold for each $J \subset T \subset (0,1)$. (However, since $Sg = -2|Dg|^{\frac{1}{2}} D^2 \sqrt{\frac{1}{|Dg|}}$, $Sg < 0$ is equivalent to $x \mapsto \sqrt{\frac{1}{|Dg(x)|}}$ being convex.)

Exercise 1.3. Below we shall see that if a diffeomorphism $g\colon (0,1) \to (0,1)$ satisfies $B(g,T,J) > 1$ for each $J \subset T \subset (0,1)$ that then $|Dg(x)| > \min(|Dg(a)|, |Dg(b)|)$ for each $x \in [a,b] \subset (0,1)$. Show that the reverse is not true.

Exercise 1.4. (Following Świątek (1990)) Let $f\colon (0,1) \to (0,1)$ be a homeomorphism and let

$$\mathcal{D}(f) = \sup_{x\in(0,1)} \left\{ \left| \log \frac{|f(L)|/|L|}{|f(R)|/|R|} \right| ; \text{ where } L = (0,x) \text{ and } R = (x,1) \right\}.$$

Furthermore, let $\phi\colon (0,1) \to (0,\infty)$ be the Möbius transformation sending $(0,1)$ injectively onto $(0,\infty)$ and let $\psi = \log \circ \phi$. Define $P(f) = \psi \circ f \circ \psi^{-1}$. Show that

$$||P(f) - id||_{C^0} = \mathcal{D}(f).$$

(Hint: $\sup_x |P(f) - x| = \sup_y |\psi \circ f(y) - \psi(y)|$. Furthermore, because $\phi(t) = t/(1-t)$,

$$\psi \circ f(y) - \psi(y) = \log \frac{\phi(f(y))}{\phi(y)} = \log \frac{|f(L)|}{|L|} \frac{|f(R)|}{|R|}$$

where $L = (0,y)$ and $R = (y,1)$.)

Exercise 1.5. Prove Corollary 1. (Hint: by considering intervals $J \subset T$ such that J, L or R is a point one can easily see that Dg must exist everywhere on T and be bounded and bounded away from zero. In order to show that $\log Dg$ is Hölder let $x, y \in J$ and define $J' = [x, y]$. We may assume that $|J'| \leq |T|/2$ because otherwise the Corollary follows trivially from the Koebe Principle. Furthermore, choose $C_1 > 0$ and an interval T' with $J' \subset T' \subset J$ such that T' is a $C_1 \left[|T|/|J'|\right]^{1/2}$-scaled neighbourhood of J'. We can even choose $C_1 > 0$ independently of $J' = [x, y]$ because we have assumed that $|J'| \leq |T|/2$ and because (by the Koebe Principle) T contains a τ'-scaled neighbourhood of J. One gets that

$$|T'| \leq 3C_1 \left[|T|/|J'|\right]^{1/2} \cdot |J'| \leq 3C_1 (|J'||T|)^{1/2} \quad (\leq 3C_1 |T|).$$

Hence, from the assumption, for each $J^*, T^* \subset T'$ one has

$$|B(g, T^*, J^*) - 1| \leq D \cdot |T'|/|T| \leq D \cdot 3C_1 \cdot \left[|J'|/|T|\right]^{1/2}.$$

Therefore, using the notation from the previous theorem applied to the pair of intervals $J' \subset T'$,

$$|C - 1| \leq D \cdot 3C_1 \cdot \left[|J'|/|T|\right]^{1/2}.$$

Furthermore, again by the Koebe Principle, $g(T')$ is a $C_1' \left[|T|/|J'|\right]^{1/2}$-scaled neighbourhood of $g(J')$ and so we get $\tau \geq C_1' \left[|T||J'|\right]^{1/2}$. From the proof of the previous theorem we get

$$C^6 \cdot (\frac{\tau}{1+\tau})^2 \leq \frac{|Dg(x)|}{|Dg(y)|} \leq \frac{1}{C^6} \cdot (\frac{1+\tau}{\tau})^2$$

and since $\dfrac{1+\tau}{\tau} - 1$ and $|C - 1|$ are at most

$$C_2 \cdot \left[|J'|/|T|\right]^{1/2} \leq C_2 \cdot \left[\frac{|x - y|}{|T|}\right]^{1/2},$$

the result follows.)

Exercise 1.6. Prove Corollary 2. (Hint: let $J = [x, b]$ and simply use the strategy as used in the proof of the Macroscopic Koebe Principle. The only place where the space to the right of J was used in that proof is to obtain the estimate (1.3). If we do not use this estimate but keep the term $Dg(b)$ everywhere we get the desired estimate.)

1.1 Some Facts about the Schwarzian Derivative

In this section we shall make some remarks about the Schwarzian derivative and its connection with projective transformations. First of all as we remarked above $Sf = 0$ implies that f is a Möbius transformation and therefore that f preserves cross-ratios. More precisely, one has the following relationship with cross-ratios:

Lemma 1.1. *Let f be a C^3 map. There exists a unique Möbius transformation ϕ such that*

$$\lim_{x \to 0} \frac{(\phi \circ f)(x) - x}{x^3}$$

is finite. This limit is equal to $Sf(0)/6$. Moreover, let $J = [h, 2h]$ and $T = [0, 3h]$ then

$$Sf(0) = -\frac{3}{2} \lim_{h \to 0} \frac{B(f, T, J) - 1}{h^2}.$$

Similarly,

$$Sf(0) = \frac{-1}{2} \lim_{h\to 0} \frac{\dfrac{1 + D(f(T), f(J))}{1 + D(T,J)} - 1}{h^2}.$$

Proof. Notice that we may assume that $Df(0) \neq 0$. Moreover, since $S(\phi \circ f) = Sf$ when ϕ is a Möbius transformation, we may even assume that $f(x) = x + ax^3 + o(|x|^3)$. Indeed, given $a_0, a_1, a_2 \in \mathbb{R}$ there exists a unique Möbius transformation ϕ for which $\phi(0) = a_0$, $D\phi(0) = a_1$ and $D^2\phi(0) = a_2$. Therefore, there exists a unique Möbius transformation ϕ for which $\phi \circ f(x) = x + ax^3 + o(|x|^3)$. Since $S(f) = 6a$, this completes the proof of the first statement. This part of the lemma is due to Martio and Sarvas (1978/79), see also Lehto (1987, p51). The second part of this lemma is due to E. Cartan (1937, p22) see also Hermann (1976, p168) and is proved as follows. Choosing L and T as above, $D(T,J) = 3$ and

$$D(f(T), f(J)) = \frac{[h + (8-1)ah^3][3h + 27ah^3]}{[h + ah^3][h + (27-8)ah^3]}$$

and so

$$B(f,T,J) - 1 = (16 - 20)ah^2 + o(h^2).$$

Hence,

$$\lim_{h\to 0} \frac{B(f,T,J) - 1}{h^2} = -4a = \frac{-2}{3} Sf(0). \quad \square$$

Next, we want to ask why one considers the invariance by the projective group and not some other group. Suppose we have a group of local diffeomorphisms on the real line generated by a Lie algebra $\mathcal{L}$ of infinitesimal generators. Suppose that this group is transitive: this implies that there exists an element $u \in \mathcal{L}$ such that $u(0) \neq 0$. Using a different coordinate system we may assume that $L(x) = \frac{d}{dx}$. Now assume that $\dim(\mathcal{L}) = m$ and take $v = f(x)\frac{d}{dx} \in \mathcal{L}$. Then $[u, v] = v'$ where $v' = f'(x)\frac{d}{dx}$. Similarly, $[u, v^k] = v^{k+1}$ where $v^k = D^k f(x)\frac{d}{dx}$. So, since $\mathcal{L}$ is m-dimensional, $v, \ldots, v^m$ are linearly dependent and, therefore, the functions $f, \ldots, D^m f$ are linearly dependent. Hence, f satisfies a linear $\leq m$-th order homogeneous differential equation with constant coefficients and so f is analytic. Now it is quite easy to see that, if f is not a polynomial of degree ≤ 2, then the Lie algebra generated by $u = f\frac{d}{x}$ is infinite-dimensional, i.e.,

$$[u, u], [u, [u, u]], [u, [u, [u, u]], \ldots$$

span an infinite-dimensional space. So the only generators which generate a finite-dimensional Lie algebra are L_1, L_2, L_3 where

$$L_p = x^{p-1}\frac{d}{dx}.$$

Let us see which group is generated by these generators. The time t map of L_1 is

$$x \mapsto x + t,$$

and, therefore, L_1 generates the group of translations. The time t map for L_2 is

$$x \mapsto xe^t$$

and generates the homotheties on the line. For L_3 this is

$$x \mapsto \frac{x}{-xt+1}$$

(which only exists for finite t) and these are projective transformations (the map $h_t(x) = \dfrac{x}{-xt+1}$ satisfies $h_t(0) = 0$, $Dh_t(0) = 1$ and $D^2h_t(0) = t$). Together these transformations generate all projective transformations

$$x \mapsto \frac{ax+b}{cx+d}.$$

It follows that the only differential invariants which correspond to finite-dimensional Lie groups are

$$D(f) = f'\ ,\ N(f) = \frac{D^2 f}{Df} \text{ and } S(f) = \frac{D^3 f}{Df} - \frac{3}{2}\left(\frac{D^2 f}{Df}\right)^2 .$$

We know all these differential operators already from the first chapter: they are the differentiation, the non-linearity operator and the Schwarzian derivative. They are invariant because one has

$$D(L_{1,t}f) = D(f),$$

$$N(L_{1,t}f) = N(L_{2,t}f) = N(f),$$

$$S(L_{1,t}f) = S(L_{2,t}f) = S(L_{3,t}f) = S(f),$$

for all t. In particular, $Df = 0$ gives the group of translations, $N(f) = 0$ the group of affine transformations and $S(f) = 0$ the group of projective transformations. For more on this we refer the reader to expositions on the work of E. Cartan on the differential invariants of S. Lie, see for example Cartan (1937, pp22), Hermann (1976, pp168), Guggenheimer (1977, pp130) and Dieudonné (1974, Exercise 7, p.143).

Exercise 1.7. If f is a polynomial of degree ≥ 2 with real coefficients and all zeros of Df are real then $Sf < 0$. (Hint: By assumption $Df(x) = A\prod_{j=1}^n (x - a_j)$ where a_j are real. Then

$$Sf(x) = 2\sum_{i<j} \frac{1}{(x-a_i)(x-a_j)} - \frac{3}{2}\left[\sum_i \frac{1}{(x-a_i)}\right]^2$$

and it easy to see that this is negative for x real.)

2 Distortion of Cross-Ratios

As we have seen in the previous section, one has some very good non-linearity estimates if bounds on the distortion of cross-ratio are known. In this section we will describe the distortion of the cross-ratio under high iterates of a smooth map. Obviously, if $Sf < 0$ then $Sf^n < 0$ and then the assumptions in the previous section on the distortion of the cross-ratios are trivially satisfied. In this case $B(f^n, T^*, J^*) > 1$ on all intervals $J^* \subset T^* \subset T$ for which $f^n|T$ is a diffeomorphism. Now we will show that C^2 maps have a similar property if one restricts to an interval whose orbit has some disjointness properties. Such disjointness was also crucial in Sections 2 and 3 in Chapter I.

Apart from these disjointness conditions we will make a smoothness and a non-flatness assumption on the maps we consider.

The smoothness assumptions

In fact, the disjointness conditions needed for the proof of Theorem A are somewhat weaker than those for Theorem B. Therefore, we will introduce three classes of maps $\mathcal{N}F^{1+Z}$, $\mathcal{N}F^{1+bv}$ and $\mathcal{N}F^{1+z}$. The last class is properly contained in the first two. Theorem A holds for each map in $\mathcal{N}F^{1+Z} \cup \mathcal{N}F^{1+bv}$ and Theorem B holds for maps in $\mathcal{N}F^{1+z}$.

Definition. Let $\mathcal{N}F^{1+bv}$ be the class of absolutely continuous maps $f\colon N \to N$ such that the conditions a) and b) below are satisfied.

a) For each $x_0 \in N$ there exist $\alpha \geq 1$, a neighbourhood $U(x_0)$ of x_0, and a homeomorphism $\phi\colon U(x_0) \to \mathbb{R}$ such that $\phi(x_0) = 0$ and

$$f(x) = \pm|\phi(x)|^\alpha + f(x_0),\ \forall x \in U(x_0).$$

b) $\log D\phi$ (which exist almost everywhere because f is absolutely continuous) coincides with the restriction of a function of bounded variation for each of these homeomorphisms ϕ.

We say that $f \in \mathcal{N}F^{1+Z}$ if f is absolutely continuous, it satisfies condition a) from above, and

b') $\log D\phi$ satisfies the Zygmund condition (this condition will be defined in Section 2a below and certainly holds if $\log D\phi$ is Lipschitz) for each of these homeomorphisms ϕ.

Next we will introduce a slightly smaller class of maps. Theorem B holds for maps from this class. Let $\mathcal{N}F^{1+z}$ be the class of C^1 maps for which the conditions c) and d) below are satisfied.

c) For each $x_0 \in N$ there exist $\alpha \geq 1$, a neighbourhood $U(x_0)$ of x_0, and a homeomorphism $\phi\colon U(x_0) \to \mathbb{R}$ such that $\phi(x_0) = 0$ and if $\alpha > 1$ then

$$f(x) = \pm|\phi(x)|^\alpha + f(x_0),\ \forall x \in U(x_0)$$

and if $\alpha = 1$ then

$$f(x) = |\phi(x)|^\alpha + f(x_0),\ \forall x \in U(x_0).$$

d) $\log D\phi$ satisfies the little Zygmund condition (this condition will be defined in Section 2a and certainly holds if $\log D\phi$ is C^1) for each of these maps ϕ.

Remark. 1. By the compactness of N, for each map f in one of the classes $\mathcal{N}F^{1+Z}$, $\mathcal{N}F^{1+bv}$ or $\mathcal{N}F^{1+z}$ there are at most a finite number of points x_0 at which either $\alpha > 1$ or at which f is not a local homeomorphism. Such points are called *critical points* of f. The set of all critical points of f is denoted by $C(f)$. If f is C^1 then $C(f) = \{x;\ Df(x) = 0\}$. The number α from assumption a) is called the *order* of the critical point.

2. $x_0 \in N$ is a critical point if and only if $\alpha > 1$ at this point. Moreover, if f is C^{k+1} in a neighbourhood of a point x_0 and $D^k f(x_0) \neq 0$ then the assumptions are satisfied near this point x_0 when we take $\alpha = k$. In particular, all C^∞ maps with non-flat critical points are contained in each of the three classes defined above.

3. The class $\mathcal{N}F^{1+bv}$ contains the set of continuous piecewise linear maps. If f is C^2 and if for each critical point x_0 there is a C^2 coordinate system ϕ and $\alpha \in \mathbb{N}$ such that $f(\phi(x)) = (x - x_0)^\alpha$ then f is contained in all three of the classes above. In particular, the results in this chapter hold for analytic maps.

The distortion of cross-ratios under iterates

In the following theorem we give a lower bound for $B(f^n, T, J)$ if the map f is as above (a similar result also holds for a different cross-ratio, see de Melo and Van Strien (1988), (1989) or Van Strien (1987), (1990) and Exercise 2.1 below). To get a lower bound for $B(f^n, T, J)$ one has to assume that the sum of the lengths of the iterates of $T, \ldots, f^{n-1}(T)$ is bounded. This occurs for example when these iterates are essentially disjoint; to formalize this we define the intersection multiplicity of a finite collection of intervals to be the maximal number of intervals in this collection whose interior has a non-empty intersection. We will come back to this notion in the next section.

Theorem 2.1. **(The distortion of cross-ratios under iterates)**

1. *If $f \in \mathcal{N}F^{1+Z} \cup \mathcal{N}F^{1+z}$ then there exists a bounded continuous function $\sigma\colon [0, \infty) \to \mathbb{R}_+$ with the following property. If $m \in \mathbb{N}$ and $J \subset T$ are intervals such that $f^m|T$ is a diffeomorphism then*

$$B(f^m, T, J) \geq \exp\{-\sigma(S) \cdot \sum_{i=0}^{m-1} |f^i(T)|\},$$

where $S = \max_{i=0,\dots,m-1} |f^i(T)|$. *If* $f \in \mathcal{N}F^{1+z}$ *then* $\sigma(0) = 0$ *whereas if* $f \in \mathcal{N}F^{1+Z}$ *then* σ *is merely bounded.*

2. *For each* $f \in \mathcal{N}F^{1+bv}$ *and each* $p \in \mathbb{N}$ *there exists* $V < \infty$ *with the following property. If* $m \in \mathbb{N}$ *and* $J \subset T$ *are intervals such that* $T, \dots, f^{m-1}(T)$ *has at most intersection multiplicity* p *and* $f^i(T) \cap C(f) = \emptyset$ *for* $i = 0, 1, \dots, m-1$ *then*
$$B(f^m, T, J) \geq \exp\{-V\}.$$

Proof of Theorem 2.1. Notice that $B(f^m, T, J) = \prod_{i=0}^{m-1} B(f, f^i(T), f^i(J))$. So for f from $\mathcal{N}F^{1+Z}$ or $\mathcal{N}F^{1+z}$ the theorem follows from the following lemma. For $f \in \mathcal{N}F^{1+bv}$ this follows from Proposition 3.1 and Lemma 3.1 in the next section. □In the next lemma we will use the following notation. $O(t)$, $o(t)$

are functions such that $O(t)/t$ is bounded and $o(t)/t \to 0$ as $t \to 0$.

Lemma 2.1. *1. If* $f \in \mathcal{N}F^{1+z}$ *then*

$$B(f, T, J) \geq \exp\{-o(|T|)\}$$

for all intervals $J \subset T$ *on which* f *is a diffeomorphism.*
2. If $f \in \mathcal{N}F^{1+Z}$ *then*

$$B(f, T, J) \geq \exp\{-O(|T|)\}$$

for all intervals $J \subset T$ *on which* f *is a diffeomorphism.*

Proof. If T is not contained in any neighbourhood of the form $U(x_0)$ as above then it is easy to see that $B(f, T, J)$ is uniformly bounded away from zero. So assume that T is contained in a neighbourhood $U(x_0)$ in which $f(x) = \pm|\phi(x)|^\alpha + f(x_0)$ where $\alpha \geq 1$ and $\phi\colon U(x_0) \to (-1, 1)$ is as in the definition of the classes $\mathcal{N}F^{1+z}$ and $\mathcal{N}F^{1+Z}$. Since $\alpha \geq 1$, the map $\phi_\alpha(x) = x^\alpha$ has Schwarzian derivative ≤ 0 and, hence, $B(\phi_\alpha, \phi(T), \phi(J)) \geq 1$. Therefore, and since f is a diffeomorphism on T,

$$B(f, T, J) = B(\phi_\alpha, \phi(T), \phi(J)) \cdot B(\phi, T, J) \geq B(\phi, T, J).$$

So we may assume that $f = \phi$ where ϕ is as before and that $|Df| \geq K$ on T. So write $T = [a, d]$ and $J = [b, c]$. Then

$$|B(f, T, J) - 1| = \left| \frac{\frac{f(c)-f(b)}{c-b} \cdot \frac{f(d)-f(a)}{d-a} - \frac{f(b)-f(a)}{b-a} \cdot \frac{f(d)-f(c)}{d-c}}{\frac{f(b)-f(a)}{b-a} \cdot \frac{f(d)-f(c)}{d-c}} \right|$$

$$\leq \frac{1}{K^2} \cdot \left| \frac{f(c)-f(b)}{c-b} \cdot \frac{f(d)-f(a)}{d-a} - \frac{f(b)-f(a)}{b-a} \cdot \frac{f(d)-f(c)}{d-c} \right|.$$

Let us first prove statement 1). First we note that if $f \in \mathcal{N}F^{1+z}$ then for each $x, y \in N$,

$$\frac{f(x)-f(y)}{x-y} = \frac{1}{x-y}\int_x^y Df(t)\,dt = \frac{Df(x)+Df(y)}{2} + o(|x-y|).$$

If f is C^2, this follows immediately from Taylor's Theorem. If Df satisfies the little Zygmund condition, this is shown in the first lemma in Section 2a. Using this, we get

$$\begin{aligned}
&\left|\frac{f(c)-f(b)}{c-b}\cdot\frac{f(d)-f(a)}{d-a} - \frac{f(b)-f(a)}{b-a}\cdot\frac{f(d)-f(c)}{d-c}\right| \\
&\le \left|\frac{(Df(b)+Df(c))(Df(a)+Df(d)) - (Df(a)+Df(b))(Df(c)+Df(d))}{4}\right| \\
&\quad + o(|T|) \\
&\le \left|\frac{(Df(a)-Df(c))(Df(b)-Df(d))}{4}\right| + o(|T|).
\end{aligned}$$

If f is C^2 then $(Df(a)-Df(c))(Df(b)-Df(d))$ is clearly at most $O(|T|^2)$ and then Statement 1) follows. If Df satisfies the little Zygmund condition then Df is Hölder with exponent $2/3$ (see the first lemma in Section 2a). Hence, in that case, $(Df(a)-Df(c))(Df(b)-Df(d))$ is also at most $O(|T|^{4/3})$. Again the result follows. This completes the proof of Statement 1).

So let us prove Statement 2) and assume that $f \in \mathcal{N}F^{1+Z}$. If $\log Df$ satisfies the Zygmund condition then we get a similar result with the difference that in the first step one has

$$\frac{f(x)-f(y)}{x-y} = \frac{1}{x-y}\int_x^y Df(t)\,dt = \frac{Df(x)+Df(y)}{2} + O(|x-y|)$$

(see the first lemma in Section 2a). Thus, we get the required result as before with $O(|T|)$ instead of $o(|T|)$. □

The Second Expansion Principle

Sometimes we need to have that $B(f,T,J)$ is strictly bigger than one. (This is for example the case if we want to apply the First Expansion Principle of the previous section.) This estimate is provided by the Second Expansion Principle. In order to state this principle, define

$$\gamma(T) = \min\left(\int_T \frac{dx}{d(x,C(f))}\quad,\quad 1\right)$$

where $d(x,C(f))$ is the distance of x to the set $C(f)$ of critical points of f, i.e., to $C(f) = \{z\,;\ Df(z)=0\}$. This quantity measures how big T is compared to its distance to $C(f)$ and could be called the 'projective length' of T in the punctured space $N \setminus C(f)$.

Theorem 2.2. **("Second Expansion Principle")** *Let $f \in \mathcal{N}F^{1+z}$. For each $\tau > 0$ there exist a constant $C_1 > 0$ such that if T is a τ-scaled neighbourhood of J and $T \cap C(f) = \emptyset$ then*

$$\log B(f,T,J) \geq C_1 \cdot \gamma(T) \,-\, o(|T|).$$

Proof. Let U be a neighbourhood of $C(f)$ as given in the definition of the class $\mathcal{N}F^{1+z}$. If T is not contained entirely in U then $\gamma(T) \to 0$ as $|T| \to 0$. From Lemma 2.1 one has

$$B(f,T,J) \geq 1 - o(|T|),$$

and, therefore, we are finishes in this case. On the other hand, if T is contained in U then f is of the form

$$f(x) = \pm|\phi(x)|^\alpha + f(x_0)$$

where $\phi\colon U(x_0) \to (-1,1)$ is so that $\log D\phi$ satisfies the little Zygmund condition and $\alpha > 1$. Now as before

$$\log B(\phi,T,J) \geq -o(|T|).$$

From this, and the fact that ϕ is a diffeomorphism it follows that it suffices to prove that

$$\log B(g,T,J) \geq C_1 \cdot \gamma(T)$$

where $g\colon (0,\infty) \to (0,\infty)$ is the map $g(x) = x^\alpha$, $J \subset T \subset (0,\infty)$ and T is a τ'-scaled neighbourhood of J. But this is easy to prove. Indeed,

$$\frac{B(g,T,J) - 1}{\gamma(T)}$$

is invariant under the multiplication $x \mapsto \lambda x$. So we may assume that the left endpoint of T is 1 and that the length of T is bounded by some universal number l. So it suffices to show that there exists a constant $C > 0$ with

$$\frac{B(g,T,J) - 1}{|T|} \geq C$$

for each such interval T of length at most l. But since $B(g,T,J)$ decreases if one increases J, we may assume that

$$T = [1, 1 + (1 + 2\tau')\epsilon]$$

and $J = [1 + \tau'\epsilon, 1 + (1 + \tau')\epsilon]$. But for these intervals one easily shows that

$$\lim_{\epsilon \to 0} \frac{B(g,T,J) - 1}{\epsilon} > 0$$

and thus the result follows. □

Exercise 2.1. If $f \in \mathcal{N}F^{1+z}$ then

$$|\log B(f,T,J)| \le o(|\gamma(T)|)$$

for each interval T on which f is a diffeomorphism where $\gamma(T)$ and $o(t)$ are as above Theorem 2.2. (Hint: write $f = g \circ \phi$ where ϕ is as before and $g(x) = x^\alpha$. In Lemma 2.1 we have shown that $|B(\phi,T,J)-1| \le o(|T|)$. So it suffices to show that

$$(*) \qquad |B(g,T,J)-1| \le o(|\gamma(T)|)$$

for all intervals $J \subset T$ not containing 0, where, in this context, $\gamma(T) = \int_T (1/x)\,dx$. Since $(*)$ is an homogeneous expression, i.e., it is invariant under multiplications $x \mapsto \lambda x$, we may assume that the left endpoint of T is at 1. But then $(*)$ follows from Lemma 2.1.)

Exercise 2.2. If in addition $D^2 f$ is Lipschitz then there exists $C_0 > 0$ such that

$$B(f^m,T,J) \ge \exp\{-C_0 \cdot \sum_{i=0}^{m-1} |f^i(T)|^2\} \ge \exp\{-C_0 \cdot S \cdot \sum_{i=0}^{m-1} |f^i(T)|\}.$$

(Hint: see de Melo and Van Strien (1989).)

Exercise 2.3. Consider the following cross-ratio:

$$D(T,J) = \frac{|J||T|}{|L \cup J||R \cup J|}$$

and define

$$A(g,T,J) = \frac{D(g(T),g(J))}{D(T,J)}.$$

Show that if f is C^3 and satisfies the non-flatness condition, then

$$A(f,T,J) - 1 \ge -C_0 \cdot |L| \cdot |R|$$

where L and R are the components of $T \setminus J$. Furthermore, show that there exist constants $\delta > 0$ and $\epsilon > 0$ such that if $f^n|T$ is a diffeomorphism, $\sum_{i=0}^{n-1} |f^i(J)| \le \delta$ and $|L| \cdot |R| < \epsilon |J|^2$ where L and R are the components of $T \setminus J$ then

$$(**) \qquad A(f^n,T,J) - 1 \ge 8\frac{|L| \cdot |R|}{|J|^2}.$$

Note that in the set-up one does not need any disjointness or bounds on the lengths of the intervals $T, \dots, f^{n-1}(T)$. In de Melo and Van Strien (1987), (1989) this cross-ratio was used to prove that no wandering intervals can exist for C^3 unimodal maps (not necessarily satisfying the negative Schwarzian derivative condition). We should note that $(**)$ is similar to Theorem I.2.2 of A. Schwartz because he also uses a bound on $\sum_{i=0}^{n-1} |f^i(J)|$ to bound the non-linearity of f^n on an interval T which is also larger than J. (Hint: see de Melo and Van Strien (1989).)

2.1 The Zygmund Conditions

In this section we will discuss the Zygmund conditions. Dennis Sullivan was the first to use these Zygmund conditions in this context. As we will see in this section these conditions are naturally related to distortions of cross-ratios. In order to define these conditions let $O(t), o(t)$ be functions such that respectively $O(t)/t$ is bounded and $o(t)/t \to 0$ as $t \to 0$.

Definition. A function $\phi\colon J \to \mathbb{R}$ on an open interval J satisfies the *Zygmund condition* if for all $x, y \in J$

$$\text{(Z)} \qquad \phi(x) + \phi(y) - 2\phi(\frac{x+y}{2}) = O(|x-y|)$$

and it satisfies the *little Zygmund condition* if

$$\text{(z)} \qquad \phi(x) + \phi(y) - 2\phi(\frac{x+y}{2}) = o(|x-y|).$$

The classes of these Zygmund functions are denoted by C^Z respectively C^z. Furthermore, we say that $\phi\colon J \to \mathbb{R}$ is *α-Hölder* at a point $y \in J$ if there exists a constant B such that

$$\frac{|\phi(x) - \phi(y)|}{|x-y|^\alpha} \le B$$

for all $x, y \in J$ with $|x - y| \le 1$.

Obviously, if ϕ is C^1 then it certainly satisfies both of these conditions and if ϕ is Lipschitz then it satisfies the Zygmund condition. The reverse is not true:

Example. The function $\phi\colon [0,1] \to \mathbb{R}$ defined by $\phi(x) = x\log(x) + Ax$ for $x > 0$ satisfies the Zygmund condition, i.e., $\phi \in C^Z$. Indeed, taking x, $y = tx$ in $\mathbb{R}^+$ one has

$$\frac{\phi(x) + \phi(y) - 2\phi(\frac{x+y}{2})}{y-x} = \frac{t\log t - (t+1)\log((1+t)/2)}{t-1}$$

which is bounded. Notice that the bound does not depend on A. Furthermore, ϕ is not Lipschitz. It is, however, α-Hölder for each $\alpha < 1$ since

$$\frac{\phi(x) - \phi(y)}{|x-y|^\alpha}$$

is bounded. This last bound depends strongly on the constant A.

If ϕ has bounded variation then it also does not necessarily satisfy the Zygmund condition and vice versa:

Example. The function $\phi(x) = \sqrt{x}$ and also any discontinuous piecewise monotone function has bounded variation but does not satisfy the Zygmund condition. On the other hand, $\phi(x) = x^2 \sin(1/x^2)$ has unbounded variation but does satisfy the Zygmund condition.

Next we will prove that any $\phi \in C^Z$ is α-Hölder for each $\alpha < 1$.

Lemma 2.2. *If $\phi \in C^Z$ and $J = (a,b)$ then*

$$\frac{1}{|J|}\int_J \phi(t)\,dt = \frac{\phi(a)+\phi(b)}{2} + O(|J|).$$

If $\phi \in C^z$ then

$$\frac{1}{|J|}\int_J \phi(t)\,dt = \frac{\phi(a)+\phi(b)}{2} + o(|J|).$$

Moreover, ϕ is α-Hölder for each $\alpha < 1$ at each interior point of J.

Proof. If $\phi \in C^Z$ then, writing $J = [m-\delta, m+\delta]$, one has

$$\begin{aligned}\frac{1}{|J|}\int_J \phi &= \frac{1}{|J|}\int_0^\delta (\phi(m-t)+\phi(m+t))\,dt \\ &= \frac{1}{|J|}\int_0^\delta (2\phi(m)+O(t))\,dt \\ &= \phi(m) + O(|J|) \\ &= \frac{\phi(m-\delta)+\phi(m+\delta)}{2} + O(|J|).\end{aligned}$$

Similarly, one proves the second statement. Let us now prove that at each interior point y of J, ϕ is α-Hölder for each $\alpha < 1$. Of course we may, without loss of generality, assume that $y = 0$, $\phi(0) = 0$ and that $\phi \in C^Z$ on $J = (-a,b) \subset (-1,1)$. Because of the Zygmund condition there exists a constant $C < \infty$ such that

$$\frac{\phi(x)}{x} - C \leq \frac{\phi(x/2)}{x/2} \leq \frac{\phi(x)}{x} + C \tag{2.1}$$

for all $x \in J$ with $x \neq 0$. Let us show that

$$\frac{|\phi(x)|}{|x|} \leq A|\log(x)| + B \tag{2.2}$$

for all $x \in J$ where $A > 0$ is so that $A\log(2) = C$ and B is so large that (2.2) holds for all $x \in (-a,-a/2) \cup (b/2,b)$. Because of (2.1), if (2.2) holds for some x then it also holds for $x/2$:

$$\frac{|\phi(x/2)|}{|x/2|} \leq \frac{\phi(x)}{x} + C \leq A|\log(x)| + B + C = A|\log(x/2)| + B.$$

In this way we have shown that (2.1) holds for all $x \in J$. Hence, ϕ is α-Hölder for each $\alpha < 1$. □

The Zygmund condition is quite natural in this context because of the following theorem.

Theorem 2.3. *Assume that* $f\colon (0,1) \to \mathbb{R}$ *is a* C^1 *diffeomorphism. Then the following statements are equivalent.*

1) *For each pair of intervals* $J \subset T \subset (0,1)$ *one has* $|B(f,T,J)-1| \le O(|T|)$*;*
2) $\log Df \in C^Z$.

Similarly, $|B(f,T,J)-1| \le o(|T|)$ *is equivalent to the* $\log Df \in C^z$.

Proof. That 2) implies 1) was already shown in Lemma 2.1. So let us prove that 1) implies 2). First note that if we take $L = [x-h,x]$, $R = [x,x+h]$ and $T = [x-h,x+h]$ all in the interior of $(0,1)$ then one has

$$
\begin{aligned}
1 - \frac{4|f(L)||f(R)|}{|f(T)|^2} &= \left(\frac{\dfrac{|f(L)|}{|L|} - \dfrac{|f(R)|}{|R|}}{\dfrac{|f(T)|}{|T|}} \right)^2 \\
&= \left(\frac{|Df(u) - Df(v)|}{|Df(w)|} \right)^2
\end{aligned}
$$

where $u,v,w \in T$. From 1) and the $C^{1+1/2}$ Koebe Principle from the previous section it follows that $\log Df$ is 1/2-Hölder on $(0,1)$; here we take in the statement of the $C^{1+1/2}$ Koebe Principle the interval $(0,1)$ to be the domain of the map g and the smaller interval to be $[x-h,x+h] \subset (0,1)$. So there exists a constant L such that

$$
\begin{aligned}
\left(\frac{|Df(u) - Df(v)|}{|Df(w)|} \right)^2 &= \left(\frac{|\dfrac{Df(u)}{Df(v)} - 1|}{\frac{|Df(w)|}{|Df(v)|}} \right)^2 \\
&\le L \cdot \left[|u-v|^{1/2} \cdot (1 + O(|w-v|^{1/2})) \right]^2 \le O(|T|).
\end{aligned}
$$

Consequently,

$$
\frac{4|f(L)||f(R)|}{|f(T)|^2} - 1 = O(|T|). \tag{2.3}
$$

But if we take the interval T as before and $J = \{x\}$ or $J = [x-h,x+h]$ then we get from 1) that

$$
\frac{|Df(x)|\dfrac{|f(T)|}{|T|}}{\dfrac{|f(L)|}{|L|}\dfrac{|f(R)|}{|R|}} = 1 + O(|T|) \tag{2.4}
$$

respectively

$$\frac{\left(\frac{|f(T)|}{|T|}\right)^2}{|Df(x-h)|\,|Df(x+h)|} = 1 + O(|T|). \tag{2.5}$$

Taking the product of the square of (2.4) and (2.5) and using $|L| = |R|$ gives

$$\frac{|Df(x)|^2|f(T)|^4}{16\,|Df(x-h)|\,|Df(x+h)|\,|f(L)|^2\,|f(R)|^2} - 1 = O(|T|).$$

Using (2.3),

$$\frac{|Df(x)|^2}{|Df(x-h)|\,|Df(x+h)|} - 1 = O(|T|).$$

and, therefore, $\log Df$ satisfies the Zygmund condition. □

Remark. From the lemma in Section 1a, it follows that $|B(f,T,J) - 1| \leq o(|T|^2)$ implies that f is a Möbius transformation.

3 Koebe Principles on Iterates

In this section we will present three Koebe Distortion Principles for iterates of a map.

The Koebe Distortion Principle

The first Koebe Principle holds for iterates f^m of a map $f \in \mathcal{N}F^{1+Z} \cup \mathcal{N}F^{1+z}$ which are diffeomorphic on some interval T and for which $\sum_{i=0}^{m-1} |f^i(T)|$ is bounded.

Theorem 3.1. **("Koebe Principle")** *For each $S, \tau > 0$ and each map $f \in \mathcal{N}F^{1+Z} \cup \mathcal{N}F^{1+z}$ there exists a constant $K(S,\tau)$ with the following property. If T is an interval such that $f^m|T$ is a diffeomorphism and if $\sum_{i=0}^{m-1} |f^i(T)| \leq S$ then for each interval $J \subset T$ for which $f^m(T)$ contains a τ-scaled neighbourhood of $f^m(J)$ one has*

$$\frac{1}{K(S,\tau)} \leq \frac{Df^m(x)}{Df^m(y)} \leq K(S,\tau), \quad \forall x, y \in J$$

where $K(S,\tau) = \dfrac{(1+\tau)^2}{\tau^2} \cdot e^{C \cdot S}$ and $C \geq 0$ only depends on f.

Proof. This is simply Theorem 2.1 and the previous Koebe Principle (Theorem 1.2) put together. □

A $C^{1+\alpha}$ Koebe Principle

Next we formulate a $C^{1+\alpha}$ Koebe Principle. (We shall not use this principle until the last chapter.) This principle holds for an iterate f^m of $f \in \mathcal{N}F^{1+z}$ which is diffeomorphic on some interval T and for which $\sum_{i=0}^{m-1} |f^i(T)|^s$ is bounded for some $s < 1$. (So in some sense the s-Hausdorff size of the orbit of T is bounded for some $s < 1$; in the previous theorem we had $s = 1$.) Under these additional conditions we get even C^α control on the distortion. Compare this principle also with the $C^{1+1/2}$ Koebe Principle from Section IV.1. A somewhat similar result for Misiurewicz maps can be found in Jiang (1991).

Theorem 3.2. **("$C^{1+\alpha}$ Koebe Principle")** *For each $\alpha < 1$, $\tau > 0$, $S < \infty$ and $f \in \mathcal{N}F^{1+Z} \cup \mathcal{N}F^{1+z}$ there exists $K < \infty$ such that for each interval T such that $f^m|T$ is a diffeomorphism for which*

$$\sum_{i=0}^{m-1} |f^i(T)|^\alpha \le S,$$

one has the following. For each interval $J \subset T$ for which $f^m(T)$ contains a τ-scaled neighbourhood of $f^m(J)$ one has

$$\left| \frac{Df^m(x)}{Df^m(y)} - 1 \right| \le K \cdot \left(\frac{|x-y|}{|T|} \right)^\alpha$$

for all $x, y \in J$.

Proof. Let $J \subset J' \subset T$ be so that $f^m(J')$ is a $\tau/2$-scaled neighbourhood of $f^m(J)$. By the previous Koebe Principle, the distortion of f^m is uniformly bounded on J'. It follows that $B(f^m, J', J)$ is uniformly bounded (from above). But

$$\log B(f^m, J', J) = \sum_{i=1}^{m-1} \log B(f, f^i(J'), f^i(J)).$$

According to the second Expansion Principle from Section IV.2, one has that

$$\sum_{i=1}^{m-1} \log B(f, f^i(J'), f^i(J)) \ge \sum_{i=1}^{m-1} \left[C_1 \cdot \gamma(f^i(J')) \; - \; C_2 \cdot |f^i(J')| \right].$$

As before, $\gamma(I)$ is the 'projective length' of an interval I,

$$\gamma(I) = \min \left(\int_I \frac{d\,x}{d(x, C(f))} \, , \, 1 \right).$$

Since the sum of the length of $f^i(J')$ is bounded, the sum of the first factors is bounded from below. So we necessarily have that $\sum_{i=1}^{m-1} \gamma(f^i(J'))$ is uniformly bounded and therefore

$$\sum_{i=1}^{m-1} \gamma(f^i(J)) \tag{3.1}$$

is uniformly bounded. Furthermore, by assumption,

$$\sum_{i=0}^{m-1} |f^i(J)|^\alpha \leq S. \tag{3.2}$$

Now we claim that if $f \in \mathcal{N}F^{1+z}$ and f is a diffeomorphism on an interval I then

$$\left|\frac{Df(x)}{Df(y)} - 1\right| \leq K\left(\gamma(I) + |I|^\alpha\right)\left(\frac{|x-y|}{|I|}\right)^\alpha \tag{3.3}$$

for all $x, y \in I$, where K is some constant which only depends on f, α, $\gamma(I)$ and $|I|$. Of course, the theorem follows from this claim: since $f^i|J$ has universally bounded distortion (3.3) gives (by taking $I = f^i(J)$)

$$\left|\frac{Df(f^i(x))}{Df(f^i(y))} - 1\right| \leq K'\left(\gamma(f^i(J)) + |f^i(J)|^\alpha\right)\left(\frac{|x-y|}{|J|}\right)^\alpha$$

for all $x, y \in J$. This, (3.1) and (3.2) imply the theorem.

So let us prove the claim. If I is far away from the critical points of f then it simply follows from the fact that Df is at least C^α. (If $f \in \mathcal{N}F^{1+z}$ then Df satisfies the little Zygmund condition and is therefore C^α for each $\alpha < 1$.) If I is close to a critical point then one uses that f is near these points the composition of a map as before and a polynomial map. From this the claim easily follows. □

The Macroscopic Koebe Principle

The last Koebe Principle holds for maps $f \in \mathcal{N}F^{1+Z} \cup \mathcal{N}F^{1+bv}$ (and so are less differentiable) on intervals T on which some iterate f^n is monotone (and so not necessarily diffeomorphic). To deal with this, we will have to rely on disjointness assumptions: otherwise the orbit of T could contain many critical points. Therefore we need to split up the iterates of T into collections of disjoint intervals. We start by discussing some disjointness properties of families of intervals.

Definition. The *intersection multiplicity* of a finite collection of intervals in N is the maximal number of intervals in this collection whose interior has a non-empty intersection.

Theorem 3.3. **("Macroscopic Koebe Principle")** *Given $f \in \mathcal{N}F^{1+bv} \cup \mathcal{N}F^{1+Z}$ and $p \in \mathbb{N}$ there exists a strictly positive function $B_0 \colon \mathbb{R}^+ \to \mathbb{R}^+$ such that for any pair of intervals $J \subset T$, any $m \geq 0$ and any $0 < \tau < 1$, satisfying the following conditions*

1. *$f^m(T)$ contains a τ-scaled neighbourhood of $f^m(J)$;*

2. *the intersection multiplicity of* $\{T, f(T), \ldots, f^{m-1}(T)\}$ *is at most* p*;*
3. $f^i(J) \cap C(f) = \emptyset$, $i \in \{0, 1, \ldots, m-1\}$ *and* $f^m|T$ *is a homeomorphism (so* T *can contain an inflection point of* f *but* J *cannot).*

Then

$$T \text{ is a } B_0(\tau)\text{-scaled neighbourhood of } J.$$

Remark. If $f \in \mathcal{N}F^{1+z}$ then one can replace 1)-3) in the previous theorem by

1. $f^m(T)$ contains a τ-scaled neighbourhood of $f^m(J)$;

$\tilde{2}$. $\sum_{i=0}^{m-1} |f^i(T)| \leq p$;

$\tilde{3}$. $f^m|T$ is a diffeomorphism.

The proof of this theorem will occupy the remainder of this section. We first need a proposition and two lemmas. The first proposition is concerned with the notion of intersection multiplicity.

Proposition 3.1. *Let* $\mathcal{W}$ *be a finite collection of intervals in* N *with intersection multiplicity at most* p *then there exists a partition of the collection*

$$\mathcal{W} = A_1 \cup A_2 \cup \cdots \cup A_{2p},$$

such that A_k *consists of mutually disjoint intervals for* $k = 1, 2, \ldots, 2p$.

Proof. Clearly we may assume that N is connected. So we only need to consider the cases that N is equal to an interval or a circle.

Proof if $N = [-1, 1]$. Of course we may assume that all the intervals in $\mathcal{W}$ are open. We claim that if $N = [-1, 1]$ then there are p classes $A_1, A_2, \ldots, A_p$ which form the desired partition of $\mathcal{W}$. Indeed let $\mathcal{I}$ be some collection of intervals in $[-1, 1]$. For $I \in \mathcal{I}$ let $\text{next}\{I, \mathcal{I}\}$ be some interval $I' \in \mathcal{I}$ such that i) I' does not intersect I and is to the right of I and ii) there is no interval $J \in \mathcal{I}$ satisfying i) which is closer to I (if there is no interval to the right of I take $\text{next}\{I, \mathcal{I}\}$ to be the empty set). Similarly $\mathcal{R}(\mathcal{I})$ is some interval in $\mathcal{I}$ such that there is no interval in $\mathcal{I}$ which has points to the right of $\mathcal{R}(\mathcal{I})$. For $k = 1, 2, \ldots, p$ we define inductively A_k as follows. Let $A_0 = \emptyset$ and suppose that we have defined by induction A_{k-1}. If $k \leq p$ and $\mathcal{B}_k = \mathcal{W} \setminus \cup_{i=1,\ldots,k-1} A_i = \emptyset$ then let $A_k = \emptyset$. If $\mathcal{B}_k$ is non-empty then take an interval $I_k \in \mathcal{B}_k$ so that there is no interval in $\mathcal{B}_k$ containing points to the left of I_k. Then define

$$A_{k,1} = \{I_k\},$$

$$A_{k,n+1} = A_{k,n} \cup \text{next}\left\{ \mathcal{R}(A_{k,n}), \left(\mathcal{W} \setminus (A_0 \cup \cdots \cup A_{k-1} \cup A_{k,n})\right) \right\},$$

and

$$A_k = \bigcup_{n \geq 1} A_{k,n}.$$

Then A_k is a collection of intervals with disjoint interiors and the collections $A_1, \ldots, A_p$ are mutually disjoint. Now we will show that $\mathcal{W} = A_1 \cup \cdots \cup A_p$. Suppose this is not the case and there exists an interval I in the collection $\mathcal{W} \setminus (A_1 \cup \cdots \cup A_p)$. Then

$$I \in \mathcal{W} \setminus (A_0 \cup \cdots \cup A_{k-1} \cup A_k)$$

for each $k = 1, \ldots, p$ and since $I \notin A_k$ it follows from the inductive definition above that for $i = 1, 2, \ldots, p$ there exists $T_i \in A_i$ such that the left boundary point of T_i is to the left of (or equal to) the left boundary point x of I. But then since all these intervals are open, points just to the right of this boundary point x are contained in I as well as in T_k, $k = 1, 2, \ldots, p$. This contradicts the assumption that the intersection multiplicity of $\mathcal{W}$ is at most p.

Proof if $N = S^1$. If $N = S^1$ then choose some $x \in S^1$ and let $I_1, \ldots, I_r$ be intervals in $\mathcal{W}$ which contain x. Then $r \leq p$. Since $S^1 \setminus \{x\} \simeq (-1, 1) \subset [-1, 1]$ it follows from the previous case that $\mathcal{W} \setminus \{I_1, \ldots, I_r\}$ can be disjointly decomposed into collections $A_1, \ldots, A_p$ of disjoint intervals. Then $\mathcal{W} = A_1 \cup \cdots \cup A_p \cup \cup_{k=1,\ldots,r}\{I_k\}$ has the desired properties. □

Lemma 3.1. *Let $f \in \mathcal{N}F^{1+bv} \cup \mathcal{N}F^{1+Z}$ and $\{T_0, T_1, \ldots, T_{m-1}\}$ a collection of intervals in N with intersection multiplicity p and assume that none of these intervals contains points of $C(f)$. Then there exists $V < \infty$ such that*

$$\sum_{i=0}^{m-1} \log B(f, T_i, J_i) \geq -V$$

for all subintervals J_i of T_i.

Proof. If $f \in \mathcal{N}F^{1+Z}$ then the lemma follows immediately from Lemma 2.1. So assume that $f \in \mathcal{N}F^{1+bv}$. For each $x_i \in C(f)$, let $U(x_i)$ be an open interval neighbourhood of x_i on which f is of the form $f(x) = \phi_{\alpha_i} \circ \phi_i(x) + f(x_i)$, where $\phi_{\alpha_i}(x) = \pm|x|^{\alpha_i}$ and $\alpha_i \geq 1$. Here $\phi_i \colon U(x_i) \to (-1, 1)$ is a homeomorphism and $\log|D\phi_i|$ exists almost everywhere and has bounded variation. Hence $V' = \sum_{i\,;\,x_i \in C(f)} Var(\log|D\phi_i|)$ is finite. Let $\mathcal{U} = \cup_{x_i \in C(f)} U(x_i)$ and let $\mathcal{V}$ be a neighbourhood of $C(f)$ such that $\mathrm{int}\,(N \setminus \mathcal{V}) \cup \mathrm{int}\,(\mathcal{U}) = N$. Then the restriction of $\log|Df|$ to $N \setminus \mathcal{V}$ exists almost everywhere and $\log|Df|$ has bounded variation on this set. So let

$$V_f = Var\,(\log[|Df\,|(N \setminus \mathcal{V})\,]) < \infty$$

and let $I_1 = \{0 \leq i \leq m-1\,;\ T_i \cap \mathcal{V} = \emptyset\}$ and $I_2 = \{0 \leq i \leq m-1\,;\ T_i \subset \mathcal{U}$ and $i \notin I_1\}$ and $I_3 = \{0, 1, \ldots, m-1\} \setminus (I_1 \cup I_2)$. By definition if $i \in I_3$ the set T_i contains a component of $\mathcal{U} \setminus \mathcal{V}$.

First assume that $i \in I_1$. Let L_i and R_i be the components of $T_i \setminus J_i$. For $u_i, v_i \in T_i$ let (u_i, v_i) be the open interval connecting u_i and v_i. Because

Df exists almost everywhere there exist $m_i \in J_i$, $l_i \in L_i$, $r_i \in R_i$ and $t_i \in T_i$ such that Df exists in these points and such that $\frac{|f(J_i)|}{|J_i|} \geq |Df(m_i)|$, $\frac{|f(T_i)|}{|T_i|} \geq |Df(t_i)|$, $\frac{|f(L_i)|}{|L_i|} \leq |Df(l_i)|$ and $\frac{|f(R_i)|}{|R_i|} \leq |Df(r_i)|$. Using this in the definition of $B(f,T_i,J_i)$ we find that there exist $l_i \in L_i$, $r_i \in R_i$, $m_i \in J_i$ and $t_i \in T_i$ such that

$$\log B(f,T_i,J_i) \geq \log\left(\frac{|Df(m_i)||Df(t_i)|}{|Df(l_i)||Df(r_i)|}\right), \tag{3.4}$$

where

$$m_i \in (l_i, r_i). \tag{3.5}$$

From (3.4) and the choice of the points l_i, m_i, r_i, t_i one has

$$\begin{aligned} \log B(f,T_i,J_i) \geq -\{&|\log|Df(m_i)| - \log|Df(l_i)||\\ &+ |\log|Df(t_i)| - \log|Df(r_i)||\} \end{aligned} \tag{3.6}$$

and also

$$\begin{aligned} \log B(f,T_i,J_i) \geq -\{&|\log|Df(m_i)| - \log|Df(r_i)||\\ &+ |\log|Df(t_i)| - \log|Df(l_i)||\}. \end{aligned} \tag{3.7}$$

Rename the points l_i, m_i, r_i, t_i in increasing order a_i, b_i, c_i, d_i. From (3.5) one gets that either $(l_i, m_i) \cap (t_i, r_i) = \emptyset$ or $(t_i, l_i) \cap (m_i, r_i) = \emptyset$, and so we can use either (3.6) or (3.7) and get

$$\begin{aligned} \log B(f,T_i,J_i) &\geq -\{|\log|Df(b_i)| - \log|Df(a_i)||\\ &\qquad + |\log|Df(d_i)| - \log|Df(c_i)||\}\\ &\geq -Var(\log|Df|T_i|) \text{ where } T_i \subset N \setminus \mathcal{V}. \end{aligned} \tag{3.8}$$

Now consider $i \in I_2$. Then T_i is contained in some component $U(x_i)$ of $\mathcal{U}$ (and does not intersect $C(f)$) and so f has the form $f(x) = \phi_{\alpha_i} \circ \phi_i + f(x_i)$ where ϕ_{α_i} and ϕ_i are as above. Hence

$$B(f,T_i,J_i) = B(\phi_{\alpha_i}, T_i', J_i') \times B(\phi_i, T_i, J_i).$$

Here $T_i' = \phi_i(T_i)$ and $J_i' = \phi_i(J_i)$. Because $\alpha_i \geq 1$, the Schwarzian derivative of ϕ_{α_i} is less or equal to 0. Hence $B(\phi_{\alpha_i}, T_i', J_i') \geq 1$ and therefore

$$\begin{aligned} \log B(f,T_i,J_i) &\geq 0 + \log B(\phi_i, T_i, J_i)\\ &\geq -Var(\log|D\phi_i|T_i|). \end{aligned} \tag{3.9}$$

Finally, if $i \in I_3$ then T_i contains a component of $\mathcal{U} \setminus \mathcal{V}$. It is not hard to see that in this case there exists a universal constant $K < \infty$ such that

$$\log B(f,T_i,J_i) \geq -K \text{ if } T_i \text{ contains a component of } \mathcal{U} \setminus \mathcal{V}. \tag{3.10}$$

Since the intersection multiplicity of $\{T_0, \ldots, T_{m-1}\}$ is at most p, using Proposition 3.1, one can write this collection as the union of $A_1, \ldots, A_{2p}$ where A_j consists of a collection of mutually disjoint intervals. Hence from (3.8), (3.9) and (3.10) one gets

$$\sum_{i=0}^{m-1} \log B(f, T_i, J_i) \geq -2p \cdot \left(V_f + \sum_{x_j \in C(f)} \mathrm{Var}(\log |D\phi_j|) + K \right).$$

The lemma follows. □

Lemma 3.2. *For each $f \in \mathcal{N}F^{1+Z} \cup \mathcal{N}F^{1+bv}$ there exists $A_1 > 0$ with the following property. If J, T are intervals with $\mathrm{cl}\,(J) \subset \mathrm{int}\,(T)$ and L and R the components of $T \setminus J$ such that*

1. $|L| \leq |J|$ *or* $|R| \leq |J|$ *and*
2. $J \cap C(f) = \emptyset$,

then

$$B(f, T, J) \geq A_1.$$

Proof. By possibly renaming L and R, we may consider the case that $|R| \leq |J|$. It is easy to show that there exists a universal number $A_0 > 0$ such that for all such intervals

$$\frac{|f(J)|/|J|}{|f(R)|/|R|} \geq A_0. \tag{3.11}$$

Hence, in this case,

$$B(f, T, J) \geq \frac{|f(T)|}{|f(L)|} \cdot \frac{|L|}{|T|} \cdot A_0 \geq A_0 \cdot \frac{|L|}{|T|}.$$

If $|L| \geq |J|$ then it follows from this and $|R| \leq |J|$ that $B(f, T, J) \geq A_0 \cdot \frac{|L|}{|L|+|J|+|R|} \geq A_0 \cdot \frac{|L|}{|L|+|L|+|L|} = A_0 \cdot \frac{1}{3} \geq A_1$ and the lemma is proved. So assume that $|L| \leq |J|$. Then exactly as before,

$$\frac{|f(J)|/|J|}{|f(L)|/|L|} \geq A_0. \tag{3.12}$$

Writing $d = (\frac{|J||J|}{|R||L|})/(\frac{|J|}{|R|} + \frac{|J|}{|L|})$ and using (3.11) and (3.12),

$$B(f, T, J) = \frac{\frac{|f(J)|}{|f(R)|} + \frac{|f(J)|}{|f(R)|}\frac{|f(J)|}{|f(L)|} + \frac{|f(J)|}{|f(L)|}}{\frac{|J|}{|R|} + \frac{|J|}{|R|}\frac{|J|}{|L|} + \frac{|J|}{|L|}} \geq \frac{A_0 + A_0^2 \cdot d}{1 + d}. \quad \square$$

Proof of the Macroscopic Koebe Principle. Let $m_0 \leq m$ be the smallest number such that $|f^{m_0}(L)| \geq \tau |f^{m_0}(J)|$ and $|f^{m_0}(R)| \geq \tau |f^{m_0}(J)|$. Let A_1 be the number from Lemma 3.2 and assume $A_1 < 1$. Let $V > 0$ be the number from Lemma 3.1. Let

$$B'_0 = [A_1]^{3p\#C(f)+1} \cdot e^{-V}.$$

We claim that

$$B(f^{m_0}, T, J) \geq B'_0.$$

Indeed, let $t(1) < t(2) < \cdots < t(s) < m_0$ be the integers $t < m_0$ such that $f^t(T) \cap C(f) \neq \emptyset$. Since the intersection multiplicity of $\{T, f(T), \ldots, f^{m-1}(T)\}$ is at most p one gets $s \leq 3p\#C(f)$. From the choice of m either $|f^{t(i)}(L)| \leq \tau|f^{t(i)}(J)| \leq |f^{t(i)}(J)|$ or $|f^{t(i)}(R)| \leq \tau|f^{t(i)}(J)| \leq |f^{t(i)}(J)|$. So from Lemma 3.2

$$B(f, f^{t(i)}(T), f^{t(i)}(J)) \geq A_1.$$

From Lemma 3.1

$$\sum_{\{j \leq m_0 - 1, j \notin \{t(1), \ldots, t(s)\}\}} \log B(f, f^j(T), f^j(J)) \geq -V.$$

Hence the claim follows. From the claim and the definition of the operator B the theorem easily follows. □

4 Some Simplifications and the Induction Assumption

We shall prove Theorem A by induction on the number of turning points of f. Let $\mathcal{A}^d$ be the collection of all continuous endomorphisms of N with precisely d turning points. We will prove inductively that

$$(Ind_d) \qquad \text{maps in } \bigcup_{i=0}^{d} \left(\mathcal{A}^i \cap (\mathcal{N}F^{1+Z} \cup \mathcal{N}F^{1+bv})\right)$$
$$\text{have no wandering intervals.}$$

Because the proof of Theorem A goes by induction on the number of turning points we have to consider a more general situation: the manifold N is not necessarily connected (but does consist of a finite number of components). As before we say that $f \colon N \to N$ is l-modal if f has l turning points in N.

We should note that allowing N to be disconnected does not give a more general result. Indeed, if a map $f \colon N \to N$ then one of the connected components N' of N is periodic of period s and if $f \colon N \to N$ has a wandering interval then one can choose N' so that the map $f^s \colon N' \to N'$ also has a wandering interval. Therefore, if we prove the theorem for connected manifolds we also get the same theorem for non-connected manifolds. However, if f has d turning points, the map f^s may have more than d turning points. This is the main reason to we allow N to be disconnected because then Ind_{d-1} can be restated as follows: *if there exists a finite disjoint union of intervals which is invariant by f and contains a wandering interval then the union of these intervals contains* *at least d turning points.*This implies that we can 'decompose the interval' in disconnected pieces.

By extending f to a slightly bigger interval we may assume that

$$f(\partial N) \subset \partial N.$$

It suffices to prove the theorem for maps f such that none of the turning points is contained in (the closure) of a wandering interval. Indeed, otherwise f also has a wandering interval W containing a turning point c in its interior. Now modify the map f in a small neighbourhood $V \subset W$ of c to a map g such that $f = g$ on $N \setminus V$ for which $g|V$ has a unique turning point in c and $g(c)$ is not contained in a wandering interval. Then $f(W)$ is still a wandering interval for g (the forward iterates of $f(W)$ under f and g are the same because these iterates never enter W) and $g(c)$ is not contained in a wandering interval of g. Repeating this procedure for every wandering interval which contains a turning point in its closure we get a map g which still has wandering intervals but such that none of its turning points is contained in the closure of a wandering interval. So it suffices the prove the theorem for g.

Furthermore, for each turning point c of f there exists a neighbourhood S_c of c and a continuous involution $\tau \colon S_c \to S_c$ such that $f(\tau(x)) = f(x)$ and $\tau(x) \neq x$ for all $x \in S_c \setminus \{c\}$. Since the critical point is non-flat for each map in $\mathcal{N}F^{1+bv} \cup \mathcal{N}F^{1+Z}$ the involution τ is Lipschitz.

Before getting involved in the rather technical properties of the pullbacks which we will introduce in the next section, let us, as an entertainment, illustrate this notion with some pictures. This discussion is not going to be used in the next sections, so these pictures can be skipped by the reader who does not like pictures.

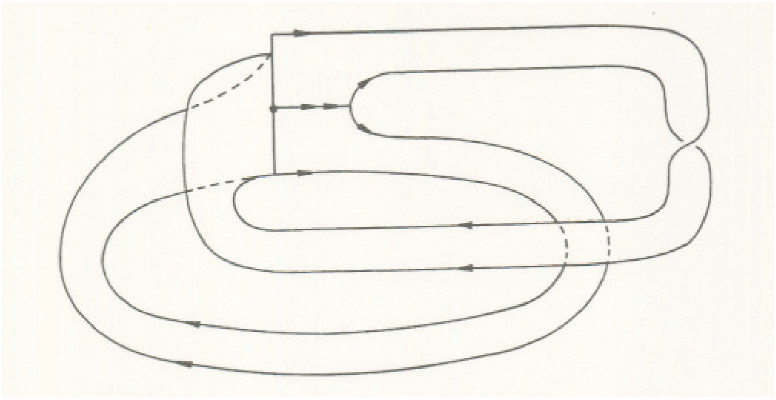

Fig. 4.1: A flow on a branched manifold.

We are going to consider a flow on a two-dimensional branched manifold, an interval I in this manifold which is a section of this flow and a map f which is the first return map to I. For example, in Figure 4.1, we represent a flow on a branched manifold with boundary whose first return map to I is a unimodal map. The interval I is precisely the set of branch points (on the left hand side of I we have two pieces of surfaces that meet tangentially at I with just one piece of surface on the right hand side). The flow has a saddle point s. The two horizontal arrows near the saddle point indicates that the contracting eigenvalue dominates the expanding eigenvalue at s so that the orbits passing near s get strongly contracted towards the unstable separatrix as indicated. Both unstable separatrices intersect I at the critical value $f(0)$ through different branches.

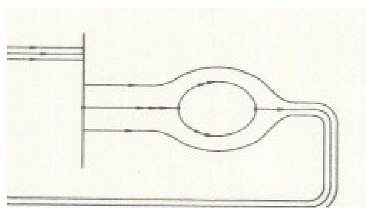

Fig. 4.2: A flow on a branched manifold which corresponds to a multimodal map.

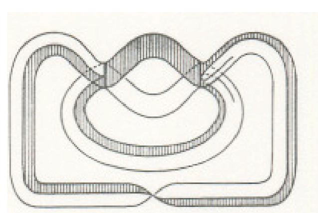

Fig. 4.3: An embedding of the flow from Figure 4.1 in a branched manifold without corners.

Notice that we can represent in the same way any map using one saddle point for each turning point and as many 'layers' meeting tangentially at I as the number of pre-images of points. The inflection points correspond to pairs of saddles as indicated in Figure 4.2. The branched manifold B_0 of Figure 4.1 has a boundary but it can be embedded in a branched manifold B without boundary as in Figure 4.2. B is consists of a cylinder and a Möbius strip which touch each other tangentially at two lines, one containing I and the other containing s. The shadowed area is part of B_0 in B. The complement is the 'imaginary part'.

As Mañé (1985) observed, Theorem III.5.1 implies that the question of density of Axiom A for interval maps can be reduced to the following conjecture:

Conjecture: Let $f\colon N \to N$ be a C^∞ map and assume that the ω-limit set of a critical point contains a critical point. Then we can increase the number of critical points in the forward orbit of some critical point by an arbitrarily small perturbation of the map in the C^∞ topology.

Note that this conjecture is in the same direction as the proof of Jakobson's C^1 Density Theorem III.2.2, see also Exercise III.2.4. We see that in terms of the flow generating the corresponding map, this corresponds to creating a saddle connection by a small perturbation of the flow. So this conjecture would be an extension to branched manifolds of Peixoto's theorem for flows on orientable surfaces, see Palis and de Melo (1982, Chapter 4).

5 The Pullback of Space: the Koebe/Contraction Principle

In this section we will start the proof of the non-existence of wandering intervals for maps in $\mathcal{N}F^{1+Z} \cup \mathcal{N}F^{1+bv}$. This will be proved by contradiction. More precisely, suppose that J is a wandering interval and that J is not contained in a larger wandering interval. The strategy of the proof is to show that in this case an interval I needs to exist which strictly contains J such that $\inf_{n\geq 0} |f^n(I)| = 0$. For simplicity we shall write

$$J_n = f^n(J).$$

Using the next principle, which also can be found in Lyubich (1989), it follows that I is a wandering interval (since $f^n(J)$ does not converge to a periodic orbit the same holds for $f^n(I)$), contradicting the maximality of J.

Contraction Principle 5.1. If I is an interval such that $\inf_{n\geq 0} |f^n(I)| = 0$ then I is a wandering interval or there exists a periodic orbit $\mathcal{O}$ such that $f^k(I) \to \mathcal{O}$ as $k \to \infty$. In particular, if I contains a wandering interval J then it is also a wandering interval.

Proof. Let $\mathcal{I} = \cup_{n\geq 0} f^n(\mathrm{int}\,(I))$. Clearly $\mathcal{I}$ is forward invariant. First suppose that there exists a component U of $\mathcal{I}$ and $n > 0$ such that $f^n(U) \cap U \neq \emptyset$. Since $\mathcal{I}$ is forward invariant this implies $f^n(U) \subset U$. There are three cases.

1. U is an interval which contains a fixed point p of $f^n\colon U \to U$ in its interior. In this case some iterate of I contains this fixed point of f^n in its closure and since $\inf_{k\geq 0} |f^k(I)| = 0$ this fixed point of f^n attracts I, $f^k(I) \to O(p)$ as $k \to \infty$. So we are finished in this case.

2. U is an interval and there exists no fixed point as in 1. Then $\mathrm{cl}\,(U)$ contains in its boundary an attracting fixed point p of $f^n\colon \mathrm{cl}\,(U) \to \mathrm{cl}\,(U)$. If $f^n(U) \neq U$ then for every $x \in \mathrm{cl}\,(U)$ one gets $f^{kn}(x) \to p$ as $k \to \infty$. If $f^n(U) = U$ then the boundary point $\{q\} = \partial U \setminus \{p\}$ is a repelling periodic point and $\inf_{k\geq 0} |f^k(I)| = 0$ implies that no iterate of I contains q in its closure.

Since every point in int (U) is asymptotic to p under iterates of f^n this implies that $f^{kn}(x) \to p$ for $x \in I$ as $k \to \infty$. Again the result follows.

3. U is a circle. Since U is a union of iterates of I, there exists a finite collection $n_1 < \cdots < n_r$ of positive integers such that $\bigcup_{i=1}^r f^{n_i}(\text{int}\,(I))$ covers S^1. But since $\inf_{n\geq 0} |f^n(I)| = 0$ this implies that there exist an integer $n > n_r$ such that $f^n(I)$ is strictly contained in $f^{n_i}(\text{int}\,(I))$ for some $i = 1, \ldots, r$. But then f^{n-n_i} has an attracting fixed point in $f^{n_i}(\text{int}\, I)$ which attracts I. So again the result follows.

Now assume that for every component U of $\mathcal{I}$ one has $f^n(U) \cap U = \emptyset$ for all $n \geq 1$. Since $\mathcal{I}$ is forward invariant and this holds for each component, this implies that $f^n(U) \cap f^m(U) = \emptyset$ for all $n > m \geq 0$. It follows that U and therefore I is a wandering interval (or asymptotic to a periodic orbit). $\square$

Definition. The *pullback* of $P_n \supset J_n$ is the sequence of intervals $\{P_i\,;\ i = 0, 1, \ldots, n\}$ where P_{i-1} is the maximal interval containing J_{i-1} such that

$$f(P_{i-1}) \subset P_i.$$

for each $i = 1, \ldots, n$. The integers i for which P_i contains a turning point in its closure are called the *cutting times*. It will turn out to be useful to call n also a cutting time. This pullback is *monotone* if $f^n|P_0$ is monotone and $f^n(P_0) = P$. Furthermore it is said to be *diffeomorphic* if $f^n|P_0$ is a diffeomorphism and $f^n(P_0) = P$ and *unimodal* if P_i contains at most one turning point for each i.

In Figure 5.1 we represent the unimodal pullback and in Figure 5.2 the monotone pullback of an interval. The horizontal line segments in these pictures represent pieces of I.

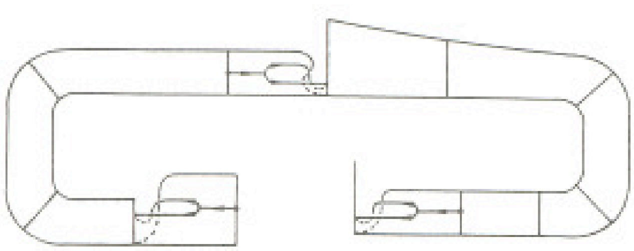

Fig. 5.1: The unimodal pullback.

All the pullbacks we will consider are unimodal because of the following lemma:

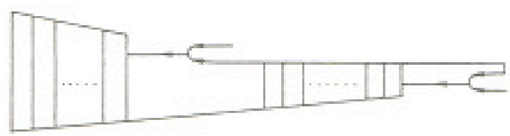

Fig. 5.2: The multiimodal pullback.

Lemma 5.1. *There exists $\eta > 0$ such that if $P_n \supset J_n$ and $|P_n| \leq \eta$ then the pullback of P_n is unimodal.*

Proof. Follows from the Contraction Principle and since by assumption no wandering interval contains two turning points. □

Next we formulate the basic principle which we shall use in the proof of Theorem A. This principle states that the pullback of an ϵ-scaled neighbourhood of J_n has high intersection multiplicity if n is large. Using some combinatorial information we shall prove that this intersection multiplicity is not always large, thereby showing that wandering intervals cannot exist.

Contraction/Koebe Principle 5.2. For each $\epsilon > 0$ and each $p \in \mathbb{N}$ there exists $N_0(\epsilon, p)$ with the following properties. Let $P_n \supset J_n$ contain a ϵ-scaled neighbourhood of J_n. If the pullback of P_n has intersection multiplicity $\leq p$ then $n \leq N_0(\epsilon, p)$.

Proof. Let $\rho = \min(\epsilon, 1/2)$ and $\hat{P}_n \subset P_n$ be a ρ-scaled neighbourhood of J_n. Let $m(0) < m(1) < \cdots < m(\ell) = n$ be the cutting times of the pullback of P_n. Since the intersection multiplicity of the pullback is $\leq p$, $\ell \leq p \cdot d$, where d is the number of turning points of f. Let $\hat{P}_i^{\pm}$ be the components of $\hat{P}_i \setminus J_i$. For $t = 0, \ldots, \ell - 1$, let $\hat{P}^+_{m(t)}$ be the component which contains the turning point. Now the map

$$f^{n-m(\ell-1)+1}\colon \hat{P}_{m(\ell-1)+1} \to \hat{P}_n$$

is monotone. Since $\hat{P}_n$ is a ρ-scaled neighbourhood of J_n it follows from the Macroscopic Koebe Principle that both components of $\hat{P}_{m(\ell-1)+1} \setminus J_{m(\ell-1)+1}$ have length at least $B_0(\rho)$. From this and the non-flatness of the turning points we get that there exists a universal constant $C \in (0, 1)$ such that

$$|\hat{P}^-_{m(\ell-1)}| \geq C \cdot B_0(\rho) \cdot |J_{m(\ell-1)}|.$$

Because $\hat{P}_{m(\ell-1)}$ is symmetric around a turning point, we also have that $|\hat{P}^{+}_{m(\ell-1)}| \geq |J_{m(\ell-1)}|$. Therefore

$$|\hat{P}^{\pm}_{m(\ell-1)}| \geq C \cdot B_0(\rho) \cdot |J_{m(\ell-1)}|.$$

Repeating this $\ell \leq p \cdot d$ times we get, letting $g(x) = C \cdot B_0(x)$, that

$$|\hat{P}^{\pm}_0| \geq g^{\ell}(\rho) \cdot |J|.$$

Hence there exists an interval P which does not depend on n and which *strictly* contains the wandering interval J such that $|f^n(P)| \leq (1+2\rho)|J_n| \leq 2|J_n|$ for each n as above. Now if there exists no upper bound $N_0(\epsilon, p)$ for n then we would get $|f^{n_i}(P)| \leq 2|J_{n_i}| \to 0$ for some sequence $n_i \to \infty$ and since P contains J this would imply from the Contraction Principle that P is also a wandering interval. But this contradicts the maximality of J. □

6 Disjointness of Orbits of Intervals

In this section we will give some upper bounds on the intersection multiplicity of orbits of intervals. These bounds are needed in order to apply the Macroscopic Koebe Principle. Let $J_n = f^n(J)$ where J is the wandering interval from above. We will define a natural neighbourhood T_n of J_n such that its monotone pullback has good disjointness properties. (For those familiar with the circle homeomorphisms these neighbourhoods will coincide with the neighbourhood $[f^{q_{k-2}}(J), f^{q_{k-2}+(a_k-1)q_{k-1}}(J)]$ of $f^{q_k}(J)$ when $n = q_k$ and with $[f^{q_{k-2}+(i-1)q_{k-1}}(J), f^{q_{k-2}+(i+1)q_{k-1}}(J)]$ of $f^{q_{k-2}+iq_{k-1}}(J)$ when $n = q_{k-2}+iq_{k-1}$ and $1 \leq i \leq a_k - 1$.)

Definition. We say that J_{n_1} and $J_{n_2} = f^k(J_{n_1})$ have the *same orientation* if f^k is orientation preserving. If $n \in \mathbb{N}$, we say that J_k is a *predecessor* of J_n if $0 \leq k < n$, if J_k and J_n have the same orientation and if $J_s \subset (J_k, J_n)$ and $0 \leq s < n$ implies that J_s and J_n have different orientations. If J_n has a predecessor to its left (right) then we denote the corresponding iterate by $L(n)$ (respectively $R(n)$). Finally, J_n has a *successor* J_{n+a} if

1. J_{n-a} is predecessor of J_n (with $0 < a \leq n$)
2. $f^a|[J_{n-a}, J_{n+a}]$ is monotone, orientation preserving, and its image contains no predecessor of J_n (if $L(n)$ and $R(n)$ both exist and if for example $n - a = L(n)$ then this implies that $f^a[J_{n-a}, J_n] \subset [J_n, J_{R(n)})$ where $a = n - L(n)$);
3. if $J_k \subset (J_n, J_{n+a})$ and $k = 0, 1, \ldots, n+a-1$ then the intervals J_k and J_{n+a} have different orientations.

Next we define the *natural neighbourhood* T_n of J_n to be the biggest open interval containing J_n which contains no neighbourhood of a predecessor or successor.

Remark. Of course J_n can have at most one predecessor on each side and it has a predecessor to its right if there exists an interval J_s with $s < n$ to the right of J_n with the *same orientation* as J_n. Moreover, as we will see in the lemma below, an interval J_n has at most one successor; so denote this successor by $J_{s(n)}$. Therefore, if J_n has two predecessors and no successor then the natural neighbourhood T_n of J_n is equal to $T_n = [J_{L(n)}, J_{R(n)}]$ and if it has a successor then it is equal to

$$T_n = [J_{L(n)}, J_{s(n)}] \text{ or } T_n = [J_{s(n)}, J_{R(n)}].$$

Lemma 6.1. *For every $n \in \mathbb{N}$, J_n can have at most one successor.*

Proof. Suppose that J_n had two successors. Then it also has two predecessors, $J_{L(n)}$ and $J_{R(n)}$. By definition $f^{n-L(n)}\colon [J_{L(n)}, J_n] \to [J_n, J_{R(n)}]$ and $f^{n-R(n)}\colon [J_n, J_{R(n)}] \to [J_{L(n)}, J_n]$ are both monotone. Hence J is attracted to a periodic point, a contradiction. □

Lemma 6.2. *Assume that the interval J_n has two predecessors $J_{L(n)}$, $J_{R(n)}$, and a successor $J_{s(n)}$. If this successor is to the right of J_n then the predecessors of $J_{s(n)}$ are J_n and $J_{R(n)}$ and if $J_{s(n)}$ has a successor then this successor must be again to the right of $J_{s(n)}$.*

Proof. The left predecessor of $s(n)$ (or rather of $J_{s(n)}$) is n (i.e., J_n) by definition of $s(n)$. The right predecessor of $s(n)$ is certainly defined because $J_{R(n)}$ and $J_{s(n)}$ have the same orientation and $R(n) < n < s(n)$. Let us show that $R(n)$ is this predecessor. If this was not the case then there exists $k < s(n)$ so that $J_k \subset (J_{s(n)}, J_{R(n)})$ and so that J_k and $J_{s(n)}$ have the same orientation. Because of the definition of $R(n)$ this implies certainly that $k > n$. Because $k < s(n)$ this implies that $0 \le k - n < s(n) - n = n - L(n)$ and $L(n) + k - n < n$. From the first of these two inequalities and the definition of $s(n)$ it follows that f^{k-n} is monotone and orientation preserving on $H = [J_{L(n)}, J_n]$. In particular $J_{L(n)+k-n}$ has the same orientation as J_n. From this, from the definition of $L(n)$ and $R(n)$ and from the second of these inequalities it follows $J_{L(n)+k-n}$ cannot be between $J_{L(n)}$ and $J_{R(n)}$. Hence,

$$f^{k-n}(H) \supset [J_{L(n)}, J_k] \supset [J_n, J_{s(n)}] = f^{n-L(n)}(H).$$

In particular, $f^{(n-L(n))-(k-n)}$ maps $f^{k-n}(H)$ monotonically into itself, and hence H and therefore J would be attracted to a periodic attractor, a contradiction.

Let us finally show that $s(n)$ cannot have a successor to its left. Indeed if it did, then by definition $f^{s(n)-R(n)}$ would map $[J_{s(n)}, J_{R(n)}]$ monotonically into $[J_n, J_{s(n)}]$. From the definition of $s(n)$, $f^{s(n)-n}$ maps $[J_n, J_{s(n)}]$ monotonically into $[J_{s(n)}, J_{R(n)}]$. Combining this would again give that J is attracted to a periodic attractor. With this contradiction the proof of this lemma is completed. □

Remark. The previous lemma implies that if J_n has a successor $J_{s(n)}$ and $J_{s(n)}$ also has a successor $J_{s(s(n))}$ then $s(n) - n = a = s(s(n)) - s(n)$ and $J_{s(n)}$ is between J_n and $J_{s(s(n))}$. Continuing this there exists a maximal integer k such that $J_{s^{i+1}(n)}$ is a successor of $J_{s^i(n)}$ for $i = 0, 1, \ldots, k-1$. In this case the intervals

$$J_{s(n)}, J_{s(s(n))}, \ldots, J_{s^k(n)}(n)$$

lie ordered and $f^a|[J_n, J_{s^k(n)}]$ is monotone. So f^a acts as a translation on these intervals.

Theorem 6.1. *Let $n \in \mathbb{N}$ and assume that J_n has two predecessors $J_{R(n)}$ and $J_{L(n)}$. Let $M_n \supset J_n$ be an open interval contained either in $[J_n, J_{R(n)}]$ or in $[J_{L(n)}, J_n]$. Assume that $\{M_{t_0}, M_{t_0+1}, \ldots, M_n\}$ is a monotone pullback of M_n. If the intersection multiplicity of this collection is at least $2p$ and $p \geq 2$ then there exists $t \in \{t_0, \ldots, n\}$ such that 1. $J_{s(t)}$, $J_{s^2(t)}, \ldots, J_{s^{2p-2}}(t)$ are defined; 2. $n = s^p(t)$ and $J_{s^j(t)}$ is contained in M_n for $j = p, \ldots, 2p-2$.*

Corollary 6.1. *If the pullback of an interval $T \supset J_n$, where T is contained in the natural neighbourhood T_n, is monotone then the intersection multiplicity of this pullback is at most* 11. *Similarly, if $T \supset J_n$ and $s^k(n)$ does not exist then the monotone pullback of T has at most intersection multiplicity* $2k+4$.

Proof of Corollary: Consider the pullback of $T \cap [J_n, J_{R(n)}]$ and $T \cap [J_{L(n)}, J_n]$ separately. If the intersection multiplicity of T is at least 12 then the pullback of either $T \cap [J_n, J_{R(n)}]$ or $T \cap [J_{L(n)}, J_n]$ has intersection multiplicity ≥ 6. So take $p = 3$ and the previous theorem implies that $p + 1 \leq 2p - 2$ and $J_{s(n)}$ is contained in $T \cap [J_n, J_{R(n)}]$, which is impossible since $T \subset T_n$. The second statement follows also immediately. (Note that $2k + 4 \geq 6$ for $k \geq 1$.) □

Proof of Theorem 6.1: In order to be definite assume that $M_n \subset [J_n, J_{R(n)}]$. By assumption there exists a point y which is contained in at least $2p$ of the intervals $M_{t_0}, \ldots, M_n$. Of course this implies that for at least p of these intervals M_i the corresponding intervals J_i lie on the same side of y. So let $N \geq p$ be the maximal number of distinct integers $t \leq i(1), \ldots, i(N) \leq n$ such that each of the intervals $M_{i(1)}, \ldots, M_{i(N)}$ contains this point y, all $J_{i(j)}$ lie on one side of y and are labeled such that

$$[J_{i(1)}, y] \supset [J_{i(2)}, y] \supset \cdots \supset [J_{i(N)}, y].$$

Since M_i has a common endpoint with J_i this implies that the intervals $J_{i(1)}, \ldots, J_{i(N)}$ all have the same orientation.

Claim 1: $i(1) < i(2) < \cdots < i(N)$ and we may assume that $i(N) = n$. Furthermore, $J_{i(1)}$ cannot be contained in $f^t(M_{i(1)})$ for $t = 1, \ldots, i(N) - i(1)$ if f^t is orientation preserving on $M_{i(1)}$. In particular if we take $a = i(N) - i(N-1)$

then f^a maps $[J_{i(1)}, J_{i(N-1)}]$ (which is contained in $M_{i(1)}$) monotonically into $(J_{i(1)}, J_{i(N)}]$.

Proof of Claim 1: For $j \in \{2, 3, \dots, N\}$ the interval $J_{i(j)}$ is contained in $[J_{i(j-1)}, y] \subset M_{i(j-1)}$ and $J_{i(j)}$ and $J_{i(j-1)}$ have the same orientation. Since $f^{n-i(j-1)}\colon M_{i(j-1)} \to M_n \subset [J_n, J_{R(n)}]$ is monotone this implies that the interval $J_{i(j)+n-i(j-1)}$ has the same orientation as J_n and is between J_n and $J_{R(n)}$. Since $J_{R(n)}$ is a predecessor of J_n this implies that $i(j) + n - i(j-1) > n$ and consequently $i(j) > i(j-1)$. Let us show that we may assume that $i(N) = n$. Indeed from what we have shown $f^{n-i(N)}$ is monotone on $M_{i(j)}$ for $j = 1, \dots, N$. In particular letting $i'(j) = i(j) + (n - i(N))$, and taking the images of these intervals under this map we get that $M_{i'(1)}, \dots, M_{i'(N)}$ all contain one point y', $J_{i'(j)}$ all lie on one side of y' and that these are labeled so that $[J_{i'(1)}, y] \supset [J_{i'(2)}, y] \supset \cdots \supset [J_{i'(N)}, y]$. Since $i'(N) = n$ we may as well assume that $i(N) = n$. The first statement of the claim follows and it follows that f^a is monotone and orientation preserving on $M_{i(1)} \supset [J_{i(1)}, J_{i(N-1)}]$. Now $J_{i(1)}$ cannot be contained in $f^t(M_{i(1)}) = M_{i(1)+t}$ for $t = 1, \dots, i(N) - i(1)$ because otherwise $J_{i(N)-t}$ would be contained in $M_{i(N)} = M_n$ and would have the same orientation as J_n. But this is impossible because $M_n \subset [J_n, J_{R(n)}]$ and $J_{R(n)}$ is a predecessor of J_n. □

Claim 2: If J_k and J_n have the same orientation where $k < n$ and if $J_k \subset [J_{i(1)}, J_n]$ then $k \in \{i(1), i(2), \dots, i(N)\}$. Furthermore, if we let $a = i(N) - i(N-1)$ then $i(j+1) - i(j) = a$ for $j = 1, 2, \dots, N-1$.

Proof. Let $j < n$ be maximal such that $J_k \subset [J_{i(j)}, y] \subset M_{i(j)}$. Then $f^{n-i(j)}$ maps J_k into M_n. Because J_k, $J_{i(j)}$ and J_n all have the same orientation, it follows that $J_{k+n-i(j)}$ also has the same orientation. This implies that $k \geq i(j)$. Suppose $k > i(j)$. If $M_k \not\supset J_n$ then $M_k \subset M_{i(j)}$ and $f^{k-i(j)}$ maps $M_{i(j)}$ monotonically into $M_k \subset M_{i(j)}$. This implies that J is attracted by a periodic orbit, a contradiction. Hence, $M_k \not\supset J_n$ and by the maximality of N, $k \in \{i(0), \dots, i(N)\}$. This proves the first statement of the claim. From Claim 1, f^a maps $[J_{i(1)}, J_{i(N-1)}]$ monotonically and orientation preservingly into $(J_{i(1)}, J_{i(N)}]$. It follows from this and the first part of this claim that $i(j) + a = i(j+1)$ for $j = 1, \dots, N-2$. Thus Claim 2 is proved.

Define $i(N+j) = n + j \cdot a$ for $j = 1, \dots, N$. As we have shown in the previous claim this formula holds for j negative. So we get

$$i(N+j) = n + j \cdot a \quad \text{for} \quad j = -N+1, -N+2, \dots, N-1, N.$$

Now consider the interval

$$H = [J_{i(1)}, J_{i(N)}].$$

The map $f^{n-i(1)} = f^{(N-1)a}$ maps $H \subset M_{i(1)}$ monotonically and orientation preservingly into $M_n \subset [J_n, J_{R(n)}]$. In particular f^a maps $[J_{i(1)}, J_{i(2N-2)}]$ monotonically and orientation preservingly onto $[J_{i(2)}, J_{i(2N-1)}]$. Therefore $J_{i(j)}$ lies between $J_{i(j-1)}$ and $J_{i(j+1)}$ for $|j| \leq N-2$. Furthermore, if $f^t(H) \cap [J_{i(1)}, J_{R(n)}] \neq$

$\emptyset$ for some $t = 1, \ldots, (N-1)a$ then $f^t|H$ is either orientation reversing or t is a multiple of a. Indeed, assume that $f^t|H$ is orientation preserving and that this intersection is non-empty. By Claim 1, $f^t(H) \subset f^t(M_{i(1)})$ does not contain $J_{i(1)}$. So $f^t(J_{i(1)})$ must be contained in $[J_{i(1)}, J_{R(n)}]$. But $f^t(J_{i(1)}) \subset f^t(H)$ cannot be contained in $[J_n, J_{R(n)}]$ since $J_{R(n)}$ is a predecessor of J_n. So $f^t(J_{i(1)}) \subset [J_{i(1)}, J_n]$ and it follows from the previous claim that t must be a multiple of a. Furthermore, consider J_k for $k < n$ and with the same orientation as J_n. This interval cannot be contained in $[J_{i(1)}, J_n]$ (see Claim 2) and neither in $[J_n, J_{R(n)}]$ from the definition of $J_{R(n)}$. Combining all this shows that the only intervals J_k inside $[J_{i(1)}, J_{R(n)}]$ for $k \leq i(2N-1)$ with the same orientation as J_n are intervals of the form $k = i(1) + j \cdot a$. It follows that $J_{i(j+1)}$ is a successor of $J_{i(j)}$ for $j = 1, \ldots, 2N-2$. □

7 Wandering Intervals Accumulate on Turning Points

In this section we are going to prove that the ω-limit set of a wandering interval contains at least one turning point and thus prove that the induction hypothesis Ind_0 holds. From this it follows in particular that maps without turning points, e.g. circle homeomorphisms in $\mathcal{N}F^{1+bv} \cup \mathcal{N}F^{1+Z}$, cannot have wandering intervals. So the proof in this section implies the classical result of Denjoy.

Suppose we have a map $f\colon N \to N$ with a wandering interval J whose iterates stay outside a neighbourhood of the turning points of f. Then we can modify the map f near these turning points without affecting the orbit of J. So change f such that each maximal interval on which f is monotone is mapped by f onto a component of N. Once we have done this we may assume that every pullback is monotone.

Proposition 7.1. *There exists $n_0 \in \mathbb{N}$ such that the following holds. If J_n where $n \geq n_0$ has two predecessors $J_{L(n)}$ and $J_{R(n)}$ and $|J_{R(n)}| > |J_n|$ and $|J_{L(n)}| > |J_n|$ then J_n has a successor and $|J_{s(n)}| < |J_n|$.*

Remark. Suppose for example that $J_{s(n)}$ is to the right of J_n. Then Lemma 6.2 implies that the interval $J_{s(n)}$ has two predecessors, namely J_n and $J_{R(n)}$ and that it has no left successor. From the proposition above it follows that the assumptions of this proposition are again satisfied for $J_{n'}$ where $n' = s(n)$. So one can apply the proposition infinitely often!

Proof. Let $n_0 = N_0(1; 11)$ be as in the Koebe/Contraction Principle and let $n \geq n_0$. Suppose $s(n)$ is not defined and let $T_n = [J_{L(n)}, J_{R(n)}]$ be the natural neighbourhood of J_n. By the Corollary of Theorem 6.1 the monotone pullback of T_n has intersection multiplicity bounded by 11. Because $|J_{L(n)}|, |J_{R(n)}| \geq$

$|J_n|$ this contradicts the Koebe/Contraction Principle. Hence $s(n)$ is defined. From Lemma 6.2 $L(s(n)) = n$ and $R(s(n)) = R(n)$. Furthermore, $s(n)$ has no left successor. Now the natural neighbourhood of J_n is $T_n = [J_{L(n)}, J_{s(n)}]$ and again by the Corollary to Theorem 6.1 the monotone pullback of T_n has intersection multiplicity bounded by 11. Since $|J_{L(n)}| > |J_n|$ it follows from the Koebe/Contraction Principle that $|J_{s(n)}| < |(J_n, J_{s(n)}]| < |J_n|$. □

Theorem 7.1. *If $f \in \mathcal{N}F^{1+Z} \cup \mathcal{N}F^{1+bv}$ and f has a wandering interval J, then the ω-limit set of J contains a turning point.*

Proof. Let us first show that $\omega(J)$ cannot have finite cardinality. Indeed otherwise $\omega(J)$ would contain a periodic point p of, say, period k and there would exist a neighbourhood U of p such that $U \cap \omega(J) = \{p\}$. Furthermore, there exist a neighbourhood $V \subset U$ of p such that $f^k(V) \subset U$ and an integer n' for which $J_{n'} \subset V$ and such that $J_n \subset V$ whenever $n \geq n'$ and $J_n \cap U \neq \emptyset$. Since $f^k(V) \subset U$ this implies by induction that $J_n \subset V \cup f(V) \cup \cdots \cup f^{k-1}(V)$ for all $n \geq n'$ and therefore that $\omega(J) = O(p)$. Hence $\omega(J)$ would be attracted to a periodic orbit, a contradiction.

Therefore, and since all intervals J_i are disjoint, there exist arbitrarily large integers $l, r < n$ such that J_l, J_r, J_n are in the same component of N, have the same orientation, $J_n \subset (J_l, J_r)$, such that

$$|J_n| \leq \min(|J_l|, |J_r|)$$

and, for $i = 0, 1, \ldots, n-1$,

$$J_i \subset (J_l, J_r) \text{ implies that } J_i \text{ has a different orientation}$$

(from the intervals J_l, J_r, J_n). Assume that l, r, n are bigger than the number n_0 from above. It follows that J_l and J_r are the predecessors of J_n. So we can apply Proposition 7.1 and hence J_n has infinitely many successors $J_{s^k(n)}$, $k = 1, 2, \ldots$. From the description in Lemma 6.2, all these successors are contained in $[J_l, J_r]$. They either all lie to the right of the previous one or all to the left. Moreover $s^k(n) - s^{k-1}(n)$ is independent of k. It follows that as k tends to infinity these intervals $J_{s^k(n)}$ converge to a fixed point of f^a where $a = s(n) - n$. Hence this fixed point is an attracting fixed point with J in its basin. But this shows that J was not a wandering interval after all, a contradiction. □

Now we know that iterates of wandering intervals cannot stay away from turning points. In order to analyze the metric properties of iterates of a wandering interval we will pay special attention to the moments where the iterates get closest to some turning point. This is formalized in the following definition. Let, as in Section 4, S_c be a neighbourhood of a turning point c of f on which the involution τ is defined.

Definition. Let c be a turning point in $\omega(J)$, let $m \geq n$ and assume that $J_n \subset S_c$. Then we say that J_n is the *m-closest approach* to c if $(J_n, \tau(J_n)) \cap (\cup_{i \leq m} J_i) = \emptyset$. Similarly, J_n is the *closest approach* to c if it is the n-closest approach to c. Now fix a turning point c in $\omega(J)$ and let $N(c)$ be the collection of integers $i \in \mathbb{N}$ with

$$J_i \subset S_c \quad \text{and} \quad [J_i, \tau(J_i)] \bigcap \left(\bigcup_{0 \leq j < i} J_j \right) = \emptyset.$$

From now on let

$$N(c) = \{n(1), n(2), \dots\}$$

where $n(1) < n(2) < \dots$. We call $J_{n(1)}, J_{n(2)}, \dots$ the *sequence of closest approach to c.*

Lemma 7.1. *If $J_{n(k)}$ has a successor $J_{s(n(k))}$ then $s(n(k)) = n(k+1)$ and $J_{n(k+1)}$ is between $J_{n(k)}$ and c. Furthermore, if there exists an integer j such that $s(j)$ and $s^2(j) = s(s(j))$ are both defined and such that $j < n(k) < s(j)$ then*

$$n(k+1) = s(n(k)).$$

Proof. In order to be definite assume that $J_{n(k)}$ is to left of c. Since $J_{n(k)}$ is a closest interval to c, any predecessor of $J_{n(k)}$ to its right must also be to the right of c. Therefore $c \in [J_{n(k)}, J_{R(n(k))}]$ and there can be no successor of $J_{n(k)}$ to its left. Hence if $J_{n(k)}$ has a successor then it must be to its right and because $f^{s(n(k))-n(k)}|[J_{L(n)}, J_{s(n)}]$ is monotone it even must be between $J_{n(k)}$ and c. So if $s(n(k)) \neq n(k+1)$ then $n(k+1) < s(n(k))$. Consider $H = [J_{L(n(k))}, J_{n(k)}]$ and let $a = n(k) - L(n(k)) = s(n(k)) - n(k)$. Then f^{2a} is monotone on H and $a > n(k+1) - n(k)$. Since $L(n(k)) + n(k+1) - n(k) < n(k)$ the interval $J_{L(n(k))+n(k+1)-n(k)} \subset f^{n(k+1)-n(k)}(H)$ is not contained in $[J_{n(k)}, \tau(J_{n(k)})]$ whereas by assumption $J_{n(k+1)} \subset f^{n(k+1)-n(k)}(H)$ is contained in this interval $[J_{n(k)}, \tau(J_{n(k)})]$. It follows that either $J_{n(k)}$ or $\tau(J_{n(k)})$ is contained in $f^{n(k+1)-n(k)}(H)$. Hence for $t = a - (n(k+1) - n(k))$ one has $0 < t < a$ and

$$(*) \qquad f^t(J_{n(k)}) \subset f^a(H) = [J_{n(k)}, J_{s(n(k))}] \subset [J_{n(k)}, c].$$

Since $J_{s(n(k))}$ is the successor of $J_{n(k)}$ this implies that f^t is orientation reversing on H. So $J_{L(n(k))+t}$ is to the right of $J_{n(k)}$. But since $L(n(k)) + t < n(k)$ this interval cannot be contained in $[J_{n(k)}, \tau(J_{n(k)})]$. Therefore, and because of $(*)$, $f^t(H)$ contains c; this contradicts the monotonicity of $f^a|H$ and proves the first statement of the lemma.

Let us now prove the second statement. According to the first part it is enough to show that $s(n(k))$ is defined. Now let $a = s(j) - j$ then $L(j) = j - a$, $s(j) = j + a$ and $s(s(j)) = j + 2a$. Since f^a is monotone on $[L_{j-a}, L_{j+2a}]$ and $n(k) - j < s(j) - j = a$ it follows that $J_{n(k)}$ is contained in $T = [J_{n(k)-a}, J_{n(k)+a}]$ and that f^a is monotone on T. Furthermore, there is no predecessor of $J_{n(k)}$ in

$f^a(T)$ because otherwise there would be a predecessor of J_{j+a} in $f^a([L_j, L_{j+2a}])$, contradicting that $s^2(j)$ exists. So Properties 1 and 2 of the definition of the successor of $J_{n(k)}$ hold. Finally there is also no interval J_t in $[J_{n(k)}, J_{n(k)+a}]$ with $t < n(k) + a$ and with the same orientation as $J_{n(k)}$ because $J_{t+(s(j)-n(k))}$ would have the same orientation as $J_{s(j)}$ and be contained in $[J_{s(j)}, J_{s(j)+a}]$, contradicting the definition of $s^2(j)$. □

8 Topological Properties of a Unimodal Pullback

As before let $\mathcal{A}^d$ be the collection of all endomorphisms of N in $\mathcal{A}$ with $f(\partial N) \subset \partial N$ and with precisely d turning points. From now on we will assume

$$(Ind_{d-1}) \qquad \text{maps in } \bigcup_{i=0}^{d} \left(\mathcal{A}^i \cap (\mathcal{N}F^{1+bv} \cup \mathcal{N}F^{1+Z})\right)$$

have no wandering intervals.

and try to show that this implies Ind_d. So assume that f contains a wandering interval J. As we have shown in the previous section we may assume that J accumulates onto some turning point c. Throughout this section we will consider properties of the unimodal pullback of two intervals. The first of these is the interval $Q_{n(k)} \supset J_{n(k)}$ not containing a turning point of f such that

$$f(Q_{n(k)}) = [f(J_{n(k-1)}), f(J_{n(k+1)})].$$

The second, and larger one contains c and is defined by

$$\hat{Q}_{n(k)} = Q_{n(k)} \cup [J_{n(k+1)}, \tau(J_{n(k+1)})].$$

Let us give an outline of the remainder of the proof of Theorem A first. The reader might at this point also find it helpful to read Section II.6 where a proof of a unimodal version of Theorem A is given. As in the proof of Theorem II.6.2, it will first be shown that $k \mapsto |J_{n(k)}|$ is monotone and decreasing. The main tools for this are the Koebe/Contraction Principle and Theorem 8.3 below which shows that the intersection multiplicity of the unimodal pullback of $Q_{n(k)}$ is universally bounded. Once we know this, we get that $\hat{Q}_{n(k)}$ contains a universally scaled neighbourhood of $J_{n(k)}$ for each k sufficiently large. If the pullback of the interval $\hat{Q}_{n(k)}$ only meets the turning points finitely often then it will follow from Theorem 8.4 below that one has enough disjointness. Therefore the Koebe/Contraction Principle will again give a contradiction in this case. So the difficult case is when the number of visits of the pullback to turning points is unbounded as k goes to infinity. Fortunately, in this case we will have

1. the turning points are visited in a periodic way, see Theorem 8.1 below (this follows from the induction hypothesis);

2. there exists a monotone pullback on a rather big interval around $J_{n(k)}$, see Theorem 8.2.

From this Theorem A will follow. The results in this section are essentially due to Lyubich (1989) and Blokh and Lyubich (1989d).

So let us first show that the intervals from the pullback of $\hat{Q}_{n(k)}$ meet the turning points in a periodic way. This fact is based on the induction assumption that maps with less than turning points have no wandering intervals.

Structure Theorem 8.1. Let f be a map as above with d turning points and with a wandering interval J. Let $n(k)$ be as above and let $P_{n(k)} \supset J_{n(k)}$ be an interval which is contained in $\hat{Q}_{n(k)}$. Let $m(0) < m(1) \cdots < m(\ell) = n(k)$ be the cutting times of its unimodal pullback $P_0, \dots, P_{n(k)}$ and let c_j denote the turning point in $P_{m(j)}$. Then we have the following properties.

1. If $i \in \{0, \dots, \ell - d\}$ then $J_{m(i)}$ is a $m(i+d) - 1$ closest approach to c_i;
2. If $i \in \{\ell - d + 1, \dots, \ell\}$ then $J_{m(i)}$ is a $n(k)$ closest approach to c_i;
3. $c_i = c_{i+d}$ for $i = 0, \dots, \ell - d$ and $\{c_{\ell-d+1}, \dots, c_\ell\}$ are distinct;
4. If $J_{m(i)+j} \subset (J_{m(i)}, \tau(J_{m(i)}))$, $i \in \{0, \dots, \ell-1\}$ and $m(i) < m(i)+j \le n(k)$ then $P_{m(i)+j} \subset [J_{m(i)}, \tau(J_{m(i)})]$.
5. $P_{m(i+d)} \subset [J_{m(i)}, \tau(J_{m(i)})] \subset P_{m(i)}$ for $i = 0, 1, \dots, \ell - d$ and therefore $f^{m(i+d)-m(i)}$ maps $P_{m(i)}$ into itself;

Proof of 4: If Statement 4 does not hold then the closure of $J_{m(i)}$ is contained in the interior of $P_{m(i)+j}$. So the closure of $J_{m(i)+n(k)-(m(i)+j)}$ is contained in the interior of $P_{n(k)}$. From the definition of $P_{n(k)}$ this implies $m(i)+n(k)-m(i)-j \ge n(k)$ and therefore $j \le 0$, a contradiction.

Proof of 1 and 2: Let us just prove Statement 1. Statement 2 is proved in exactly the same way. Suppose by contradiction that there exists $l \in \{0, \dots, m(i+d) - 1\}$ such that $J_l \subset (J_{m(i)}, \tau(J_{m(i)}))$. Then $J_{l+n(k)-m(i)} = f^{n(k)-m(i)}(J_l) \subset f^{n(k)-m(i)}(P_{m(i)}) \subset P_{n(k)}$ and $l \neq m(i)$. Hence $l > m(i)$. From Statement 4 we know that $P_l \subset P_{m(i)}$. So $f^{l-m(i)}$ maps $P_{m(i)}$ into $P_l \subset P_{m(i)}$. Because $l < m(i+d)$ the map $f\colon \cup_{t=0}^{l-m(i)-1} f^t(P_{m(i)}) \to \cup_{t=0}^{l-m(i)-1} f^t(P_{m(i)})$ has at most $d-1$ turning points. Since $J_{m(i)}$ is a wandering interval of this map, we get a contradiction with the induction hypothesis.

Proof of 3: Suppose there are $i - d < j \le i \le s$ with $c_i = c_j$. From Statement 1 we get that $J_{m(i)}$ and $J_{m(j)}$ are both $m(j+d)$ closest approaches to $c_i = c_j$. Hence because $m(i), m(j) < m(j+d)$ this implies $i = j$. Since f has precisely d turning points one gets that $c_{\ell-d+1}, \dots, c_\ell$ are distinct and that $c_i = c_{i+d}$ for $i \in \{0, \dots, \ell - d\}$.

Proof of 5: The proof of Statement 5 follows immediately from the other statements. □

The following theorem shows that we can even take monotone pullbacks of intervals which contain topologically rather large sets. A unimodal version of this theorem was already used in Guckenheimer (1979).

Monotone Extension Theorem 8.2. Let $P_{n(k)} = \hat{Q}_{n(k)}$ and consider the unimodal pullback $\{P_0, P_1, \dots, P_{n(k)}\}$ of $P_{n(k)} \supset J_{n(k)}$. Let $m(0) < m(1) < \dots < m(\ell) = n(k)$ be the cutting times of this pullback (i.e., the integers when the pullbacks meet a turning point) and let c_i be the turning point in $P_{m(i)}$. Let $H_{m(i)} \supset J$ be the maximal interval such that $f^{m(i)}$ is monotone on $H_{m(i)}$. Let $R_{m(i)}$ be the component of $P_{m(i)} \setminus J_{m(i)}$ which contains c_i and let $L_{m(i)}$ be the other component. If the number ℓ of cutting times of the pullback is at least $d+1$ then

$$f^{m(i)}(H_{m(i)}) \supset [L_{m(i)}, c_i],$$

$$f^{m(i-1)}(H_{m(i)}) \supset [L_{m(i-1)}, c_{i-1})$$

and

$$f^{m(i)-m(i-1)}(L_{m(i-1)}) \subset L_{m(i)}$$

for $i = 0, 1, \dots, \ell - d$.

Proof. Let us first show that

$$(*) \qquad f^{m(i+1)-m(i)}(P_{m(i)}) \supset [J_{m(i+1)}, c_{i+1}]$$

for $i = 0, 1, 2, \dots, \ell - d$. Suppose by contradiction that there exists $i \in \{0, 1, \dots, \ell - d\}$ with $P_{m(i+1)} \supset f^{m(i+1)-m(i)}(P_{m(i)}) \not\ni c_{i+1}$. By Statement 5 of the previous theorem $f^{m(i+d)-m(i)}$ maps $P_{m(i)}$ into itself. Now $f^{m(i+1)-m(i)}(P_{m(i)}) \not\ni c_{i+1}$ implies that

$$T = \bigcup_{t=0}^{m(i+d)-m(i)-1} f^t(P_{m(i)})$$

does not contain c_{i+1}. Hence f maps T into itself and has at most $d-1$ turning points. Since $J \subset P_{m(i)}$ it follows from the induction hypothesis that J is not a wandering interval, a contradiction. This proves $(*)$. Furthermore

$$(**) \qquad f^{m(i+1)-m(i)} \colon L_{m(i)} \to L_{m(i+1)} \text{ is monotone and onto}$$

for $i = 0, 1, 2, \dots, \ell - d$ because otherwise there exists such an integer i with

$$f^{m(i+1)-m(i)}(R_{m(i)}) = L_{m(i+1)}$$

and then as before

$$T = \bigcup_{t=0}^{m(i+d)-m(i)-1} f^t(R_{m(i)} \cup J_{m(i)})$$

contains as most $d-1$ turning points and f maps this interval into itself. Since J is contained in T this contradicts the induction hypothesis. It follows from $(*)$

and (∗∗) that $f^{m(i+1)-m(i)}$ maps $[L_{m(i)}, c_i]$ monotonically over $[L_{m(i+1)}, c_{i+1}]$. The theorem clearly follows. □

Next we give two results about the disjointness of unimodal pullbacks of intervals in $Q_{n(k)}$ and in $\hat{Q}_{n(k)}$. The first result deals with the unimodal pullback of $Q_{n(k)}$.

Theorem 8.1. *Let $m(0) < m(1) < \cdots < m(\ell) = n(k)$ be the cutting times of the unimodal pullback of $Q_{n(k)}$. Then*

1. *$\ell \leq d - 1$;*
2. *for every $0 \leq j \leq \ell$ either $s(m(j))$ or $s^2(m(j))$ is not defined;*
3. *the intersection multiplicity of the unimodal pullback of $Q_{n(k)}$ is universally bounded (in fact by $12d$).*

Proof. Suppose by contradiction that $\ell \geq d$. By Statement 5 of Theorem 8.1 it follows that $Q_{m(\ell)} = Q_{n(k)}$ is contained in $Q_{m(\ell-d)}$. Hence f maps

$$\cup_{t=0}^{m(\ell)-m(\ell-d)-1} f^t(Q_{m(\ell-d)})$$

into itself and since $Q_{n(k)}$ contains no turning point this map has at most $d-1$ turning points. This contradicts Ind_{d-1}. So let us prove Statement 2 by assuming by contradiction that there exists $j \in \{0, 1, \ldots, \ell\}$ for which $s(m(j))$ and $s(s(m(j))$ are defined. Because $m(j)$ is $n(k)$-closest we get $s(m(j)) > n(k)$. Hence from Lemma 7.1 we get that $n(k+1) = s(n(k))$ and $s(n(k)) - n(k) = s(j) - j$. Because the closure of $J_{s(m(j))}$ is contained in $Q_{m(j)}$ we get that the closure of $J_{n(k+1)} = f^{n(k)-m(j)}(J_{s(m(j))})$ is contained in $f^{n(k)-m(j)}(Q_{m(j)}) \subset Q_{n(k)}$ which contradicts the definition of $Q_{n(k)}$. So let us prove Statement 3. From Statement 2 and the Corollary of Theorem 6.1 it follows that the intersection multiplicity of $\{P_{m(j)}, \ldots, P_{m(j+1)}\}$ for $j = -1, 0, 1, \ldots, \ell-1$ (where we let $m(-1) = 0$) is bounded by 11. Since $\ell \leq d-1$ the theorem follows. □

Theorem 8.2. *Assume $n(k) > n(k-1) + (n(k-1) - n(k-2))$. Let $m(0) < m(1) < \cdots < m(\ell) = n(k)$ be the cutting times of the unimodal pullback of $\hat{Q}_{n(k)}$ and let $\ell \geq 2d$. Then $s(n(k-1)) = s(m(\ell-d))$ and $s^l(m(j))$ are both not defined if $j \in \{\ell - ld, \ldots, \ell - ld + d - 1\}$ and $l \geq 2$. Furthermore, if $j \in \{\ell - ld, \ldots, \ell - ld + d - 1\}$, the intersection multiplicity of*

$$\{\hat{Q}_{m(j)}, \ldots, \hat{Q}_{m(j+1)}\}$$

is bounded by $4 + 2l$. Similarly the intersection multiplicity of

$$\{\hat{Q}_0, \hat{Q}_1, \ldots, \hat{Q}_{m(0)}\}$$

is bounded by $4 + 2[\ell/d]$.

Proof. Let us show that $s(n(k-1))$ is not defined. Indeed, otherwise we would have $J_{n(k-2)} \subset [J_{L(n(k-1))}, J_{n(k-1)})$ and $n(k) = s(n(k-1))$. Therefore, for $a = n(k) - n(k-1) = s(n(k-1)) - n(k-1)$,

$$J_{n(k-2)+a} \subset [J_{n(k-1)}, J_{n(k)}).$$

But since $n(k-2) + a < n(k-1) + a = n(k)$ this implies that $n(k-2) + (n(k) - n(k-1)) = n(k-2) + a$ must be equal to $n(k-1)$ and this contradicts the assumption of the theorem.

Let us show that if $l \geq 2$ and $j \in \{\ell - ld, \ldots, \ell - ld + d - 1\}$ then $s^l(m(j))$ is not defined. So suppose by contradiction $s^l(m(j))$ exists. We claim that then $s^{i+1}(m(j)) > n(k-1) > s^i(m(j))$ for some $i = 0, 1, \ldots, l-1$. Indeed, since $m(0) < n(k-1)$ we may otherwise assume that $s^{l-1}(m(j)) < n(k-1)$. But then $J_{m(j+(l-1)d)}$ is the $(l-1)$-th successor of $J_{m(j)}$ and, again by Lemma 7.1, the successor of $J_{m(j+(l-1)d)}$ must be between $J_{m(j+(l-1)d)}$ and c_j. By Statement 2 of Theorem 8.1 this implies that $s^l(m(j)) > n(k-1)$. This proves the claim. Hence Lemma 7.1 and $s^{i+1}(m(j)) > n(k-1) > s^i(m(j))$ imply that $s(n(k-1))$ exists and so we get a contradiction. The disjointness statements immediately follow from the Corollary of Theorem 6.1. □

9 The Non-Existence of Wandering Intervals

In Section 7 we have proved Ind_0 and so Theorem A follows from

$$Ind_{d-1} \Rightarrow Ind_d.$$

So let us assume that Ind_{d-1} holds and by contradiction assume that there exists a map $f \in \left(\mathcal{A}^d \cap (\mathcal{N}F^{1+bv} \cup \mathcal{N}F^{1+Z})\right)$ which has a wandering interval J. Moreover, assume that J is maximal in the sense that J is not contained in any strictly larger wandering interval. From the Contraction Principle this implies that

$$\frac{|H_n|}{|J|} \to 1 \quad \text{if } n \to \infty$$

where H_n is the maximal interval containing J on which f^n is monotone.

From Section 7 we know that J accumulates at a turning point, say c. Consider the sequence of closest approach to c, $\{J_{n(k)}\}_{k \geq 0}$.

Theorem 9.1. *There exists k_0 such that for all $k \geq k_0$*

$$n(k) - n(k-1) \leq n(k-1) - n(k-2).$$

Corollary 9.1. *J is not a wandering interval.*

Proof of Corollary: *Since $n(k) - n(k-1) \leq n(k-1) - n(k-2)$ it follows that $n(k) - n(k-1)$ is eventually equal to some integer a for all k sufficiently large. In particular, since $J_{n(k-1)}$ and $J_{n(k)}$ tends to c it follows that c is an attracting fixed point of f^a and J is in the basin of this fixed point. Hence J is not a wandering interval.* □

Proof of Theorem 9.1: Since the intervals $J_{n(k)}$ are disjoint there exist arbitrarily large k such that $|J_{n(k-1)}| > |J_{n(k)}|$. Let $Q_{n(k)}$ be as before. If $|J_{n(k)}| < |J_{n(k+1)}|$ then $|J_{n(k\pm 1)}| > |J_{n(k)}|$. Because f is non-flat at the critical point c this implies that $Q_{n(k)}$ is a (1/2)-scaled neighbourhood of $J_{n(k)}$ (if k is large). By Theorem 8.3 the intersection multiplicity of the unimodal pullback of $Q_{n(k)}$ is at most $12d$ and therefore the Koebe/Contraction Principle shows that $n(k)$ is bounded. This gives a contradiction. So we have shown that $k \mapsto |J_{n(k)}|$ is monotone decreasing for k large.

Let us now show that $n(k) \leq n(k-1) + (n(k-1) - n(k-2))$ for k large. So assume by contradiction $n(k) > n(k-1) + (n(k-1) - n(k-2))$. Consider the unimodal pullback of $\hat{Q}_{n(k)}$. If $\ell \leq 2d$ then from Theorem 8.4 the intersection multiplicity of $\{\hat{Q}_0, \ldots, \hat{Q}_{n(k)}\}$ is uniformly bounded. Because $\hat{Q}_{n(k)}$ contains a 1/2-scaled neighbourhood of $J_{n(k)}$ this gives a contradiction with the Koebe/Contraction Principle.

This implies that for k large the number of cutting times ℓ is at least $2d+1$. Now let L_i and R_i be as in Theorem 8.2. From Theorem 8.2 and for $j = 1, 2, \ldots, \ell - d$, there exists an interval $H \supset J$ which is mapped by $f^{m(j)}$ monotonically onto $[L_{m(j)}, c_{j-d}]$. We claim that $f^{m(j)}(H)$ does not contain $\tau(J_{m(j)})$. Indeed, since Theorem 8.2 gives

$$f^{m(j)}(H) \supset [L_{m(j)}, c_{m(j)}]$$

one would otherwise have

$$f^{m(j)}(H) \supset [L_{m(j)}, \tau(J_{m(j)})].$$

If we take $j \in \{\ell - 2d, \ldots, \ell - d\}$ then $s^2(m(j))$ does not exist by Theorem 8.4, and since $f^{m(j)}$ is monotone on H we can apply the Corollary of Theorem 6.1 and the intersection multiplicity of $H, f(H), \ldots, f^{m(j)}(H)$ is at most 8. Furthermore, from the definition $\hat{Q}_{n(k)}$, $[L_{m(j)}, c_{j-d}]$ contains $[J_{m(j-d)}, c_j]$ or it contains $[\tau(J_{m(j-d)}), c_j]$. Since the length of the closest approach intervals decreases one has $|J_{m(j-d)}| > |J_{m(j)}|$ and therefore $f^{m(j)}(H)$ contains a 1/2-scaled neighbourhood of $f^{m(j)}(J)$. Therefore we get a contradiction with the Koebe/Contraction Principle and this proves the claim.

Now let $F_{m(j)} = [J_{m(j)}, \tau(J_{m(j)}]$. By definition $f^{m(j)}|H$ is monotone and by Theorem 8.2, $f^{m(j-1)}(H) \supset [J_{m(j-1)}, c_{j-1})$ and $f^{m(j)-m(j-1)}$ maps $[J_{m(j-1)}, c_{j-1})$ monotonically over $[J_{m(j)}, c_j]$. By the previous claim, the image of $[J_{m(j-1)}, c_{j-1})$ under this map is contained in $F_{m(j)}$ and therefore we get

$$f^{m(j)-m(j-1)}(F_{m(j-1)}) \subset F_{m(j)}$$

for all $j \in \{\ell - 2d, \ldots, \ell - d\}$. In particular,

$$f^{n(k-1)-n(k-2)}(F_{n(k-2)}) \subset F_{n(k-1)}.$$

Hence

$$J_{n(k-1)+(n(k-1)-n(k-2))} \subset [J_{n(k-1)}, \tau(J_{n(k-1)})].$$

Since $n(k-1) + (n(k-1) - n(k-2)) > n(k-1)$ and $J_{n(k)}$ is a closest approach this gives $n(k-1) + (n(k-1) - n(k-2)) \geq n(k)$, a contradiction. □

10 Finiteness of Attractors

In this section we will prove Theorem B. If $f : N \to N$ is a diffeomorphism then the period of periodic orbits of f is bounded. So Theorem B holds trivially in this case. So from now on assume that f is not a diffeomorphism and that $f \in \mathcal{N}F^{1+z}$. Of course some points in $C(f)$ may be attracted by periodic orbits so let $\hat{n}$ be an upper bound for the periods of these attracting orbits.

In this section we have to show that one has some expansion near periodic orbits. For this it will be convenient to consider the orientation preserving period of a periodic orbit. More precisely let $\mathcal{O}$ be a periodic orbit of period $k > \hat{n}$ and larger than, say, 300. Then $p \in \mathcal{O}$ implies $Df^k(p) \neq 0$. Let $n = 2k$ if $Df^k(p) < 0$ and $n = k$ otherwise.

The main idea of the proof of Theorem B is to choose $p \in \mathcal{O}$ and get points θ^1 and θ^2 on both sides of p very close to p, with $Df^n(\theta^i) \geq 1 + 2\rho$. Using the Minimum Principle we will get $Df^n(p) \geq 1 + \rho$ for large n and we are done.

For $p \in \mathcal{O}$ let us define T_p to be the maximal open interval such that both components of $T_p \setminus \{p\}$ contain at most one point of $\mathcal{O}$. (So the closure of T_p contains at most 5 points of $\mathcal{O}$.) The interval T_q, $q \in \mathcal{O}$, is a *direct neighbour* of T_p if $T_q \cap T_p = \emptyset$ and $\mathrm{cl}\,(T_p)$ and $\mathrm{cl}\,(T_q)$ have one point in common).

Lemma 10.1. *There exists a number $\tau > 0$ such that for each periodic orbit $\mathcal{O}$ of period $\geq \max(\hat{n}, 300)$ there exists $p \in \mathcal{O}$ such that: 1) T_p has direct neighbours on both sides; 2) $|T_{q_1}| \geq 2\tau|T_p|$ and $|T_{q_2}| \geq 2\tau|T_p|$ where T_{q_1} and T_{q_2} are the two direct neighbours of T_p.*

Proof. Let $s \in \mathcal{O}$ be such that i) $\#(\mathrm{cl}\,(T_s) \cap \mathcal{O}) = 5$ and ii) $|T_s| \leq |T_q|$ for all q with $\#(\mathrm{cl}\,(T_q) \cap \mathcal{O}) = 5$. If T_s has neighbours on both sides then we take $p = s$ and we are done. Otherwise $N = [-1, 1]$, and then let T^l and T^r be the smallest open intervals containing points of respectively $\{-1\}$ and $\{1\}$ such that $\#(\mathrm{cl}\,(T^l) \cap \mathcal{O}) = \#(\mathrm{cl}\,(T^r) \cap \mathcal{O}) = 5$. Because T_s has no two neighbours we have either $T_s \subset T^l$ or $T_s \subset T^r$. Since the interval $\mathrm{cl}\,(T_s)$ contains five points of $\mathcal{O}$, and n is at most twice the period of p, there are at most $5 \times 2 \times (5 + 5) = 100$ integers t, $0 \leq t \leq n$ such that f^t maps a point in $\mathrm{cl}\,(T_s) \cap \mathcal{O}$ into a point of $\mathrm{cl}\,(T^r) \cup \mathrm{cl}\,(T^l) \cap \mathcal{O}$. So there exists a $0 \leq t \leq 101$ such that $f^t(\mathrm{cl}\,(T_s) \cap \mathcal{O})$ is between T^l and T^r. Let p be the middle point of

$f^t(\mathrm{cl}\,(T_s) \cap \mathcal{O})$. Let $S = \max\{1, \sup_{x \in N} |Df(x)|\}$. Since $T_p \subset f^t(T_s)$, we get

$$|T_p| \leq |f^t(T_s)| \leq \left(\sup_{x \in N} |Df(x)|\right)^t |T_s| \leq S^t |T_s|.$$

As $t \leq 101$ and $|T_s| \leq |T_{q_i}|, i = 1, 2$ the lemma follows. □

Let $p \in \mathcal{O}$ be the point from Lemma 10.1 and as before let n be the period or twice the period of p. Let U_n be the 2τ-scaled neighbourhood around T_p. So $U_n \subset T_{q_1} \cup T_p \cup T_{q_2}$. Let J be the maximal interval around p for which f^n maps J orientation preservingly into U_n. Let J^l and J^r be the components of $J \setminus \{p\}$ and let U_n^i be the component of $U_n \setminus \{p\}$ which contains J^i for $i \in \{l, r\}$. Let $\hat{n}$ be as in the beginning of this section.

Lemma 10.2. *If $n > \hat{n}$ then $J \subset T_p$ and $f^n(J^i) \supset J^i$ for $i = l, r$.*

Proof. We claim that $J \cap P$ consists of at most 2 points. Indeed, take I to be the maximal interval containing p such that $\partial I \subset P$ and such that $f^n|I$ is a diffeomorphism. Since f is not a diffeomorphism, I is an interval (i.e. not equal to S^1). Then $f^k|I$ is diffeomorphism for all $k \geq 0$ and since I is maximal, $f^i(I) \cap I \neq \emptyset$ implies that $f^i(I) = I$. In particular the boundary points of I (which are in P) cannot be mapped into $\mathrm{int}\,(I)$ and so I contains at most two points of P. (This argument also shows that if I contains two points of P then f^n interchanges these two points.) This proves the first inclusion. If the second inclusion does not hold then $f^n(J^i) \subset J^i$ and a critical point of f is attracted by a periodic orbit of period n. This implies $n \leq \hat{n}$. □

The main step in the proof of Theorem B is the following

Proposition 10.1. *There exist a number $\rho > 0$ and an integer n_0 such that if the number n corresponding to p is greater than n_0 then there exist $\theta^i \in J^i, i = l, r$, such that*

$$Df^n(\theta^i) \geq 1 + 2\rho.$$

Remark. Let τ be the number from Lemma 10.1. If there exists $\theta^i \in J^i$ with $Df^n(\theta^i) \geq 1 + \tau$ then we are done with J^i. So from now on we may assume that

$$(*) \qquad 0 < Df^n(x) < 1 + \tau, \ \forall x \in J^i.$$

By the previous lemma, for $n > \hat{n}$ and $i = l, r$ we know that $J^i \subset f^n(J^i) \subset U^i$. Then $f^n(J^i) = U^i$ is impossible because otherwise

$$\frac{|f^n(J^i)|}{|J^i|} \geq \frac{2\tau |T_p| + |J^i|}{|J^i|} \geq (1 + 2\tau),$$

a contradiction with $(*)$. Hence in this case

$$f^n(J^i) \underset{\neq}{\to} \subset U^i.$$

Let $\{U_0, \ldots, U_n\}$ be the diffeomorphic pullback of $U_n \supset f^n(J)$: U_i is the maximal interval containing $f^i(J)$ which is mapped by f diffeomorphically into U_{i+1}. Since $f^n|J$ is a diffeomorphism this is well defined and $U_k \supset f^k(J)$. Furthermore from the maximality of J one has $U_0 = J$.

Lemma 10.3. 1. *There is a universal upper bound for the intersection multiplicity of the diffeomorphic pullback $\{U_0, U_1, \ldots, U_n\}$ of $U_n \supset f^n(J)$ (in fact it is at most 74);*

2. *For every $\epsilon > 0$ there exists $n_0 \in \mathbb{N}$ such that if $n > n_0$ then $|U_k| \le \epsilon$ for all $k = 0, 1, \ldots, n$.*

Proof. Let $U_{k(1)} \cap \cdots \cap U_{k(r)} \ni x$ with $k(1) < \cdots < k(r) \le n$. Because $f^{n-k}(U_k) \subset \operatorname{int}[T_{q_1}, T_{q_2}]$ and $f(U_k) \subset U_{k+1}$ we get that

$$f^{n-k(r)}(x) \in U_{k(1)+n-k(r)} \cap \cdots \cap U_n.$$

Therefore $\#\{k\,;\ U_k \cap \operatorname{int}[T_{q_1}, T_{q_2}] \ne \emptyset\} \ge r$. Hence statement 1) follows from

$$\#\{\,k\,;\ U_k \cap \operatorname{int}[T_{q_1}, T_{q_2}] \ne \emptyset\,\} \le 74$$

which we will prove now. Note that $\operatorname{int}[T_{q_1}, T_{q_2}]$ contains 11 points of $\mathcal{O}$. So there are at most 22 integers $0 \le k < n$ such that $f^k(p) \in \operatorname{int}[T_{q_1}, T_{q_2}]$. Hence there are at most 22 integers $0 \le k < n$ such that $U_k \subset \operatorname{int}[T_{q_1}, T_{q_2}]$. Let we need to show that there are at most 52 integers $0 \le k < n$ such that $U_k \cap \partial[T_{q_1}, T_{q_2}] \ne \emptyset$. So let a and b be the boundary points of $[T_{q_1}, T_{q_2}]$, so $a, b \in \mathcal{O}$. If U_k is not contained in $\operatorname{int}[T_{q_1}, T_{q_2}]$ but has a non-empty intersection with this set then $f^{n-k}(a)$ or $f^{n-k}(b) \in \operatorname{int}[T_{q_1}, T_{q_2}]$. Because $\operatorname{cl}([T_{q_1}, T_{q_2}])$ contains only 13 points of $\mathcal{O}$, there exist at most $2 \times 2 \times 13 = 52$ integers $0 \le k < n$ with this property. This completes the proof of the inequality and finishes the proof of Statement 1). Statement 2) follows from the Contraction Principle and the fact that U_k contains at most 5 points of $\mathcal{O}$. □

Proof of Proposition 10.1: Let $C > 0$ be so that if $T \supset I$ are intervals, $f|T$ is a diffeomorphism and $f(T)$ is a ϵ-scaled neighbourhood of $f(I)$ then T is a $(C \cdot \epsilon)$-scaled neighbourhood of I. Since f is non-flat at its critical points such a constant exists.

Let $m(1) < m(2) < \cdots < m(\ell)$ be the 'cutting' times of the pullback, i.e., the integers for which U_j contains a turning point in its closure. Let L_j and R_j be the components of $U_j \setminus f^j(J^i)$ and $J^i_j = f^j(J^i)$ and let $R_{m(k)}$ be the component which contains c_k in its boundary. Now U_n contains a τ scaled neighbourhood of $f^n(J^i)$. Since $f^{n-m(\ell)+1}\colon U_{m(\ell)+1} \to U_n$ is a diffeomorphism it follows from the Macroscopic Koebe Principle that there exists a positive function $B_0\colon \mathbb{R} \to \mathbb{R}^+$

(which only depends on f and the intersection multiplicity 74 from Lemma 10.3) such that $U_{m(\ell)+1}$ contains a $B_0(\tau)$-scaled neighbourhood of $J^i_{m(\ell)+1}$. If $U_{m(\ell)}$ also contains a $C \cdot B_0(\tau)$-scaled neighbourhood of $J^i_{m(\ell)}$ we repeat this procedure and we get from the Macroscopic Koebe Principle that $U_{m(\ell-1)+1}$ contains a $B_0(g(\tau))$-scaled neighbourhood of $J^i_{m(\ell-1)+1}$ where g is the function $g(x) = C \cdot B_0(x)$. $U_{m(\ell-1)}$ contains a $C \cdot B_0(g(\tau)) = g^2(\tau)$-scaled neighbourhood of $J^i_{m(\ell-1)}$ then we repeat this procedure. Since $U = J$ this procedure must stop however. Say it stops at $m(r)$ where $r \le \ell$ and then (by the definition of C above)

$$|R_{m(r)}| < g^r(\tau)|J^i_{m(r)}|,$$

where one has $r \le 74 \cdot \#K_f$, since intersection multiplicity is at most 74.

Now let M' be the middle third interval of $J^i_{m(r)}$ and $M \subset J^i$ be such that $f^{m(r)}(M) = M'$. From the Macroscopic Koebe Principle, see Theorem 3.3, J^i is a δ-scaled neighbourhood of M. From Lemma 10.3, $|U_k|$ is small if n is large. So we can apply the Second Expansion Principle, see Theorem 2.2, and get some universal constant $\xi > 0$ such that

$$B(f, f^{m(r)}(J^i), M') \ge 1 + \xi.$$

Let λ be such that $(1-\lambda)^2(1+\xi) \ge 1 + \frac{1}{2}\xi$. From the disjointness property of the orbit of J^i we get

$$B(f^{m(r)}, J^i, M) \ge 1 - \lambda,$$

$$B(f^{n-m(r)-1}, f^{m(r)+1}(J^i), f(M')) \ge 1 - \lambda,$$

for n large enough. Hence,

$$B(f^n, J^i, M) \ge 1 + \frac{1}{2}\xi,$$

for n large enough. Because of the First Expansion Principle, see Theorem 1.3, it suffices to show that the length of both components of $J^i \setminus M$ is at least $\delta \cdot |J^i|$. But since $f^{n-m(r)-1}$ has bounded distortion on $f(J^i_{m(r)})$, since f is non-flat at critical points and since the components of $J^i_{m(r)} \setminus M'$ have the same length as M', there exists a universal constant β such that the length of both components of $f^n(J^i \setminus M) = f^{n-m(r)}(J^i_{m(r)} \setminus M')$ is at least $\beta \cdot |f^n(J^i)|$. However, by Lemma 10.2, $f^n(J^i) \supset J^i$ and by assumption $|Df^n| \le 1 + \tau$ on J^i. It follows that both components of $J^i \setminus M$ have at least size $\dfrac{\beta}{1+\tau} \cdot |J^i|$. □

Proof of Theorem B

Let n be the period or twice the period of the periodic orbit $\mathcal{O}$ as before. Assume that $n > n_0$ where n_0 is as in Proposition 10.1. So there exist two points θ^1, θ^2 such that $p \in T = [\theta^1, \theta^2]$,

$$Df^n(\theta^i) \ge 1 + 2\rho, \ i = l, r,$$

For n large, $|f^i(T)|$ is small for all $i \in \{0, 1, \ldots, n\}$ and the orbit has intersection multiplicity ≤ 74. Therefore we get

$$[B(f^n, T^*, J^*)]^3 \geq \frac{1+\rho}{1+2\rho}$$

for all intervals $J^* \subset T^* \subset T$, provided n is large. Now applying the Minimum Principle,

$$|Df^n(x)| \geq [\inf B(f^n, T^*, J^*)]^3 \cdot (1+2\rho) \geq 1+\rho$$

for all $x \in T$. So $Df^n(p) \geq 1+\rho$. This proves Theorem B. □

11 Some Further Remarks and Open Questions

In this chapter we have shown that wandering intervals do not exist for $\mathcal{N}F^{1+Z}$ maps satisfying some non-flatness conditions. Furthermore, the examples by Denjoy (1932), Hall (1981), Sarkovskii and Ivanov (1983) and de Melo (1987) show that for continuous piecewise monotone maps this result cannot be improved much more.

Using Theorem A one sees that the result of Schwartz (1963) on minimal sets of flows on surfaces holds also for vector fields for which the holonomy (the translation along leaves) of the foliations is only C^{1+Z}. A more general result of Sacksteder (1965) extends the result of Schwartz to pseudo-groups and this gives a basic result on foliations with codimension-one leaves. The analytical part of Sacksteder's proof is exactly the same as the one in Schwartz's proof and therefore it requires the generators of the pseudo-group to be C^2 or $C^{1+\text{Lipschitz}}$. So a natural question is whether Sackstedter's result holds if the generators of the pseudo-group are merely C^{1+Z}, see also Hurder (1991).

The results in the chapter give a complete description in the smooth case. Some questions in the non-smooth case remain.

Conjecture: Suppose that $f : S^1 \to S^1$ is a homeomorphism such that both f and f^{-1} are smooth except in a finite number of points where f can locally be written in the form x^δ with $\delta > 0$. Then f has no wandering intervals.

The problem with such maps is that the cross-ratio distortion is not bounded from below at points where f is of the form x^δ with $\delta \in (0, 1)$. Therefore the method of proof given in this chapter breaks down. Similarly Blokh has asked

whether wandering intervals can exist for unimodal maps f which are smooth except at their critical point where $f(x)$ is locally of the form $f(x) = (x-c)^a$ for $x < c$ and $f(x) = (x-c)^b$ for $x > c$ with $a \neq b$. In the proof of the non-existence of wandering interval a local symmetry condition is needed (the involution τ needs to be Lipschitz), so the proof of Theorem A given above completely breaks down in this case.

Of course, for maps which are not continuous the situation is quite different. For example, as we have seen in Exercise I.2.2, there exists a piecewise affine interval exchange transformation which has wandering intervals! On the other hand, an extension of Theorem A might hold for all maps which are analytic except at, say, two discontinuities. Moreover, it is not clear whether the analogue of Theorem B of this chapter holds (i.e., the period of attractors is bounded). Partial results in this direction can be found in Berry and Mestel (1991) and Martens and de Melo (1992).

In higher dimensions the situation is far more complicated. On a 2-torus, a $C^{3-\epsilon}$ diffeomorphism can have a wandering domain, see McSwiggen (1992). By contrast, Norton and Velling (1991) have shown that there exists no C^1 diffeomorphism of the torus such that all iterates on some wandering domain are uniformly quasiconformal. Of course, this last condition is rather strong. We believe that partial results on the non-existence of wandering domains can also be obtained for C^3 diffeomorphisms.

Chapter V.

Ergodic Properties and Invariant Measures

In this chapter we want to study the typical behaviour of orbits. Up till now we have seen that the topological behaviour of interval maps is quite well understood: for example if f is a unimodal interval map with negative Schwarzian derivative and such that the fixed point on the boundary is repelling, then by Guckenheimer's theorem, see Theorem III.4.1, there are three possibilities:

1. f has a periodic attractor and then the basin of this attractor is a dense set in the interval;
2. f is infinitely renormalizable and then there exists a corresponding solenoidal Cantor set on which f acts as an adding machine, and, furthermore, a dense set of points is attracted by this Cantor set;
3. f is finitely often renormalizable and f is transitive on some finite union of intervals Λ: there exists a dense orbit in Λ. A dense set of points is attracted to Λ and periodic points appear densely in Λ.

Now we want to analyze whether we can say something similar about typical points in the Lebesgue sense and describe the asymptotics of typical orbits in more metrical detail. The metric analogue of the topological transitivity condition is the notion of ergodicity. As before let N be either the unit circle or a compact interval. Let $\mathcal{B}$ denote the Borel σ-algebra of N, i.e., the smallest σ-algebra that contains all the open subsets of N. We say that $f : N \to N$ is *ergodic* with respect to a measure $\mu \colon \mathcal{B} \to [0, \infty]$ if for each Borel set A such that $f^{-1}(A) = A$ we have either $\mu(A) = 0$ or $\mu(N \setminus A) = 0$. A measure $\mu \colon \mathcal{B} \to [0, \infty]$ is called f-*invariant* if $\mu(f^{-1}(A)) = \mu(A)$ for every $A \in \mathcal{B}$. If $\mu(N) = 1$ we say that μ is a *probability measure*. As we will see later, such invariant measures often describe the relative frequency certain parts of the space are visited by typical orbits. Since a non-invertible one-dimensional dynamical system usually has infinitely many periodic points, it also has infinitely many invariant measures: if p is a periodic point of period n of $f \colon N \to N$ and δ_x denotes the

Dirac measure at x (i.e., $\delta_x(A) = 1$ if $x \in A$ and $\delta_x(A) = 0$ otherwise) then $\mu = \frac{1}{n}\sum_{i=0}^{n-1} \delta_{f^i(p)}$ is an invariant probability measure. So it is necessary to look for further properties in order to select the relevant invariant measures. One property that reveals, as we will see below, some chaotic or stochastic behavior of the dynamics is the existence of an invariant measure which is absolutely continuous with respect to the Lebesgue measure λ. A measure $\mu\colon \mathcal{B} \to [0,\infty]$ is *absolutely continuous* with respect to the Lebesgue measure λ, if for every $A \in \mathcal{B}$ with $\lambda(A) = 0$ we have that $\mu(A) = 0$. From the Radon-Nikodym theorem, μ is a finite absolutely continuous measure if and only if it has a density function $\phi \in L^1(N)$, i.e., $\mu(A) = \int_A \phi \, d\lambda$.

In this chapter we will almost always confine ourselves to unimodal maps which satisfy the negative Schwarzian derivative condition. In many cases the extension to the general case can be achieved with the techniques from Chapter IV.

In Section 1 we will describe some work of Blokh and Lyubich with extensions by Martens (1990) about the ergodicity and the (metric) attractors of unimodal maps. It will be shown that such maps are ergodic and that for each such map there exists a set C such that the ω-limit of Lebesgue almost every x is equal to C. This set can be of four types. Whether all these types actually occur is unknown.

The next question to be discussed in this chapter is the existence of absolutely continuous invariant measures for one-dimensional dynamical systems. Many papers have addressed this question: Adler (1973), Benedicks and Carleseon (1985), (1991), Bowen (1977b), (1979), Collet and Eckmann (1983), Jakobson (1981), Johnson (1986), (1987), Hofbauer and Keller, (1982b), (1990a), Keller (1987), (1989), Lasota and Yorke (1973) Misiurewicz (1981), Nowicki and Van Strien (1988), (1991a), Piangiani (1979), (1980), Renyi (1957), Ruelle (1977), Van Strien (1990), Rees (1984), (1986), Rychlik (1983), (1988) and Szlenk (1979). In these papers it was shown that many maps possess such invariant measures. In particular, it is proved in Jakobson (1981) that in the quadratic family, the set of parameter values for which the corresponding map has an absolutely continuous invariant measure has positive Lebesgue measure. In this sense, the situation here is similar to the case of one parameter families of circle diffeomorphisms where, as we have seen in Chapter I, in many families the set of parameter values corresponding to diffeomorphisms with such an invariant measure has positive Lebesgue measure.

One motivation for studying the existence of an absolutely continuous invariant probability measure is that it implies chaotic behavior. In fact, Ledrappier, see also Section 3, has shown that if the interval map $f\colon I \to I$ has an absolutely continuous invariant probability measure with positive metric entropy then there exists a subset $A \subset I$ of positive Lebesgue measure, such that the *Lyapunov exponent* $\lim \frac{1}{n} \log |Df^n(x)|$ exists and is equal to some positive constant for each $x \in A$. By a result of Ruelle (1979) this implies that f exhibits *sensitive dependence on initial conditions*, i.e., the orbits through nearby points from a set of positive Lebesgue measure get separated exponentially fast. So, for these points, an error on the initial condition propagates exponentially fast under iteration.

Moreover, as we will show in Section 1 these measures are *Sinai-Bowen-Ruelle*. We say that a f-invariant measure $\mu\colon \mathcal{B} \to [0,1]$ is a Sinai-Bowen-Ruelle measure if there exists a set $B \subset I$ of positive Lebesgue measure such that for any continuous function $\phi\colon I \to \mathbb{R}$ one has

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} \phi(f^n(x)) = \int \phi\, d\mu \text{ for every } x \in B.$$

Any measure whose support is an attracting periodic orbit is a Sinai-Bowen-Ruelle measure. Moreover, as we will show in Section 1, a unimodal map which is infinitely renormalizable has a Sinai-Bowen-Ruelle invariant probability measure whose support is the attracting Cantor set.

In Sections 2, 3 and 4 we shall study under what conditions maps have absolutely continuous invariant probability measures. In Section 2 we will prove a folklore theorem about the existence of such measures for expanding Markov maps. In Section 3 we will prove the results of Ledrappier mentioned above and its converse due to Keller. In Section 4 another very weak condition is given for the existence of these measures for unimodal maps. In Section 5 we will give a result due to Hofbauer and Keller (extending work of Johnson) which shows that the support of Bowen-Ruelle-Sinai measures can be quite unexpected. Finally, in Section 6 we will give a detailed proof of the result of Benedicks and Carleson (1991) that for many one-parameter families of unimodal maps, for a positive set of parameters the corresponding maps have positive Lyapunov exponents. Combining this with Section 4 we get a result which implies the theorem of Jakobson mentioned above.

1 Ergodicity, Attractors and Bowen-Ruelle-Sinai Measures

In this section we will consider unimodal maps $f\colon [-1,1] \to [-1,1]$ with a unique critical point, such that this critical point is non-flat, f has negative Schwarzian derivative and such that the fixed point of f on the boundary of $[-1,1]$ is repelling. (This last condition is in order to avoid silly problems.) The first result we will prove is that such maps are ergodic. This result is due to Blokh and Lyubich (1986), (1987), (1989a,b), (1990) and (1991). Recently in Blokh and Lyubich (1989c) and Lyubich (1990) the same result in the multimodal case is proved. Since we will confine ourself throughout this chapter to unimodal maps we will not present the multimodal proof here. As before, we say that f is *ergodic* with respect to the Lebesgue measure if each completely invariant set X (by this we mean $f^{-1}(X) = X$ and certainly not $f^{-1}(X) \subset X$) has either zero or full Lebesgue measure. An alternative way to define this notion of ergodicity goes as follows: f is ergodic if for each two forward invariant sets X and Y such that $X \cap Y$ has Lebesgue measure zero, at most one of these sets has positive Lebesgue measure. (Here X is called forward invariant if $f(X) \subset X$.)

Exercise 1.1. Prove that these two definitions are equivalent. (Hint: since $X \cap Y$ has Lebesgue measure zero, the same holds for $B(X) = \{x\,;\ f^n(x) \in X \text{ for some } n \geq 0\}$ and $B(Y)$ and since these last two sets are completely invariant the equivalence of these two definitions follows.)

Furthermore, we will give several results about attractors of interval maps. Following Milnor (1985) we say that a closed forward invariant set A is a *metric attractor* if the basin $B(A) = \{x\,;\ \omega(x) \subset A\}$ satisfies

1. the measure of $B(A)$ is positive;
2. each closed forward invariant subset A' which is strictly contained in A has a smaller basin of attraction: $B(A) \setminus B(A')$ has positive Lebesgue measure.

Next we will show that such a unimodal map can have at most one attractor. This result was also first proved by Blokh and Lyubich, see Blokh and Lyubich (1987). We will prove all these results using a distortion result of Martens (1990) and also prove his result that the attractor has either Lebesgue measure zero, or contains intervals. We should point out that a similar result was also obtained by Guckenheimer and Johnson (1990).

A distortion result for unimodal maps with recurrence

Given a unimodal map f, we say that an interval U is *symmetric* if $\tau(U) = U$ where $\tau\colon [-1,1] \to [-1,1]$ is so that $f(\tau(x)) = f(x)$ and $\tau(x) \neq x$ if $x \neq c$. Furthermore, for each symmetric interval U let

$$D_U = \{x\,;\ \text{ there exists } k > 0 \text{ with } f^k(x) \in U\};$$

for $x \in D_U$ let $k(x,U)$ be the minimal positive integer with $f^k(x) \in U$ and let

$$R_U(x) = f^{k(x,U)}(x).$$

We call $R_U\colon D_U \to U$ the *Poincaré map* or *transfer map* to U and $k(x,U)$ the *transfer time* of x to U. The distortion result states that one can find a sequence of symmetric neighbourhoods of the turning point such that the Poincaré maps to these intervals have a distortion which is universally bounded:

Theorem 1.1. *Let $f\colon [-1,1] \to [-1,1]$ be a unimodal map with one non-flat critical point with negative Schwarzian derivative and without attracting periodic points. Then there exists $\rho > 0$ and a sequence of symmetric intervals $U_n \subset V_n$ around the turning point which shrink to c such that V_n contains a ρ-scaled neighbourhood of U_n and such that the following properties hold.*

1. *The transfer time on each component of* D_{U_n} *is constant.*

2. *Let* I_n *be a component of the domain* D_{U_n} *of the transfer map to* U_n *which does not intersect* U_n*. Then there exists an interval* $T_n \supset I_n$ *such that* $f^k|T_n$ *is monotone,* $f^k(T_n) \supset V_n$ *and* $f^k(I_n) = U_n$*. Here* k *is the transfer time on* I_n*, i.e.,* $R_{U_n}|I_n = f^k$*.*

Corollary 1.1. *There exists* $K < \infty$ *such that*

1. *for each component* I_n *of* D_{U_n} *not intersecting* U_n*, the transfer map* R_{U_n} *to* U_n *sends* I_n *diffeomorphically onto* U_n *and the distortion of* R_{U_n} *on* I_n *is bounded from above by* K*;*

2. *on each component* I_n *of* D_{U_n} *which is contained in* U_n*, the map* $R_{U_n}: I_n \to U_n$ *can be written as* $(f^{k(n)}|f(I_n))\circ f|I_n$ *where the distortion of* $f^{k(n)}|f(I_n)$ *is universally bounded by* K*.*

Proof. Proof of the Corollary Follows immediately from the previous theorem and the Koebe Principle, see Theorem IV.1.1. □

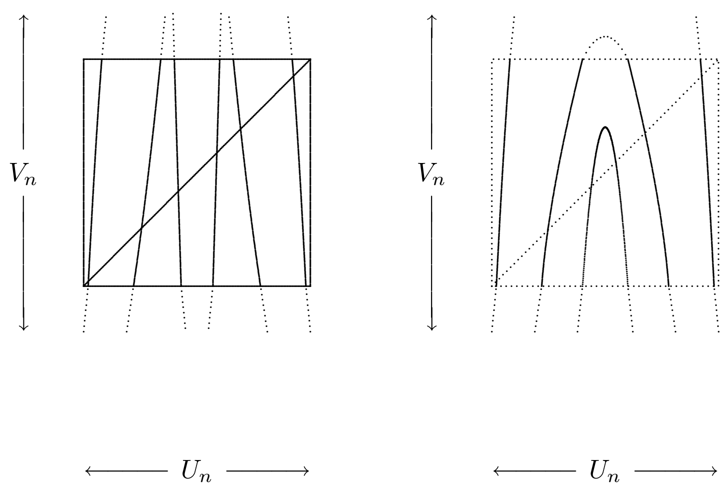

Fig. 1.1: If the forward orbit of c does not enter U_n then the transfer map to U_n is as drawn on the left, otherwise it is as the map on the right. The extension of the branches is also drawn. Outside U_n each branch extends monotonically to V_n. Inside U_n each branch, except perhaps a central branch containing c, is mapped by the transfer map monotonically to U_n.

Remarks. 1. The Contraction Principle of Section IV.5 tells us that if U_n and I_n are sequences of intervals with $|U_n| \to 0$ and such that some iterate $f^{k(n)}$ of f maps I_n into U_n, then $|I_n| \to 0$. We will use this property several times in this section.

2. The collection of maps of the form $f^{k+1}|U$ where U is a neighbourhood of c, $f^{k+1}|U$ has one extremum and for which $f^k|f(U)$ has universally bounded distortion is called *polynomial-like*. 3. Because the intervals U_n shrink to c, for each point x in the set $\{x\,;\;\omega(x)\ni c\}$ and which is not in the backward orbit of c (i.e., for each x from a set of full measure) the transfer time $k(n)$ tends to infinity as $n\to\infty$. 4. Martens calls the property described in the previous theorem the *weak Markov property*.

All our proofs in this section are based on this result which is due to Martens (1990). Related results were already proved by Blokh and Lyubich (1986)-(1990), but they only got a one-sided inequality for the distortion of f^n. Guckenheimer and Johnson (1990) also have a result similar in spirit to Theorem 1.1. However, in their result the intervals U_n and I_n are constructed using metric decision rules, whereas the collection of intervals U_n of Theorem 1.1 are determined by topological properties (in particular if two unimodal maps are conjugate then the conjugacy sends the intervals U_n onto corresponding intervals).

Ergodicity and description of attractors

Before proving Theorem 1.1, we will first state some corollaries of this theorem. All these corollaries, except the last statement in Theorem 1.3 below, were first proved by Blokh and Lyubich (1986)-(1990). However, Keller (1987), (1989) and Guckenheimer and Johnson (1990) also obtained similar results.

Theorem 1.2. **(Blokh and Lyubich)** *Let $f\colon[-1,1]\to[-1,1]$ be a unimodal map with a non-flat critical point with negative Schwarzian derivative and without an attracting periodic points. Then f is ergodic with respect to the Lebesgue measure.*

Proof that Theorem 1.2 follows from Theorem 1.1. Suppose that X and Y are forward invariant sets such that $X\cap Y$ has zero Lebesgue measure and both of these sets have positive Lebesgue measure. We will show that this is impossible. By Theorem III.3.2, the set $\mathcal{C}$ of points whose ω-limit contains c has full Lebesgue measure. Therefore one can take a density point x of $X\cap\mathcal{C}$. By the previous theorem there exist intervals $I_n\ni x$, $U_n\ni c$ and integers $k(n)$ such that $f^{k(n)}$ maps I_n with bounded distortion to U_n. One has therefore

$$\frac{|U_n\setminus X|}{|U_n|}\le\frac{|f^{k(n)}(I_n\setminus X)|}{|f^{k(n)}(I_n)|}\le K\frac{|I_n\setminus X|}{|I_n|}.$$

Since x is a density point of X, one has $\frac{|I_n\setminus X|}{|I_n|}\to 0$ (here we use that $|I_n|\to 0$ which holds because of the Contraction Principle, see Remark 1 above). Therefore, the last inequality implies that

$$\frac{|U_n\cap X|}{|U_n|}\to 1 \text{ as } n\to\infty.$$

So the upper density of X at c is one (in Exercise 1.3 below it is shown that c is in fact a density point of X). Similarly, one has

$$\frac{|U_n \cap Y|}{|U_n|} \to 1 \text{ as } n \to \infty.$$

But then $X \cap Y$ can certainly not have Lebesgue measure zero and thus we get a contradiction. □

Exercise 1.2. Let f be a non-renormalizable unimodal map and X a forward invariant set. Suppose there are symmetric neighbourhoods $U_n \subset V_n$ of c such that V_n contains a δ-scaled neighbourhood of U_n and $V_n \cap X \subset U_n$ and such that $|V_n| \to 0$. Show that X has Lebesgue measure zero. (Hint: as before let $\mathcal{C}$ be the set of points whose ω-limit contains c. This set has full Lebesgue measure. For each point of $x \in \mathcal{C} \cap X$, there exists k such that $f^k(x) \in V_n$. Let k be the minimal such integer. Let us first show this implies that there exists an interval $T_n \ni x$ such that f^k maps T_n diffeomorphically onto V_n. Indeed, assume by contradiction that the maximal interval T_n on which f^k is a diffeomorphism and for which $f^k(T_n) \subset V_n$ would be mapped strictly inside V_n. Then there exists some integer $i < k$ such that $f^i(T_n)$ contains c in its boundary. Since $f^i(x) \notin V_n$ this implies that $f^i(T_n)$ contains a component of $V_n \setminus \{c\}$. So $f^{k-i}(V_n) \subset f^{k-i}(f^i(T_n)) \subset V_n$ which contradicts the assumption that V_n is non-renormalizable. Consequently, T_n as above exists. Let $S_n \subset T_n$ be so that $f^k(S_n) = U_n$. From the Macroscopic Koebe Principle T_n contains a δ'-scaled neighbourhood of S_n. From the forward invariance of X one gets that $T_n \cap X \subset S_n$ and S_n contains x. Since $|T_n| \to 0$, as we saw in Remark 1 above, it follows that x is not a density point of $X \cap \mathcal{C}$. Since $\mathcal{C}$ has full Lebesgue measure this implies that almost no point of X is a density point. Therefore, X has Lebesgue measure zero.)

Exercise 1.3. Let U_n be the sequence of intervals from Theorem 1.1. In the proof of Theorem 1.2 it is shown that any forward invariant set X which has positive Lebesgue measure has the property that

$$(*) \qquad \frac{|U_n \cap X|}{|U_n|} \to 1 \text{ as } n \to \infty.$$

So the upper density of X in c is one. Now let τ be the involution such that $f(\tau(x)) = f(x)$ and $\tau(x) \neq x$. Show that if X is a symmetric set, $\tau(X) = X$ then c is a density point of X. So prove that $(*)$ holds for *any* sequence of intervals U_n with $c \in U_n$ and $|U_n| \to 0$. This result is due to Blokh and Lyubich. (Hint: For an interval I let $\rho(I) = |X \cap I|/|I|$. Since X has positive Lebesgue measure there exists a density point x of X whose ω-limit contains c. Write $x_n = f^n(x)$ and let $x_{n(k)}$ be the sequence of closest approach. This means that $n(0) = 1$ and $n(k+1)$ is inductively defined to be the smallest positive integer l such that $x_l \in (x_{n(k)}, \tau(x_{n(k)}))$. Let $V_n = (x_{n(k)}, \tau(x_{n(k)}))$. As before since $x_l \notin V_n$ for $l < n(k+1)$ there exists an interval $T_n \ni x$ such that $f^{n(k+1)}$ maps T_n diffeomorphically onto V_n. Since x is a density point of X, $|T_n^\pm \cap X|/|T_n^\pm|$ tends to 1 for both components $T_n^\pm$ of $T_n \setminus \{x\}$. By the one-sided Koebe Principle from Section IV.1 this implies that the shorter component V_n' of $V_n \setminus f^{n(k+1)}(x)$ also satisfies $|V_n' \cap X|/|V_n'|$ tends to 1. Now V_n' is one of the components of $V_n \setminus V_{n+1}$. Since X is symmetric and τ smooth (because the critical point is non-flat) one therefore gets that $\rho(V_n \setminus V_{n+1}) \to 1$ as $n \to \infty$. It follows that $\rho(V_n) \to 1$ as $n \to \infty$ and therefore c is a density point.)

Exercise 1.4. Show that any absolutely continuous invariant probability measure μ of a unimodal map as above has positive metric entropy. This result is due to Blokh and Lyubich (1990). (Hint: if the metric entropy of the measure μ were zero, then it is a general result that f would be invertible μ-almost everywhere (i.e., there exists a set D of full μ measure so that $f|D$ is injective). But this would imply that the support of μ could have at most density $1/2$ at the critical point of f, in contradiction to the previous exercise.)

Let us now describe the attractors of unimodal maps. In the next result it is shown that maps as above have a unique attractor and that this set is either an attracting periodic orbit, or a finite union of intervals or a Cantor set of zero Lebesgue measure. In order to be more specific we remind the reader that each such map f can be of three topological types, see Section III.4:

1. *f has an attracting periodic orbit;*
2. *f is infinitely often renormalizable and the closure of the orbit of the critical point is a Cantor set;*
3. *f is only finitely often renormalizable and f is transitive on some finite union of intervals (i.e., there exist orbits which are dense in these intervals).*

The next result states that this topological classification coincides with the metric one in many ways. The first part of this result is due to Blokh and Lyubich (1986)-(1990) and the last statement is due to Martens (1990).

Theorem 1.3. *Let $f\colon [-1,1] \to [-1,1]$ be a unimodal map with a non-flat critical point c and with negative Schwarzian derivative. Then f has a unique attractor A, $\omega(x) = A$ for almost all x and A either consists of intervals or has Lebesgue measure zero. Furthermore, one has the following:*

1. *if f has an attracting periodic orbit then A is this periodic orbit;*
2. *if f is infinitely often renormalizable then A is the attracting Cantor set $\omega(c)$ (in which case it is called a solenoidal attractor);*
3. *f is only finitely often renormalizable then either*
   a. *A coincides with the union of the transitive intervals, or,*
   b. *A is a Cantor set and equal to $\omega(c)$.*

If $\omega(c)$ is not a minimal set then f is as in case 3.a and each closed forward invariant set either contains intervals or has Lebesgue measure zero. Moreover, if $\omega(c)$ does not contain intervals, then $\omega(c)$ has Lebesgue measure zero.

Here a forward invariant set X is said to be *minimal* if the closure of the forward orbit of a point in X is always equal to X.

It is not known whether maps with attractors as in case 3.b really do exist. Such an attractor is called a *non-renormalizable attracting Cantor set* or *absorbing Cantor attractor*. (This last terminology was introduced in Guckenheimer and Johnson (1990).) If such an attractor really can exist then because of the previous result one has the following strange phenomenon: there exist many orbits which are dense in some finite union of intervals and yet almost all points tend to a minimal Cantor set of Lebesgue measure zero (this Cantor set is $\omega(c)$). In Section II.5a we have seen the example of the Fibonacci map which is non-renormalizable and for which $\omega(c)$ is a Cantor set. Recently, it was shown by Milnor and Lyubich (1991) that the quadratic map with this dynamics has no absorbing Cantor sets. More generally, Jakobson and Świątek (1991a) proved that maps with negative Schwarzian derivative and which are close to the map $f(x) = 4x(1-x)$ do not have such Cantor attractors. Moreover, Lyubich (1992) has shown that these absorbing Cantor attractors cannot exist if the critical point is quadratic:

Theorem 1.4. **(Lyubich)** *If $f\colon [-1,1] \to [-1,1]$ is C^3 unimodal, has a quadratic critical point, has negative Schwarzian derivative and has no periodic attractors, then each closed forward invariant set K which has positive Lebesgue measure contains an interval.*

On the other hand, computer experiments done by Keller and Nowicki and also by Lyubich and Milnor indicate that such absorbing Cantor sets can exist if the critical point is of order 6 or larger.

The next result, which is due to Martens (1990), shows that if these absorbing Cantor attractors do not exist then one has a lot of 'expansion'. This expansion will later on, in Section 3, imply that f induces a Markov map; for this reason Martens calls Property 3 in the theorem below a Markov property. Let x not be in the preorbit of c and define $T_n(x)$ to be the maximal interval on which $f^n|T_n(x)$ is monotone. Let $R_n(x)$ and $L_n(x)$ be the components of $T_n(x) \setminus x$ and define $r_n(x)$ be the minimum of the length of $f^n(R_n(x))$ and $f^n(L_n(x))$. In the next section we shall show that Markov maps have absolutely continuous invariant probability measures.

Theorem 1.5. **(Martens)** *Let f be a C^3 unimodal map with negative Schwarzian derivative whose critical point is non-flat. Then the following three properties are equivalent.*

1. *f has no absorbing Cantor attractor;*
2. $\limsup_{n\to\infty} r_n(x) > 0$ *for almost all x;*
3. *there exist neighbourhoods $U \subset V$ of c with $\mathrm{cl}\,(U) \subset \mathrm{int}\,(V)$ such that for almost every x there exists a positive integer m and an interval neighbourhood T of x such that $f^m|T$ is monotone,*

$$f^m(T) \supset V \text{ and } f^m(x) \in U.$$

Remark. 1. Note the difference of 3) in this theorem with the situation described in Theorem 1.1. In contrast to the situation above it is not claimed in Theorem 1.1 that one can find arbitrary small neighbourhoods of most points which are mapped with bounded distortion onto a fixed neighbourhood of c. In fact, in Theorem 1.1 one uses the first return map and in general small intervals near c are not mapped monotonically onto U_n by the first return map (this only happens if $c \notin D_{U_n}$).

2. Martens (1991), has shown that if f is non-renormalizable and $\omega(c)$ is a Cantor set then it has a σ-finite absolutely continuous invariant measure. If this measure is finite, then this implies that f has no absorbing Cantor attractor, see Theorem 1.5 below.

Proofs of Theorem 1.3 and 1.4

Proof that Theorem 1.3 follows from Theorem 1.2 (except the statement that $|\omega(c)| = 0$): If f is as in Case 1 or 2 then the theorem follows from the main result in Section III.4. So let us assume that f is only finitely often renormalizable and let Λ be the union of the transitive intervals of f. We want to show that if Λ is not an attractor (in the sense defined above) then $\omega(c)$ must be an attractor. For this we use the Macroscopic Koebe Principle. Let us quickly recall this principle. Since f has negative Schwarzian derivative we don't need to assume any disjointness conditions as in Chapter IV. Therefore this principle simply states that for every $\epsilon > 0$ there exist $\delta > 0$ and $K < \infty$ such that the following holds. Let $M \subset T$ be intervals in $[-1, 1]$. If $f^n|T$ is monotone and $f^n(T)$ contains an ϵ-scaled neighbourhood of $f^n(M)$ then T contains a δ-scaled neighbourhood of M and $|Df(x)|/|Df(y)| \le K$ for each $x, y \in M$. The idea of the proof of this theorem is to use this Principle to show that if x is a density point of a forward invariant set X then its iterates are also density points of X (in a way made precise below).

So let J_n be a countable collection of open intervals with $|J_n| \to 0$ and such that $\cup_{n \ge N} J_n$ covers Λ for each $N \in \mathbb{N}$. Furthermore, let $C_n = \{x \in \Lambda\,;\ f^k(x) \notin \mathrm{cl}\,(J_n)$ for all $k \ge 0\}$. Clearly C_n is forward invariant. If we take $C = \cup_{n \ge 0} C_n$ and $D = [-1, 1] \setminus C$ we get that the orbit of each point in D is dense in Λ, i.e., $D = \{x\,;\ \omega(x) = \Lambda\}$. From the description of C and D it follows that C and D are completely invariant and therefore from the ergodicity of f, either C or D has full Lebesgue measure. If D has full measure we are finished. So assume C has full Lebesgue measure. Then one of the sets C_n has positive Lebesgue measure. We claim that for each density point of $x \in C_n$ one has that $\omega(x)$ is contained in the closure $\overline{O_+(c)}$ of the forward orbit of c. Indeed, otherwise there exist iterates $k(i) \to \infty$ and $\delta > 0$ such that $\mathrm{dist}(f^{k(i)}(x), \overline{O_+(c)}) \ge \delta$. Let T_i be the maximal interval containing x such that $f^{k(i)}|T_i$ is monotone. Because all critical values of $f^{k(i)}$ are contained in $\overline{O_+(c)}$, one has that $f^{k(i)}(T_i)$ contains a δ neighbourhood of $f^{k(i)}(x)$ for i large. Let T_i' be intervals containing x such that $f^{k(i)}(T_i')$ contains a $\delta/2$ neighbourhood of $f^{k(i)}(x)$ for i large. Since C_n

is forward invariant and x is a density point of C_n it follows from the Koebe Principle that there exists a universal constant $K < \infty$ such that

$$\frac{|f^{k(i)}(T_i') \setminus C_n|}{|f^{k(i)}(T_i')|} \leq \frac{|f^{k(i)}(T_i' \setminus C_n)|}{|f^{k(i)}(T_i')|} \leq K\frac{|T_i' \setminus C_n|}{|T_i'|} \to 0$$

as $i \to \infty$. Since the intervals $f^{k(i)}(T_i')$ have length $\geq \delta/2$ for i large, one can find an interval H of length $\delta/3$ such that one has $f^{k(i)}(T_i') \supset H$ for infinitely many i. The previous limit therefore implies that $|H \setminus C_n| = 0$ and since C_n is closed this gives $H \subset C_n$. But as we saw in Section III.4, since H is an interval, $\cup_{i\geq 0} f^i(H) \supset \Lambda$. Therefore $\Lambda \subset C_n$. So C_n contains intervals. But this is impossible for the same reason because $\cup_{i\geq 0} f^i(U) \supset J_n$ for each interval U, a contradiction. This proves the claim: $\omega(x)$ is contained in the closure of the forward orbit of c for each density point $x \in C$. It is easy to show that this implies that $\omega(x) \subset \omega(c)$ for Lebesgue almost every x, using the ergodicity of f. But since the ω-limit of almost every point contains c, we also have $\omega(c) \subset \omega(x)$ for almost every point x of C. It follows that either 3.a or 3.b. holds.

Let us finally show that any closed forward invariant set X which does not contain intervals has Lebesgue measure zero if $\omega(c)$ is not minimal. Since $\omega(c)$ is not minimal, it contains a point x whose forward orbit stays outside a neighbourhood U of c. As before this implies that there exists a sequence of intervals $I_n \ni x$ shrinking down to x and intervals $J_n \subset I_n$ with $x \in J_n$ such that $I_n \setminus J_n$ contains no points of X and such that I_n contains a δ-scaled neighbourhood of J_n. Indeed, let H_n be the maximal interval containing x for which $f^n|H_n$ is monotone. Since $c \notin \omega(x)$, the length $|f^n(H_n)|$ does not tend to zero, By taking some appropriate subintervals $I_n \subset H_n$ we may assume that $f^{n(i)}(I_{n(i)})$ converges to some interval V for some subsequence $n(i)$ and that the boundary points of V are not contained in X. Let U be some closed interval which is contained in the interior of V such that $V \cap X \subset U$ and let $J_{n(i)} \subset I_{n(i)}$ so that $f^{n(i)}(J_{n(i)}) = U$. From the Koebe Principle, there exists a universal constant $\delta > 0$ (which only depends on f, U and V) such that $I_{n(i)}$ is a δ-scaled neighbourhood of $J_{n(i)}$. Now let $k(n)$ be the smallest integer such that $f^{k(n)}(c_1) \in I_n$ where $c_1 = f(c)$. It follows that there exists an interval $T_n \ni c_1$ such that $f^{k(n)}|T_n$ is monotone and $f^{k(n)}(T_n) = I_n$. Let $S_n \subset T_n$ be so that $f^{k(n)}(S_n) = J_n$. From the Macroscopic Koebe Principle T_n is a δ'-scaled neighbourhood of S_n. Furthermore, the forward invariance of X gives $T_n \cap X \subset S_n$ and therefore $f^{k(n)}(c_1) \in J_n$. In particular, $c_1 \in S_n$. Since c is non-flat this implies that there are symmetric neighbourhoods $U_n \subset V_n$ of c such that V_n is a δ''-scaled neighbourhood of U_n and $V_n \cap X \subset U_n$. By Exercise 1.2 this implies that X has Lebesgue measure zero. Applying the above reasoning to the sets C_n from the first part of the proof yields $|C_n| = 0$ and hence that D has full measure. Hence the non-minimality of $\omega(c)$ implies that f is as in case 3.a. □

The proof of the last part of Theorem 1.3, that A either contains intervals or has Lebesgue measure zero, will require some additional information. This information will be obtained during the proof of Theorem 1.1.

Proof of Theorem 1.4: Our proof differs with the one given in Martens (1990) and is much simpler. Without loss of generality we may assume that f is not renormalizable and that f has no periodic attractors. Throughout the proof we shall use the following. By the Contraction Principle for each $\delta > 0$ there exists a constant $\kappa(\delta) > 0$ such that the length of $f^k(T)$ is at least $\kappa(\delta)$ for each interval T with $|T| \geq \delta$ and each integer $k \geq 0$ so that $f^k|T$ is monotone.

So assume that Property 1 holds. If $\omega(c)$ is a Cantor set then, since this set is not an absorbing Cantor attractor (and f is not renormalizable), we are as in case 3.b of Theorem 1.3. Hence $\omega(x)$ is equal to the interval $[f^2(c), f(c)]$ for almost all x. Hence for each $z \in [f^2(c), f(c)] \setminus \omega(c)$, there is a sequence $k(n) \to \infty$ with $f^{k(n)}(x) \to z$. Since all extremal values of $f^{k(n)}$ are forward iterates of c and since $\omega(c)$ is a Cantor set, this implies $\limsup_{n\to\infty} r_n(x) > 0$ for almost all x. So Property 2 holds.

Now assume again that Property 1 holds but that $\omega(c)$ is not a Cantor set. Then $\omega(c) = [f^2(c), f(c)]$. Take an open neighbourhoods U of c such that $f^n(\partial U) \cap U = \emptyset$ for $n = 1, 2, \ldots$. Since $\omega(c) = [f^2(c), f(c)]$, there exists a nested decreasing sequence of intervals Z_n whose endpoints are periodic points p_n and q_n with $O(p_n) \cap U = O(q_n) \cap U = \emptyset$ and with $|Z_n| \to 0$. We will first prove that there are disjoint intervals $I_n^i \subset Z_n$ and integers $k_i(n)$ for each $i \in Z$ such that

1. $f^{k_i(n)}$ maps I_n^i with bounded distortion onto U;
2. I_n^i lies to the left of I_n^j if $i < j$;
3. $\lim_{i\to\infty} I_n^i = p_n$ and $\lim_{i\to-\infty} I_n^i = q_n$;
4. the size of the 'gap' components of $Z_n \setminus \bigcup_j I_n^j$ neighbouring I_n^j have the same order as this interval.

To construct these intervals, let $k(n)$ be the smallest integer such that $f^{k(n)}(Z_n) \cap U \neq \emptyset$. This integer exists because otherwise Z_n would be a homterval and because f has no periodic attractors and no wandering intervals this is impossible. From the choice of p_n and q_n and from the minimality of $k(n)$, it follows that $f^{k(n)}$ maps Z_n diffeomorphically over U. Moreover, $f^{k(n)}|Z_n$ has bounded distortion by Theorem III.3.3. It follows that there is an interval I_n^0 in the interior of Z_n such that $f^{k(n)}(I_n^0) = U$. Since $f^{k(n)}|Z_n$ has bounded distortion, I_n^0 and the components of $Z_n \setminus I_n^0$ have the same size up to some multiplicative factor which is universally bounded from above and below. Next let $H_n^{\pm 1}$ be the components of $Z_n \setminus I_n^0$ and let $k_{\pm 1}(n)$ be the smallest integer such that $f^{k_{\pm 1}(n)}(H_n) \cap U \neq \emptyset$. Note that by the choice of $k(n)$ we have that $k_{\pm 1}(n) > k(n)$ and that the interval $f^i(H_n^{\pm 1})$ is outside U for $i = 0, 1, \ldots, k(n)$. Moreover, $f^{k(n)}$ maps one endpoint of $H_n^{\pm 1}$ into ∂U and the other endpoint outside U. Since forward iterates of ∂U never enter U it follows that $f^{k_{\pm 1}(n)}(H_n^{\pm 1}) \supset U$. Hence, we get two intervals $I_n^{\pm 1}$ one on each side of I_n^0 such that $f^{k_{\pm 1}(n)}(I_n^{\pm 1}) = U$ and such that $I_n^{\pm 1}$ and the components of $H_n^{\pm 1} \setminus I_n^{\pm 1}$ have the same size up to some factor which is again universally bounded from above and below. Continuing in this way we get a sequence of disjoint intervals

I_n^i in Z_n with $i \in \mathbb{Z}$ such that the size of each component of $Z_n \setminus \bigcup_i I_n^i$ is of the same order as the size of the neighbouring components of $\bigcup_i I_n^i$. This collection has the required properties 1)-4). Next let $s(n) > 0$ be the smallest integer such that $f^{s(n)}(c) \in Z_n$. By the minimality of $s(n)$ there exists an interval $\tilde{S}_n \ni c$ such that $f^{s(n)-1}$ maps $\tilde{S}_n$ diffeomorphically onto Z_n. Let $\tilde{I}_n^i$ be the preimage of I_n^i under this map. By the Koebe Principle, the size of each gap of $\bigcup \tilde{I}_n^i$ in $\tilde{Z}_n$ is again of the same order as its neighbouring components: use two neighbouring intervals to show that the pullback of the gap between them is not too large; similarly, use two neighbouring gaps to show that the pullback of an interval I_n^i between them is not too large. Note that $f^{k_i(n)+s(n)-1}$ maps $\tilde{I}_n^i$ diffeomorphically onto U. Now let $\hat{Z}_n = f^{-1}(\tilde{Z}_n)$ and let $\hat{I}_n^i = f^{-1}(\tilde{I}_n^i)$. Note that the preimage of some intervals $\tilde{I}_n^i$ consists of two components, some of one and some of none at all (if they are at the 'wrong' side of $f(c)$). Because f is non-flat at the critical point, the gaps of $\bigcup_i \hat{I}_n^i$ are again of the same order as the neighbouring intervals. So if we disregard the interval $\hat{I}_n^i$ containing c together with the two components from the collection $\bigcup \hat{I}_n^i$ nearest to c, then we get that each such component $\hat{I}_n^i$ is mapped by $f^{k_i(n)+s(n)}$ diffeomorphically and with universally bounded distortion onto U. Let $\bigcup_{i\in A} I_n^i$ be the resulting collection. Now take x with $\omega(x) \ni c$ and let $t(n) > 0$ be minimal such that $f^{t(n)}(x) \in \hat{Z}_n$. Then there exists an interval $\bar{Z}_n \ni x$ such that $f^{t(n)}$ maps $\bar{Z}_n$ diffeomorphically onto $\hat{Z}_n$. Let $\bar{I}_n^i = f^{-t(n)}(\hat{I}_n^i) \cap \bar{Z}_n$ for $i \in A$. Again, from the Koebe Principle, the gaps and the neighbouring intervals from $\bigcup_{i\in A} \bar{I}_n^i$ have roughly the same size and $f^{k_i(n)+s(n)+t(n)}$ maps $\bar{I}_n^i$ with bounded distortion onto U for $i \in A$. Finally, denote the middle third of the intervals $\hat{I}_n^i$ by J_n^i. Since $f^{k_i(n)+s(n)+t(n)}$ maps $\bar{I}_n^i$ with bounded distortion onto U, there exists a universal constant $\delta > 0$ such that $r_{k_i(n)+s(n)+t(n)}(x) \geq \delta$ for $x \in J_n^i$. Moreover, there exists $\epsilon > 0$ such that $|\bigcup_i \bar{I}_n^i|/|\bar{Z}_n| \geq \epsilon$. Since this holds for each n, it follows that the lower density in x of the set $\{z\,;\ \limsup r_n(z) \geq \delta\}$ is bounded from below by ϵ. Because this is true for each x with $\omega(x) \ni c$, the ergodicity of f implies that $Z(\delta) = \{z\,;\ \limsup r_n(z) \geq \delta \text{ and } c \in \omega(z)\}$ has full Lebesgue measure.

Let us now show that Property 2 implies Property 3. So take symmetric neighbourhoods U and V of c with $U \subset \mathrm{cl}\,(U) \subset \mathrm{int}\,(V)$ as in Theorem 1.1. (We should note that we do not need the full strength of Theorem 1.1 here. Only the topological properties of U and V which are proved in Lemma 1.1 below are essential.) The set

$$Z(\delta) = \{x\,;\ \limsup_{n\to\infty} r_n(x) \geq \delta \text{ and } c \in \omega(x)\}$$

is completely invariant. By ergodicity, for each given δ this set therefore has either zero or full Lebesgue measure. From Property 2 and since $c \in \omega(x)$ for almost all x, it has full Lebesgue measure for some $\delta > 0$. Choose $U \subset V$ so that $|V| < \kappa(\delta/2)/2$. Take a point $x \in Z(\delta)$ and choose n so that $r_n(x) \geq \delta/2$. If $f^n(x) \in U$ then $f^n(T_n(x)) \supset V$ and we are done. If $f^n(x) \notin U$ then take $m > n$ minimal so that $f^m(x) \in U$. By Theorem 1.1,

$$(*) \qquad f^{m-n}(T_{m-n}(f^n(x))) \supset V.$$

If $f^n(R_n(x))$ contains a component of $T_{m-n}(f^n(x)) \setminus f^n(x)$ then $(*)$ implies that $f^m(R_m(x))$ contains a boundary point of V. The other possibility is that $f^n(R_n(x))$ is contained in a component of $T_{m-n}(f^n(x)) \setminus f^n(x)$. Then $f^m|R_n(x)$ is monotone and therefore $R_n(x) = R_m(x)$ and $|f^m(R_m(x))| \geq \kappa(\delta/2)$. Since $f^m(x) \in U$ and because of the size of V this again implies that $f^m(R_m(x))$ contains a boundary point of V. Since the same statements hold for $L_m(x)$, we get $f^m(T_m(x)) \supset V$ and therefore Property 3.

From Property 3 it easily follows as in the proof of Theorem 1.3 that $\omega(x) = [f^2(c), f(c)]$ for almost all x. Hence, Property 1 holds. □

The proof of Theorem 1.1

So let us start with the proof of Theorem 1.1. For each $x \in I$, let $U_x = (x, \tau(x))$ and

$$\mathcal{N} = \{x \in I\,;\ U_x \cap \operatorname{orbit}(x) = \emptyset\}$$

where $\operatorname{orbit}(x) = \cup_{n \geq 0} f^n(x)$. The points in $\mathcal{N}$ are called *nice*. Every periodic orbit contains a nice point, and therefore $\mathcal{N}$ is certainly not empty. Moreover, since the turning point is accumulated by periodic points, the set $\mathcal{N}$ accumulates on c. Clearly, $\mathcal{N}$ is also closed.

First we prove Theorem 1.1 in the case that c is not recurrent, i.e., when $\omega(c)$ does not contain c. In this case let $V \supset U$ be some fixed neighbourhoods of c such that $V \cap \operatorname{orbit}(c) = \{c\}$ and with $\partial U, \partial V \subset \mathcal{N}$. Take any point x with $c \in \omega(x)$. Then there exists a sequence $k(n) \to \infty$ such that $f^{k(n)}(x) \in U$. Since all critical values of $f^{k(n)}$ are contained in the forward orbit of c there exist intervals $I_n \subset J_n$ such that $f^{k(n)}$ maps J_n diffeomorphically onto V and such that $f^{k(n)}(I_n) = U$. By the Koebe Principle there exists a universal constant K (which only depends on the size of the components of $V \setminus U$) such that $|Df^{k(n)}(x)|/|Df^{k(n)}(y)| \leq K$ for all $x, y \in I_n$. Theorem 1.1 follows and in fact one even has that the sequence U_n does not depend on n.

So in the remainder of the proof of Theorem 1.1 we may assume that $\omega(c)$ is recurrent (and that f has no attracting periodic points). Fix $x \in \mathcal{N}$ and let $D_x = D_{U_x}$, i.e.,

$$D_x = \{y\,;\ \text{there exists } k > 0 \text{ with } f^k(y) \in U_x\}.$$

As before, let $k(y, x)$ be the minimal integer with this property and let $R_x(y) = f^{k(y,x)}(y)$. We call $k(y, x)$ the *transfer time* of y to U_x and $R_x\colon D_x \to U_x$ the *Poincaré map* to U_x. The transfer time k is constant on each component of D_x: if I is a maximal subinterval of D_x on which the transfer time is constant then, by maximality, some iterate of ∂I is mapped into ∂U_x. But since $x \in \mathcal{N}$ this implies that this boundary point of I never enters U_x and therefore $\partial I \cap D_x = \emptyset$ and $R_x(\partial I) \subset \partial U_x$. Hence I is a component of D_x. On each component I of D_x which does not contain c the map R_x is monotone and $R_x(I) = U_x$. Indeed, let $R_x = f^k$ on I; if $R_x|I$ is not monotone and I does not contain c, then

$f^i(I)$ contains c for some $0 < i < k$, but then k is not the first transfer time, a contradiction.

We want to show that this Poincaré map has a universally bounded distortion on each interval where it is defined. This is done by arranging it in such a way that on each component of D_x the Poincaré map R_x has a monotone extension to an interval T such that $R_x|T = f^k$ for some k and such that $f^k(T)$ contains a δ-scaled neighbourhood of U_x where $\delta > 0$ is a universal constant. By applying the Koebe Principle the bound on the distortion of R_x will follow.

Lemma 1.1. *Let $z \in \mathcal{N}$ and assume that c is recurrent. Let S_z be the component of D_z containing $c_1 = f(c)$. Next let x be the point (which we also denote by $\psi(z)$) so that $U_x = (x, \tau(x)) = f^{-1}(S_z)$ is the component of D_z containing c. Then one has the following properties:*

1. *x is contained in the closure of U_z, $f^{-1}(S_z) = U_x$ and $x \in \mathcal{N}$ (one even has that $f^i(x) \notin U_z$ for $i \geq 1$);*
2. *for each component I of D_x not contained in U_x there exists an interval $T \supset I$ such that f^k (where k is the transfer time on I) maps T monotonically onto U_z.*

Proof. Let $R_z = f^n$ on S_z and therefore $R_z = f^{n+1}$ on U_x. One has that x is contained in the closure of U_z because otherwise $z \in U_x$ and then $f^{n+1}(z) \in f^{n+1}(U_x) \subset U_z$, contradicting that z is nice. Moreover, since n is the transfer time to U_z on S_z, one has that $f^{k+1}(x) \in f^k(\partial S_z)$ is outside U_z for $k = 0, 1, \ldots, n-1$ and since $f^{n+1}(x) \in f^n(\partial S_z) \subset \partial U_z$ and $z \in \mathcal{N}$, this implies Statement 1. In order to prove Statement 2, let I be a component of D_x, $R_x = f^k$ on I and T the largest interval containing I such that f^k maps T monotonically into U_z. If $f^k(T)$ is not equal to U_z, then the maximality of T implies that there exists $0 \leq i < k$ such that one of the boundaries of $f^i(T)$ contains c. Since $f^i(I) \cap U_x = \emptyset$ this implies that $f^i(T)$ contains x. But then $f^k(T) \subset U_z$ contains $f^{k-i}(x)$, contradicting Statement 1. □

From Statement 2 of the previous lemma for each component I of D_x we get an interval $T \supset I$ such that if k is the return time on I to U_x (i.e., $R_x = f^k$ on I) then f^k maps T diffeomorphically onto U_z. If U_z is a δ-scaled neighbourhood of U_x then we can apply the Koebe Principle to get a bound on the distortion of R_x in terms of δ. In the next lemma we shall show that there exists a monotone extension of R_x to an even larger interval in some special cases.

and that both these situation will give some 'Koebe space'.

Lemma 1.2. *There exist $\delta, \rho > 0$ such that for any x, z, U_x and S_z as in the previous lemma one has the following properties.*

1. *Assume that $c \in R_z(U_x)$. If I is a component of D_x not intersecting U_x, $R_x|I = f^k$ and T is the maximal interval containing I for which $f^k|T$ is monotone then $f^k(T)$ is a δ-scaled neighbourhood of U_x.*

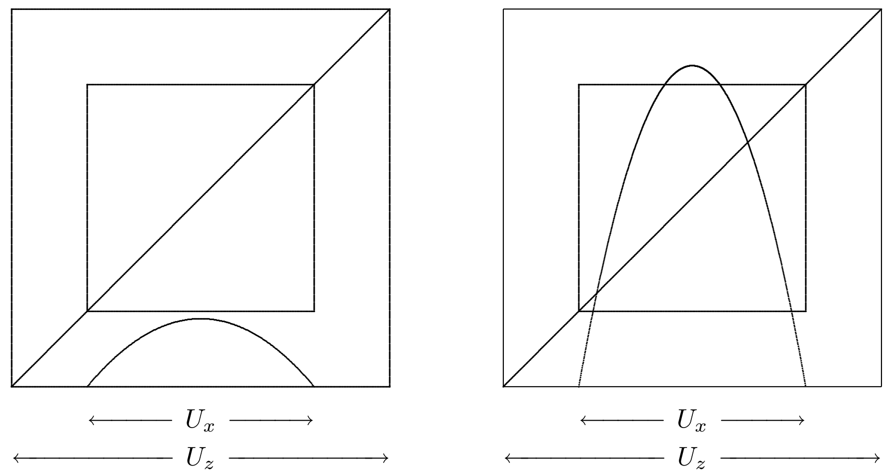

Fig. 1.2: On the left $R_z : U_x \to U_z$ is shown when $c \notin R_z(U_x)$ and $R_z(c) \notin U_x$. On the right when $c \in R_z(U_x)$. The basic idea of the proof of Theorem 1.1 is that one may always assume to be in one of these situations and that both these situation will give some 'Koebe space'.

2. *Assume that $c \notin R_z(U_x)$ and $|U_z| \leq (1+\rho)|U_x|$. If $R_z|S_z = f^n$ and $T \supset S_z$ is the maximal interval for which $f^n|T$ is monotone then $f^n(T)$ contains a δ-scaled neighbourhood of $[R_z(U_x), c]$.*

Proof. Let $R_z|S_z = f^n$ and $M = f(U_x)$ (M is contained in S_z). The intervals $M, \dots, f^n(M)$ are all pairwise disjoint. Indeed, otherwise there would be some $k < n$ such that $f^k(M) \cap f^n(M) \neq \emptyset$ and since $f^n(M) = U_z$ this would imply that $f^k(M) \cap U_z \neq \emptyset$ but then $f^{n-k}(z) \in U_z$ and one gets a contradiction with the assumption that z is nice. So let $m' \in \{0, 1, \dots, n\}$ be equal to the index i for which $|f^i(M)|$, $i = 0, \dots, n-1$ is minimal. Since for $i \geq 2$, $f^i(M)$ is between M and $f(M)$ we can take $m \in \{m', m'+1, m'+2\}$ such that both to the left and to the right of $f^m(M)$ at least one of the intervals $M, f(M), \dots, f^n(M)$ can be found. If $||Df||$ denotes the supremum of $|Df(x)|$ on I then one has $|f^m(M)| \leq ||Df||^2 \cdot |f^j(M)|$ for all $i \in \{0, 1, \dots, n\}$ and $f^m(M)$ has on both sides neighbours (by this we mean intervals of the form $f^i(M)$ with $j = 0, 1, \dots, n$). Letting $f^l(M)$ and $f^r(M)$ be the immediate neighbours of $f^m(M)$ one has therefore

$$|f^i(M)| \geq \frac{1}{||Df||^2} |f^m(M)| \text{ for } i = l, r. \tag{1.1}$$

Let H be the maximal interval containing M for which $f^m|H$ is monotone and such that $f^m(H) \subset [f^l(M), f^r(M)]$. We claim that

$$f^m(H) = [f^l(M), f^r(M)].$$

Indeed, assume H does not satisfy this property. Then let H^+ and H^- be the components of $H \setminus M$. By the maximality of H there exists $i \in \{0, 1, \dots, m-1\}$

with $c \in \partial f^i(H^-)$ and because $f^i(M) \cap U_x = \emptyset$ this implies that $f^i(H^-)$ contains a component of $U_x \setminus \{c\}$ and therefore that $f^m(H^-) \supset f^{m-i}(U_x) = f^{m-i-1}(M)$. Similarly $f^m(H^+)$ contains an iterate of M. Hence $f^m(H)$ contains the two neighbours (in the sense defined above) of $f^m(M)$, a contradiction. From (1.1) and the Macroscopic Koebe Principle we get a universal constant $\tilde{\delta} > 0$ (which only depends on $||Df||^2$) such that H contains a $\tilde{\delta}$-scaled neighbourhood of M. Because the critical point is non-flat there exists a universal constant $\tilde{\delta}' > 0$ such that for the preimage $H' = f^{-1}(H)$ and $U_x = f^{-1}(M)$,

$$H' \text{ contains a } \tilde{\delta}'\text{-scaled neighbourhood of } U_x. \tag{1.2}$$

Proof of Statement 1: Assume that $c \in f^n(M) = f^{n+1}(U_x) = R_z(U_x)$. Take a component I of D_x which is not contained in U_x, let k be the transfer time to U_x on I and let T be the maximal interval containing I for which $f^k|T$ is monotone and $f^k(T) \subset H'$. We claim that $f^k(T) = H'$. Indeed, let L and R be the components of $T \setminus I$. From the maximality of T there exists $0 \le i < k$ such that $f^i(L)$ contains c in its boundary. But $f^i(I) \cap U_x = \emptyset$ and therefore $f^i(L)$ contains one component of $U_x \setminus \{c\}$. Consequently, $f^{i+1}(L)$ contains M and $f^k(L)$ contains $f^{k-i-1}(M)$. Now suppose by contradiction that the closure of $f^k(L)$ is contained in H'. Then, because $f^{k-i-1}(M) \subset f^k(L)$ and because f^{m+1} restricted to each component of $H' \setminus \{c\}$ is monotone, this implies that $f^{m+1}|f^{k-i-1}(M)$ is monotone. Since $f^{k-i-1}|M$ is monotone this gives that $f^{k+m-i}|M$ is monotone. In particular, because $c \in f^n(M)$ this implies that $k+m-i \le n$. Furthermore, $f^{k+m-i}(M) \subset f^{k+m+1}(L) \subset [f^l(M), f^r(M)]$. But then $f^{k+m+1}(L)$ contains a neighbour of $f^m(M)$. Similarly, $f^{k+m+1}(R)$ contains a neighbour of $f^m(M)$. But this is only possible if $f^{k+1}(L)$ and $f^{k+1}(R)$ both contain a component of $H \setminus M$. Thus, we have proved that $f^k(T) \supset H'$. Because of (1.2), this completes the proof of Statement 1 of this lemma.

Proof of Statement 2: Assume that $c \notin f^n(M)$. Let $T \supset S_z$ be the maximal interval for which $f^n|T$ is monotone and let L and R be the components of $T \setminus M$. Since $T \supset S_z$ one has $f^n(T) \supset U_z$. Since $c \notin f^n(M)$ for one of the components of $T \setminus M$, which we denote by R, the interval $f^n(R)$ contains c. Therefore $f^n(R)$ contains a component of $U_z \setminus \{c\}$ and $f^n(M)$ is contained in the other component. Hence there exists a constant $\hat{\delta} > 0$ with

$$|f^n(R)| \ge \hat{\delta}|U_z|. \tag{1.3}$$

(Indeed, when z is sufficiently close to c then we can take $\hat{\delta}$ close to 1 because c is non-flat and therefore the two components of $U_z \setminus \{c\}$ have roughly the same length.) Let us show that $f^n(L)$ contains a component of $H' \setminus U_z$. So assume by contradiction that the closure of $f^n(L)$ is strictly contained in the interior of $H' \setminus U_z$. As before there exists $i \in \{0, 1, \ldots, n-1\}$ such that $f^i(L)$ contains c in its boundary. Because $f^i(M) \cap U_z = \emptyset$, this implies that $f^i(L)$ contains U_z, $f^{i+1}(L) \supset M$ and hence $f^n(L) \supset f^{n-i-1}(M)$. Since $f^n(M \cup L)$ lies on one side of c, the intervals $f^{n-i-1}(M)$ and $f^n(M)$ are both on the same side of c. Since, by assumption, $f^n(L) \subset H'$ one has $f^n(L) \subset [f^{n-i-1}(M), c]$ and we have assumed that the closure of $f^n(L)$ is contained in a component of the interior of $H' \setminus \{c\}$, the map f^{m+1} is monotone on these intervals. Hence the closure of $f^{m+1+n-i-1}(M)$ is contained in the interior of $[f^l(M), f^r(M)]$. Since $f^l(M)$ and $f^r(M)$ are the nearest neighbours (in the sense defined above) of $f^m(M)$ this implies $m+1+n-i-1 > n$ and therefore $m > i$. In particular, since, by assumption, $[f^{n-i-1}(M), c] \subset H'$, since f^{m+1} is monotone on the components of $H' \setminus \{c\}$ and since $i < m$, we get

$$f^{i+1} \text{ maps } [f^{n-i-1}(M), c] \text{ monotonically into } [f^i(M), f^n(M)] \tag{1.4}$$

and the last interval does not contain c. So we can conclude that $f^{n-i-1}(M)$, $f^n(M)$ and $f^i(M)$ are all on one side of c. Since $f^i(M) \cap U_z = \emptyset$ and $f^n(M) \subset U_z$ one has

$$[f^i(M), c] \supset [f^n(M), c]. \tag{1.5}$$

Furthermore, f^{n-i} maps $[f^i(M), c]$ (which is contained in $f^i(L \cup M)$) monotonically onto the interval $[f^n(M), f^{n-i-1}(M)]$. Hence

$$[f^n(M), f^{n-i-1}(M)] \not\subset [f^i(M), c]$$

because otherwise f has a periodic attractor, contradicting our assumption. Therefore, because all these intervals are on the same side of c and from (1.5) one gets

$$f^i(M) \subset (f^{n-i-1}(M), f^n(M)). \tag{1.6}$$

Because of (1.4) and (1.6), the map f^{i+1} has an attracting fixed point, a contradiction. Thus we have shown by contradiction that $f^n(L)$ contains a component of $H' \setminus U_z$. Since $|U_z| \le (1+\rho)|U_x|$ combining this with (1.2) gives that the length of $f^n(L)$ is at least $(\tilde{\delta}'/2)|U_z|$ provided ρ is small enough. Together with (1.3) this implies that $f^n(T)$ contains a δ-scaled neighbourhood of $R_z(U_x)$ when $\delta > 0$ is sufficiently small. □

Corollary 1.2. *There exists a constant $\rho > 0$ (not depending on x) such that if $c \notin R_z(U_x)$ and $R_z(c) \notin U_x$ then U_z contains a ρ-scaled neighbourhood of U_x.*

Proof. Let ρ be as in the previous lemma and let ρ' be so that $|U_z| = (1+\rho')|U_x|$. If $\rho' \ge \rho$ then we are finished. So assume that $\rho' < \rho$ and let us show that ρ' cannot be too small. Let $R_z = f^{n+1}$ on U_x. Note that the assumption is equivalent to

$$f^n(M) = f^{n+1}(U_x) \subset U_z \setminus U_x.$$

Let $T \supset S_z$ be the maximal interval on which f^n is monotone. From the previous lemma $f^n(T)$ contains a δ-scaled neighbourhood of $[R_z(U_x), c]$ (this is one component of $U_z \setminus \{c\}$). So take $T' \supset M$ so that $f^n(T')$ contains a $\delta/2$-scaled neighbourhood of $[R_z(U_x), c]$. Then $f^n|T'$ has bounded distortion by the Koebe Principle. Hence, if we define $U = f^{-1}(T')$ then $f^{n+1}(U)$ contains an interval of size $(\delta/2)|U_x|$ and $f^{n+1}|U$ is polynomial-like in the sense define below Theorem 1.1. Since $f^{n+1}(U_x) \subset U_z \setminus U_x$, $|U_z \setminus U_x| \le \rho'|U_x|$ and since $f^{n+1}|U$ is polynomial-like it follows that the derivative of f^{n+1} is at most a universal constant times ρ' on U_x. Since $f^{n+1}|U$ is polynomial-like and $|f^{n+1}(U)| \ge (\delta/2)|U_x|$ the interval U is much larger than U_x when ρ' is small. In particular, there exists a neighbourhood $\hat{U} \subset U$ of U_x of size $2|U_x|$. On such an interval $\hat{U}$ the derivative of f^{n+1} is also at a universal constant K times ρ'. Hence

$$(*) \qquad |f^{n+1}(\hat{U})| \le K \cdot \rho' \cdot |\hat{U}|.$$

Since $f^{n+1}(U_x) \subset U_z$, the interval U_z is contained in a ρ'-scaled neighbourhood of U_x and $\hat{U}$ is a 1-scaled neighbourhood of U_x, inequality $(*)$ implies that $f^{n+1}(\hat{U}) \subset \hat{U}$ when ρ' is very small. So this would imply that f has an attracting periodic orbit, a contradiction. So ρ' cannot be too small. □

Proof of Theorem 1.1: First we are going to define the sequence of closest approach to c. Let $c_n = f^n(c)$, and

$$q(1) = 1 \text{ and } q(n+1) = \min\{i \in \mathbb{N}\,;\ f^i(c) \in U_{c_{q(n)}}\}.$$

Because $c \in \omega(c)$ and c is not periodic the sequence $q(1), q(2), \dots$ is well defined. Since c is an accumulation point of $\mathcal{N}$ there are infinitely many $n > 0$ for which $(U_{c_{q(n-1)}} \setminus U_{c_{q(n)}}) \cap \mathcal{N} \neq \emptyset$. For those $n \in \mathbb{N}$, let

$$z(n) = \sup\{y < c\,;\ y \in \mathcal{N} \cap (U_{c_{q(n-1)}} \setminus U_{c_{q(n)}})\}$$

(because $\mathcal{N}$ is symmetric this intersection is non-empty) and let

$$x(n) = \psi(z(n))$$

where ψ is as in Lemma 1.1.

Because $z(n) \in U_{c_{q(n-1)}} \setminus U_{c_{q(n)}}$ one has that $f^i(c) \notin U_{z(n)}$ for $i < q(n)$ and $c_{q(n)} \in U_{z(n)}$. In particular, $R_{z(n)} = f^{q(n)}$ on the component $U_{x(n)}$. We distinguish two cases.

Case 1: $c \in R_{z(n)}(U_{x(n)})$. In this case we can apply Statement 1 of Lemma 1.2. Therefore, for each component I of $D_{x(n)}$ there exists an interval $T \supset I$ such that if $R_x|I = f^k$ then $f^k|T$ is monotone and $f^k(T)$ contains a definite neighbourhood of $U_{x(n)}$. (Note that we did not need to use the special definition of $z(n)$ in this case.)

Case 2: $c \notin R_{z(n)}(U_{x(n)})$. This case is more complicated. First of all, if $x(n) = z(n)$ then $R_{z(n)}$ maps $U_{x(n)}$ into $U_{z(n)}$, i.e., into itself. Moreover, the map $R_{z(n)} \colon U_{z(n)} \to U_{z(n)}$ sends the boundary points of $U_{z(n)}$ into itself. Therefore, this map is unimodal. Since $c \notin R_{z(n)}(U_{x(n)})$ this unimodal map has an attracting fixed point, contradicting our assumption. So we may assume that $x(n) \neq z(n)$. Now by the definition of $z(n)$ there exists no $z' \in \mathcal{N}$ such that $z' \in (U_{z(n)} \setminus U_{c_{q(n)}})$. In particular, since $x(n) \in U_{z(n)} \cap \mathcal{N}$ one has $x(n) \in U_{c_{q(n)}}$, i.e.,

$$R_{z(n)}(c) = c_{q(n)} \in U_{z(n)} \setminus U_{x(n)}.$$

Hence we can apply the Corollary to Lemma 1.2 and consequently $U_{z(n)}$ contains a ρ-scaled neighbourhood of $U_{x(n)}$, where $\rho > 0$ is a universal constant. Moreover, for any component I of $D_{x(n)}$ not intersecting $U_{x(n)}$ there exists an interval $T \supset I$ such that $f^k|T$ is monotone (where k is the transfer time to U_x on I, i.e., $R_x|I = f^k$) and $f^k(T) = U_{z(n)}$. Since $U_{z(n)}$ contains a ρ-scaled neighbourhood of $U_{x(n)}$, the theorem follows. □

The Lebesgue measure of the ω-limit of the critical point

Let us now prove the last part of Theorem 1.3:

Theorem 1.6. **(Martens)** *Let $f\colon [-1,1] \to [-1,1]$ be a unimodal map with a non-flat critical point c and negative Schwarzian derivative. If $\omega(c)$ contains no intervals then*

$$|\omega(c)| = 0.$$

Proof. Let us prove that $\omega(c)$ has Lebesgue measure zero if $\omega(c)$ contains no intervals. If f is infinitely often renormalizable then $f|\omega(c)$ is injective. Therefore for each $x \in \mathcal{N}$, one has that $|\omega(c) \cap U_x|$ is either less than $|U_x^+|$ or less than $|U_x^-|$, where $U_x^{\pm}$ are the components of $U_x \setminus \{c\}$. Hence the limit superior of $|\omega(c) \cap U_x|/|U_x|$ is at most to 1/2 as x tends to c because c is non-flat. This contradicts the statement made in the proof of Theorem 1.2.

So let us deal with the situation when f cannot be renormalized. If $\omega(c)$ is not minimal then we have already shown in the proof of Theorem 1.3 that each closed forward invariant set which is not equal to the union of transitive intervals has Lebesgue measure zero. So we will assume that $\omega(c)$ is minimal. Let $z(n)$ be the sequence from the proof of Theorem 1.1 and as before let $x(n) = \psi(z(n))$. Let us define related sequences $v(n) \in \mathcal{N}$ and $u(n) = \psi(v(n))$ such that

(∗) $\qquad U_{v(n)}$ contains a ρ-scaled neighbourhood of $U_{u(n)}$.

We distinguish three cases.

1. If $c \notin R_{z(n)}(U_{x(n)})$ then let $v(n) = z(n)$ and $u(n) = x(n)$. As we saw in the proof of the previous theorem, in this case one has $R_{v(n)}(c) \in U_{v(n)} \setminus U_{u(n)}$ and the Corollary to Lemma 1.2 implies that (∗) holds.
2. On the other hand if $c \in R_{z(n)}(U_{x(n)})$ and $U_{z(n)}$ contains a ρ-scaled neighbourhood of $U_{x(n)}$ then we take again $v(n) = z(n)$ and $u(n) = x(n)$ and again (∗) holds.
3. Finally if $c \in R_{z(n)}(U_{x(n)})$ and $U_{z(n)}$ does not contain a ρ-scaled neighbourhood of $U_{x(n)}$ then we do the following. $R_{z(n)}\colon U_{x(n)} \to U_{z(n)}$ has one orientation preserving fixed point p. Because f is not renormalizable $R_{z(n)}(U_p)$ strictly contains U_p. Hence there exists a point $q \in U_p$ such that $R_{z(n)}(q) = \tau(p)$. Since $R_{z(n)}\colon U_{x(n)} \to U_{z(n)}$ is polynomial-like and $|U_{z(n)}|$ is not much bigger than $|U_{x(n)}|$, this implies that U_p contains a ρ-scaled neighbourhood of U_q where ρ

is a universal constant. So defining $v(n) = p$ and $u(n) = q = \psi(p)$ we get again that $(*)$ holds.

So in any case $(*)$ is satisfied. Now we get to the main idea of the proof. Because $\omega(c)$ is minimal, the forward orbit of each point $w \in \omega(c)$ accumulates onto c, and so for each $u(n) \in \mathcal{N}$, the set $\omega(c)$ is contained in $D_{u(n)}$. Because $\omega(c)$ is compact, there are only finitely many components of $D_{u(n)}$ which intersect $\omega(c)$. Let I be a component of $D_{u(n)}$ with the largest transfer time k to $U_{u(n)}$. Furthermore, let T be the interval containing I such that $f^k|T$ is monotone and $f^k(T) = U_{v(n)}$ (this interval exists by the second statement of Lemma 1.1). We claim that $T \setminus I$ contains no points from $\omega(c)$. Indeed, because of the maximality of k, there exists otherwise a point r in $T \setminus I$ which returns to $U_{u(n)}$ in time $k' < k$. But since $f^{k'}(I) \cap U_{u(n)} = \emptyset$ and $f^{k'}(r) \in U_{u(n)}$ there exists $s \in T$ with $f^{k'}(s) = u(n)$. But then $f^k(s) = f^{k-k'}(u(n)) \notin U_{v(n)}$ since $u(n) = \psi(v(n))$ is nice, see Statement 1 of Lemma 1.1, contradicting that $f^k(T) \subset U_{v(n)}$.

Let us now show that the fact that $T \setminus I$ contains points of ω (which because of $(*)$ means that $\omega(c)$ is not 'too thick' in T) implies that c cannot be a density point of $\omega(c)$. So let i be the smallest integer such that $f^i(c_1) \in T$. Because $I \subset T$ contains points of $\omega(c)$ such an integer exists. By the minimality of i, there exists an interval $S_n \ni c_1$ such that f^i maps S_n diffeomorphically onto T and f^{i+k} maps S_n diffeomorphically onto $U_{v(n)}$. So let $M_n \subset S_n$ be so that $f^i(M_n) = I$, i.e., such that $f^{i+k}(M_n) = U_{u(n)}$. Since $T \setminus I$ contains no points of $\omega(c)$ and $\omega(c)$ is forward invariant,

$$T_n \setminus M_n \text{ contains no points of } \omega(c).$$

Since $U_{v(n)}$ contains a ρ-scaled neighbourhood of $U_{u(n)}$ we get from the Macroscopic Koebe Principle that

$$T_n \text{ contains a } \delta\text{-scaled neighbourhood of } M_n$$

for some universal constant $\delta > 0$. From these statements one gets that the symmetric neighbourhood $U_n = f^{-1}(T_n)$ of c contains a δ'-scaled neighbourhood of $V_n = f^{-1}(M_n)$ and that $\omega(c) \cap U_n \subset V_n$. From Exercise 1.2 it follows that $\omega(c)$ has Lebesgue measure zero. $\square$

The existence of Bowen-Ruelle-Sinai measures

The next results deal with Bowen-Ruelle-Sinai measures of unimodal maps. Before defining this notion let us remind the reader of Birkhoff's Ergodic Theorem. This theorem deals with an *ergodic invariant measure* μ of $f\colon N \to N$. This is an invariant measure such that for any measurable set A for which $f^{-1}(A) = A$ either $\mu(A) = 0$ or $\mu(N \setminus A) = 0$. Birkhoff's Ergodic Theorem states that for any continuous function ϕ and for any ergodic invariant probability measure μ one has

$$(*) \qquad \lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \phi(f^i(x)) = \int \phi \, d\mu$$

for μ-almost all x. This means that if we want to compute the average value of the function ϕ along the orbit of a μ-typical point x then this is equal to the μ-average of ϕ (over the whole the space). Now this theorem is only useful if the support of the measure μ is rather large or if that is not the case $(*)$ holds for at least a large set of points x. Therefore one defines the following notion.

Definition. We say that μ is a *Bowen-Ruelle-Sinai measure*, or a *B.R.S. measure* if for each continuous function ϕ the set of points x for which $(*)$ holds has positive Lebesgue measure. For this it is enough that the sequence of measures

$$\nu_n = \frac{1}{n}\sum_{i=0}^{n-1} \delta_{f^i(x)}$$

tends weakly to μ.

In the following corollary of the previous result it is shown that there exists at most one Bowen-Ruelle-Sinai measure and that each absolutely continuous invariant probability measure is such a measure. As before, we say that an invariant measure μ is *absolutely continuous with respect to the Lebesgue measure* or simply *absolutely continuous* if for each measurable set A with zero Lebesgue measure one has $\mu(A) = 0$. This measure is *equivalent* to the Lebesgue measure if $\mu(A) = 0$ if and only if the Lebesgue measure of A is zero. By the Radon-Nikodym Theorem this implies that there exists a L^1 function p for which

$$\mu(A) = \int_A p(t)\, d\lambda(t)$$

where λ is the Lebesgue measure. The next theorem states that each absolutely continuous invariant measure is a Bowen-Ruelle-Sinai measure and also that it is equivalent to the Lebesgue measure on a finite union of intervals.

In Section 5 of this chapter we will use this theorem to show that there exist quadratic maps which are transitive on certain intervals but which do not have finite absolutely continuous invariant measures.

Theorem 1.7. *Let $f\colon [-1,1] \to [-1,1]$ be as before. Then f has at most one Bowen-Ruelle-Sinai measure. Furthermore, if f has an absolutely continuous invariant probability measure then*

a. *this absolutely continuous measure is a Bowen-Ruelle-Sinai measure and $(*)$ holds for Lebesgue almost all x;*

b. *f is only finitely often renormalizable and has no periodic attractors;*

c. *f has no absorbing Cantor attractor: the attractor A of f is the finite union of the transitive intervals;*

d. *the support of this measure is equal to A and in particular the measure is equivalent (in the sense defined above) to the Lebesgue measure restricted to this finite union of intervals.*

Remark. Let $r_i(x)$ be defined as above the statement of Theorem 1.4. Then $\limsup r_n(x) > 0$ for Lebesgue almost all x if f has an absolutely continuous invariant probability measure. This follows from Statement 1.4 and Statement c) of Theorem 1.5. In Theorem V.3.2 we shall see that f has an absolutely continuous invariant probability measure if and only if $\limsup_n \frac{1}{n}\sum_{i=0}^{n-1} r_i(x) > 0$ for almost all x. Moreover, that theorem implies that the density of such a measure is bounded from below on the set A.

Proof. Let μ be an f-invariant probability measure. Let B_μ be the set of points $x \in I$ such that $(*)$ holds for all continuous functions ϕ. The set B_μ is clearly totally invariant. Therefore, from the ergodicity of f this set has either zero or full Lebesgue measure. So assume that μ and $\bar{\mu}$ are two Bowen-Ruelle-Sinai measures (which are not necessarily absolutely continuous). Let us show that $\mu = \bar{\mu}$. Since μ and $\bar{\mu}$ are Bowen-Ruelle-Sinai measures, B_μ and $B_{\bar{\mu}}$ both have non zero and therefore full Lebesgue measure. Hence $B_{\bar{\mu}} \cap B_\mu$ has full measure and $(*)$ implies that $\int \phi \, d\mu = \int \phi \, d\bar{\mu}$ for all continuous functions ϕ. It follows that the two measures coincide.

Let us next show that if μ is absolutely continuous that then B_μ has full Lebesgue measure (and therefore μ is a Bowen-Ruelle-Sinai measure). In fact, let $\phi_n \in C^0(I, \mathbb{R})$ be a sequence which is dense in $C^0(I, \mathbb{R})$. By Birkhoff's Ergodic Theorem, see the Appendix, there exists a set B_n of μ-measure one, such that $(*)$ holds for each $x \in B_n$ and $\phi = \phi_n$. Taking $B = \cap B_n$ and using that any continuous function ϕ is the uniform limit of some subsequence of ϕ_n we get that $(*)$ holds for ϕ and any $x \in B$ and this proves the claim.

Since μ is absolutely continuous, it follows that B_μ has positive Lebesgue measure. From Theorem 1.2 we get that B_μ has full Lebesgue measure.

If f is infinitely renormalizable or has a periodic attractor then for almost all points x one has $\omega(x) \subset \omega(c)$ and $\omega(c)$ is a Cantor set. So in this case, any Bowen-Ruelle-Sinai measure must have its support in $\omega(c)$. But by Theorem 1.3, $\omega(c)$ has zero Lebesgue measure if it is a Cantor set and so f has no absolutely continuous invariant probability measure in this case.

So let us show now that if f has an absolutely continuous invariant measure μ then the support of μ is equal to a finite union of transitive intervals. By Theorem 1.3, if the attractor A of f is a Cantor set then for almost all x one has $\omega(x) \subset \omega(c)$ and $\omega(c)$ is a finite set or a Cantor set of Lebesgue measure zero. Therefore the support K of this measure is contained in $\omega(c)$ and consequently has Lebesgue measure zero, a contradiction because K is the support of an absolutely continuous measure. So A is a union of intervals and therefore f has no absorbing Cantor attractor. So it remains to show that the support K of the measure μ is equal to A. Indeed, $\omega(x) = A$ for almost all x. In particular, $\omega(x) = A$ for Lebesgue almost all $x \in K$. Moreover, by Birkhoff's Theorem one has $\omega(x) = K$ for almost all $x \in K$. Combining this gives $A = K$. □

In the previous theorem we have observed that an absolutely continuous invariant probability measure for a map $f\colon [-1,1] \to [-1,1]$ is a Bowen-Ruelle-

Sinai measure. Let us now show that even if one has no absolutely continuous invariant measure one may still have a Bowen-Ruelle-Sinai measure. If, for example, f satisfies the Axiom A it also has a Bowen-Ruelle-Sinai measure: the invariant probability measure supported on an attracting periodic orbit is such a measure. (If this map has negative Schwarzian derivative and a periodic attractor then $(*)$ holds again for almost all x.) Similarly, as we will show now there exists a Bowen-Ruelle-Sinai measure in the infinite renormalizable case:

Theorem 1.8. *Let $f\colon [-1,1] \to [-1,1]$ be a C^2 unimodal map with a non-flat critical point. If f is infinitely renormalizable then f has a unique invariant probability measure μ. It is a Bowen-Ruelle-Sinai measure which is supported in the closure of the forward orbit of the critical point. This set is an attracting Cantor set Γ and if x is in the basin $B(\Gamma)$ of Γ then*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \phi(f^i(x)) = \int \phi \, d\mu$$

for every continuous function $\phi \in C^0([-1,1],\mathbb{R})$.

Proof. Since f is infinitely renormalizable, there exists a decreasing sequence of symmetric intervals around the critical point c, $I_1^0 \supset I_2^0 \supset I_3^0 \dots$ and an increasing sequence of integers $n(1) < n(2) < n(3) \dots$ such that $f^{n(j)}$ maps I_j^0 in a unimodal way into itself. Let $F_j = \cup_{i=0}^{n(j)-1} f^i(I_j^0)$. From Corollary 2 of Theorem AB of Chapter IV, the intersection

$$\Gamma = \cap_{j=1}^{\infty} F_j$$

is a forward invariant Cantor set and its basin has positive Lebesgue measure. (If all the periodic points of f are hyperbolic and f has no attracting periodic orbits then $B(\Gamma)$ has full Lebesgue measure.) Let $x_j \in I_j$ and let μ_j be the probability measure

$$\mu_j = \frac{1}{n(j)} \sum_{i=0}^{n(j-1)} \delta_{f^i(x_j)}$$

where δ_x denotes the Dirac measure at x. The μ_j-measure of each component of F_j is equal to $\frac{1}{n(j)}$ because each such component contains one and only one point of the set $\{x_j, f(x_j), \dots, f^{n(j)-1}(x_j)\}$. It follows that for each open interval J, the measures $\mu_j(J)$ are all the same for j sufficiently large (if $k > j$ then any two components of F_j contain the same number of components of F_k and therefore any two components of F_j have the same μ_k-measure). Hence μ_j converges weakly to an invariant measure μ.

Let $x \in B(\Gamma)$ and let ν_k be the probability measure

$$\nu_k = \frac{1}{k} \sum_{i=0}^{k-1} \delta_{f^i(x)}.$$

As we have proved in Theorem III.5.1 and its corollary, for each j, $f^i(x) \in I_j^0$ for some sufficiently large i. It follows as before that for each component I of F_j, $\lim_{k\to\infty} \nu_k(I) = \frac{1}{n(j)}$. This proves the theorem. □

2 Invariant Measures for Markov Maps

Let us start with a very simple and general remark. If μ is an invariant probability measure of a map $f\colon N \to N$, then the support of μ is contained in the set of non-wandering points of f. Here, as before, a point x is non-wandering if for each neighbourhood U of x there exists $n > 0$ with $f^n(U) \cap U \neq \emptyset$. In fact, if x is a wandering point and V is a neighbourhood of x such that $f^n(V) \cap V = \emptyset$ for all $n \in \mathbb{N}$ then $f^{-(n+m)}(V) \cap f^{-n}(V) = \emptyset$ for all $n, m > 0$, because, otherwise, $V \cap f^{n+m}(f^{-n}(V)) = V \cap f^m(V) \neq \emptyset$. By invariance, $\mu(f^{-n}(V)) = \mu(V)$ and, therefore, since all the sets $f^{-n}(V)$ are mutually disjoint, $\sum_{n=0}^{\infty} \mu(f^{-n}(V)) \leq 1$ implies that $\mu(V) = 0$. It follows from this remark that if the non-wandering set of f has zero Lebesgue measure then f cannot have an absolutely continuous invariant probability measure. In particular, from Theorem III.2.3, we get that if a $C^{1+\alpha}$ map $f\colon N \to N$ satisfies the Axiom A and has an absolutely continuous invariant probability measure then f is an expanding map of the circle. Therefore, no C^2 interval map satisfying the Axiom A has an absolutely continuous invariant measure and in this sense it is not chaotic.

The simplest example of a one-dimensional smooth dynamical system possessing an absolutely continuous invariant probability measure is the quadratic map $f\colon [-1,1] \to [-1,1]$ defined by $f(x) = 1 - 2x^2$ which was first considered by von Neuman and Ulam (1947). In fact, as we have seen in Section II.3 the homeomorphism $h\colon [-1,1] \to [-1,1]$ defined by $h(x) = \dfrac{2}{\pi}\sin^{-1}x$ is a conjugacy between f and the tent map $T(x) = 1 - 2|x|$. Since the Lebesgue measure λ is T-invariant, the push-forward measure $h^*\lambda$, defined by $h^*\lambda(A) = \lambda(h(A))$ is f-invariant. It is clear that $h^*\lambda$ is an absolutely continuous invariant measure with density $\phi(x) = \dfrac{2}{\pi}\dfrac{1}{\sqrt{1-x^2}}$.

As we will see in the later sections of this chapter in general smooth interval maps do not have nice invariant measures. There are several ways to prove the existence of invariant measures. Many of these use the Perron-Frobenius operator. In this section we will illustrate this operator for special interval maps $f\colon [-1,1] \to [-1,1]$. These maps will be called Markov maps, because for each such map there is an associated partition of the interval $[-1,1]$ into a collection of intervals I_i (these are the states of the 'Markov process') such that f is expanding on each of these intervals and such that the closure of the image of f of one of the intervals is equal to the closure of some union of these intervals (so from each state one can have some 'proper' transitions to other states). In this section we will show that such maps have absolutely continuous measures with nice properties. In the next section we will show how to obtain such Markov maps from certain smooth interval maps. In Section 3 we will give a somewhat different approach to obtain invariant measures.

The first important result on the existence of an absolutely continuous invariant measure is now considered to be a folklore theorem which originated with the following basic result due to Renyi (1957).

Theorem 2.1. **(Folklore Theorem)** *Let $f\colon M \to M$ be a C^1 expanding map of a compact (n-dimensional) manifold M, whose derivative is Hölder continuous (f is $C^{1+\alpha}$). Then f has an absolutely continuous invariant probability measure μ. Furthermore, μ is ergodic, its density is bounded and bounded away from zero and $\mu(A) = \lim_{n\to\infty} \lambda(f^{-n}(A))$ for each measurable set A.*

It is not very hard to prove that if a C^2 covering map $f\colon S^1 \to S^1$ has derivative bigger than one at every point except at the fixed point p where the derivative is equal to one, then f has no finite absolutely continuous invariant measure, see Bowen (1979). In the same spirit one can prove the existence of an absolutely continuous invariant measure for a class of expanding map of N (an interval or a circle) which are not continuous but have some Markov-like properties (they send intervals of continuity onto unions of such intervals).

Definition. A C^1 map $f\colon N \to N$ is called *Markov* if there exists a finite or countable family I_i of disjoint open intervals in N such that

a) $N \setminus \cup_i I_i$ has Lebesgue measure zero and there exist $C > 0$ and $\gamma > 0$ such that for each $n \in \mathbb{N}$ and each interval I such that $f^j(I)$ is contained in one of the intervals I_i for each $j = 0, 1, \dots, n$ one has

$$\left|\frac{Df^n(x)}{Df^n(y)} - 1\right| \le C \cdot |f^n(x) - f^n(y)|^\gamma \text{ for all } x, y \in I.$$

b) if $f(I_k) \cap I_j \ne \emptyset$ then $f(I_k) \supset I_j$;

c) there exists $r > 0$ such that $|f(I_i)| \ge r$ for each i.

Remark. 1. As we shall see in the next section, Assumption a) often follows immediately from the Koebe Principle if the map has negative Schwarzian derivative. Indeed, if there exists an interval T for each of the intervals I from Assumption a) such that f^n can be extended diffeomorphically to T and $f^n(T)$ is some uniformly scaled neighbourhood of $f^n(I)$ then we can apply the last part of the Koebe Principle, see Theorem IV.1.1, and get Assumption a) with $\gamma = 1$. We shall come back to this in the next section.

2. Instead of Assumption a) one can require that the following two conditions are met
2.a. there exists $\gamma, C > 0$ such that $f|I_i$ is a $C^{1+\gamma}$ diffeomorphism for each i and that for each i and each $x, y \in I_i$, the following Hölder condition holds

$$\left|\frac{Df(x)}{Df(y)} - 1\right| \le C \cdot |f(x) - f(y)|^\gamma$$

(so we only require Assumption a) from above for $n = 1$); 2.b. it is expanding: there exist $K > 0$ and $\beta > 1$ such that $|Df^n(x)| \ge K \cdot \beta^n$ for any $n \in \mathbb{N}$ and $x \in N$ for which $f^j(x) \in \cup_i I_i$ for all $0 \le j \le n$.

Let us show that these two conditions imply Assumption a). Take $x, y \in I$ as in Assumption a). By the Chain Rule and the above assumptions

$$\frac{|Df^n(x)|}{|Df^n(y)|} = \prod_{m=0}^{n-1} \frac{|Df(f^m(x))|}{|Df(f^m(y))|} \leq \prod_{m=0}^{n-1} \left(1 + C \cdot |f^{m+1}(x) - f^{m+1}(y)|^\gamma\right)$$

and

$$|f^n(x) - f^n(y)| \geq K\beta^{n-m-1}|f^{m+1}(x) - f^{m+1}(y)|.$$

Therefore,

$$\begin{aligned}\frac{|Df^n(x)|}{|Df^n(y)|} &\leq \prod_{m=0}^{n-1} \left\{1 + C \cdot \left(\frac{|f^n(x) - f^n(y)|}{K}\beta^{m+1-n}\right)^\gamma\right\} \\ &\leq \prod_{m=0}^{\infty} \left\{1 + C \cdot \left(\frac{|f^n(x) - f^n(y)|}{K}\beta^{-m}\right)^\gamma\right\} < \infty.\end{aligned}$$

Assumption a) follows from this and by interchanging the role of x and y.

3. Sometimes it is more natural to require instead of Assumption a) that

$$\frac{|D^2f(y)|}{|Df(z)|^2} \leq C$$

for each I_i and each $y, z \in I_i$.

4. Assumption b) prohibits that certain intervals are only mapped partly over some other intervals and is the main reason such a map is called Markov. Assumption c) prohibits that images of the intervals I_i can be too small. This last Assumption can be somewhat weakened but it is certainly not possible to dispense with it altogether if we want to have that Markov maps necessarily have an absolutely continuous invariant probability measure, see for example Lasota and Yorke (1973) and Blank (1991). Often we shall assume that one has an additional Assumption which guarantees that one can get from any interval to any other interval in a finite number of steps: see the theorem below. This additional assumption implies that the system is 'transitive'. Of course, the Hölder condition as in Remark 2 above also follows from the expanding condition and $\left|\frac{Df(x)}{Df(y)} - 1\right| \leq C' \cdot |x-y|^\gamma$ (and, therefore, holds if for example $\log|Df|$ is Lipschitz); however, the condition in Remark 2 sometimes holds even when this Hölder condition does not:

Exercise 2.1. Consider the Gauss map $G\colon (0,1) \to (0,1)$ defined by $G(x) = \frac{1}{x} - \left[\frac{1}{x}\right]$. Show that this map has a finite absolutely continuous invariant measure $d\mu = \frac{1}{\log 2}\frac{dx}{1+x}$. Show that G also satisfies Properties b)-c) of the definition Markov maps. Show that it does not satisfy the following Hölder Property: there exists $\gamma > 0$,

$$\left|\frac{DG(x)}{DG(y)} - 1\right| \leq C \cdot |x-y|^\gamma.$$

However, it does satisfy Assumption a). Note that there are infinitely many branches of G near 0 and, therefore, that G is 'infinitely expanding near this point'. (Compare this exercise with the last exercise in this section.)

The main result of this section shows that such Markov maps have very good mixing properties. Not only is there an absolutely continuous invariant measure for such maps, this measure even has some very strong mixing properties. As before we say that an f-invariant probability measure μ is *ergodic* if any Borel set A which is (completely) invariant, $f^{-1}(A) = A$, either has full or zero measure with respect to μ. Furthermore, we say that μ is *exact* if there exist no Borel sets $A, A_1, A_2, \dots$ with $0 < \mu(A) < 1$ such that $A = f^{-n}(A_n)$ for any $n \geq 0$. Moreover, we say μ is *mixing* if for any Borel sets A, B,

$$\lim_{n\to\infty} \mu(f^{-n}(A) \cap B) \to \mu(A)\mu(B).$$

It is not hard to show that any exact measure is mixing (although it is not so easy to see this from the definitions) and, therefore, also ergodic.

Exercise 2.2. Show that a f-invariant probability measure μ cannot be exact if this map is almost everywhere invertible on the support of the measure. It follows from this that the Bowen-Ruelle-Sinai measure on the attracting Cantor set in the infinite renormalizable case is not exact. Show that this measure is not mixing either.

Exercise 2.3. Let $f\colon [0,1] \to [0,1]$ be defined by $f(x) = 2x \mod 1$. Show directly from the definition that the Lebesgue measure is an exact invariant measure. (Hint: if A is of the form $f^{-n}(A_n)$ then this means that $x \in A$ if and only if $x + \frac{1}{2^n} \in A$ (modulo 1). Suppose that A has positive Lebesgue measure. Then A has a density point d. Letting I_n be the interval of the form $[\frac{p}{2^n}, \frac{p+1}{2^n})$ containing d one has $\frac{|A \cap I_n|}{|I_n|} \to 1$. Because of the translation invariance one has $\frac{|A \cap [0,1]|}{|[0,1]|} = \frac{|A \cap I_n|}{|I_n|}$ and, therefore, A has full Lebesgue measure in $[0,1]$.)

Exercise 2.4. Show that μ is ergodic if it is mixing. (Hint: let B be the complement of A and apply the definitions.)

As usual, let λ be the Lebesgue measure on N. We may assume that λ is a probability measure, i.e., $\lambda(N) = 1$. Often we will denote the Lebesgue measure of a Borel set A by $|A|$.

Theorem 2.2. *Let $f\colon N \to N$ be a Markov map and let $\cup I_i$ be corresponding partition. Then there exists a f-invariant probability measure μ on the Borel sets of N which is absolutely continuous with respect to the Lebesgue measure. This measure satisfies the following properties:*

a) *its density* $\frac{d\mu}{d\lambda}$ *is uniformly bounded and Hölder continuous. Moreover, for each* i *the density is either zero on* I_i *or uniformly bounded away from zero.*

If for every i *and* j *one has* $f^n(I_j) \supset I_i$ *for some* $n \geq 1$ *then*

b) *the measure is unique and its density* $\frac{d\mu}{d\lambda}$ *is strictly positive;*

c) f *is exact with respect to* μ*;*

d) $\lim_{n\to\infty} |f^{-n}(A)| = \mu(A)$ *for every Borel set* $A \subset N$*. If* $f(I_i) = N$ *for each interval* I_i *then*

e) *the density of* μ *is also uniformly bounded from below.*

Let us first sketch the main idea of the proof of this result. Let $\mathcal{P}_n$ be the partition of N defined by iterating the partition $\cup I_i$ of N. More precisely, $I \in \mathcal{P}_n$ if and only if I is a maximal interval such that $I, \dots, f^{n-1}(I)$ are all contained in the union of the open intervals I_i. (If each point in $N \setminus \cup I_i$ is a discontinuity point or a turning point of f then I is a maximal interval on which f^n is a homeomorphism.) Let λ_n be the measure on N by pulling back the Lebesgue measure by f^n, i.e., take

$$\lambda_n(A) = |f^{-n}(A)|.$$

Since the space of measures on N is compact with respect to the weak topology, there exist certainly convergent subsequences of the measures $\mu_n := \frac{1}{n}\sum_{i=0}^{n-1} \lambda_i$ (in this topology). Let μ be a limit of $\mu_{n(k)}$. Then μ is an invariant measure. Indeed, for each Borel set A,

$$\begin{aligned}
\mu(f^{-1}(A)) &= \lim_{k\to\infty} \frac{1}{n(k)} \sum_{i=0}^{n(k)-1} |f^{-i}(f^{-1}(A))| \\
&= \lim_{k\to\infty} \left\{ \frac{1}{n(k)} \sum_{i=0}^{n(k)-1} |f^{-i}(A)| - \frac{1}{n(k)}|A| + \frac{1}{n(k)}|f^{-n(k)}(A)| \right\} \\
&= \lim_{k\to\infty} \frac{1}{n(k)} \sum_{i=0}^{n(k)-1} |f^{-i}(A)| \\
&= \mu(A).
\end{aligned}$$

Let us explain why μ will be absolutely continuous. Take a measurable set A. By Assumption a), $f^n|I$ has bounded distortion for each $I \in \mathcal{P}_n$. Hence it will follow that

$$\frac{|f^{-n}(A) \cap I|}{|I|} \leq K \frac{|A \cap f^n(I)|}{|f^n(I)|} \leq K \frac{|A|}{|f^n(I)|}.$$

Now it follows from Assumptions b) and c) that there exists $r > 0$ such that for each $n \geq 1$, $|f^n(I)| \geq r$. Therefore, we get

$$|f^{-n}(A) \cap I| \leq \frac{K}{r} \cdot |A| \cdot |I|.$$

Taking all intervals $I \in \mathcal{P}_n$, we get

$$|f^{-n}(A)| \leq \frac{K}{r}|A|.$$

Clearly this implies that

$$\mu(A) \leq \frac{K}{r}\lambda(A)$$

and, therefore, μ is absolutely continuous and has a density which is uniformly bounded from above by K/r. If $f^n(I) = N$ for each $I \in \mathcal{I}_n$ then one also has that $A \cap f^n(I) = A$ and

$$\frac{|f^{-n}(A) \cap I|}{|I|} \geq \frac{1}{K}\frac{|A \cap f^n(I)|}{|f^n(I)|} = \frac{1}{K}\frac{|A|}{|N|}$$

and therefore $|f^{-n}(A)| \geq \dfrac{1}{K|N|}|A|$; clearly this implies that $\mu(A) \geq \dfrac{1}{K|N|}\lambda(A)$ and so in this case the density of μ is also bounded from below.

Let us now give a proof of this theorem which also gives the additional properties of the density of μ with respect to λ. First of all we claim that λ_n is absolutely continuous and that its density $S_n = \dfrac{d\lambda_n}{d\lambda}$ with respect to λ is equal to

$$S_n(x) = \sum_{z \in f^{-n}(x)} \frac{1}{|Df^n(z)|}.$$

Indeed, it is enough to show that for each $A \subset I \in \mathcal{P}_n$ one has

$$\lambda_n(A) = \int_A S_n d\lambda.$$

But this is easy to prove. Indeed,

$$(*) \qquad \lambda_n(A) = |f^{-n}(A)| = \sum_{I \in \mathcal{P}_n} |f^{-n}(A) \cap I|.$$

Because the restriction of f^n to each interval $I \in \mathcal{P}_n$ is a diffeomorphism we get from the transformation rules for integration

$$(**) \qquad \begin{aligned} |f^{-n}(A) \cap I| &= \int_{f^{-n}(A)\cap I} d\lambda = \int_A |D(f^n|I)^{-1}(x)|\, d\lambda(x) \\ &= \int_A \frac{1}{|Df^n(z(x))|}\, d\lambda(x) \end{aligned}$$

where $(f^n|I)^{-1}$ is the inverse of $f^n|I$ and $z(x) = f^{-n}(x) \cap I$. Putting $(*)$ and $(**)$ together one gets

$$\begin{aligned} \lambda_n(A) &= \sum_{I \in \mathcal{P}_n} \int_A |D(f^n|I)^{-1}(x)|\, d\lambda(x) = \int_A \sum_{z \in f^{-n}(x)} \frac{1}{|Df^n(z)|}\, d\lambda(x) \\ &= \int_A S_n(x)\, d\lambda(x). \end{aligned}$$

In order to show that μ is absolutely continuous with respect to the Lebesgue measure we will estimate the functions $|Df^n(x)|$ and $S_n(x)$.

But before we do this let us first connect what we have done so far with the *Perron-Frobenius operator*. This operator associates to a density function u on $[0,1]$ the function

$$Pu(x) := \frac{d}{dx} \int_{f^{-1}([0,x])} u(t)\,dt.$$

This means that if μ is the measure with density u then the measure $f_*\mu$ has density Pu. In order to calculate $Pu(x)$ one has to consider all inverse images of x and in this way one gets

$$Pu(x) := \sum_{z \in f^{-1}(x)} \frac{u(z)}{|Df(z)|}.$$

So in other words we have simply that

$$S_n = P^n\,1$$

where 1 is the constant function with value equal to 1 (so the density function of the Lebesgue measure). In other words, we are iterating the Perron-Frobenius operator and determining its limits. In Section 4 of this chapter we will also iterate the Perron-Frobenius operator but apply some 'sliding' as well. Then it will turn out that it may not always be advantageous to really use the Perron-Frobenius operator, but to think more of pulling back mass.

So let us come back to the estimates on S_n.

Lemma 2.1. *There exists $K < \infty$ such that for every $n \in \mathbb{N}$ and every $x, y \in I_0 \in \mathcal{P}_0$,*

$$S_n(x) \le K,$$

$$|S_n(x) - S_n(y)| \le K \cdot S_n(x) \cdot |x-y|^\gamma.$$

Proof. Because of Properties b) and c), $|f^n(I)| \ge r$ for all $n \ge 0$ and all $I \in \mathcal{P}_n$ and from Assumption a) it follows that there exists $K < \infty$ such that $|f^n(I)| \le K|Df^n(z)| \cdot |I|$ for each $z \in I \in \mathcal{P}_n$. Therefore,

$$S_n(x) = \sum_{z \in f^{-n}(x)} \frac{1}{|Df^n(z)|} \le \sum_I \frac{K \cdot |I|}{|f^n(I)|} \le \sum_I \frac{K \cdot |I|}{r} \le \frac{K}{r}$$

where the last two sums run over all intervals $I \in \mathcal{P}_n$ with $f^{-n}(x) \in I$ since each element $I \in \mathcal{P}_n$ can only contain at most one element of $f^{-n}(x)$. Next notice that from the Markov Assumption b) one has that if $x, y \in I_0 \in \mathcal{P}_0$, then $f^{-n}(x) \cap I$ consists of at most one point for each $I \in \mathcal{P}_n$ and this set is non-empty if and only if $f^{-n}(y) \cap I$ is non-empty. Enumerate the points $z_1, z_2, \dots$ in $f^{-n}(x)$ and the points in $z_1', z_2', \dots$ in $f^{-n}(y)$ so that z_k, z_k' are both contained in the same element of $\mathcal{P}_n$. Then, using Assumption a),

$$\begin{aligned}|S_n(x)-S_n(y)| &\le \sum \left|\frac{1}{|Df^n(z_k)|}-\frac{1}{|Df^n(z_k')|}\right| \\ &\le \sum \frac{1}{|Df^n(z_k)|}\cdot\left|1-\frac{|Df^n(z_k)|}{|Df^n(z_k')|}\right| \\ &\le \sum \frac{1}{|Df^n(z_k)|}\cdot K\cdot|f^n(z_k)-f^n(z_k')|^\gamma\end{aligned}$$

Hence

$$|S_n(x)-S_n(y)| \le \sum \frac{1}{|Df^n(z_k)|}\cdot K\cdot|x-y|^\gamma \le S_n(x)\cdot K\cdot|x-y|^\gamma. \quad \square$$

Lemma 2.2. *There exists a sequences $n(k)\to\infty$ such that the measures $\mu_{n(k)} := \frac{1}{n}\sum_{i=0}^{n(k)-1}\lambda_i$ converge (in the weak topology) to an invariant measure μ with a density which is Hölder continuous and bounded (with respect to the Lebesgue measure). If for every i and j one has $f^n(I_j)\supset I_i$ for some $n\ge 1$ then i) the restriction of the density is uniformly positive on each interval I_i and ii) the measure is exact.*

Proof. According to the previous lemma the density $\hat{S}_n = \frac{1}{n}\sum_{i=0}^{n-1}S_i$ of μ_n satisfies

$$\hat{S}_n(x)\le K,$$
$$|\hat{S}_n(x)-\hat{S}_n(y)|\le K\cdot\hat{S}_n(x)\cdot|x-y|^\gamma$$

for each $x,y\in I_0\in\mathcal{P}_0$. It follows that $\hat{S}_n\colon \cup I_i\to\mathbb{R}$ is bounded and equicontinuous on each interval I_i, and consequently there exists a subsequence of $\hat{S}_n$ which converges uniformly to a function S such that

$$S(x)\le K,$$

$$|S(x)-S(y)|\le K\cdot S(x)\cdot|x-y|^\gamma$$

for each $x,y\in I_0\in\mathcal{P}_0$. Let μ be the corresponding measure. From the definition of $\hat{S}_n$ it follows as before that S and therefore μ is invariant. Let us first show that the density S is bounded from below on each interval I_i. Indeed, if $\inf_{x\in I_i}S(x)=0$ for some $I_i\in\mathcal{P}_0$ then $|S(x)-S(y)|\le K\cdot S(x)\cdot|x-y|^\gamma$ implies that $S(y)=0$ for all $y\in I_i$. But then $\mu(I_i)=0$. Now take an arbitrary $I_j\in\mathcal{P}_0$. If the last additional assumption is satisfied then there exists $n\ge 0$ such that $f^n(I_j)\supset I_i$ then But then $I_j\cap f^{-n}(I_i)\ne\emptyset$ and $\mu(I_j\cap f^{-n}(I_i))\le\mu(f^n(I_j)\cap I_i)=0$. From the previous argument it follows that $\mu(I_j)=0$. By assumption this would hold for all j and therefore we would have $\mu(N)=0$, a contradiction.

It remains to show that f is exact with respect to μ. So assume by contradiction that there exist Borel sets $A,A_1,A_2,\ldots$ with $0<\mu(A)<1$ and such that

$A = f^{-n}(A_n)$. But $f^n|I$ has bounded distortion on $I \in \mathcal{P}_n$. More specifically, there exists a constant $K < \infty$ such that for each $I \in \mathcal{P}_n$,

$$\frac{|f^n(I)|}{|I|} \geq \frac{1}{K} \frac{|A_n|}{|f^{-n}(A_n) \cap I|}$$

and since $|f^n(I)| \geq r$ for all $n \geq m$, we get

$$\frac{|A \cap I|}{|I|} = \frac{|f^{-n}(A_n) \cap I|}{|I|} \geq \frac{1}{Kr}|A_n|$$

for each $I \in \mathcal{P}_n$. If $\mu(A_n) = \mu(A) > 0$ then, because the density of μ is strictly positive on each $I \in \mathcal{P}_0$ and μ is absolutely continuous, there exists a constant $c > 0$ such that $|A_n| > c$ and the last inequality implies that $N \setminus A$ has no density points, and, therefore, that $|N \setminus A| = 0$, i.e., $\mu(A) = 1$. □

Proof of Theorem 2.2: Statement a) follows from the previous lemmas. So let us prove Statements b)-d) under the additional assumption made in these statements. As we have seen in the previous lemmas, each invariant measure μ has a strictly positive and bounded density. Moreover, μ is exact and therefore ergodic. Birkhoff's Ergodic Theorem then implies that for Lebesgue almost every x and for each open interval U one has $\mu(U) = \lim_{n\to\infty} \#\{0 \leq i < n;\ f^i(x) \in U\}/n$. It follows that μ is unique. Moreover, because μ is exact it Statement d) holds, see Mañé (1987, Proposition II.8.3). Statement e) follows from the previous lemma.

Exercise 2.5. In this exercise we will show that piecewise monotone expanding maps which do not satisfy the Markov Assumption also have absolutely continuous invariant measures. This result is due to Lasota and Yorke (1973). A generalization of this result, using the ideas from Section 4 of this chapter appeared in Kondah and Nowicki (1990). So let $f\colon [0,1] \to [0,1]$ be C^2 except possibly in $0 = a_0 < a_1 < a_2 < \cdots < a_r = 1$, such that i) $|Df(x)| \geq \lambda > 1$ and ii) $\dfrac{|D^2f(x)|}{|Df(x)|^2} \leq c < \infty$ on $I_i = (a_{i-1}, a_i)$ for each i. Then f has an absolutely continuous invariant probability measure. (A related and stronger theorem can be proved using the techniques from Section 4 of this chapter.) The result will proved in a few steps.
a) Let $u\colon [0,1] \to \mathbb{R}$ have bounded variation and let $[a,b] \subset [0,1]$. Show that

$$\mathrm{Var}_0^1(u1_{[a,b]}) \leq 2\mathrm{Var}_a^b(u) + \frac{2}{b-a}\int_a^b |u(x)|\,dx.$$

(Hint: let $1_{[a,b]}$ be the indicator function on $[a,b]$. Then clearly $\mathrm{Var}_0^1(u1_{[a,b]}) \leq \mathrm{Var}_a^b(u) + |u(a)| + |u(b)|$. So for each $c \in [a,b]$, $\mathrm{Var}_0^1(u1_{[a,b]}) \leq \mathrm{Var}_a^b(u) + |u(a) - u(c)| + |u(b) - u(c)| + 2|u(c)| \leq 2\mathrm{Var}_a^b(u) + 2|u(c)|$. Choosing c suitably one has $|u(c)| \leq 1/(b-a)\int_a^b |u(x)|\,dx$ and this gives the result.) b) Let u be as in a) and v be C^1. Show that

$$\mathrm{Var}(uv) \leq \sup|v|\,\mathrm{Var}(u) + \int |u(t)v'(t)|\,dt.$$

(Hint: since $\sum |a_ib_i - a_{i-1}b_{i-1}| = \sum |b_i(a_i - a_{i-1}) + a_{i-1}(b_i - b_{i-1})|$, one has

$$\begin{aligned}\sum |u(x_i)v(x_i) - u(x_{i-1})v(x_{i-1})| &\leq \\ &\leq \sum \{|v(x_i)||u(x_i) - u(x_{i-1})| + |u(x_{i-1})||v(x_i) - v(x_{i-1})|\} \\ &\leq \sup |v| \operatorname{Var}(u) + \sum |u(x_{i-1})||v'(\xi_i)||x_i - x_{i-1}|\end{aligned}$$

which in the limit is at most $\sup |v| \operatorname{Var}(u) + \int |u(t)v'(t)|\,dt$.) c) If k_n is a sequence of real numbers such that for some $\rho < 1$ one has $k_{n+1} \leq \rho \cdot k_n + L$ for all $n \in \mathbb{N}$. Then k_n is bounded. d) Let u be a positive function of bounded variation such that its integral over $[0,1]$ is one and let P be the Perron-Frobenius operator. Furthermore, let f be as above. Then there exists a universal constant L such that

$$\operatorname{Var}_0^1(Pu) \leq \frac{2}{\lambda} \operatorname{Var}(u) + L.$$

(Hint: as we have seen before

$$Pu(x) = \sum_{z \in f^{-1}(x)} \frac{u(z)}{|Df(x)|} = \sum_{i=1}^{r} \frac{u(f^{-1}(x) \cap I_i)}{|Df(x)|} 1_{I_i}(x)$$

(if $f^{-1}(x) \cap I_i = \emptyset$ then the corresponding term is zero). Using this and the second inequality from part a) of this exercise,

$$\operatorname{Var}_0^1(Pu) \leq 2 \sum_{i=1}^{r} \operatorname{Var}_{I_i} \frac{u(f^{-1}(x))}{|Df(x)|} + \sum_{i=1}^{r} \frac{2}{|I_i|} \int_{I_i} \frac{u(f^{-1}(x))}{|Df(x)|}.$$

Using part b) of this exercise, $|Df(x)| \geq \lambda$ and $|D\frac{1}{Df}| \leq c$ (which follows from the second assumption one gets

$$\operatorname{Var}_{I_i} \frac{u(f^{-1}(x))}{|Df(x)|} \leq \frac{1}{\lambda} \operatorname{Var}(u(f^{-1}(x))) + c \int_{I_i} u(f^{-1}(x))$$

and hence

$$\operatorname{Var}_0^1(Pu) \leq \frac{2}{\lambda} \sum \operatorname{Var}(u(f^{-1}(x))) + 2 \sum \left[c + \frac{1}{|I_i|}\right] \int_{I_i} u(f^{-1}(x)).$$

So one gets the required inequality with $L = 2\sum \left[c + \frac{1}{|I_i|}\right]$ since $\int u = 1$. Note that this step fails if f is not piecewise monotone, as is for example the case for the Gauss map. e) Prove the result stated at the beginning of this exercise. (Hint: since $\lambda > 1$, some iterate of f satisfies $|Df^k| > 2$. So we may assume that $\lambda > 2$. But then one gets from parts c) and d) of this exercise that $\operatorname{Var}(P^n u)$ is bounded for all n. Since $\int P^n u = 1$, it follows that some subsequence of $P^n u$ has a limit which also has bounded variation and, therefore, is in L^1. This limit is the density of an invariant absolutely continuous probability measure.)

Exercise 2.6. Show that the measure from the previous exercise does not need to be ergodic. (Hint: construct a map $f\colon [0,1] \to [0,1]$ which maps both $[0,1/2]$ and $[1/2,1]$ into itself.)

Exercise 2.7. Show that the measure is ergodic if $\sup_{n>0} |f^n(U)| = 1$ for each open interval U. In Bowen (1977) it is shown that in this case f is even weakly mixing.

3 Constructing Invariant Measures by Inducing

In the presence of critical points the expanding conditions in the definition of Markov maps fails: we no longer have expansion and control of the non-linearity. However, as we have seen in Chapter III, we may get expansion if we stay long enough away from the critical points to compensate the contraction we get near the critical point. Also, as in Chapter IV, we may recover some control of non-linearity by using the Koebe Principle. Using these ideas we will show below that in many situations we can associate to some maps a Markov map and, using Theorem 2.2, we will prove the existence of absolutely continuous invariant measures for such maps. In the next section we shall construct invariant measures using an alternative strategy.

Definition. We say that a map $f\colon I \to I$ *induces a Markov map* if there is an interval $J \subset I$ and a Markov map F on J, defined on a subset $\cup_{j=1}^{\infty} I_j$ of J with full Lebesgue measure such that, for each $j \in \mathbb{N}$, the restriction of F to I_j is an iterate $f^{k(j)}$ of f with $f^{k(j)}(I_j) \subset J$. We say that F is induced by f on J.

As we have seen in the previous section, the Markov map F has an absolutely continuous invariant probability measure. In general this need not imply that f also has such a measure. However,

Lemma 3.1. *Let $f\colon I \to I$ induce the Markov map $F\colon \cup_{j=1}^{\infty} I_j \to J$, let $k(j)$ be so that $F|I_j = f^{k(j)}$ and let ν be the absolutely continuous invariant probability measure of F. If*

$$\sum_{j=1}^{\infty} k(j)\nu(I_j) < \infty \tag{3.1}$$

then f has an absolutely continuous invariant probability measure. If

$$\sum_{j=1}^{\infty} k(j)|I_j| < \infty$$

then (3.1) holds.

Proof. By Theorem 2.2, F has an invariant probability measure ν which is absolutely continuous with respect to the Lebesgue measure. Let ν_j denote the measure $\nu|I_j$, namely, $\nu_j(A) = \nu(A \cap I_j)$. Hence $\nu = \sum_{j=1}^{\infty} \nu_j$. Let us define a measure μ by the formula

$$\mu = \sum_{j=1}^{\infty} \sum_{i=0}^{k(j)-1} f_*^i \nu_j,$$

where $\phi_* \nu$ denotes the measure obtained by pushing forward ν by ϕ, i.e., $\phi_*\nu(A) = \nu(\phi^{-1}(A))$. Clearly, μ is absolutely continuous because $|A| = 0$ implies that $\nu(A) = 0$ and hence $f_*^i \nu_j(A) = \nu_j(f^{-i}(A)) = 0$. It remains to prove that μ is f-invariant. In fact,

$$f_*\mu = \sum_{j=1}^{\infty} \sum_{i=0}^{k(j)-1} f_*^{i+1} \nu_j = \sum_{j=1}^{\infty} \left(\sum_{i=1}^{k(j)-1} f_*^i \nu_j + f_*^{k(j)} \nu_j \right).$$

Since $f_*^{k(j)} \nu_j(A) = \nu_j(f^{-k(j)}(A)) = \nu(f^{-k(j)}(A) \cap I_j) = \nu(F^{-1}(A) \cap I_j)$ and $\nu(F^{-1}(A)) = \sum_{j=1}^{\infty} \nu(F^{-1}(A) \cap I_j)$, $\nu(A) = \sum_{j=1}^{\infty} \nu(A \cap I_j)$, we get, by the F-invariance of ν, that $\sum_{j=1}^{\infty} f_*^{k(j)} \nu_j = \sum_{j=1}^{\infty} \nu_j$. Therefore, $f_*\mu = \mu$. Since

$$\mu(I) = \sum_{j=1}^{\infty} \sum_{i=0}^{k(j)-1} f_*^i \nu_j(I) = \sum_{j=1}^{\infty} k(j) \nu(I_j) \tag{3.2}$$

one has $f_*\mu = \mu$ if (3.1) holds. The last statement also follows from this because the density of ν is bounded from above. □

Let us give some examples of this strategy.

The Misiurewicz case

First we shall apply this idea to Misiurewicz maps. Misiurewicz (1981) has shown the existence of an absolutely continuous invariant measures for C^3 maps $f: I \to I$ having a finite number of critical points, negative Schwarzian derivative, no attracting periodic orbit and satisfying the so called Misiurewicz condition: no critical point is in the ω-limit set of a critical point. This result was extended in Van Strien (1990) to C^2 interval maps. More precisely,

Theorem 3.1. *Let $f: I \to I$ be a mapping satisfying the following conditions:*

1. *f is C^2 and is non-flat in each critical point;*
2. *all periodic points of f are hyperbolic repelling;*
3. *f satisfies the Misiurewicz condition: the forward orbit of a critical point of f does not accumulate onto critical points.*

Then f has an invariant probability measure which is absolutely continuous with respect to the Lebesgue measure.

We have already indicated in Exercise III.6.1 how this result follows from the quasi-polynomial behaviour of Misiurewicz maps. Here we will follow an alternative strategy and deduce it from the following results, see Theorem III.6.1 and Theorem III.5.1.

Proposition 3.1. *If $f\colon I \to I$ is a C^2 map satisfying conditions 1), 2) and 3) of Theorem 3.1 then there exists $C > 0$ such that*

$$\sum_{i=0}^{n-1} |f^i(T)| < C$$

for each $n \in \mathbb{N}$ and each interval T for which $f^n|T$ is a diffeomorphism.

Proposition 3.2. *If $f\colon I \to I$ is a C^2 map satisfying conditions 1) and 2) from Theorem 3.1 then for each neighbourhood U of $C(f)$ there exists $K < \infty$ with the following property. For each $n \in \mathbb{N}$ and each interval T such that $f^i(T) \cap U = \emptyset$ for all $i = 0, 1, \ldots, n-1$, the distortion of $f^n|T$ is uniformly bounded by K.*

In Van Strien (1990), see also Section III.6, it was shown that the iterates f^n are quasi-polynomial on each branch and for each $n \in \mathbb{N}$. This easily implies the existence of an invariant measure, see Exercise 6.1 in Section III.6. Here we shall use Proposition 3.2 to show that the induced map satisfies good distortion properties.

Using the above propositions we will prove the following lemma, see also Lemma 3.3 of Vargas (1991).

Lemma 3.2. *Let f be a unimodal map satisfying the hypothesis of Theorem 3.1. Then there exists an interval J containing the critical point of f such that the first return map of f to J is a Markov map R.*

Proof. Let W be an open interval neighbourhood of the critical point c such that $f^n(c) \notin W$ for all $n = 1, 2, \ldots$. It follows that if x is so that $f^n(x) \in W$ then there exists an interval neighbourhood $T_n(x)$ of x such that $f^n|T_n(x)$ is monotone and $f^n(T_n(x)) \supset W$. Now take a periodic point $p \in W$ such that $f^i(p) \notin (p, \tau(p))$ for all $i \geq 0$ and let $J = (p, \tau(p))$. Let R be the return map to J. Since $c \in \omega(x)$ for almost every x, the domain of R has full Lebesgue measure in J. Let I_j be the intervals from the domain of R. This map clearly satisfies Properties b)-d) of the definition of a Markov map. Indeed, J is contained in W and therefore when $x \in J$ and $n > 0$ is minimal so that $f^n(x) \in J$ then $f^n(T_n(x)) \supset W \supset J$ where $f^n|T_n(x)$ is monotone. So by the choice of the endpoints of J, we get that the interval $I \ni x$ such that $f^n(I) = J$ is contained in J. It follows that $R(I_j) = J$ on each component I_j of the domain of R. So let us prove that it also satisfies Property a). If $n \in \mathbb{N}$ and an interval I is so that $R^i(I)$ is contained in $\bigcup_j I_j$ for each $i = 0, 1, \dots, n$, then there exists k such that $R^{i+1}|I = f^k|I$ and $f^k(x) \in J$. Hence f^k is monotone on the interval neighbourhood $T_k \supset I$ from above and $f^k(T_k) \supset W$. From Proposition 3.1 and the Koebe Principle (Theorem IV.3.1) one immediately gets that

$$\left| \frac{Df^k(x)}{Df^k(y)} - 1 \right| \le C \cdot |f(x) - f(y)|$$

for all $x, y \in I$. □

Lemma 3.3. *Assume that f are as above and J is as in Lemma 3.2. Let R be the first return map of f to J. Let $\cup_j I_j$ be the domain of R and let $R|I_j = f^{k(j)}$. Then $\sum_{j=1} k(j)|I_j| < \infty$.*

Proof. Take J as in the proof of Lemma 3.2. Let

$$\begin{aligned} \Lambda_n &= J \setminus \bigcup_{\{j\,;\,k(j)<n\}} I_j \\ &= \{x \in J\,;\; f^i(x) \notin J \text{ for all } i = 1, 2, \dots, n-1\}. \end{aligned}$$

We claim that the Lebesgue measure of Λ_n goes exponentially fast to zero as $n \to \infty$. The lemma follows from this claim because

$$\sum_j k(j) \cdot |I_j| = \sum_n \sum_{\{j\,;\,k(j)=n\}} n|I_j| \le \sum_n n|\Lambda_{n-1}|$$

and by the claim this last expression is bounded from above.

So let us prove the claim and take a component E of Λ_n. Both endpoints of E are contained in intervals of the form I_j with $k(j) < n$ or in ∂J and $f^i(E) \cap J = \emptyset$ for $i = 1, \dots, n-1$. $f^n|E$ is monotone except if E contains the critical point. Since forward iterates of ∂J never enter J it follows that E is contained in the interior of J and that the endpoints of E hit ∂J in less than n iterates. Let us first assume that E does not contain c. Since $f^n|E$ is monotone, both endpoint of $f^n(E)$ are distinct points in the forward orbit of the periodic point p and therefore there exists $\delta > 0$ such that $|f^n(E)| \ge \delta$. Hence there exists a universal constant $N < \infty$ (which does depend on J but not on n) such that $c \in \operatorname{int}(f^k(E))$ for some (minimal) k with $n \le k \le n + N$. Since $f^k|E$ is monotone, $f^k(E) \supset J$. Because $f^i(E) \cap J = \emptyset$ for $i = 1, 2, \dots, n-1$, $f^{k-1}|f(E)$ has universally bounded distortion by Proposition 3.2. Combining this gives that the Lebesgue measure of $\{x \in E\,;\; f^k(x) \in J\}$ is at least a universal constant times the Lebesgue measure of E. Since

$$(\Lambda_n \setminus \Lambda_{n+N}) \cap E \subset \{x \in E\,;\; f^k(x) \in J\},$$

the Lebesgue measure of $\Lambda_{n+N} \cap E$ is at most the Lebesgue measure of $\Lambda_n \cap E$ times a universal factor $\tau < 1$.

Since f is Misiurewicz, this argument can also be applied to the unique component $E_{n,0}$ of Λ_n which contains the critical point: let (x_n, c) be one of the components of $E_{n,0} \setminus \{c\}$. Then x_n is mapped to ∂J in $i < n$ iterates and $f^i(c)$ is outside W. So we get as before a universal integer N and constant $\tau < 1$ such that $|E_{n+N,0}| \leq \tau \cdot |E_{n,0}|$. It follows that the Lebesgue measure of Λ_n tends exponentially fast to zero. □

Proof of Theorem 3.1: Follows from Lemmas 3.1-3.3. A similar proof also works for the general multimodal case. The difference is that we have to consider a Markov map defined on a disjoint collection of intervals. □

Keller's Theorem

Let us strengthen this result: we shall derive a necessary and sufficient condition for the existence of an absolutely continuous invariant probability measure due to Keller (1990a). As before, let $T_n(x)$ be a maximal interval containing x on which f^n is monotone. (If x is not in the backward orbit of c then there exists only one such interval.) Furthermore, let $R_n(x)$ and $L_n(x)$ be the components of $T_n(x) \setminus x$ and define $r_n(x)$ to be the minimum of the lengths of $f^n(R_n(x))$ and $f^n(L_n(x))$.

This result states that f has an absolutely continuous invariant probability measure if and only if f has a positive Liapounov exponent in almost every point:

Theorem 3.2. **(Keller)** *Let $f\colon [-1,1] \to [-1,1]$ be a unimodal map with one non-flat critical point with negative Schwarzian derivative. Then the following statements are equivalent:*

1. f has an absolutely continuous invariant probability measure;
2. for almost all x,

$$\limsup_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} r_i(x) > 0; \tag{3.3}$$

3. for almost all x,

$$\limsup_{n\to\infty} \frac{1}{n} \log |Df^n(x)| > 0. \tag{3.4}$$

If such a measure exists then f has no absorbing Cantor set and the density of this measure is uniformly bounded from below on the attractor of f (which, as we saw in Theorem 1.3, consists of a finite union of intervals in this case).

We shall prove this theorem using the Folklore Theorem similarly as before. Let us first deduce the following corollary from it.

Corollary 3.1. *Let $f\colon [-1,1] \to [-1,1]$ be a unimodal map with one non-flat critical point with negative Schwarzian derivative. Then there exists a constant λ_f such that for almost all x,*

$$\limsup_{n\to\infty} \frac{1}{n} \log |Df^n(x)| = \lambda_f.$$

Moreover,

1. *$\lambda_f > 0$ if and only if f has an absolutely continuous invariant probability measure; in this case $\lim_{n\to\infty} \frac{1}{n} \log |Df^n(x)| = \lambda_f$ for almost every x.*
2. *$\lambda_f < 0$ if and only if f has a hyperbolic periodic attractor.*

Proof. If f has a hyperbolic periodic attractor then, as we have seen in the first section, almost every point is in the basin of this attractor. Therefore, $\lim_{n\to\infty} \frac{1}{n} \log |Df^n(x)|$ exists and is equal to some negative constant λ_f (the eigenvalue of this periodic attractor).

So assume that f has no hyperbolic periodic attractor. Then, as we have seen in Section 1, for almost every x, we have $c \in \omega(x)$. For each such x we can define a sequence $k(i) \to \infty$ such that $k(0) = 0$ and so that $k(i+1) > k(i)$ is the smallest integer such that $f^{k(i+1)}(x) \in (f^{k(i)}(x), \tau(f^{k(i)}(x)))$. Therefore, $f^{k(i)}(x) \in V_{k(i+1)-k(i)}$ where

$$\begin{aligned} V_k = \{y\,;\, f^i(y) \notin [y, \tau(y)] \text{ for all } 0 < j < k \text{ and} \\ f^k(y) \in [y, \tau(y)]\}. \end{aligned}$$

This set was also used in Guckenheimer's proof of Theorem II.6.3, see Exercise II.6.1. As was shown in that exercise, if $T = [u, v]$ is a component of V_k then $f^k|T$ is monotone and $f^k(u) = u$, $f^k(v) = \tau(v)$ (or vice versa). Since f has no repelling periodic attractors, this implies $|Df^k(u)| \geq 1$ and $|Df^k(\tau(v))| \geq 1$. Because of the non-flatness of c this implies that there exists a universal constant $C \in (0,1)$ so that $|Df^k(u)| \geq 1$ and $|Df^k(v)| \geq C$. Hence, by the Minimum Principle, $|Df^k(y)| \geq C$ for all $y \in T$. It follows that

$$|Df^{k(j)}(x)| \geq C^j.$$

Since $k(j+1) - k(j) \to \infty$ and therefore $k(j)/j \to \infty$ it follows that

$$\limsup_{n\to\infty} \frac{1}{n} \log |Df^n(x)| \geq 0.$$

The previous theorem implies that this measure μ exists if and only if $\limsup \frac{1}{n} \log |Df^n(x)| > 0$ for almost all x. The remainder of the corollary follows: if f has an absolutely continuous invariant probability measure μ then by Birkhoff's Ergodic Theorem,

$$\lim_{n\to\infty} \frac{1}{n} \log |Df^n(x)| = \int \log |Df|\, d\mu$$

for almost every x. □

Remark. We say that the set of periodic points of f has a *hyperbolic structure* if there exists $\lambda > 1$ such that for each periodic point p of period n one has $|Df^n(p)| > \lambda^n$. The previous proof shows that this implies $\lim_{n\to\infty} \dfrac{1}{n} \log |Df^n(x)| \geq \lambda$ for almost all x.

Proof of Theorem 3.2: Our proof is quite different from Keller's. He uses a classification of positive L^1 operators and the tower construction from Section II.3.b. Our proof is based on using Theorem 2.2 from the previous section. First note that each of the Statements 1), 2) and 3) is impossible if f has an attracting periodic point.

First we observe that if $\limsup_{n\to\infty} r_n(x) > 0$ for almost all x then by the ergodicity of f, there exists $\delta > 0$ such that

$$\limsup_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} r_i(x) \geq \delta$$

for almost all x. As before, let $T_n(x)$ the maximal interval containing x on which f^n is a diffeomorphism.

Statement 1 implies 3: Assume that Statement 1 holds, i.e., f has an absolutely continuous invariant probability measure μ. As we have seen in Exercise V.1.4 the metric entropy $h_\mu(f)$ of such a measure is positive (we refer to Walters (1982) or Mañé (1987) for the definition of the definition of the metric entropy of a measure). As is well known we have the Rochlin formula

$$h_\mu(f) = \int \log |Df| \; d\mu,$$

see for example Ledrappier (1981). Moreover, by Birkhoff's Ergodic Theorem,

$$\lim_{n\to\infty} \frac{1}{n} \log |Df^n(x)| = \lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \log |Df(f^i(x))| = \int \log |Df| \; d\mu.$$

Since the last term is positive, the limit on the left is also positive. This proves that Statement 3 holds. We should note that one can also prove Statement 2 quite easily using Keller (1989a). Indeed, in that paper it is shown that such a measure can be lifted to an absolutely continuous probability measure $\hat{\mu}$ which is defined on the tower from Section II.3.b and leaves invariant the lift $\hat{f}$ to the tower of the map f. Then pick some interval D_k in the tower and take the middle third interval U_k in D_k. Because of Birkhoff's Ergodic Theorem for almost all x one has that $\#\{0 \leq i < n\,;\; \hat{f}^i(x) \in U_k\}/n$ tends to $\hat{\mu}(U_k) > 0$ as $n \to \infty$. When $\hat{f}^i(x) \in U_k$ then $r_i(x) \geq |D_k|/3$ because by definition of the tower, f^i maps a neighbourhood of x diffeomorphically to the interval then D_k and so it follows that

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} r_i(x) \geq \hat{\mu}(U_k) \cdot |D_k|/3$$

for almost all x.

Statement 2 implies Statement 1: So assume that there exists $\delta > 0$ so that the set

$$X = \{x\,;\ \limsup_{n\to\infty} \frac{1}{n}\sum_{i=0}^{n-1} r_i(x) \geq \delta\}$$

has full Lebesgue measure. We start by taking a finite forward invariant set $P \subset [-1,1]$ so that each component of $[-1,1]\setminus P$ has length $\leq \delta/4$. We can do this by taking P to be a finite union of periodic orbits. Let δ' be the length of the smallest of these components. We shall use the partition generated by P to show that we can define a Markov map F if the set $X' = \{x\,;\ \limsup r_n(x) \geq \delta\}$ has full Lebesgue measure. As we will show, this Markov map F has an absolutely continuous invariant measure. This measure can be used to get an absolutely continuous invariant measure for f provided Statement 2 holds. So let $J(x)$ denote the interval of the partition $[-1,1]\setminus P$ containing x (we add the right endpoints to each component of $[-1,1]\setminus P$ to get a covering of $[-1,1]$). Let N be so large that each interval of monotonicity of f^N has length $\leq \delta'/2$. Take $x \in X'$ and let $k(x) \geq N$ be minimal so that $f^{k(x)}(T_{k(x)}(x))$ contains $J(f^{k(x)}(x))$ and its two neighbours from the partition. By the choice of the partition such an integer exists for each $x \in X'$. Let $I(x) \subset T_{k(x)}(x)$ be so that $f^{k(x)}(I(x))$ is equal to $J(f^{k(x)}(x))$. By definition, for each $y \in I(x)$ one has $I(y) = I(x)$ and $k(x) = k(y)$. So it makes sense to define $F\colon \bigcup_{x\in X} I(x) \to \bigcup J_i$ by $F|I(x) = f^{k(x)}$ and let $k(j) = k|I_j$. Let us show that F is a Markov map. By definition F maps $I(x)$ into one interval from the partition. Moreover, $I(x)$ never contains a point of P in its interior because the points of P are mapped again into P. This implies that F maps $I(x)$ into a union of intervals $I(y_i)$. It satisfies the Markov Assumption b) from the previous section. Moreover, $|F(I(x))| \geq \delta'$ and therefore it satisfies Assumption c) from the definition of Markov maps. So let us show that it also satisfies the Markov Assumption a) that for each $k \geq 0$ and the restriction of F^k to each branch is not too non-linear. So take $s \geq 0$ and for $x \in X'$ let $I_s(x)$ be the domain of F^s containing x. Let $m(s,x)$ be so that $F^s(x) = f^{m(s,x)}(x)$. By the choice of N, the interval $T_{m(s,x)}(x)$ is contained in the union of at most two elements of the partition. Moreover, by definition $f^{m(1,x)}$ maps $T_{m(1,x)}$ diffeomorphically over $J(F(x))$ and its two neighbours from the partition. From this the following assertion can be proved by induction on s: for each $x \in X'$ the interval $T_{m(s,x)}(x)$ is mapped diffeomorphically by $f^{m(s,x)}$ over $J(F^s(x))$ and also over its two neighbours from the partition. Since $f^{m(s,x)}|T_{m(s,x)}(x)$ is an extension of $F^s|I_s(x)$ and since $Sf^{m(s,x)} < 0$ it follows from the Koebe Principle that F^s satisfies

$$\left|\frac{DF^s(x)}{DF^s(y)} - 1\right| \leq C\cdot |F^s(x) - F^s(y)| \text{ for all } x, y \in I.$$

It follows from Theorem 2.2 that F has an absolutely continuous invariant probability measure ν. From Lemma 3.1, the map f has also an absolutely continuous invariant probability measure if $\sum k(j)\nu(I_j) < \infty$. We shall use the assumption to show that this condition is met. So assume by contradiction that $\sum k(j)\nu(I_j)=\infty$. Because of Birkhoff's Ergodic Theorem for ν-almost all x,

$$\frac{n(s)}{s}=\frac{k(x)+k(F(x))+\cdots+k(F^s(x))}{s}$$

tends to

$$\int_{[-1,1]} k(x)d\nu(x)=\sum k(j)\nu(I_j)=\infty$$

as $s\to\infty$. Let $F^i(x)=f^{n(i)}(x)$. By construction if $n(i)\le n\le n(i+1)$ and $r_n(x)\ge\delta$ then either $n(i)\le n\le n(i)+N$ or $n=n(i+1)$. Indeed, if $r_n(x)\ge\delta$ then f^n maps some interval T containing x diffeomorphically to a δ neighbourhood of $f^n(x)$. So, in particular, $f^{n-n(i)}$ maps $f^{n(i)}(T)$ diffeomorphically to $J(f^n(x))$ and its two neighbours from the partition. Because of the minimality of $k(i)$ either $n-n(i)=k(i)$ (and therefore $n=n(i+1)$) or $n-n(i)\le N$). So take n and let s be so that $n(s)\le n<n(s+1)$. Then

$$\frac{\#\{0\le i<n\,;\ r_i(x)\ge\delta\}}{n}\le\frac{(N+1)s}{n(s)}.$$

Because of the previous limit, this last ratio converges to zero as $s\to\infty$. It follows that $\limsup_{n\to\infty}\frac{1}{n}\sum_{i=0}^{n-1} r_i(x)<\delta$ for ν-almost all x. Since ν is absolutely continuous, this contradicts the assumption. It follows that f has also an absolutely continuous invariant probability measure μ. From Theorem 2.2, the density of ν is uniformly bounded from below in one of the intervals I_j. So the density of μ is also uniformly bounded from below on I_j. Let L be the finite union of transitive intervals in the non-wandering set of f. Take a component L' of L. Then there exists k such that $f^k(I_j)\supset L'$. It follows from this, the invariance and since Df^k is bounded that the density of μ is also uniformly bounded from below on each component of L.

Statement 2 implies 3: Because f is ergodic, 2) implies that there exists $\delta>0$ such that

$$\limsup_{n\to\infty}\frac{1}{n}\sum_{i=0}^{n-1} r_i(x)\ge\delta$$

for almost all x. Because $r_i(x)\le 1$ for each x and each i, this implies that there exists arbitrarily large integers n such that

$$\begin{aligned}\frac{3}{4}\delta &\le \frac{1}{n}\sum_{0\le i<n} r_i(x)\\ &=\frac{1}{n}\sum_{\{0\le i<n\,;\ r_i(x)\le\delta/2\}} r_i(x)+\frac{1}{n}\sum_{\{0\le i<n\,;\ r_i(x)>\delta/2\}} r_i(x)\\ &\le\frac{\delta}{2}+\frac{\#\{0\le i<n\,;\ r_i(x)\ge\delta/2\}}{n}.\end{aligned}$$

So for almost all x there exist infinitely many integers n such that

$$\frac{1}{n}\#\{0\le i<n\,;\ r_i(x)\ge\delta/2\}\ \ge\ \delta/4.$$

From this it easily follows that there exist arbitrarily large integers n with

$$\frac{1}{n}\#\{0 \le i < n\,;\; r_i(x) \ge \delta/2 \text{ and } r_n(x) \ge \delta/2\} \;\ge\; \delta/4 \tag{3.5}$$

Take a point x and an integer $n \in \mathbb{N}$ for which (3.5) is satisfied. For each $\delta > 0$ there exists an integer N which only depends on δ such that each interval of length $\delta/8$ in $[-1, 1]$ contains a point of $\cup_{m=0}^{N} f^{-m}(c)$. So f^N is never monotone on an interval of length $\ge \delta/8$. In particular, if $r_i(x) \ge \delta/2$ then $f^i(T_i(x))$ contains a $\delta/2$ neighbourhood of $f^i(x)$. Because of the previous sentence this implies that $f^i(T_i)$ must contain a $\delta/8$-scaled neighbourhood of $f^i(T_{i+N}(x))$. Hence from the Macroscopic Koebe Principle, there exists a constant $\kappa < 1$ which only depends on δ such that for each integer i with $r_i(x) \ge \delta/2$,

$$|T_{i+N}(x)| \le \kappa \cdot |T_i(x)|.$$

Because of (3.5), for at least $\delta/4$ of the integers $i = 0, 1, 2, \ldots, n$ one has $r_i(x) \ge \delta/2$ and therefore $|T_{i+N}(x)| \le \kappa \cdot |T_i(x)|$. Hence there exists $\rho < 1$ and $C > 0$ depending only on δ with

$$|T_n(x)| \le C \cdot \rho^n.$$

Since $r_n(x) \ge \delta/2$, the Koebe Principle implies that there exist $K < \infty$ and $C' > 0$ with

$$|Df^n(x)| \ge \frac{1}{K} \cdot \frac{|f^n(T_n(x))|}{|T_n(x)|} \ge \frac{1}{K} \cdot \frac{\delta}{|T_n(x)|} \ge \frac{C'}{\rho^n}.$$

Statement 3 follows.

Statement 3 implies 2: Assume that $\limsup \frac{1}{n} \log |Df^n(x)| > 0$ for almost all x. As before, the ergodicity of f implies that there exists a constant $\lambda_f > 0$ such that $\limsup \frac{1}{n} \log |Df^n(x)| = \lambda_f$ for almost all x. Let $\lambda \in (0, \lambda_f)$. Let $l\colon \mathbb{R}^+ \to \mathbb{R}^+$ be maximal so that

$$|x - c| \le \delta \text{ and } f^i(x) = c \text{ implies } i \ge 2l(\delta). \tag{3.6}$$

Note that we can choose the function l so that $l(\delta) \to \infty$ as $\delta \to 0$. Let $\delta > 0$, $a_i \in \{0, 1\}$ and define

$$S_\delta(a_0, a_1, \ldots, a_s) = \{x\,;\; r_i(x) \ge \delta \text{ if } a_i = 1 \text{ and } 0 < r_i(x) < \delta \text{ if } a_i = 0\}.$$

Note that if I is a component of this set then $f^i|I$ is a diffeomorphism for $0 \le i \le s$ (because $r_i > 0$ on such a component) and $|f^i(I)| \le 2\delta$ if $a_i = 0$. Moreover, the partition of $S_\delta(a_0, \ldots, a_{s+1})$ refines that of $S_\delta(a_0, \ldots, a_s)$: each component of the first set is contained in one of the second set and each component of the last set is a union of components of the first set (with their boundary points). Note that f^s is monotone on each component of $S_\delta(a_0, \ldots, a_s)$ and so f^{s+1} can have at most two monotone branches over such a component. Furthermore, $S_\delta(a_0, \ldots, a_s, a_{s+1})$ may have at most three parts in a component of $S_\delta(a_0,\ldots,a_s)$ on which f^{s+1} is monotone. Let $\#S$ denote the number of components of the set S. Then this kind of argument gives:

$$\#S_\delta(a_0,\ldots,a_s,0)+\#S_\delta(a_0,\ldots,a_s,1)\le 6\#S_\delta(a_0,\ldots,a_s). \tag{3.7}$$

Proof of (3.7): Let I be a component of the last set with endpoints u and v. Let us consider what may happen to this interval.

Case 1: $a_s=0$ and $f^{s+1}|I$ is a diffeomorphism. Because $a_s=0$, the map f^s has a critical point at one endpoint u of I and either

a. v is not a turning point of f^s. In this case I is split up into at most two intervals. It is split into two intervals precisely if $r_{s+1}(v)\ge\delta$ and so 'space' is created on this side of I: in this case the interval containing u in its boundary is in $S_\delta(a_0,\ldots,a_s,0)$ and the other one is in $S_\delta(a_0,\ldots,a_s,1)$.

b. v is a turning point of f^s. In this case, it is not split if $|f^{s+1}(I)|\le\delta$ and otherwise into three intervals I_l,I_m,I_r: the closures of the components I_l,I_r contain the endpoints of I and are contained in $S_\delta(a_0,\ldots,a_s,0)$ and a middle one I_m which is contained in $S_\delta(a_0,\ldots,a_s,1)$.

Case 2: If $f^{s+1}|I$ is a diffeomorphism and $a_s=1$ then I might be split up into at most three intervals (because one might 'loose' the space on the sides of I).

Case 3: If $f^{s+1}|I$ is not a diffeomorphism then I is split up in at most six similar intervals: it is cut in two because of the critical point and each of these intervals can be split in at most three intervals for the same reasons as before. Combining this proves (3.7). □Next we want to prove that for any $l\le l(\delta)$,

$$\#S_\delta(a_0,\ldots,a_s,0^{l+1})\le 6\#S_\delta(a_0,\ldots,a_s,0). \tag{3.8}$$

where 0^i is a string of i consecutive 0's.

Proof of (3.8): Consider a component I of $S_\delta(a_0,\ldots,a_s,0)$. Then $|f^{s+1}(I)|\le 2\delta$. Because of the choice of l then $f^{s+i}(I)\ni c$ for at most one $i\in\{1,\ldots,l+1\}$. Now $f^{s+i}|I$ is a diffeomorphism. If $I'\in S_\delta(a_0,\ldots,a_s,0^i)$ and $I'\subset I$ then I' has at least one common endpoint with I. So I is split up in at most two subintervals which are contained in $S_\delta(a_0,\ldots,a_s,0^i)$. Because $f^{s+i+1}|I$ is not a diffeomorphism, one of these subintervals is split up in at least two pieces and at most six pieces of $S_\delta(a_0,\ldots,a_s,0^{i+1})$ (as in Case 3 above). Continuing one gets (3.8). □Combining (3.7) and (3.8) gives that

$$\sum\#S_\delta(a_0,\ldots,a_n)\le 6^{l\delta n}6^{n(1-\delta)/l} \tag{3.9}$$

if we sum over all $(a_0,\ldots,a_n)$ with $a_0+\cdots+a_n\le\delta n$. This holds because there are most δn of 0's which appear in a block of less than l zeros and of 1's. Therefore, if we define

$$X_n(\delta)=\{x\,;\ \frac{1}{n}\sum_{i}^{n-1}r_i(x)<\delta,\quad r_n(x)>0\}$$

then

$$\#X_n(\delta) \le 6^{l\delta n} 6^{n(1-\delta)/l}. \tag{3.10}$$

Now choose l so that

$$6^{1/l} < e^{\lambda/4}.$$

Next choose $\delta > 0$ so small so that $l < l(\delta)$ and so that

$$6^{\delta l} 6^{1/l} \le e^{\lambda/2}. \tag{3.11}$$

Furthermore, let

$$Y_n = \{x\,;\ |Df^n(x)| \ge e^{n\lambda}\}.$$

By definition,

$$\cup_{n \ge n_0} Y_n \text{ has full Lebesgue measure for each } n_0. \tag{3.12}$$

We will show that for $\delta > 0$ sufficiently small,

$$|Y_n \cap X_n(\delta)| \le e^{-n\cdot\lambda/4}. \tag{3.13}$$

Let us show that this implies that there exists $\delta > 0$ such that

$$\{x\,;\ \limsup_{n\to\infty} \frac{1}{n}\sum_{i=0}^{n-1} r_i(x) \ge \delta\}$$

has full Lebesgue measure. Indeed, by ergodicity otherwise this set would have zero Lebesgue measure. Hence for each $\epsilon > 0$ there exists a set A of Lebesgue measure $\le \epsilon$ such that

$$X_n(\delta) \cup A \supset [-1,1] \text{ for all } n \ge n_0$$

provided n_0 is sufficiently large. Therefore

$$|Y_n \setminus A| \le |Y_n \cap X_n(\delta)| \le e^{-n\cdot\lambda/4}.$$

So

$$|\cup_{n\ge n_0} Y_n| \le |\cup_{n \ge n_0} (Y_n \setminus A)| + |A| \le \sum_{n \ge n_0} e^{-n\cdot\lambda/4} + \epsilon$$

which by (3.12) is smaller than 2ϵ if n_0 is sufficiently small. This contradicts (3.12) because we can choose $\epsilon > 0$ as small as we like. Hence, we only need to prove (3.13).

Let us show that (3.10) and (3.11) imply (3.13). So take a component I from the collection $Y_n \cap X_n(\delta)$. But this implies that for each such component I,

$$|f^n(I)|/|I| \ge e^{(3/4)\cdot\lambda\cdot n} \tag{3.14}$$

provided $\delta > 0$ is sufficiently small. Indeed, first note that $f^n|I$ is monotone and that most of the intervals $I, \dots, f^{n-1}(I)$ are small. For example,

$$\#\{i \in \{0,1,\dots,n-1\}\,;\ |f^i(I)| \ge \sqrt{2\delta}\} \le \sqrt{\delta}\ n. \tag{3.15}$$

Moreover,

$$|Df^n(x)| \geq e^{\lambda \cdot n} \tag{3.16}$$

for each $x \in I$. Indeed, there exists for each $\kappa \in (0,1)$ and $\alpha > 0$ some $\delta > 0$ such that

$$|f^{i+1}(I)|/|f^i(I)| \geq \kappa \cdot |Df(f^i(x))| \text{ if } \operatorname{dist}(f^i(I), c) \geq \alpha, \ |f^i(I)| \leq \sqrt{2\delta}. \tag{3.17}$$

Moreover, by the non-flatness of the critical point of f, there exists $C > 0$ such that

$$|f^{i+1}(I)|/|f^i(I)| \geq C \cdot |Df(f^i(x))| \tag{3.18}$$

for any $i = 0, 1, \ldots, n-1$ and any $x \in I$ (because $f^i(I)$ does not contain the critical point). On the other hand, since c is not periodic, it takes a small interval near c very long to come back close to c. So for each $\xi > 0$ one has for each $\alpha, \delta > 0$ sufficiently small,

$$\#\{0 \leq i < n\, ;\ d(f^i(I), c) \leq \alpha\} \leq \xi\, n$$

because, as we explained above, most of the iterates $I, \ldots, f^n(I)$ are small, see (3.15). Using this and (3.15) it follows that we for 'most' i we can use (3.17) and for the others we still have (3.18):

$$|f^n(I)|/|I| \geq |Df^n(x)| \cdot \kappa^{(1-\xi-\sqrt{\delta})n} \cdot C^{(\xi+\sqrt{\delta})\cdot n}$$

Because of (3.16) this is at least

$$e^{\lambda n} \kappa^{(1-\xi-\sqrt{\delta})n} \cdot C^{(\xi+\sqrt{\delta})\cdot n}.$$

Notice that we can take $\kappa < 1$ arbitrarily close to one and $\xi > 0$ arbitrarily close to zero by taking $\delta > 0$ sufficiently close to zero. Hence one gets (3.14). Thus we get that for each component I of $Y_n \cap X_n(\delta)$,

$$|I| \leq e^{-(3/4)\cdot\lambda\cdot n}.$$

Because the number of components of $X_n(\delta)$ is at most $e^{-\lambda n/2}$ (and each component of $X_n(\delta)$ contains at most one component of $Y_n \cap X_n(\delta)$ because of the Minimum Principle) one gets (3.12). □

Exercise 3.1. Assume that f and $\tilde{f}$ are C^3 maps with negative Schwarzian derivative and without absorbing Cantor attractors. Show that any absolutely continuous conjugacy h between these maps is, in fact, $C^{1+\text{Hölder}}$. (Hint: because of Theorem 1.4 there exists a neighbourhood U of c such that the return map R is defined almost everywhere. Moreover, R has bounded distortion on each branch. Because of Theorem 2.2, R has an absolutely continuous invariant probability measure μ with a density u which is strictly positive and Hölder. Similarly, one has $\tilde{R}$, $\tilde{\mu}$ and $\tilde{u}$ for $\tilde{f}$. Since h is absolutely continuous and since U and $\tilde{U}$ can be defined topologically, $h_*\mu = \tilde{\mu}$. Therefore, $Dh(x) = \tilde{u}(h(x))/u(x)$ is Hölder.)

4 Constructing Invariant Measures by Pulling Back

The aim of this section is to show that for unimodal maps of the interval a very weak condition guarantees the existence of an invariant probability measure which is absolutely continuous with respect to the Lebesgue measure.

We will do this by an inductive pullback argument. In general, the existence of an absolutely continuous measure of some interval map is related to the amount of expansion this map has. Indeed, if a map $f\colon [0,1] \to [0,1]$ is everywhere expanding, C^2 and the f-image of each maximal interval of monotonicity is not 'too small' then it has an absolutely continuous invariant probability measure as we saw in Section 2 of this chapter. However, if the map has a critical point there is no universal expansion. Nevertheless, by some analytical means one can estimate the counter play between the contraction ruled by the derivative near the critical point and the expansion ruled by the derivative near the critical value. We have already seen this in the case of Misiurewicz maps in the previous section. However, in the Misiurewicz case the turning point is not recurrent and this is a very special situation. (One can show that in the family $f_a(x) = ax(1-x)$ the parameters a for which these maps are Misiurewicz maps and have no hyperbolic attractor has Lebesgue measure zero.) So let us get to weaker conditions which imply the existence of invariant measures. Let $f\colon [-1,1] \to [-1,1]$ be a C^3 unimodal map with negative Schwarzian derivative and a non-flat critical point c, i.e., there exists $l \geq 1$ and constants O_1, O_2 such that

(NF) $$O_1|x-c|^{\ell-1} \leq |Df(x)| \leq O_2|x-c|^{\ell-1}.$$

Moreover, let $c_1 = f(c)$ and assume that the growth-rate of $|Df^n(c_1)|$ is exponential, i.e, there exists $K > 0$ and $\lambda > 1$ such that

(CE1) $$|Df^n(c_1)| \geq K \cdot \lambda^n \text{ for all } n \geq 0.$$

Combining a result of Collet and Eckmann (1983) and Nowicki (1985b) and (1988), it follows that maps of this type have a unique absolutely continuous invariant probability measure μ which is ergodic and of positive entropy.

Remark. More precisely, it was shown by Nowicki (1985b), (1988) that (CE1) implies that there exists $K' > 0$ and $\lambda' > 1$ such that for each z with $f^n(z) = c$ and each $n \geq 1$ one has

(CE2) $$|Df^n(z)| \geq K \cdot \lambda^n.$$

Collet and Eckmann proved that (CE1) and (CE2) imply the existence of such an invariant measure.

In this section we will give a stronger result from Nowicki and Van Strien (1991a), see also (1988), which shows that a much weaker growth rate of $|Df^n(c_1)|$

is sufficient for the existence of such invariant measures. The proof of this result is no harder than the proof given in Collet and Eckmann (1983).

Theorem 4.1. *Let f be a unimodal C^3 map with negative Schwarzian derivative and assume that the critical point c of f is of finite order $l \geq 1$, i.e., assume that there are constants O_1, O_2 so that*

(NF) $$O_1|x-c|^{\ell-1} \leq |Df(x)| \leq O_2|x-c|^{\ell-1}.$$

Furthermore, assume that the growth-rate of $|Df^n(c_1)|$ is so fast that

(SC) $$\sum_{n=0}^{\infty} |Df^n(c_1)|^{-1/\ell} < \infty.$$

Then f has a unique absolutely continuous invariant probability measure μ which is ergodic and of positive entropy. Furthermore, there exists a positive constant K such that

$$\mu(A) \leq K|A|^{1/\ell},$$

for any measurable set $A \subset (0,1)$.

As we saw in the previous section, Keller (1987), (1989) has given other non-uniform conditions equivalent to the existence of an absolutely continuous invariant probability measure.

Remark. 1. As we will show below the density ρ of the measure μ with respect to the Lebesgue measure is a $L^{\tau-}$ function where $\tau = \ell/(\ell-1)$, $L^{\tau-} = \cup_{1\leq t<\tau} L^t$ and $L^t = \{\rho \in L^1\,;\ \int |\rho|^t\,dx < \infty\}$. In Nowicki (1991), an alternative proof of Theorem 4.1 and sharp estimates for the density of μ are given.

2. It is not hard to show that there exist many parameters a for which the quadratic map $f(x) = ax(1-x)$ satisfies condition (SC) but not the condition (CE1) (this can be shown easily with the techniques of Section 6 of this chapter). In fact, Lyubich and Milnor (1991) have shown that the quadratic Fibonacci map from Section II.3 satisfies (SC) and definitely not (CE). So the condition (SC) is much weaker than the well-known Collet-Eckmann condition. Benedicks and Young (1990) proved the existence of absolutely continuous invariant measures for maps with a non-flat critical point for which $|Df^n(c_1)|$ is at least $e^{\alpha\sqrt{n}}$ and for which, moreover, the distance of $f^n(c_1)$ to c is at least of the form $e^{-\alpha n}$, see also Benedicks and Carleson (1985). Clearly, Theorem 4.1 implies their result.

3. Of course the estimate $\mu(A) \leq K|A|^{1/\ell}$ shows that the poles of the invariant measure μ are at most of the form $|x-x_0|^{\mathrm{el}-1}$. It is not hard to show that any absolutely continuous invariant probability measure has a pole of this order at the critical values $f^n(c)$, $n \geq 1$, and, therefore, this estimate is optimal. Even for maps for which $|Df^n(c_1)|$ grows exponentially fast this result does not follow from Collet and Eckmann (1983) or Nowicki and Van Strien (1988). In those

papers only some bounds for the order of the poles are given. Notice that the density of the invariant measure is always a $L^{\tau -}$ function (where $\tau = \ell/(\ell - 1)$), independently of the size of $\sum_{n=0}^{\infty} |Df^n(c_1)|^{-\text{el}} < \infty$. In Nowicki and Van Strien (1991a) it is conjectured that these maps f either have an absolutely continuous invariant probability measure with a $L^{\tau -}$ density or do not have a finite absolutely continuous invariant measure at all.

4. In Kondah and Nowicki (1991) the ideas from this section were used to generalize the results of Lasota and Yorke (1973) for expanding maps.

5. Contrary to what was conjectured in Nowicki and Van Strien (1991a) the condition (SC) is not optimal. Indeed, Bruin (1992b) has given an example of a unimodal map with negative Schwarzian derivative and a non-flat critical point which has an absolutely continuous invariant probability measure but for which (SC) fails. In fact, his example is topological: each combinatorially equivalent map with negative Schwarzian derivative and non-flat critical point has the same properties.

Step 1: A reformulation of Theorem 4.1 and an outline of its proof

As we saw in Section 1 any unimodal map with negative Schwarzian derivative is ergodic (w.r.t. the Lebesgue measure) and in Exercise 1.4 of that section it was shown that any absolutely continuous invariant probability measure μ has positive metric entropy. Therefore, in order to prove Theorem 4.1 it is enough to establish the existence of an absolutely continuous invariant probability measure μ.

Often invariant measures are constructed by considering iterations of the Perron-Frobenius operator. This operator associates to the density of a measure ν the density of $f_*\nu$. Of course $f_*\nu$ will have poles at the critical values of f even if ν does not. Therefore, in order to show that iterations of the Perron-Frobenius operator (i.e. the densities $f_*^n\nu$) have a nice limit density, one has to choose a good 'topology' on a space of densities with infinitely many poles. In some cases one chooses L_p spaces, in other cases spaces with weighted norms.

Rather than to look at the densities of $f_*^n\nu$, in Nowicki and Van Strien (1988) it was proposed to compare the measures $f_*^n\nu$ with the Lebesgue measure. More precisely, it is easy to see that f has an absolutely continuous invariant probability measure provided that for any $\epsilon > 0$ there exists $\delta > 0$ such that for any measurable set A with $|A| < \delta$ one has that $|f^{-n}(A)| < \epsilon$ for all $n > 0$. In fact we will show that there exists a constant K such that for every n and every measurable set A,

$$|f^{-n}(A)| < K|A|^{\text{el}}.$$

Let us first explain why this inequality implies that f has an absolutely continuous invariant probability measure with a $L^{\text{el}\,-}$ density. For simplicity assume that $|I| = 1$ and let λ be the Lebesgue measure on I. Define $\lambda_n(A) = |f^{-n}(A)|$ (which is nothing but the probability measure $f_*^n\lambda$ from above when λ is the

Lebesgue measure) and let $\mu_n = \dfrac{1}{n}\Sigma_{i=0}^{n-1}\lambda_i$, i.e., $\mu_n(A) = \dfrac{1}{n}\Sigma_{i=0}^{n-1}|f^{-i}(A)|$. Since the space of probability measures on I is compact (with respect to the weak topology), there exist a sequence $n_i \to \infty$ and a probability measure μ such that μ_{n_i} converges weakly to μ. From the definition of μ_n it follows easily as in Section 2 that μ is invariant, $\mu(f^{-1}(A)) = \mu(A)$, and from $|f^{-n}(A)| < K|A|^{\mathrm{el}}$ one has that $\mu(A) \le K|A|^{\mathrm{el}}$ for each measurable set A. Hence, μ is absolutely continuous.

Let ρ be the density of μ with respect to the Lebesgue measure, i.e, $\mu(A) = \int_A \rho(x)\,dx$. Informally speaking, the inequality $\mu(A) \le K|A|^{\mathrm{el}}$ implies that the poles of the density can be at worse of the form $x^{-(\ell-1)/\ell}$: if $d\mu \approx x^{-(\ell-1)/\ell}\,dx$ then integrating gives $\int_0^\epsilon x^{-(\ell-1)/\ell}\,dx = \epsilon^{\mathrm{el}}$. So one expects the density to be in the space $L^{\tau-}$ where $\tau = \ell/(\ell-1)$. Let us make this argument precise. Take $t \ge 1$ and $C_k = \{x\,;\ k \le \rho^t(x) \le k+1\}$ and $D_k = \cup_{l=k}^{\infty} C_l = \{x\,;\ \rho^t(x) \ge k\}$. Since $\mu(A) \le K|A|^{\mathrm{el}}$ one has $k^{1/t}|D_k| \le \int_{D_k}\rho\,dx = \mu(D_k) \le K\cdot|D_k|^{\mathrm{el}}$. Therefore,

$$|D_k| \le K' \cdot k^{-\ell/((\ell-1)\cdot t)} = K' \cdot k^{-\tau/t}.$$

Hence,

$$\int \rho^\tau\,dx \le \Sigma_{k=0}^{\infty}(k+1)|C_k| = \Sigma_{k=0}^{\infty}|D_k|,$$

i.e.,

$$\int \rho^\tau\,dx \le \Sigma_{k=0}^{\infty}K' \cdot k^{-\tau/t} < \infty$$

whenever $0 < t < \tau$. This shows that $\rho \in L^{\tau-}$.

Let us now come to the main result of this step: in order to show

$$|f^{-n}(A)| < K|A|^{\mathrm{el}}$$

it is sufficient to estimate the size of preimages of the interval $(c-\epsilon, c+\epsilon)$. The reason this simplifies the situation is because it allows us to consider only preimages of the special intervals $(c-\epsilon, c+\epsilon)$. If the image under some iterate of a branch has 'some space' around this interval then we will be able to get good estimates for the preimage of $(c-\epsilon, c+\epsilon)$ under this iterate using the Koebe Principle (or rather a one-sided version of it). In this way we will be able to get an estimate for $f^{-n}(c-\epsilon, c+\epsilon)$ by induction.

Proposition 4.1. *Let f be a unimodal C^3 map with negative Schwarzian derivative and which satisfies (NF) and (SC). Suppose that there exists a constant K' such that for any n and any $\epsilon > 0$,*

$$|f^{-n}(c-\epsilon, c+\epsilon)| < K'\epsilon.$$

Then there exists a constant K such that for any measurable set A,

$$|f^{-n}(A)| \le K \cdot |A|^{\mathrm{el}}.$$

For the proof of this proposition we need some lemmas. First of all we need a one-sided version of the Koebe Principle, see Corollary 2 from Section IV.1.

Lemma 4.1 (A one-sided version of the Koebe Principle)**.** *Let g have negative Schwarzian derivative. Then for each $\rho \in (0,1)$ there exists $K < \infty$ with the following property. Assume that g is a diffeomorphism on an interval $T = [a,b]$ and choose $x \in [a,b]$ so that*

$$\frac{|g(x)-g(a)|}{|g(T)|} \geq \rho$$

then

$$|Dg(x)| \geq \frac{1}{K}|Dg(b)|.$$

Next we want to show that one can transport mass to the top or bottom of branches: let us show that in our estimation of $|f^{-n}(A)|$ we can assume A to be an interval which contains some extremal values of f^n.

Lemma 4.2. *Let g have negative Schwarzian derivative. Then there exists a constant $K < \infty$ with the following property. Let A be a measurable set and assume that g is a diffeomorphism on an interval $I = (\alpha, \beta)$. Let I_α and I_β be the maximal interval of length at most $|A|$ which are contained in $g(I)$ and which contains $g(\alpha)$ respectively $g(\beta)$. Then*

$$|g^{-1}(A) \cap I| \leq K \cdot \left\{ | \left(g^{-1}(I_\alpha \cup I_\beta) \right) \cap I| \right\}.$$

Proof. Follows immediately from the Minimum Principle. □

It follows that restricted to each branch of f^n one can compare the size of $f^{-n}(A)$ with the preimage under f^n of some special interval which contain a critical value $c_s = f^s(c)$ of f^n. However, each branch has different critical values and so one has to take different intervals. In the next lemma they are all transported to 'canonical intervals' of the form $(c-\epsilon, c+\epsilon)$. Here ϵ depends on how large s is. For each $\epsilon > 0$ define

$$A_i(\epsilon) = f^{-i}(c_1 - \epsilon, c_1)$$

for $i \in \mathbb{N}$. Then we have the following

Lemma 4.3. *Let f be as above. Then there exists a constant $K < \infty$ such that for any interval I for which (i) $f^n|I$ is monotone, (ii) f^n has a critical point at one of the endpoints of I and (iii) $|f^n(I)| \leq \epsilon$, one has*

$$I \subset A_i\left(\frac{K \cdot \epsilon}{|Df^{n-i}(c_1)|}\right) \text{ for some } 0 \leq i \leq n.$$

Proof. Let a be the (an) endpoint of I which is a critical point of f^n and let $0 \le i < n$ be such that $c_i = f^n(a)$; here we use the notation that $c_i = f^i(c)$. If $|f^n(a) - c_1| \le 3\epsilon$ then $I \subset A_n(4\epsilon)$ and the required inequality follows by taking $K = 4$. So assume that $|f^n(a) - c_1| > 3\epsilon$. Let D_n be the component of $f^{-n}[c_i - 3\epsilon, c_i + 3\epsilon]$ containing I and $C_n = D_n \cap f^{-n}[c_i - \epsilon, c_i + \epsilon]$. Let index$(D_n)$ be the minimum of $\{j \ge 1\,;\ c_j$ is a critical value of $f^n|D_n\}$. From $|f^n(a) - c_1| > 3\epsilon$ and the fact that a is a critical point of f^n it follows that index(D_n) is well-defined and that $1 <$ index$(D_n) \le n$. Let $s =$ index$(D_n) - 1$. From the way s is defined

1. $f^{n-s}(D_n)$ contains c_1 (and is contained in $(0, c_1]$);
2. $f^s|f^{n-s}(D_n)$ has some values in $[c_i - \epsilon, c_i + \epsilon]$ and also assumes the value $c_i - 3\epsilon$ or $c_i + 3\epsilon$;
3. $f^s|f^{n-s}(D_n)$ has no critical values.

From this and Lemma 4.1 one gets a universal constant $K < \infty$ such that

$$f^{n-s}(C_n) \subset [c_1 - \frac{K \cdot \epsilon}{|Df^s(c_1)|}, c_1].$$

Hence,

$$I \subset C_n \subset f^{-(n-s)}\left[c_1 - K \cdot \frac{\epsilon}{|Df^s(c_1)|}, c_1\right] \subset A_{n-s}\left(\frac{K \cdot \epsilon}{|Df^s(c_1)|}\right). \quad \square$$

Now we can prove Proposition 4.1 by simply adding all the contributions. One gets in this way:

Proof of Proposition 4.1: Let A be a measurable set of Lebesgue measure ϵ and consider a branch I of f^n. As before, let I_α and I_β be the maximal intervals of length $\le \epsilon$ each containing one of the endpoint of $f^n(I)$. From Lemma 4.2 one gets a constant $K_1 < \infty$ such that

$$|f^{-n}(A) \cap I| \le K_1 \cdot \left|\left(f^{-n}(I_\alpha) \cup f^{-n}(I_\beta)\right) \cap I\right|.$$

Using Lemma 4.3 one gets a constant $K_2 < \infty$ such that

$$\left(f^{-n}(I_\alpha) \cup f^{-n}(I_\beta)\right) \cap I \subset \bigcup_{i=0}^{n} A_i(\frac{K_2 \cdot \epsilon}{|Df^{n-i}(c_1)|}).$$

Since this is true for any branch one gets

$$|f^{-n}(A)| \le K_1 \cdot \sum_{i=0}^{n} |A_i(\frac{K_2 \cdot \epsilon}{|Df^{n-i}(c_1)|})|.$$

Using the assumption that the turning point of f has order ℓ one concludes from the assumption

$$|f^{-n}(c - \epsilon, c + \epsilon)| < K_3\epsilon$$

of the proposition that

$$|A_i(\epsilon)| \le K_4 \epsilon^{\mathrm{el}} \quad \text{for all} \quad i \ge 0.$$

Combining the last two inequalities one gets that

$$|f^{-n}(A)| \le K_5 \cdot \sum_{i=0}^{n} \left(\frac{\epsilon}{|Df^{n-i}(c_1)|} \right)^{\mathrm{el}}.$$

Using the summability condition (SC) the proposition follows. □

Step 2: Branches which will be 'slided' later on

The main idea in the proof of

$$|f^{-n}(c-\epsilon, c+\epsilon)| < K''\epsilon$$

is to show that each component of $f^{-n}(c-\epsilon, c+\epsilon)$ is either contained in or at least can be compared in size (this process we will call *sliding*) with a set of the form

$$f^{-(n-k)}\left(c - \frac{C\epsilon}{|Df^k(c_1)|^{\mathrm{el}}}, c + \frac{C\epsilon}{|Df^k(c_1)|^{\mathrm{el}}} \right).$$

Using this and the summability condition, inequality $|f^{-n}(c-\epsilon, c+\epsilon)| < K''\epsilon$ will then be proved by induction.

Let us explain the idea of sliding first in words. Suppose that we have proved by induction that $|f^{-k}(c-\epsilon, c+\epsilon)| < K''\epsilon$ for $k < n$. Consider a component I of $f^{-n}(c-\epsilon, c+\epsilon)$ such that f^n is monotone on I. Then let T be the largest interval containing I on which f^n is monotone and let n_1 be the largest integer $< n$ such that $f^{n_1}(T)$ contains c in its closure. Furthermore, let R_1 be the interval in $f^{n_1}(T)$ between $I_1 = f^{n_1}(I)$ and c and let L_1 the other component of $f^{n_1}(T) \setminus I_1$. Let us explain why there are three different cases. The first case (which we will later subdivide into the transport case and the regular case) is that I_1 is quite close the c in the sense that $|R_1| \le |I_1|$. In this case we will estimate the length of I (which is a component of $f^{-n_1}(I_1)$) from above by

$$|f^{-n_1}(I_1 \cup R_1)| \le |f^{-n_1}(c-\epsilon', c+\epsilon')| < K''\epsilon'$$

where $\epsilon' > 0$ is so that $I_1 \cup R_1 \subset (c-\epsilon', c+\epsilon')$.

However, if $|I_1|$ is much smaller than $|R_1|$ the estimate we would obtain is extremely bad and so we have to find a different strategy. The second and third possibilities are related to sliding. The second possibility is that $|R_1| \ge |I_1|$, so I_1 is far away from c, but, on the other hand, $|L_1|$ (or rather the length of some related interval $A_1 \supset L_1$ which we will introduce below) is large compared to $|R_1|$. In this case we have a lot of space on both sides of I_1 and, using a Koebe argument, we will be able to 'slide' the interval I_1 closer to c. So we will compare the size of I with the size of $f^{-n_1}(J)$ where J is some interval of roughly the same size as I_1 but which does have c in its boundary. This sliding will be explained in Step 2.b below. The third and final possibility is that $|R_1| \le |I_1|$ but that there is not much space on the other side. In that case we cannot 'slide' and in the present step we will analyze this situation. We show that all the pullbacks $f^n(I)$ can be estimated until we reach a moment when we can 'slide'. At that moment we are again in the second case. The stopping rule for this is given by $(***a)$ and $(***b)$ below.

Let us be more precise now. Let f^n be monotone on an interval I and assume that $f^n(I) = (c - \epsilon, c + \epsilon)$. Let T be the largest interval containing I on which f^n is monotone and label the endpoints α and β of T so that $|f^n(\alpha) - c| \le |f^n(\beta) - c|$. Denote the endpoints of I by γ and δ so that either $\alpha < \gamma < \delta < \beta$ or $\alpha > \gamma > \delta > \beta$. In this section we will assume that

$$(*) \qquad |f^n(\alpha) - f^n(\gamma)| \ge 2\epsilon.$$

Let

$$A_0 = f^n(\alpha, \gamma), I_0 = f^n(I) = f^n(\gamma, \delta), R_0 = f^n(\delta, \beta).$$

By $(*)$ one has

$$(**) \qquad |R_0| \ge |A_0| \ge |I_0|.$$

The reader should think here of the case that A_0 is much bigger than I_0. This means that although we could apply the Macroscopic Koebe Principle because there is space on both sides of I_0 and $|I_0| = 2\epsilon$, the ratio $|I|/|T|$ is not necessarily smaller than some universal constant times ϵ. (In order to get this we would need to assume that there exists a universal constant $\sigma > 0$ such that $|R_0|, |A_0| \ge \sigma$. We will treat this case separately in Step 3 below.)

Later on we shall show that if $|R_0|$ is not too large compared to $|A_0|$ then the set $I_0 = f^n(I)$ can be 'slided'. In this section we will show that if I_0 cannot be 'slided' at least some smaller iterate $f^{k_s}(I)$ of I can be 'slided'. If $|A_0 \cup I_0| \ge |R_0|$ then set $s = 0$ and we are finished. Otherwise we shall define inductively a finite sequence of intervals $T^i = [\alpha_i, \alpha_{i-1}]$ and integers n_i as follows. Let $n_0 = n$, $\alpha_0 = \alpha$, $\alpha_{-1} = \beta$ and $T^0 = [\alpha_0, \alpha_{-1}] = [\alpha, \beta]$ (i.e. $T^0 = T$). By maximality of T one can choose n_1 such that $0 < n_1 < n$ and $f^{n_1}(\alpha_0) = c$. Now choose α_1 such that T^1 is the maximal interval of the form $T^1 = [\alpha_1, \alpha_0]$ which contains T^0 and on which f^{n_1} is monotone (of course one may have $T^1 = T$). Now assume that n_{i-1} and $T^{i-1} = [\alpha_{i-2}, \alpha_{i-1}]$ are defined. Then simply define $n_i < n_{i-1}$ such that $f^{n_i}(\alpha_{i-1}) = c$, and let T^i be the maximal interval of the form $[\alpha_i, \alpha_{i-1}]$ which contains $T^{i-1} = [\alpha_{i-2}, \alpha_{i-1}]$ and on which f^{n_i} is monotone. It follows that for $i \ge 1$, T^i and T^{i-1} have precisely one common boundary point and that

$$I \subset T^0 \subset \cdots \subset T^i.$$

Figure 4.1: The intervals R_i, I_i, L_i and A_i.

Let us now define the integers k_i and intervals I_i, R_i, A_i, L_i as follows:

$$k_i = n_i - n_{i+1},\ I_i = f^{n_i}(I),$$

$$R_i = f^{n_i}(\alpha_{i-1}, \gamma) \setminus I_i\ ,\ A_i = f^{n_i}(\alpha_i, \gamma) \setminus I_i\ ,\ L_i = f^{n_i}(\alpha_{i-2}, \gamma) \setminus I_i.$$

In other words, R_i is the component of $f^{n_i}(T^i \setminus I)$ which contains c in its closure and A_i is the other component. Furthermore, L_i is contained in A_i and

$$f^{k_i}(I_{i+1}) = I_i,\ f^{k_i}(R_{i+1}) = A_i,\ f^{k_i}(L_{i+1}) = R_i,$$

for all $i = 0, \ldots, s-1$. We stop the construction at $i = s$, when

$$(***a) \qquad |A_s \cup I_s| \geq |R_s|$$

or when $n_s = 0$. In particular,

$$(***b) \qquad |A_i \cup I_i| \leq |R_i|,\ \text{for } i = 0, 1, \ldots, s-1.$$

That $(***b)$ holds implies that $f^{k_i}(I_i)$, $i = 1, 2, \ldots, s-1$ is very far from c compared to its size. This is bad: if we wanted to get back into induction we would need an interval containing the critical point. The smallest interval containing I_i with the property is $R_i \cup I_i$ and so this interval could be very large compared to I_i for $i = 1, 2, \ldots, s-1$. If $n_s > 0$ then $f^{k_s}(I_s)$ is relatively close to c and, as we will later see, this means that we can 'slide' I_s towards c. Instead of I_s we will consider another interval J which contains c and which will allow us to get back into induction. This interval J is the 'slided' one. If $n_s = 0$ then there will be no need to do this sliding.

The main result of this section is that as long as one has $(***b)$ one can estimate the effects of this pulling back completely. So although we would rather be in case $(***a)$, condition $(***b)$ allows us to apply Koebe and get a very good estimate on the size of I_s. Note that the intervals I_i are not of the 'canonical form' $(c-\epsilon, c+\epsilon)$ and we will, therefore, not use induction to get back to the canonical form till the very end. As before, let ℓ be the order of the critical point. More precisely we will prove the following

Proposition 4.2. *There exist constants K', K'' such that*

$$|I_s| \leq K' \frac{|f^n(I)|}{\prod_{i=0}^{s-1} K'' |Df^{k_i}(c_1)|^{\text{el}}}\ .$$

In order to prove this proposition we need some lemmas:

Lemma 4.4.

$$|I_s| \leq |I_1| \frac{|A_1 \cup I_1 \cup R_1|}{|A_1|} \prod_{i=1}^{s-1} \frac{|R_{i+1}|}{|R_i|}\ .$$

Proof. Since cross-ratio's are expanded by f^{k_i} we have

$$\frac{|f^{k_i}(L_{i+1})|}{|L_{i+1}|}\frac{|f^{k_i}(R_{i+1})|}{|R_{i+1}|} < \frac{|f^{k_i}(I_{i+1})|}{|I_{i+1}|}\frac{|f^{k_i}(L_{i+1}\cup I_{i+1}\cup R_{i+1})|}{|L_{i+1}\cup I_{i+1}\cup R_{i+1}|} .$$

By definition of the sequences of intervals A_i, L_i, I_i, R_i this is equivalent to

$$|I_{i+1}| \le |I_i|\frac{|L_{i+1}|}{|A_i|}\frac{|R_i\cup I_i\cup A_i|}{|R_{i+1}\cup I_{i+1}\cup L_{i+1}|}\frac{|R_{i+1}|}{|R_i|} .$$

By induction we get

$$|I_s| \le |I_1|\frac{|A_1\cup I_1\cup R_1|}{|A_1|}\times\prod_{i=1}^{s-1}\frac{|R_{i+1}|}{|R_i|}\times$$

$$\times\prod_{i=2}^{s-1}\left(\frac{|L_i|}{|A_i|}\frac{|R_i\cup I_i\cup A_i|}{|R_i\cup I_i\cup L_i|}\right)\times\frac{|L_s|}{|L_s\cup I_s\cup R_s|}.$$

The last factor is clearly less than 1, and we can say the same about the last but one factor because these terms are of the form $\frac{l(a+w)}{a(l+w)}$ and because $l(a+w) < a(l+w)$ for positive l, a, w and $l < a$. □

Next we need two distortion lemmas. In order to clarify the proof we use the notation $O(\tau)$ for a constant which only depends on τ and $O(KL)$ for a universal constant which comes from the Koebe Principle. $O(NF)$ stands for a universal constant which depends on the non-flatness condition.

Lemma 4.5. *Assume that f^k is a diffeomorphism on (c, w) and that for some $z \in (c, w)$ one has $f^k(z) = c$ and $|f^k(c, z)| < \tau|f^k(z, w)|$ for some $\tau \in (0, 1)$. Then*

$$\frac{|f^k(c,z)|}{|(c,z)|} \ge O(\tau)|Df^k(c_1)|^{\mathrm{el}} .$$

Proof. Using the chain-rule, the non-flatness condition, the one sided Koebe Principle (see Lemma 4.1), and $f^k(z) = c$ one has

$$\begin{aligned}|Df^k(c_1)| &= |Df(f^{k-1}(c_1))||Df^{k-1}(c_1)|\\ &\le O(NF)|f^{k-1}(c_1)-c|^{\ell-1}|Df^{k-1}(c_1)|\\ &\le O(NF)O(KL,\tau)|f^{k-1}(c_1)-c|^{\ell-1}\frac{|f^{k-1}(f(c,z))|}{|f(c,z)|}\\ &= O(NF)O(KL,\tau)\frac{|f^k(c,z)|^{\ell}}{|f(c,z)|}.\end{aligned}$$

Using again the non-flatness condition gives the required estimate. □

Lemma 4.6. *Assume that f^k is a diffeomorphism on (c, z), that $f^k(z) = c$ and that for some $y \in (c, z)$ one has $|f^k(c, y)| < \tau|f^k(y, z)|$ for some $\tau \in (0, 1)$. Then*

$$\frac{|f^k(y,z)|}{|(c,y)|} \ge O(\tau)|Df^k(c_1)|^{\mathrm{el}} .$$

Proof. From the non-flatness condition and since $|f^k(y,z)| \geq \frac{1}{\tau}|f^k(c,y)|$ one has

$$\left(\frac{|f^k(y,z)|}{|c,y|}\right)^\ell \geq O(NF)\frac{|f^k(y,z)|^{\ell-1}|f^k(y,z)|}{|f(c,y)|}$$
$$\geq \frac{O(NF)}{\tau}|f^k(y,z)|^{\ell-1}\frac{|f^k(c,y)|}{|f(c,y)|} .$$

Since $|f^k(c,y)| < \tau|f^k(y,z)|$ one gets from the one-sided Koebe Lemma 4.1 that the last factor is at least $O(\tau, KL)|Df^{k-1}(c_1)|$. Moreover, one has $|f^k(y,z)| \geq \frac{1}{1+\tau}|f^k(c,z)| = \frac{1}{1+\tau}|f^k(c) - c|$. From all this

$$\left(\frac{|f^k(y,z)|}{|(c,y)|}\right)^\ell \geq O(1)|f^k(c) - c|^{\ell-1}|Df^{k-1}(c_1)| .$$

By the non-flatness condition $|f^k(c) - c|^{\ell-1} \geq O(NF)|Df(f^k(c))|$, and, therefore, the lemma follows. □

Lemma 4.7. *If $s > 0$ then there exists a constant $K < \infty$ such that*

$$|I_1|\frac{|A_1 \cup I_1 \cup R_1|}{|A_1|} \leq K\frac{|f^n(I)|}{|Df^{k_0}(c_1)|^{\text{el}}} ,$$

$$\frac{|f^{k_0}(R_1)|}{|R_1|} \geq \frac{1}{K}|Df^{k_0}(c_1)|^{\text{el}},$$

$$\frac{|R_i|}{|R_{i+1}|} \geq \frac{1}{K}|Df^{k_i}(c_1)|^{\text{el}} \text{ for } i = 1, \ldots, s-1$$

Proof. Since cross-ratio's are expanded by f^{k_i} we have

$$|I_1|\frac{|A_1 \cup I_1 \cup R_1|}{|A_1|} \leq |f^{k_0}(I_1)|\frac{|R_1|}{|f^{k_0}(R_1)|}\frac{|f^{k_0}(A_1 \cup I_1 \cup R_1)|}{|f^{k_0}(A_1)|} \leq$$

$$\leq |f^n(I)|\frac{|R_1|}{|f^{k_0}(R_1)|}\left(1 + \frac{|I_0 \cup A_0|}{|f^{k_0}(L_1)|}\right)$$
$$\leq |f^n(I)|\frac{|R_1|}{|f^{k_0}(R_1)|}\left(1 + \frac{|I_0 \cup A_0|}{|R_0|}\right).$$

By $(**)$ this gives

$$|I_1|\frac{|A_1 \cup I_1 \cup R_1|}{|A_1|} \leq |f^n(I)|\frac{|R_1|}{|f^{k_0}(R_1)|} \cdot 3.$$

Therefore, in order to prove the second and third inequality it is enough to prove that $|f^{k_0}(R_1)|/|R_1| > O(1)|Df^{k_0}(c_1)|^{\text{el}}$. So let us prove this. One has $f^{k_0}(R_1) = A_0$. Let $R_1' \subset R_1 \cup I_1$ be the smallest interval containing R_1 such that $f^{k_0}(R_1')$ contains c. We want to apply Lemma 4.5 by taking (c, w) to be the interval T_1 and (c,z) the interval R_1'. Since $f^{k_0}(T_1 \setminus R_1') \supset f^{k_0}(L_1) = R_0$ and since $s > 0$ we get from $(***b)$,

$$|f^{k_0}(T_1 \setminus R_1')| \geq |R_0| \geq |A_0 \cup I_0| \geq |f^{k_0}(R_1')|.$$

From all this it follows that we can apply Lemma 4.5 and get that

$$|f^{k_0}(R_1')|/|R_1'| > O(1)|Df^{k_0}(c_1)|^{\mathrm{el}}.$$

But since $|f^{k_0}(R_1')| = |A_0| + \epsilon < 2|A_0| = 2|f^{k_0}(R_1)|$, this implies

$$\frac{|f^{k_0}(R_1)|}{|R_1|} \geq \frac{1}{2}\frac{|f^{k_0}(R_1')|}{|R_1'|} > \frac{O(1)}{2}|Df^{k_0}(c_1)|^{\mathrm{el}}.$$

So the first two statements of the lemma follow.

To prove the third statement let as before $R_{i+1}' \subset R_{i+1} \cup I_{i+1}$ be the smallest interval containing R_{i+1} such that $f^{k_i}(R_{i+1}')$ contains c. Because the construction did not stop at i, $|f^{k_i}(L_{i+1})| = |R_i| \geq |A_i \cup I_i| \geq |f^{k_i}(R_{i+1}')|$. Therefore, one can apply Lemma 4.6 and we get

$$\frac{|R_i|}{|R_{i+1}|} \geq \frac{|R_i|}{|R_{i+1}'|} = \frac{|f^{k_i}(L_{i+1})|}{|R_{i+1}'|} \geq \frac{|f^{k_i}(R_{i+1}')|}{|R_{i+1}'|} \geq O(1)|Df^{k_i}(c_1)|^{\mathrm{el}}.$$

This proves the third statement of the lemma. □

Proof of Proposition 4.2: This follows immediately from Lemmas 4.4 and 4.7. □

Step 3: A subdivision into three cases

In this section we shall prepare the estimates for the preimages of the intervals around the critical point c. So consider the set $E(\delta) = (c - \delta, c + \delta)$ and its preimages $E_n(\delta) = f^{-n}(E(\delta))$. For a given $\epsilon > 0$ we shall subdivide the collection of components of $E_n(\epsilon)$ into three subcollections.

Let σ be some positive number and let $\epsilon \in (0, \frac{1}{2}\sigma)$. We define $\nu(\sigma)$ as $\inf\{k > 0\,;\ f^k(c) - c| < \sigma\}$. Clearly $\nu(\sigma)$ is monotone and $\nu(\sigma)$ tends to infinity as $\sigma \to 0$. Later on we shall choose σ appropriately.

Let I be a component of $E_n(\epsilon)$. Suppose that $I \subset I' \subset I''$, where I' is a component of $E_n(2\epsilon)$ and I'' is a component of $E_n(\sigma)$. If $Df^n_{|I''} \neq 0$ then I belongs to the collection $\mathcal{R}_n$. If $I \notin \mathcal{R}_n$ but $Df^n_{|I'} \neq 0$ then I belongs to the collection $\mathcal{S}_n$. All the other components form the collection $\mathcal{T}_n$. These are the three cases mentioned above.

Step 3.a: The collection $\mathcal{R}_n$, the regular case

If $I \in \mathcal{R}_n$, then f^n is a diffeomorphism on I and there exists $\gamma \in I$ such that $f^n(\gamma) = c$. Let (α, β) be the maximal interval containing I on which f^n is a diffeomorphism. By definition of $\mathcal{R}_n$ we have $|f^n(\alpha,\gamma)|$, $|f^n(\beta,\gamma)| \geq \sigma$. Therefore, we can use the one-sided Koebe Lemma on I and obtain:

Proposition 4.3. *There exists a constant K_R such that for $\epsilon < \sigma/2$ and any regular component I as above, one has*

$$\frac{|I|}{|(\alpha,\beta)|} \le K_R \frac{\epsilon}{\sigma}.$$

Proof. If $y \in I$ then $f^n(\alpha, y)$ and $f^n(y, \beta)$ have at least length σ. Therefore, by the one-sided Koebe Principle, one has

$$|Df^n(x)| \le O(KL)|Df^n(y)|$$

for any $x \in (\alpha, \beta)$ and $y \in I$. Hence,

$$\frac{\sigma}{|(\alpha,\beta)|} \le \frac{|f^n(\alpha,\beta)|}{|(\alpha,\beta)|} \le O(1)\frac{|f^n(I)|}{|I|} \le O(1)\frac{\epsilon}{|I|}. \quad \square$$

Corollary 4.1. *For $I \in \mathcal{R}_n$ let $\Delta_n(I)$ be the maximal interval on which f^n is a diffeomorphism. By the previous proposition we obtain*

$$\sum_{I\in\mathcal{R}_n} |I| \le \sum_{I\in\mathcal{R}_n} \frac{|I|}{|\Delta_n(I)|}|\Delta_n(I)| \le K_R \frac{\epsilon}{\sigma} \sum_{I\in\mathcal{R}_n} |\Delta_n(I)| \le K_R \frac{\epsilon}{\sigma}\ .$$

Step 3.b: The collection $\mathcal{S}_n$, the case to slide

Let $I \in \mathcal{S}_n$. Then there exists $s \ge 0$ and a sequence $n_s < n_{s-1} < \cdots < n_0 = n$ as in Step 2 such that if $n_s > 0$ (in the terminology used there) $|A_s \cup I_s| > |R_s|$. Since f^{n_s} is a diffeomorphism on $T^s \supset I$, since $f^{n_s}(T^s) \supset R_s$ and since R_s contains c in its closure there exists an interval $J \subset T^s$ such that $G = f^{n_s}(J)$ contains c and such that

$$|I| = |J|.$$

In other words, by choosing an appropriate $x_s \in [x'_s, c] = I_s \cup R_s$ one can assure that the preimage J of $G = [x_s, c]$ has the same size as I. This process we call sliding. Note that $f^{n_s}(J)$ contains c. Because $|A_s \cup I_s| > |R_s|$ we can use the one-sided Koebe Lemma 4.1 and obtain

$$\frac{|f^{n_s}(J)|}{|J|} \le O(KL)\frac{|f^{n_s}(I)|}{|I|}$$

and, therefore, $|G| \le O(1)|I_s|$. Therefore, by Proposition 4.2,

$$|G| \le \frac{O(1)K'|f^n(I)|}{\prod_{j=0}^{s-1} K''|Df^{k_j}(c_1)|^{\mathrm{el}}} \le \frac{O(1)K'2\epsilon}{\prod_{j=0}^{s-1} K''|Df^{k_j}(c_1)|^{\mathrm{el}}}\ .$$

So for each such component I of $E_n(\epsilon) = f^{-n}(E(\epsilon))$, there exists an interval J as above such that $|I| = |J|$ and, therefore, such that $|I|$ is at most the size of the f^{n_s}-preimage of

$$\left(c - \frac{O(1)K'2\epsilon}{\prod_{j=0}^{s-1} K''|Df^{k_j}(c_1)|^{\mathrm{el}}}, c + \frac{O(1)K'2\epsilon}{\prod_{j=0}^{s-1} K''|Df^{k_j}(c_1)|^{\mathrm{el}}}\right)$$

that is contained in T^s. Now even for a given sequence of $0 < n_s < n_{s-1} < \cdots < n_0 = n$, there may be several such components I in T^s. Even worse, some of these may give the same interval J (or at least overlapping intervals). But for every given sequence of $n_s < n_{s-1} < \cdots < n_0 = n$, there exist at most 2^s different components I of $E_n(\epsilon)$ of type $\mathcal{S}_n$, such that the corresponding intervals J overlap. Indeed, at the first step of the construction two intervals I and $\tilde{I}$ can only slide onto overlapping intervals J^1 and $\tilde{J}^1$ if there is precisely one turning point of f^{n_1} between these two intervals. Similarly at the i-th step two intervals J^{i-1} and $\tilde{J}^{i-1}$ can only slide onto overlapping intervals J^i and $\tilde{J}^i$ if there is precisely one turning point of f^{n_i} between these two intervals. So at each step the number of intervals $I \in \mathcal{S}_n$ which correspond to overlapping intervals J can at most double. Thus we get

$$\sum_{\{I\in\mathcal{S}_n\,;\,n_s>0\}} |I| \leq \sum_{\sum_{j=0}^{s-1} k_j \leq n} 2^s |f^{-n_s}(c-|G|, c+|G|)|.$$

Clearly also

$$\sum_{I\in\mathcal{S}_n\,;\,n_s=0} |I| \leq \sum_{\sum_{j=0}^{s-1} k_j \leq n} |f^{-n_s}(c-|G|, c+|G|)|\ .$$

So

$$(*) \qquad \sum_{I\in\mathcal{S}_n} |I| \leq \sum_{\sum_{j=0}^{s-1} k_j \leq n} 2^s \left| f^{-n_s}(E\left(K_S \frac{\epsilon}{\prod_{j=1}^{s-1} K''|Df^{k_j}(c_1)|^{\mathrm{el}}}\right)) \right|.$$

Lemma 4.8. *There exists $\sigma_0 > 0$ such that $k_0, \ldots, k_{s-1} \geq \nu(\sigma)$ for each $\sigma \in (0, \sigma_0)$.*

Proof. Choose $\sigma_0 > 0$ so small that for each $\sigma \in (0, \sigma_0)$ and each $k > \nu(\sigma)$ one has $|Df^k(c_1)|^{\mathrm{el}} > 2K$ where K is the constant from Lemmas 4.5. Because $\nu(\sigma) \to \infty$ as $\sigma \to 0$ and $|Df^k(c_1)| \to \infty$ as $k \to \infty$ this is possible.

By definition of k_0, $f^{k_0}(c)$ is contained in the closure of A_0. Since $I \in \mathcal{S}_n$, at least one critical value of $f^n|T_0$ is contained in $(c-\sigma, c+\sigma)$. As $T_0 = [\alpha, \beta]$ and as $|f^n(\alpha) - c| \leq |f^n(\beta) - c|$, this implies $f^n(\alpha) = f^{k_0}(c) \in (c-\sigma, c+\sigma)$. Hence, $k_0 \geq \nu(\sigma)$ and, observing the definition of A_0, we have $A_0 \subset (c-\sigma, c+\sigma)$. Next notice that

$$(**) \qquad |R_i \cup I_i \cup A_i| \leq \sigma \text{ implies } k_i \geq \nu(\sigma), \text{ for } i = 1, \ldots, s-1.$$

Indeed, c is contained in the closure of $R_i \cup I_i \cup A_i$, this interval has at most length σ and $f^{k_i}(c)$ is contained in the closure of A_i. Therefore, $k_i \geq \nu(\sigma)$ follows from the definition of $\nu(\sigma)$.

Let us show by induction that for $\sigma \in (0, \sigma_0)$, $|R_i \cup I_i \cup A_i| \leq \sigma$ for $i = 1, \ldots, s-1$. Assume $s \geq 2$ and let us first prove that this inequality holds for $i = 1$. From Lemma 4.7, $|f^{k_0}(R_1)| \geq \frac{1}{K}|Df^{k_0}(c_1)|^{\mathrm{el}}\,|R_1|$. Therefore,

$$|R_1 \cup I_1 \cup A_1| \leq 2|R_1| \leq 2K \frac{|f^{k_0}(R_1)|}{|Df^{k_0}(c_1)|^{\mathrm{el}}}$$

$$= 2K \frac{|A_0|}{|Df^{k_0}(c_1)|^{\mathrm{el}}} \leq |A_0|,$$

where the last inequality holds provided $\sigma \in (0, \sigma_0)$ because $k_0 \geq \nu(\sigma)$, and $|Df^k(c_1)| > 2K$ for $k \geq \nu(\sigma)$. Since $|A_0| \leq \sigma$, the induction assertion is proved for $i = 1$. Similarly, we get for $i < s$, using the third inequality of Lemma 4.7, that

$$|R_i \cup I_i \cup A_i| \leq 2|R_i| \leq 2K \frac{|R_{i-1}|}{|Df^{k_{i-1}}(c_1)|^{\mathrm{el}}}.$$

From the inductive assumption we know that $|R_{i-1} \cup I_{i-1} \cup A_{i-1}| \leq \sigma$ and from $(**)$ this implies $k_{i-1} > \nu(\sigma)$ and so we get again

$$|R_i \cup I_i \cup A_i| \leq |R_{i-1}| \leq |R_{i-1} \cup I_{i-1} \cup A_{i-1}| \leq \sigma$$

when $\sigma \in (0, \sigma_0)$. Therefore, one has $k_j \geq \nu(\sigma)$. □

Using this, $(*)$ and the estimate for $|G|$ one gets the following

Proposition 4.4. *There exists a constant K_S such that for $\sigma \in (0, \sigma_0)$,*

$$\sum_{I \in \mathcal{S}_n} |I| \leq \sum_{\substack{k_j \geq \nu(\sigma) \\ \sum_{j=0}^{s-1} k_j \leq n}} 2^s \left| f^{-n_s}(E\left(K_S \frac{\epsilon}{\prod_{j=1}^{s-1} K'' |Df^{k_j}(c_1)|^{\mathrm{el}}}\right)) \right| .$$

Step 3.c: The collection $\mathcal{T}_n$, the case to transport

We shall now reduce the estimation of the n-th preimage $I \in \mathcal{T}_n$ to the estimation of some k-th preimage, with $k < n$. Let $I \in \mathcal{T}_n$ and let I' (resp. I'') be the component of $E_n(2\epsilon)$ (resp. $E_n(\sigma)$) containing I. By definition f^n has at least one critical point in $I' \supset I$.

Since f^n has a critical point in I' there exists an integer $k < n$ such that $c \in f^k(I')$. Let k be the largest such integer. For simplicity we say that I belongs to the subcollection $\mathcal{T}_n^k$ of $\mathcal{T}_n$. From the properties just stated one has

$$f^{n-k}(c) \in (c - 2\epsilon, c + 2\epsilon).$$

Since $c \notin f^i(I')$ for $i = k+1, \ldots, n-1$, the map f^{n-k-1} is clearly a diffeomorphism on $f^{k+1}(I')$.

Proposition 4.5. *There exists a constant K_T such that for every n*

$$\sum_{I \in \mathcal{T}_n} |I| = \sum_k \sum_{I \in \mathcal{T}_n^k} |I| \leq \sum_k \left| f^{-k}(E\left(\frac{K_T \epsilon}{|Df^{n-k}(c_1)|^{\mathrm{el}}}\right)) \right| .$$

Proof. Let $f^{k+1}(I') = (x, c_1] \supset f^{k+1}(I)$. As we saw f^{n-k-1} is a diffeomorphism on (x, c_1). Moreover, $f^n(I) \subset (c - \epsilon, c + \epsilon)$, $f^{n-k-1}(x) = c \pm 2\epsilon$ and $f^n(I') \subset$

$(c-2\epsilon, c+2\epsilon)$. Therefore, one gets from the one sided Koebe Principle (see Lemma 4.1) immediately that

$$\frac{|f^n(I)|}{|f^{k+1}(I)|} \geq O(KL)|Df^{n-k-1}(c_1)| \ .$$

Hence,

$$|f^{k+1}(I)| \leq O(1) \cdot \frac{\epsilon}{|Df^{n-k-1}(c_1)|} \ .$$

From the non-flatness condition this gives

$$|f^k(I)| \leq O(NF) \cdot |f^{k+1}(I)|^{\mathrm{el}} \leq O(1) \left(\frac{\epsilon}{|Df^{n-k-1}(c_1)|} \right)^{\mathrm{el}} \ .$$

Since $f^{n-k-1}(c_1) \in (c-2\epsilon, c+2\epsilon)$, the non-flatness condition implies that

$$\left(\frac{\epsilon}{|Df^{n-k-1}(c_1)|} \right)^{\mathrm{el}} \leq O(1) \cdot \frac{\epsilon}{|Df^{n-k}(c_1)|^{\mathrm{el}}} \ ,$$

i.e.,

$$|f^k(I)| \leq O(1) \cdot \frac{\epsilon}{|Df^{n-k}(c_1)|^{\mathrm{el}}} \ .$$

Since $f^n(I) \subset (c-\epsilon, c+\epsilon)$ it follows that there exists a constant K_T such that

$$I \subset f^{-k}(f^k(I)) \subset f^{-k}E\left(\frac{K_T\epsilon}{|Df^{n-k}(c_1)|^{\mathrm{el}}} \right) \ .$$

The proposition follows. □

Step 4: The proof of the Theorem 4.1

Let σ be so small that for every n,

$$(*) \qquad \sum_{\substack{k_j \geq \nu(\sigma), 1 \leq j \leq s \\ \sum k_j \leq n}} 3K_S \prod_{j=0}^{s-1} \frac{2}{K''|Df^{k_j}(c_1)|^{\mathrm{el}}} \leq 1 \ ,$$

and

$$(**) \qquad \sum_{k > \nu(\sigma)} 3K_T |Df^k(c_1)|^{-\,\mathrm{el}} \leq 1 \ .$$

That one can choose σ so that $(**)$ holds simply follows from the summability condition and because $\nu(\sigma)$ tends to infinity as $\sigma \to 0$. That $(*)$ is possible follows also from this and because of the following

Lemma 4.9. *Suppose that $d_k \geq 0$ and $\sum_{k=0}^{\infty} d_k < \infty$. Then for any $\eta, \zeta > 0$ there exists ν_0 such that*

$$P = \sum_n \prod_{\substack{k_j \geq \nu_0 \\ \sum_{j=0}^{s-1} k_j \leq n}} (\eta d_{k_j}) < \zeta \ .$$

Proof. Consider $S_{k_0} = \sum_{k>k_0} \eta d_k$. Then both S_{k_0} and $S = \sum_{s\in\mathbb{N}} S_{k_0}^s$ tend to zero when k_0 tends to infinity. Clearly $P \le S$ which proves the assertion. □

We shall now prove Theorem 6.1 in the following formulation:

Theorem 4.2. *For any n and $\epsilon \le \sigma/2$,*

$$|f^{-n}(c-\epsilon, c+\epsilon)| \le 3K_R \frac{\epsilon}{\sigma} \ .$$

Proof. With the notations from the previous section we have

$$f^{-n}(c-\epsilon, c+\epsilon) = \bigcup_{I\in\mathcal{R}_n} I \cup \bigcup_{I\in\mathcal{S}_n} I \cup \bigcup_{I\in\mathcal{T}_n} I \ ,$$

and

$$|f^{-n}(c-\epsilon, c+\epsilon)| \le \sum_{I\in\mathcal{R}_n} |I| + \sum_{I\in\mathcal{S}_n} |I| + \sum_{I\in\mathcal{T}_n} |I| \ .$$

Therefore,

$$\begin{aligned} |f^{-n}(c-\epsilon, c+\epsilon)| \le \quad & K_R\frac{\epsilon}{\sigma} \quad + \\ (***) \qquad & \sum_{\substack{k_j\ge\nu(\sigma), 0\le j\le s-1 \\ \sum k_j \le n}} 2^s \left| f^{-n_s}(E\left(K_S \frac{\epsilon}{\prod_{j=0}^{s-1} K''|Df^{k_j}(c_1)|^{\mathrm{el}}}\right))\right| \\ & + \sum_{k\ge\nu(\sigma)} \left| f^{-k}(E\left(K_T\frac{\epsilon}{|Df^{n-k}(c_1)|^{\mathrm{el}}}\right))\right| \ . \end{aligned}$$

We shall now prove the theorem by induction. For n small only the first term in $(***)$ is non-zero and, consequently, the assertion of the theorem is true. Suppose that it is true for any $\epsilon < \sigma/2$ and any $n < N$. Then by the choice of σ, and by $(*)$, $(**)$, $(***)$ and the induction assumption we have

$$\begin{aligned} |f^{-N}(c-\epsilon, c+\epsilon)| &\le K_R\frac{\epsilon}{\sigma} \\ &\quad + \sum_{\substack{k_j\ge\nu(\sigma), 0\le j\le s-1 \\ \sum k_j \le N}} 2^s K_S \frac{3K_R\epsilon/\sigma}{\prod_{j=0}^{s-1} K''|Df^{k_j}(c_1)|^{\mathrm{el}}} \\ &\quad + \sum_{\nu(\sigma)\le k\le N} K_T \frac{3K_R\epsilon/\sigma}{|Df^{N-k}(c_1)|^{\mathrm{el}}} \\ &\le 3K_R\frac{\epsilon}{\sigma} \ , \end{aligned}$$

which completes the proof. □

5 Transitive Maps Without Finite Continuous Measures

In this section we will show that many maps from the quadratic family $x \mapsto ax(1-x)$ do not have an absolutely continuous invariant probability measure, even if they are transitive on some interval. This result is due to Johnson (1987). In fact, we will prove a more general result due to Hofbauer and Keller (1990a) and (1990b).

In Section 1 of this chapter we have seen that unimodal maps $f\colon [0,1] \to [0,1]$ with negative Schwarzian derivative and with a repelling fixed point in 0 can be of three types.

1. f has a periodic attractor;
2. f is infinitely often renormalizable;
3. f has no periodic attractor and is not infinitely often renormalizable. In this case f is ergodic and transitive on a finite union $I_1 \cup \dots I_p$ intervals.

Furthermore, either

a) $\omega(c)$ is a minimal Cantor set and the attractor A of f is $\omega(c)$, i.e., for almost all points x one has $\omega(x) = \omega(c) = A$, or,

b) the attractor A of f is the finite union of intervals: for almost all points $x \in [0,1]$ one has $\omega(x) = A$.

We should emphasize that a non-renormalizable map for which $\omega(c)$ is a minimal Cantor set can still be as in 3b). In fact, the Fibonacci map from Example 2 in Section II.3.b is such a map: if such a map has negative Schwarzian derivative and a non-degenerate critical point then it has an absolutely continuous invariant probability measure, see Lyubich and Milnor (1991). As we saw in Section 1 this implies that $\omega(x)$ is the interval $[f^2(c), f(c)]$ for almost all $x \in [0,1]$ and so this map is as in 3b). For a long time it was believed that any map as in 3) would have a finite absolutely continuous invariant measure. (Because, as we have seen in Section 1, the set $\omega(c)$ has Lebesgue measure zero if it contains no intervals and because the support of an invariant measure is contained in the attractor this would have implied that case 3a) does not occur.) However, Johnson (1987) proved that in general such measures do not always exist.

Theorem 5.1. **(Johnson)** *Let $f_a\colon [0,1] \to [0,1]$ be a full family of unimodal maps with negative Schwarzian derivative, non-flat critical point and depending continuously on the parameter. Then there exist parameters a such that f_a is ergodic and transitive on an interval and f_a has no finite absolutely continuous invariant measure.*

The proof of Johnson has the same flavour as the construction of the counter-examples of Arnol'd from Section I.5.

Hofbauer and Keller improved this result in a rather remarkable way. They consider Bowen-Ruelle-Sinai measures. More precisely they define the set

$$\hat{\omega}_f(\mu) = \{\nu\,;\ \nu \text{ is an accumulation point of } \frac{1}{n}\sum_{k=0}^{n-1} \mu \circ f^{-k}\}$$

where we take the weak topology on the space of probability measures. Let δ_x be the Dirac measure in x. Then an element of $\hat{\omega}_f(\delta_x)$ is a limit of the measures

$$\mu_n = \frac{1}{n}\sum_{i=0}^{n-1} \delta_{f^i(x)}.$$

(Hence the density of μ_n is simply given by the 'histogram' of the relative frequency of visits of $x, \ldots, f^{n-1}(x)$ to parts of the interval $[0,1]$.) Therefore, one can think of this set as the set of *physical measures*. Hofbauer and Keller (1990a) proved that these physical measures can be rather unexpected. For example the physical measure can be a point mass in a repelling fixed point. In other words, typical orbits can be almost all the time near a repelling fixed point! Indeed, letting λ be the Lebesgue measure on $[0,1]$ we have

Theorem 5.2. **(Hofbauer and Keller)** *Let f_a be a full family of unimodal maps with negative Schwarzian derivative, non-flat critical point and depending continuously on the parameter. Let p_a be the repelling orientation reversing fixed point of f_a. Then there exist parameters a such that*

1. *f_a is ergodic and transitive on an interval;*
2. *f_a has no absolutely continuous invariant probability measure;*
3. *$\hat{\omega}_f(\delta_x) = \hat{\omega}_f(\delta_c) = \hat{\omega}_f(\lambda) = \delta_p$ for almost all x.*

Moreover, one can choose f_a such that $\omega(c)$ is a (non-minimal) Cantor set.

Remark. 1. If for almost all $x \in [0,1]$ the set $\hat{\omega}_f(\delta_x)$ consists of a unique measure then this measure is of course a Bowen-Ruelle-Sinai measure and vice versa each Bowen-Ruelle-Sinai measure is of this form. Furthermore, as we have seen in Section 1 any finite absolutely continuous invariant measure is Bowen-Ruelle-Sinai and, therefore, unique. Thus Theorem 5.1 follows immediately from Theorem 5.2.

2. Such a map f is certainly not infinitely renormalizable because otherwise the measure $\hat{\omega}_f(\delta_c)$ would have its support on the Cantor set $\omega(c)$ which has Lebesgue measure zero. It is not even once renormalizable. Indeed, if I is a central interval such that $f^n(I) \subset I$ and $I, \ldots, f^{n-1}(I)$ are disjoint then the orbit of c visits each of the intervals $I, f(I), \ldots, f^{n-1}(I)$ with frequency $1/n$ and, therefore, $\hat{\omega}_f(\delta_c)(f^i(I)) = 1/n$ for each $i = 1, 2, \ldots, n-1$. So $\hat{\omega}_f(\delta_c) \neq \delta_p$.

3. Hofbauer and Keller (1990a), (1990b) also have examples of quadratic maps which are ergodic and transitive on an interval, such that $\hat{\omega}_f(\delta_x) = \hat{\omega}_f(\delta_c) = \hat{\omega}_f(\lambda)$ for almost all x and such that the metric entropy of the ergodic measures in $\hat{\omega}_f(\lambda)$ is not constant. (We refer to Walters (1982) and Mañé (1987) for the definition of the metric entropy of a measure.)

Of course, Theorem 5.2 follows immediately from the following two theorems. The first one states that f has an absolutely continuous invariant probability measure if for any x from a set of positive measure the orbit of x does not shadow c too much. More precisely,

Theorem 5.3. **(Keller)** *Let $f\colon [0,1] \to [0,1]$ be a unimodal map with negative Schwarzian derivative, non-flat critical point and without attracting periodic points. If f has no absolutely continuous invariant probability measure then for Lebesgue almost all points x one has that $\hat{\omega}_f(\delta_x)$ is contained in the convex hull of $\hat{\omega}_f(\delta_c)$.*

The idea of the proof of the next result is somewhat similar to that of the proof of Johnson's example (1987) but gives more precise information about the support of any limit measure.

Theorem 5.4. **(Keller)** *In every full family f_a of unimodal maps with negative Schwarzian derivative there exists a parameter $a(\infty)$ such that if we let p be the orientation reversing fixed point of $f = f_{a(\infty)}$ then $\delta_p \in \hat{\omega}_f(\delta_x)$ for almost all x and $\hat{\omega}_f(\delta_c) = \delta_p$.*

Before giving the proof of Theorems 5.3 and 5.4 let us show that they imply Theorem 5.2.

Proof of Theorem 5.2: The map $f = f_a$ from the previous theorem has no absolutely continuous invariant probability measure. Indeed, if f has such a measure μ then Theorem I.5 would imply that $\hat{\omega}_f(\delta_x) = \mu$ for almost all x. This would contradict $\hat{\omega}_f(\delta_x) \ni \delta_p$ for almost all x because μ is absolutely continuous. Therefore, from Theorem 5.3 we get that $\hat{\omega}_f(\delta_x) = \hat{\omega}_f(\delta_c)$ for almost all x. □

Let us start with the proof of Theorem 5.3.

Proof of Theorem 5.3: Assume that f is as above and has no absolutely continuous invariant probability measure. As before, let $T_n(x)$ be a maximal interval containing x on which f^n is monotone. (If x is not in the backward orbit of c then there exists only one such interval.) Let $R_n(x)$ and $L_n(x)$ be the components of $T_n(x) \setminus x$ and define $r_n(x)$ to be the minimum of the length of $f^n(R_n(x))$ and $f^n(L_n(x))$. Since f has no absolutely continuous invariant probability measure, we get by Theorem 3.2 that for each $\rho > 0$ and almost all x,

$$\lim_{n\to\infty} \frac{1}{n} \#\{i\,;\ r_i(x) \geq \rho \text{ and } 0 \leq i \leq n-1\} = 0. \tag{5.1}$$

So long branches occur quite infrequently. Let us show that (5.1) implies that $\hat{\omega}_f(\delta_x)$ is contained in the convex hull of $\hat{\omega}_f(\delta_c)$ for Lebesgue almost all points x. We do this by showing that for any neighbourhood U of $\hat{\omega}_f(\delta_c)$ each element of $\hat{\omega}_f(\delta_x)$ is a convex combination of elements of U. Choose n_0 so large and $\rho > 0$ so small that for each $n \geq n_0$, the measure

$$\frac{1}{n} \sum_{k=0}^{n-1} \delta_{x_i} \tag{5.2}$$

is contained in U when $n \geq n_0$ and $|x_i - f^i(c)| < \rho$ for all $i = 0, 1, \ldots, n$. Moreover, choose these numbers such that $c \notin f^k(J)$ when $0 \leq k \leq n_0$ and such that J is a neighbourhood of c of size $\leq \rho$. Furthermore, take m_0 such that f^{m_0} is not a homeomorphism on J whenever J is an interval with $|J| \geq \rho$.

Now choose a point x which satisfies (5.1). If x is a preimage of c then clearly $\hat{\omega}_f(\delta_x)$ has the same set of limits as $\hat{\omega}_f(\delta_c)$. So assume x is not a preimage of c and let I_n be the maximal interval containing x on which $f^n|I_n$ is monotone. Let $\sigma_1 < \sigma_2 < \sigma_3 \ldots$ be the integers σ for which $c \in f^{\sigma}(I_n)$. Furthermore, let τ_n be the maximal integer $\sigma_n \leq \tau_n \leq \sigma_{n+1}$ such that $r_k(x) \leq \rho$ for all $k = \sigma_n, \ldots, \tau_n - 1$. (Of course if $r_{\sigma_n}(x) \geq \delta$ then $\tau_n = \sigma_n$.) From the choice of m_0 it follows that for all $n \in \mathbb{N}$,

$$|\sigma_{n+1} - \tau_n| \leq m_0.$$

Since $r_{\tau_n}(x) \geq \rho$, it follows from this and (5.1) that

$$\frac{\sum_{\tau_i \leq m} |\sigma_{i+1} - \tau_i|}{m} \leq m_0 \cdot \frac{\#\{i\,;\ r_i(x) \geq \rho \text{ and } 0 \leq i \leq m-1\}}{m} \tag{5.3}$$

tends to 0 as $m \to \infty$. Similarly, we get from (5.1) that

$$\frac{1}{m} \sum \{|\tau_j - \sigma_j|\,;\ \tau_j \leq m \text{ and } |\tau_j - \sigma_j| \leq n_0\} \to 0 \tag{5.4}$$

as $m \to \infty$. Indeed, $c \notin f^k(J)$ when $0 \leq k \leq n_0$ and J is a neighbourhood of c of size $\leq \rho$. So if $\tau_j = \sigma_{j+1}$ and $|\tau_j - \sigma_j| \leq n_0$ then $r_{\sigma_j}(x) \geq \rho$. Moreover, by definition, if $\tau_j < \sigma_{j+1}$ then $r_{\tau_j+1}(x) \geq \rho$. Hence, the expression on the left hand side of (5.4) is at most n_0 times $\#\{i\,;\ r_i(x) \geq \rho \text{ and } 0 \leq i \leq m-1\}/m$. Thus (5.4) follows from (5.1). Let us take $m \in \mathbb{N}$ and choose $j(m)$ so that $\sigma_{j(m)} \leq m \leq \sigma_{j(m)+1}$. In order to be definite assume that $\tau_{j(m)} \leq m$ (the other case goes similarly). Then

$$\frac{1}{m} \sum_{i=0}^{m} \delta_{f^i(x)} =$$
$$= \sum_{j=1}^{j(m)} \left[\frac{1}{m} \sum_{i=\sigma_j}^{\tau_j - 1} \delta_{f^i(x)} + \frac{1}{m} \sum_{i=\tau_j}^{\sigma_{j+1}-1} \delta_{f^i(x)} \right] + \frac{1}{m} \sum_{i=\tau_j}^{m} \delta_{f^i(x)}.$$

The total mass contribution of the middle and the last term is at most

$$\sum_{j=1}^{j(m)} \frac{|\sigma_{j+1}-\tau_j|}{m}$$

and, therefore, tends by (5.3) to 0 as $m \to \infty$. Hence $\frac{1}{m}\sum_{i=0}^{m} \delta_{f^i(x)}$ has the same limits as

$$\sum_{j=1}^{j(m)} \frac{1}{m} \sum_{i=\sigma_j}^{\tau_j-1} \delta_{f^i(x)} \quad = \sum_{j=1}^{j(m)} \frac{\tau_j-\sigma_j}{m} \frac{1}{\tau_j-\sigma_j} \sum_{i=\sigma_j}^{\tau_j-1} \delta_{f^i(x)}. \tag{5.5}$$

By definition, for each j with $|\tau_j - \sigma_j| \geq n_0$ the measure

$$\frac{1}{\tau_j-\sigma_j} \sum_{i=\sigma_j}^{\tau_j-1} \delta_{f^i(x)}$$

is contained in U. Furthermore, by (5.4),

$$\frac{1}{m} \sum \{|\tau_j-\sigma_j|\,;\ j=1,\ldots,j(m) \text{ and } |\tau_j-\sigma_j| \geq n_0\}$$

tends to one. It follows that (5.5) converges (in the weak topology) to a convex combination of measures in U. This completes the proof.

□

Proof of Theorem 5.4: In the proof of this theorem we will inductively construct a nested decreasing sequence of parameter intervals $J(n)$, parameters $a(n) \in \partial J(n)$ and integers $k(n) \leq l(n) \to \infty$ such that for each $n \geq 1$,

a) $f_{a(n)}$ has a restrictive interval I_n of period $k(n)$: this means that the intervals $I_n, \ldots, f_{a(n)}^{k(n)-1}(I_n)$ are disjoint, $f_{a(n)}^{k(n)}(\partial I_n) \subset \partial I_n$ and $f_{a(n)}^{k(n)}(I_n) \subset I_n$. So $f_{a(n)}$ is renormalizable to I_n with period $k(n)$;

b) most of the iterates of I_n are near the orientation reversing repelling fixed point p_n of $f_{a(n)}$: for each $n \in \mathbb{N}$, each m with $k(n) \leq m \leq nl(n)$ and each $a \in J(n)$,

$$\frac{\#\{0 \leq i < m-1\,;\ f_a^i(I_n) \subset (p_n - \frac{1}{n}, p_n + \frac{1}{n})\}}{m} \geq 1 - \frac{1}{n}; \tag{5.6}$$

c) for each $a \in J(n)$, the set

$$B_n = \left\{ x\,;\ \text{there exists } l'(n) < l(n) \text{ with } f_a^{l'(n)}(x) \in I_n \right\}$$

has at least size $1 - \dfrac{1}{n}$.

Let us first show that the theorem follows from a)-c). Take $a \in \bigcap_{n\geq 0} J(n)$ and let $f = f_a$. Because of (5.6) and since $f^i(c) \in f^i(I_n)$ we know that

$$\mu(p_n - 1/n, p_n + 1/n) \geq (n-1)/n \text{ for all} n \in \mathbb{N}$$

and each measure $\mu \in \hat{\omega}_f(\delta_c)$. Therefore, $\hat{\omega}_f(\delta_c)$ contains just the measure δ_p. Similarly, since B_n has at least Lebesgue measure $1 - \dfrac{1}{n}$ for each $n \in \mathbb{N}$, for almost all x there exists an integer $n(x)$ such that $x \in B_n$ for all $n \geq n(x)$. (This follows from Borel-Cantelli: the set of points x for which there exists such an integer $n(x)$ is the set

$$\bigcap_{m\geq 0}\bigcup_{n\geq m} B_n = \lim_{m\to\infty}\bigcup_{n\geq m} B_n.$$

This set has full Lebesgue measure because the complement of $\cup_{n\geq m} B_n$ is contained in $\cap_{n\geq m}(B_n)^{\text{compl}}$ and, therefore, has at most Lebesgue measure $\dfrac{1}{n} \to 0$.) Thus, if $n(x) < \infty$ there exists $l'(n) < l(n)$ with $f^{l'(n)}(x) \in I_n$ and, therefore, for each ϵ and each $n \geq n(x)$ large enough, we get by (5.6),

$$\begin{aligned}
&\frac{\#\left\{0 \leq i < nl(n)\,;\ |f^i(x) - p| < \epsilon\right\}}{nl(n)} \\
&\quad\geq \frac{\#\left\{0 \leq j < nl(n) - l'(n)\,;\ \text{dist}(f^j(I_n), p) < \epsilon\right\}}{nl(n) - l'(n)} - \frac{1}{n} \\
&\quad\geq \frac{\#\left\{0 \leq j < nl(n) - l'(n)\,;\ f^j(I_n) \subset (p_n - 1/n, p_n + 1/n)\right\}}{nl(n) - l'(n)} - \frac{1}{n} \\
&\quad\geq \frac{n-1}{n} - \frac{1}{n} \to 1 \text{ as } n \to \infty.
\end{aligned}$$

Since ϵ is arbitrary, it follows that for such points x, $\hat{\omega}_f(\delta_x)$ contains the measure δ_p. Now δ_p is a Dirac measure and as we have seen in Section 1, if f has an absolutely continuous invariant probability μ measure then $\hat{\omega}_f(\delta_x) = \mu$ for almost all x. But since $\hat{\omega}_f(\delta_x) \ni \delta_p$ for almost all x, it follows that f has no absolutely continuous invariant probability measure.

It remains to construct a map satisfying conditions a)-c). In order to do this we will inductively choose parameters $a(n)$ such that also the following condition is satisfied:

d) There exist parameters a arbitrarily near $a(n)$ such that f_a is not renormalizable of period $k(n)$. Therefore, $f^{k(n)}$ either maps I_n *onto* itself or $f^{k(n)}$ has a one-sided attractor in the boundary of I_n.

To start this construction choose $t \in (\sqrt{2}, 2]$ such that the turning point of the tent map $F_t\colon [0,1] \to [0,1]$ with slope $\pm t$ has period $k(1)$. Since f_a is a full family, there exists a parameter $a(1)$ such that $f_{a(1)}$ is semi-conjugate to F_t. Furthermore, as we saw in Section III.4, $f_{a(1)}$ has a restrictive interval I_1 of period $k(1)$ such that $f_{a(1)}^{k(1)}$ is a unimodal map from I_1 into itself. Of course we can choose $a(1)$ so that there exists a parameter a arbitrarily near $a(1)$ such that

f_a is not renormalizable of period $k(1)$ and so $f^{k(1)}$ either maps I_1 onto itself or $f^{k(1)}$ has a one-sided attractor in the boundary of I_1. Since I_1 is forward invariant and contains the orbit of the critical point, the set of points which always stays outside I_1 has Lebesgue measure zero. So there exists $l(1) \in \mathbb{N}$ such that the Lebesgue measure of

$$\left\{x\,;\ \text{there exists } 0 \le l'(1) < l(1) \text{ with } f_{a(1)}^{l'(1)}(x) \in I_1\right\}$$

is at least 2/3. Since f_a depends continuously on the parameter, there exists therefore an interval $J(1)$ containing $a(1)$ such that the Lebesgue measure of

$$\left\{x\,;\ \text{there exists } 0 \le l'(1) < l(1) \text{ with } f_{a(1)}^{l'(1)}(x) \in I_1\right\}$$

is at least 1/2 for each $a \in J(1)$. This proves statements a)-d) for $n = 1$.

Let us now assume that a)-d) holds for some integer n. By d) there exists a parameter a arbitrarily close to $a(n)$ for which f_a is not renormalizable with period $k(n)$. One of the endpoints of I_n has period $k(n)$ and it is easy to see that the preimages of p_n accumulate onto this periodic point. (Indeed, f_a is semi-conjugate to a non-renormalizable tent map T and this statement holds for T because the inverse iterates of each point in the interval $[T^2(c), T(c)]$ form a dense set since the slope of T is $\ge \sqrt{2}$, see Exercise III.4.1.) So we can find $a'(n+1)$ arbitrarily close to $a(n)$ such that the critical point of $f_{a'(n+1)}$ is eventually mapped into the orientation reversing fixed point p_{n+1}. Furthermore, preimages of the critical point accumulate onto p_{n+1} and, therefore, there exists $\hat{a}(n+1)$ arbitrarily close to $a'(n+1)$ such that the turning point is periodic under $f_{\hat{a}(n+1)}$ with period $k(n+1)$. This period $k(n+1)$ becomes arbitrarily large when we choose $\hat{a}(n+1)$ sufficiently close to $a'(n+1)$: by choosing $\hat{a}(n+1)$ sufficiently close to $a'(n+1)$ we get that more and more of the iterates of c stay close to p_n. More precisely, we can choose $\hat{a}(n+1)$ so that

$$\frac{1}{k(n+1)}\#\left\{0 \le j < k(n+1)\,;\ f_{\hat{a}(n+1)}^{j}(c) \subset (p_{n+1} - \frac{1}{n}, p_{n+1} + \frac{1}{n})\right\} \tag{5.7}$$

is as close to 1 as we like. Since the critical point of $f_{\hat{a}(n+1)}$ is periodic of some period $k(n+1)$, there exists a maximal parameter interval $\tilde{J}(n+1)$ containing $\hat{a}(n+1)$ such that $f_{\hat{a}(n+1)}$ is renormalizable of period $k(n+1)$. Since we can find such parameters $\hat{a}(n+1)$ as close as we like to $a'(n+1)$, and for each such parameter there exists such an interval $J(n+1)$ (which are mutually disjoint when we take different parameters $\hat{a}(n+1)$), we can also find $a(n+1)$ in the boundary of such a maximal interval $J(n+1)$ arbitrarily close to $a'(n+1)$. But if $a(n+1) \in \partial J(n+1)$ then f_a is not renormalizable of period $k(n+1)$ for some a arbitrarily near $a(n+1)$. Letting I_{n+1} be the restrictive interval of $f_{a(n+1)}$ of period $k(n+1)$ we have proved statements a) and d). Because the set of point which stay outside I_{n+1} has Lebesgue measure zero, taking $l(n+1)$ sufficiently large, we get that the Lebesgue measure of

$$\left\{x\,;\ \text{there exists } 0 \le l'(n+1) < l(n+1) \text{ with } f_{a(n+1)}^{l'(n+1)}(x) \in I_{n+1})\right\}$$

is at least $1-\dfrac{1}{n+1}$. By choosing $a(n+1)$ sufficiently close to $a'(n+1)$, we get as in (5.7) for $k(n+1) \le m \le (n+1)l(n+1)$ that

$$\frac{\#\left\{0 \le j < m-1\,;\ f^j_{a(n+1)}(c) \subset (p_n - \frac{1}{n+1}, p_n + \frac{1}{n+1})\right\}}{m} \ge 1 - \frac{1}{3n}$$

is as close to 1 as we like. Since all the iterates $I_{n+1}, \ldots, f^{k(n+1)-1}_{a(n+1)}(I_{n+1})$ are disjoint, we can even make sure that for each $k(n+1) \le m \le (n+1)l(n+1)$,

$$\frac{\#\left\{0 \le j < m-1\,;\ f^j_{a(n+1)}(I_{n+1}) \subset (p_n - \frac{1}{n+1}, p_n + \frac{1}{n+1})\right\}}{m} \ge 1 - \frac{1}{2n}.$$

Now choosing $J(n+1)$ to be a sufficiently small one-sided neighbourhood of $a(n+1)$ (such that f_a is not renormalizable of period $k(n+1)$ for $a \in J(n+1)$) we get that the Lebesgue measure of

$$\left\{x\,;\ \text{there exists } l'(n+1) < l(n+1) \text{ with } f_a^{l'(n+1)}(x) \in I_{n+1})\right\}$$

is at least $1-\dfrac{1}{n+1}$ and that

$$\frac{\#\left\{0 \le j < m-1\,;\ f^j_{a(n+1)}(I_{n+1}) \subset (p_n - \frac{1}{n+1}, p_n + \frac{1}{n+1})\right\}}{m}$$

is at least $1-\dfrac{1}{n+1}$ for all $a \in J(n+1)$. This proves b) and c) and completes the proof of the theorem: for the proof that we can make sure that $\omega(c)$ is a Cantor set we simply choose all the maps f in the construction so that no forward iterate of c enters a neighbourhood of the periodic points of period 2. Since f is not renormalizable, this implies that the closure of $\omega(c)$ contains no intervals and is, therefore, a Cantor set. □

6 Frequency of Maps with Positive Liapounov Exponents in Families and Jakobson's Theorem

As we saw in the previous section one can construct unimodal maps with negative Schwarzian derivative whose attractors consist of intervals and which have no finite absolutely continuous invariant measures. In this section we shall prove Jakobson's result that this last phenomenon is not typical: for a large set of parameters a the quadratic map $Q_a(x) = ax(1-x)$ has an absolutely continuous finite invariant measure, see Jakobson (1981). More precisely, we will give the proof of Benedicks and Carleson (1991) that the set of parameters a for which the quadratic map $Q_a(x) = ax(1-x)$ has a positive Liapounov exponent has positive Lebesgue measure. Using Theorem 4.1, for each such parameter the corresponding map certainly has a finite absolutely continuous invariant measure.

In fact, we will consider a more general situation. Let I be some interval, and let $\mathcal{FU}$ be the class of families of C^2 unimodal maps $f_a\colon I \to I$ which depend on a real parameter a and for which

1. f_a has a quadratic critical point;
2. the fixed point of f_a on the boundary of I is repelling;
3. the map $(x,a) \mapsto (f_a(x), Df_a(x), D^2 f_a(x))$ is C^1.

(We should emphasize that we do not require that f_a has negative Schwarzian derivative.) Denote the critical point of f_a by c. Without loss of generality we may assume that c does not depend on a.

Now suppose that f_{a_*} is a Misiurewicz map, i.e., the forward iterates of $f_{a_*}(c)$ stay outside a neighbourhood U of the turning point and assume that f_{a_*} has no periodic attractors. Let K_a be the maximal invariant set of points whose forward orbit stays outside U. As we saw in Chapter III this set K_a is hyperbolic and therefore persists for a near a_*. In particular, there exists a point $x_a \in K_a$ such that the kneading sequence of x_a (with respect to f_a) is the same for each a near a_* and such that $x_{a_*} = f_{a_*}(c)$. Since K_a is hyperbolic, $a \mapsto x_a$ is differentiable for a sufficiently close to a_*. An example of this situation is when $f_{a_*}(c)$ is a repelling periodic point of f_{a_*} and x_a the periodic point of f_a near this orbit.

Of course the family Q_a satisfies these conditions when $a = 4$ but also for each parameter when the critical point is eventually periodic. Furthermore, let $D_n(a)$ be the expansion at the critical value $c_1 = f_a(c)$ (this point depends on a):

$$D_n(a) = Df_a^n(c_1).$$

Then one has

Theorem 6.1. **(Benedicks and Carleson)** *Consider a family $f_a \in \mathcal{FU}$ and a_* such that f_{a_*} is a Misiurewicz map and has no periodic attractors. Moreover, assume that*

$$\frac{d}{da}(x_a - f_a(c)) \neq 0$$

at $a = a_*$, where x_a is defined as above. Then there exist $C_0, \gamma > 0$ and a subset E of parameters having a_* as a density point such that for each $a \in E$,

$$|D_n(a)| \geq C_0 \cdot e^{\gamma n} \text{ for all } n \geq 1.$$

A sketch of the proof of this result was given by Benedicks and Carleson (1991) for the case that $f_a = ax(1-x)$ and $a_* = 4$. An outline of the proof of this result is given in Mora and Viana (1989). Since the proof of Theorem 6.1 is rather technical and the available proofs are rather sketchy, we will prove Theorem 6.1 in detail. Thieullen, Tresser and Young (1992) also have a complete proof of Theorem 6.1. Tsujii (1992b) has given an alternative proof of Theorem 6.1. His proof is shorter than the one given here because he considers a slightly smaller set of parameters where the parameter dependence will be uniform by definition. In Benedicks and Carleson's proof (and ours) the fact that this dependence on parameters is uniform follows from the inductive assumption on the growth of spatial derivatives, see the remark at the end of Step 4.

Before going into the proof of this result let us discuss its history and some of its ramifications. In an earlier paper of Benedicks and Carleson (1985), it was proved that the expansion at the critical value grew subexponentially fast for a set E as above: $|D_n(a)| \geq e^{\sqrt{n}\gamma}$. (For an outline of the proof of this result and results on stochastic perturbations of these maps, see Benedicks and Young (1992).) Related to the above is the following important result of Rees (1984) on rational maps: in the space of rational maps of a given degree, there exists a subset $\mathcal{E}$ of positive Lebesgue measure such that for every $R \in \mathcal{E}$ the following properties are satisfied: i) the forward orbit of each critical point is dense in the Riemann sphere; ii) there exists $\gamma > 0$ such that $|DR^n(R(c))| > e^{\gamma n}$ for every critical point c; iii) R is ergodic with respect to the Lebesgue measure, i.e., every totally invariant subset has Lebesgue measure equal to either one or zero.

From Theorem 4.1 and Theorem 6.1 one gets an alternative proof of the following remarkable result of Jakobson (1981):

Corollary 6.1. (Jakobson) *Let $Q_a\colon [0,1] \to [0,1]$, $a \in (0,4]$, be the quadratic family $Q_a(x) = ax(1-x)$. There exists a subset $\mathcal{C} \subset (0,4]$ of positive Lebesgue measure with the following properties:*

1. *If $a \in \mathcal{C}$ then Q_a has an absolutely continuous invariant probability measure with positive entropy.*

2. *The parameter value $a = 4$ is a Lebesgue density point of $\mathcal{C}$, namely,*

$$\lim_{\epsilon \to 0} \frac{\lambda(\mathcal{C} \cap [4-\epsilon, 4])}{\epsilon} = 1.$$

Johnson (1986), using a construction due to Guckenheimer (1984), gave another proof of Jakobson's result by constructing an induced expanding map as in Section 3 for each map f_a for $a \in \mathcal{C}$. But, unlike in Jakobson's proof, the expanding map constructed is not Markov because it has an infinite image partition. We will not follow this approach but prove Theorem 6.1 since this gives more information.

Step 1: An outline of the proof and some notation

So let us start by giving an outline of the proof of Theorem 6.1. Following Benedicks and Carleson (1991), we will make two basic assumptions on the iterates of the critical point which, when satisfied, will ensure that $|D_n(a)|$ grows exponentially. The basic tool to get this exponential growth was already proved in Chapter III and will be stated in Step 2. In Steps 3 and 4 we will use this tool to get the exponential growth up to the n-th iterate provided the critical orbit satisfies the two basic assumptions up to this iterate. In the remainder of the proof it is shown that there is a rather large set of parameters a such that the critical orbit of the corresponding map f_a satisfies the basic assumptions for all iterates. This is done by removing for each n the parameter values for which the n-th iterate of the critical point fails these assumptions. To do this, we will prove in Step 5 that the position of the n-th iterate of the critical point depends very uniformly on the parameter inside a parameter interval for which assumptions hold up to the n-th iterate. This, and some exponential growth, allows us to prove in Step 6 that the proportion of the parameters which has to removed at the n-th iterate is exponentially small in terms of n. Thus we get left with a positive set of parameters for which one has exponential growth for all iterates.

To simplify the notation in the proof, we will use the letter C_0 (respectively C) for all positive constants that appear in estimates from below (respectively above) that do not depend on n and also not on a constant Δ which we will introduce below (provided this Δ is large enough). During the proof we will decrease C_0 (respectively increase C), but every inequality will continue to hold. Moreover, these constants will depend on their previous values and we will sometimes write expressions like $C = 2C$. If in a formula we get an estimate in which a large power (say the power is s) of a previous constant C appears, we will explicitly show this by writing C^s. Furthermore, we denote by $O(1)$ a function which is uniformly bounded from above and below.

For simplicity, let

$$D_n(a) = Df_a^n(c_1)$$

and

$$\xi_n(a) = f_a^n(c).$$

Step 2: Basic results and definitions

As long as the orbit of the critical point stays away from a neighbourhood of the critical point, $D_n(a)$ increases exponentially. This is shown in Proposition 6.1. However, for almost all parameter values the orbit of the critical point will return arbitrarily close to c. But forward iterates of such a return shadow a previous piece of the orbit of the critical point. This information will be used in an inductive argument. To make full use of this shadowing, we will classify these returns according to which interval I_r, defined presently, it falls into. This classification is used throughout the entire proof. So let U_r be an exponentially fast shrinking sequence of neighbourhoods of c, i.e.,

$$U_r = (c - e^{-r}, c + e^{-r}).$$

Furthermore, we introduce the pair of intervals

$$I_r = U_r \setminus U_{r+1}$$

(I_r consists of two components one on each side of c) and we will split each component of I_r into r^2 intervals,

$$I_{r,1}, I_{r,2}, \ldots, I_{r,r^2}$$

of equal length (so $I_{i,j}$ consists of two components). Throughout the proof we will only consider $r \geq \Delta$ (or sometimes $r \geq \Delta - 1$) where Δ will be some very large integer. Since f_{a_*} satisfies the Misiurewicz condition there exists a neighbourhood W of c such that

$$\xi_k(a_*) \notin W$$

for all $k > 0$. Also define

$$U = U_\Delta.$$

We only consider parameters a with $|a - a_*| < \epsilon$ where $\epsilon > 0$ will be small. Later on we shall increase Δ repeatedly and therefore shrink U. By decreasing ϵ we will always be able to keep the constants C, C_0 and so on independent of Δ.

The main tool used to get exponential expansion is the following proposition which states that a piece of the orbit is expanding as long as it stays outside $U = U_\Delta$:

Proposition 6.1. *Let $f_a \in \mathcal{FU}$ and suppose that f_{a_*} is a Misiurewicz map without periodic attractors. Then there exist $\gamma_0 > 0$, $C_0 > 0$ and a neighbourhood W of c such that for each Δ sufficiently large, there exists $\epsilon > 0$ such that if $|a - a_*| < \epsilon$ and $f_a^j(x) \notin U_\Delta$ for $0 \leq j \leq k-1$ and $f_a^k(x) \in W$ then*

$$|Df_a^k(x)| \geq C_0 \cdot e^{\gamma_0 k}. \tag{6.1}$$

If $f_a^j(x) \notin U_\Delta$ for $0 \leq j \leq k-1$ but not necessarily $f_a^k(x) \in W$, then

$$|Df_a^k(x)| \geq C_0 \cdot e^{\gamma_0 k} \inf_{j=0,\ldots,k-1} |Df_a(f_a^j(x))|. \tag{6.2}$$

Moreover, if $x, f_a(x), \ldots, f_a^{k-1}(x) \notin W$ then

$$|Df_a^k(x)| \geq C_0 \cdot e^{\gamma_0 k}. \tag{6.3}$$

Finally, for each neighbourhood V of c with $\mathrm{cl}\,(V) \subset \mathrm{int}\,(W)$ there exists a constant $K < \infty$ such that for each interval $[x, y]$ and each $n \in \mathbb{N}$ for which $f^n[x, y] \subset V$ one has

$$\frac{|Df_{a_*}^n(x)|}{|Df_{a_*}^n(y)|} \leq K.$$

Remark. The remarkable thing is that the constants C_0 and γ_0 do not depend on U (as long as one allows $\epsilon > 0$ to shrink to zero if U becomes small). This fact will turn out to be crucial in the remainder of the proof.

Proof. See Theorems III.6.3 and III.6.4 for the first statements. The last statement follows from Theorem III.6.1. □

Let $\bar{\gamma} = \sup_{a,x} |Df_a(x)|$ and take γ with $0 < \gamma < \min(\gamma_0, 1/40)$ and $\alpha, \beta, \tau > 0$ such that

$$\begin{aligned} &0 < \alpha < \beta < 10^{-4} \min\left(\gamma, \gamma_0 - \gamma, \frac{\gamma(\gamma_0 - \gamma)}{\gamma + \bar{\gamma}}, \frac{\gamma^2}{\gamma + \bar{\gamma}}\right) \\ &0\tau < \frac{\gamma_0 - \gamma - \alpha}{\gamma_0}. \end{aligned} \tag{6.4}$$

From now on $\alpha, \beta, \gamma, \gamma_0, \bar{\gamma}$ and τ will not change.

The next Proposition shows that parameter and space derivatives are of the same order provided one has exponential growth of $D_n(a)$.

Proposition 6.2. *There exist constants $C, \epsilon > 0$ and N_0 (which do not depend on Δ) such that if $|a - a_*| < \epsilon$, $n \geq N_0$ and*

$$|D_k(a)| \geq e^{\gamma k} \text{ for all } k = N_0, \ldots, n-1$$

then for all $k = N_0, \ldots, n$,

$$\frac{1}{C} \leq \frac{|\xi_k'(a)|}{|D_{k-1}(a)|} \leq C.$$

Proof. Let $x(a)$ be the point whose kneading sequence with respect to f_a is the same as the kneading sequence of $x(a_*) = c_1(a_*)$. Of course, for a close to a_*, the point $x(a)$ is contained in a hyperbolic forward invariant set K_a. Since K_a is hyperbolic, the set K_a depends smoothly on a and in particular the absolute value of $\frac{d}{da} f_a^k(x(a))$ is universally bounded for all $k \geq 0$. By assumption

$$\frac{d}{da}(x(a) - f_a(c)) \big|_{a=a_*} \neq 0.$$

Now $\xi_k(a) = f_a^{k-1}(c_1(a)) = f_a^{k-1}(x(a) + [c_1(a) - x(a)])$ and by the Chain Rule this gives,

$$\xi_k'(a_*) = \frac{d}{da}\left(f_a^{k-1}(x(a))\right) \big|_{a=a_*} + Df_{a_*}^{k-1}(c_1(a_*)) \frac{d}{da}(c_1(a) - x(a)) \big|_{a=a_*}.$$

Since $D_{k-1}(a_*) = Df_{a_*}^{k-1}(c_1(a_*))$ grows exponentially with k,

$$\frac{d}{da}(x(a) - c_1(a)) \big|_{a=a_*} \neq 0$$

and $\frac{d}{da}\left(f_a^{k-1}(x(a))\right)\big|_{a=a_*}$ is universally bounded from above and below, there exist universal constants $D \in (1,\infty)$ and N' such that

$$\frac{1}{D} \le \frac{|\xi_k'(a_*)|}{|D_{k-1}(a_*)|} \le D$$

for all $k \ge N'$. Let us now show that the same holds for a sufficiently close to a_*. By the Chain Rule, $D_{k-1}(a) = Df_a(\xi_{k-1}(a))D_{k-2}(a)$ and $\xi_k'(a) = Df_a(\xi_{k-1}(a))\xi_{k-1}'(a) + \partial_a f_a(\xi_{k-1}(a))$. Therefore,

$$\left|\frac{\xi_k'(a)}{D_{k-1}(a)} - \frac{\xi_{k-1}'(a)}{D_{k-2}(a)}\right| \le \frac{|\partial_a f_a(\xi_{k-1}(a))|}{|D_{k-1}(a)|}.$$

From this, and since $|D_j(a)| \ge e^{\gamma j}$ for $j = N_0, \ldots, n-1$,

$$\left|\frac{\xi_k'(a)}{D_{k-1}(a)} - \frac{\xi_{N_0}'(a)}{D_{N_0-1}(a)}\right| \le \frac{1}{2D}$$

provided $N_0 \ge N'$ is sufficiently large. Combining this and

$$\frac{1}{D} \le \frac{|\xi_{N_0}'(a_*)|}{|D_{N_0-1}(a_*)|} \le D,$$

the proposition follows. □

So let $N_0, \epsilon, C_0, \gamma_0$ be as in the previous Propositions and fix $N > N_0$ so large that

$$\begin{aligned} e^{-\alpha N} &\ge 10e^{-\beta N} \\ C_0 \cdot e^{\gamma_0 k} &\ge e^{\gamma k} \\ |\xi_k(a_*) - c| &\ge 10e^{-\alpha k} \end{aligned} \tag{6.5a}$$

for all $k \ge N$. Since f_{a_*} is Misiurewicz, all iterates of $\xi_k(a_*)$ stay outside a neighbourhood of c and so this is possible. So there exists a neighbourhood W of c so that

$$\lim_{a \to a_*} \inf\{k > 0\,;\ \xi_k(a) \in W\} = \infty. \tag{6.5b}$$

Next let

$$\nu_1(a) := \inf\{k > 0\,;\ \xi_k(a) \in U_\Delta\}$$

and take Δ so large and $\epsilon > 0$ so small that $\nu_1(a) \ge N$ for each $|a - a_*| < \epsilon$. By Proposition 6.1, we get that

$$|D_n(a)| \ge e^{\gamma n} \text{ for all } n = N, \ldots, \nu_1(a) - 1$$

and each $|a - a_*| < \epsilon$. We should emphasize that we shall not change N anymore.

Step 3: The two assumptions (BA) and (FA) on the parameters

We begin by taking a (possibly one-sided) interval neighbourhood ω_0 of a_* of size smaller than ϵ. Let us consider $\xi_k(\omega_0)$: this is the set of all possible images of c under f_a^k for varying parameter values $a \in \omega_0$. Clearly $\xi_k(\omega_0)$ is an interval. Furthermore, if we take as before x_a to be the point in K_a whose kneading sequence does not vary with a and such that $x_{a_*} = f_{a_*}(c)$, then

$$\frac{d}{da}(x_a - f_a(c)) \neq 0 \text{ at } a = a_*.$$

Therefore, the kneading itinerary of $c_1(a) = \xi_1(a)$ changes if a varies. But it can only change if some iterate $\xi_k(a)$ hits the turning point c and therefore for each interval ω_0, the interval $\xi_k(\omega_0)$ contains c for k large enough. So choose ν_1 minimal such that $c \in \xi_k(\omega_0)$ for $k = \nu_1$. Since ω_0 has size $< \epsilon$, we get from the definition of ϵ and the integer N, that $\nu_1 \geq N$. One end point of $\xi_k(\omega_0)$ lies on the forward orbit of the point $f_{a_*}(c_1)$. Because f_{a_*} is a Misiurewicz map the orbit of this point stays away from c. So the interval $\xi_k(\omega_0)$ does become big because it eventually will contain the critical point c. By shrinking ω_0 we can assume that $\xi_k(\omega_0) \cap U = \emptyset$ for all $k = 1, \ldots, \nu_1 - 1$ and

$$\xi_{\nu_1}(\omega_0) \supset W.$$

Later on, we shall increase Δ and therefore shrink U. We will also shrink ω_0. In both cases ν_1 increases. We will always shrink ω_0 in such a way that the property $\xi_{\nu_1}(\omega_0) \supset W$ is still satisfied.

We will show by induction that for each parameter a which satisfies two assumptions to be defined presently, one has $|D_n(a)| \geq e^{\gamma n}$ for all $n \geq N$. Let us say that ν is a *return* if $\xi_\nu(a) \in U$. The first condition which we will impose disallows returns to be too close to the critical point. This condition requires that

$$(BA_n) \qquad |\xi_k(a) - c| \geq e^{-\alpha k} \text{ for } k = N, \ldots, n.$$

Note that by assumption $\xi_k(a) \notin U$ for $a \in \omega$ and $k \leq N$. In Step 6 we shall show that 'most parameters' a near a_* satisfy (BA_n) for every n. Before stating the second condition let us go into the purpose of (BA_n). Take a parameter a which satisfies (BA_n) for each $n \geq N$. By the Chain Rule, $|D_n(a)| = |D_{n-1}(a)| \cdot |Df_a(\xi_n(a))|$. Now if $\xi_n(a) \in U$ then $|Df_a(\xi_n(a))|$ is not too small because of (BA_n). Therefore, we can hope to compensate this small term $|Df_a(\xi_n(a))|$ by a subsequent part of the orbit as follows. Consider a return $\xi_{\nu_i}(a)$. If the first p subsequent iterates, $\xi_{\nu_i+1}(a), \ldots, \xi_{\nu_i+p}(a)$, are very close (in a way to be defined below) to the corresponding iterates of c, i.e., to $\xi_1(a), \ldots, \xi_p(a)$ then $\xi_{\nu_i+1}(a), \ldots, \xi_{\nu_i+p}(a)$ is called a *bound orbit.* For such a bound orbit we will show that $|Df_a^p(\xi_{\nu_i+1}(a))|$ is close to $|D_p(a)|$. Therefore, even though several of the points $\xi_{\nu_i+1}(a), \ldots, \xi_{\nu_i+p}(a)$ could come close to c, the induction assumption that $|D_p(a)|$ is exponentially large and the definition of p will imply that $|Df_a^p(\xi_{\nu_i+1}(a))|$ is large enough to compensate for the small term $|Df_a(\xi_{\nu_i}(a))|$. Indeed, $|Df_a^{p+1}(\xi_{\nu_i}(a))| \geq e^{\gamma p/4}$. The first integer $\nu > \nu_i+p$ such that $\xi_\nu(a) \in U$ is defined to be ν_{i+1} and is called the next *free return*. By Proposition 6.1, all this implies

$$\begin{aligned}|Df_a^{\nu_{i+1}-\nu_i}(\xi_{\nu_i}(a))| &= |Df_a^{\nu_{i+1}-\nu_i-p-1}(\xi_{\nu_i+p+1}(a))| \cdot |Df_a^{p+1}(\xi_{\nu_i}(a))| \\ &\geq C_0 \cdot e^{\gamma(\nu_{i+1}-\nu_i-p-1)}.\end{aligned}$$

So let us be more precise now. For each parameter $a \in E$ and each $r \geq \Delta$ let $p(r,a)$ be the maximal integer p such that

$$|f_a^j(x) - \xi_j(a)| \leq e^{-\beta j} \text{ for } 0 \leq j \leq p,$$

for each $x \in U_r$. In other words, $p(r,a)$ is so that

$$\text{(BP)}\quad |f_a^j(U_r)| \leq e^{-\beta j} \text{ for } 0 \leq j \leq p(r,a) \quad\text{and}\quad |f_a^{p(r,a)+1}(U_r)| > e^{-\beta(p(r,a)+1)}.$$

If $\xi_\nu(a)$ is a return let $r \geq \Delta$ be so that $\xi_\nu(a) \in I_r$. In this case we say that $p(r,a)$ is the *binding period* associated to the return $\xi_\nu(a)$ and $\xi_\nu(a), \ldots, \xi_{\nu+p(r,a)}(a)$ is the *binding orbit* associated to this return. We say that $\xi_\nu(a)$ is a *free return* if it is a return and does not belong to the binding period associated to the previous free return. More precisely, define these free returns inductively as follows: first of all, $\nu_1(a) := \min\{k\,;\ \xi_k(a) \in U\} \geq N$ is a free return. If $\nu_1(a) < \cdots < \nu_i(a)$ are the first free returns of $\xi_j(a)$ for $j \leq \nu_i(a)$, then define $p_i(a)$ to be the binding period associated to the return $\xi_{\nu_i(a)}(a)$ and let $\nu_{i+1}(a)$ be the largest integer such that $\xi_j(a)$ is outside U for all $j = \nu_i(a) + p_i + 1, \ldots, \nu_{i+1}(a) - 1$.

In order to define the second assumption assume that $\nu_1(a) < \nu_2(a) < \cdots < \nu_s(a) \leq n$ are the free return times of $\xi_j(a)$ for $0 \leq j \leq n$ to U_Δ and let $p_1(a), p_2(a), \ldots, p_s(a)$ be the binding periods associated to these free returns. Furthermore, let

$$\text{(6.6)}\qquad \begin{aligned} q_0(a) &= \nu_1(a) \\ q_1(a) &= \nu_2(a) - 1 - (\nu_1(a) + p_1(a)) \\ \vdots &= \vdots \\ q_{s-1}(a) &= \nu_s(a) - 1 - (\nu_{s-1}(a) + p_{s-1}(a)). \end{aligned}$$

If $n > \nu_s(a) + p_s(a)$ we also define $q_s(a) = n - (\nu_s(a) + p_s(a))$. Furthermore, define

$$\text{(6.7)}\qquad F_n(a) = \begin{cases} q_0(a) + \cdots + q_{s-1}(a) & \text{if } n \leq \nu_s(a) + p_s(a) \\ q_0(a) + \cdots + q_s(a) & \text{otherwise.} \end{cases}$$

The second condition we will impose on parameters is that

$$(FA_n)\qquad \frac{F_k(a)}{k} \geq (1-\tau) \text{ for } k = 1, \ldots, n,$$

for each n. We require this condition because over the total bound period we only we get a growth of the derivative of the form $e^{\gamma p/4}$, see Statement c) in

Lemma 6.1 below. If we had no condition like (FA_n) then the induction we will apply below would show that $|D_k(a)| \geq e^{\gamma k}$ for $k < n$ implies something like $|D_{2n}(a)| \geq e^{\gamma n/4}$. Doing this again would imply $|D_{4n}(a)| \geq e^{\gamma n/16}$. Continuing this would therefore give no definite lower bound for the growth rate of $D_n(a)$. The condition (FA_n) will enable us to compensate for this loss, see also the remark following the proof of Theorem 6.2 below. Let BA_n and FA_n denote the sets of parameters which satisfy the corresponding conditions.

Step 4: Parameters satisfying (BA) and (FA) give rise to exponential growth

Theorem 6.2. *For each Δ sufficiently large there exist $\epsilon > 0$ such that if $|a - a_*| < \epsilon$ and $a \in BA_n \cap FA_n$ then we have exponential growth of the derivative up to time n:*

$$(EX_n) \qquad |D_k(a)| \geq e^{\gamma k} \text{ for all } k = N, \ldots, n.$$

Note that (BA_n), (FA_n) and (EX_n) all depend only on the first n iterates of c. Since $|D_k(a)| \geq e^{\gamma k}$ for all $k = N, \ldots, \nu_1(a) - 1$, the theorem holds for $n = N, \ldots, \nu_1(a) - 1$. So let us prove Theorem 6.2 by induction. For this we will use the following lemma. Assume that $a \in BA_n \cap FA_n$ and that

$$(EX_{n-1}) \qquad |D_k(a)| \geq e^{\gamma k} \text{ for all } k = N, \ldots, n-1.$$

We will show that (EX_n) also holds.

Lemma 6.1. *There exists a constant $C_0 > 0$ such that for every Δ sufficiently large there exists $\epsilon > 0$ such that for any $a \in BA_n \cap EX_{n-1}$ with $|a - a_*| < \epsilon$ the following statements hold. Suppose that $\nu \leq n$ is a return of $\xi_i(a)$ to $U_{\Delta-1}$ and $p(r, a)$ is the bound period of $\xi_\nu(a)$, where $r \geq \Delta - 1$ is the largest integer such that $\xi_\nu(a) \in U_r$. Then*

a) $p(r, a) \leq 3r/\gamma \leq 3\alpha\nu/\gamma < \frac{1}{100}n$ and in particular $p(r, a) < n$;

b) for every $x \in U_r$ and $1 \leq j \leq p(r, a)$,

$$C_0 \leq \frac{|Df_a^j(f_a(x))|}{|D_j(a)|} \leq \frac{1}{C_0};$$

in particular, $|Df_a^j(f_a(x))| \geq C_0 \cdot e^{\gamma j}$ for $1 \leq j \leq p(r, a)$.

c) $p \geq C_0 \cdot r$ and $|Df_a^{p(r,a)+1}(x)| \geq e^{\gamma p(r,a)/4}$ for each $x \in I_r$.

Proof. Let us first prove b). Consider $x \in U_r$. By the Chain Rule,

$$\frac{|Df_a^j(f_a(x))|}{|Df_a^j(f_a(c))|} = \prod_{i=1}^{j} \frac{|Df_a(f_a^i(x))|}{|Df_a(\xi_i(a))|}.$$

Since $(x,a) \to (f_a(x), Df_a(x), D^2f_a(x))$ is C^1 and the critical point of f_a is non-degenerate, $\frac{|Df_a(x)|}{|x-c|}$ is bounded from below and from above and

$$\frac{|Df_a(x)|}{|x-c|}\frac{|y-c|}{|Df_a(y)|} = 1 + O(|x-y|),$$

where $O(t)$ denotes some function for which $O(t)/t$ is bounded as $t \to 0$. By definition of $p(r,a)$, we have $|f_a^j(x) - \xi_j(a)| \le e^{-\beta j}$ for $0 \le j \le p(r,a)$ and this implies

$$\prod_{i=1}^{j} \frac{|Df_a(f_a^i(x))|}{|Df_a(\xi_i(a))|} = O(1) \cdot \prod_{i=1}^{j} \frac{|f_a^i(x) - c|}{|\xi_i(a) - c|}. \tag{6.8}$$

Now by assumption $|\xi_i(a) - c| \ge e^{-\alpha i}$ for $i = N, \ldots, n$. Hence,

$$\left|\frac{|f_a^i(x) - c|}{|\xi_i(a) - c|} - 1\right| \le \frac{|f_a^i(x) - \xi_i(a)|}{|\xi_i(a) - c|} \le \frac{e^{-\beta i}}{e^{-\alpha i}}$$

for $i = N, \ldots, \min(p(r,a), n)$. Since $\alpha < \beta$, this implies that

$$\sum_{i=N}^{j} \left|\frac{|f_a^i(x) - c|}{|\xi_i(a) - c|} - 1\right|$$

is uniformly bounded from above and below. Hence (6.8) is universally bounded from above and below provided the first N factors from the right hand side of (6.8) are universally bounded from above and below. But, since f_{a_*} is Misiurewicz $f_a^i(x)$ stays outside some fixed neighbourhood of c for all $x \in U$ and all $i \le N$ provided U and $|a - a_*|$ are sufficiently small (note that N is a fixed number). Hence, $\prod_{i=1}^{N} \frac{|f_a^i(x)-c|}{|\xi_i(a)-c|}$ is universally bounded from above and below. All this implies that the expression in (6.8) is uniformly bounded from below and above for $j = 1, \ldots, \min(p(r,a), n)$. Thus Statement b) is proved, provided we can show that $p(r,a) < n$.

The first part of the proof gives that $|f_a^{j+1}(x) - \xi_{j+1}(a)|$ is equal to $O(1) \cdot |D_j(a)| \cdot |f_a(x) - f_a(c)|$ for $j = 1, \ldots, \min(p(r,a), n)$ for $x \in U_r$. Since f_a is quadratic at c, one has $|f_a(x) - f_a(c)| = O(1) \cdot e^{-2r}$ for $x \in I_r$. Hence, taking $x \in I_r$,

$$|f_a^{j+1}(x) - f_a^{j+1}(c)| = O(1) \cdot |D_j(a)| \cdot e^{-2r} \tag{6.9}$$

for $j = 1, \ldots, \min(p(r,a), n)$. But by assumption $|D_j(a)| \ge e^{\gamma j}$ for $j = N, \ldots, n-1$ and since the left hand side in (6.9) can be no more than 1 for $j \le p(r,a)$, this implies that $\min(p(r,a), n) \le \frac{3r}{\gamma}$ when Δ is sufficiently large since $r \ge \Delta - 1$. Moreover, since $\xi_\nu(a) \in I_r$, $\nu \ge N$ and $|\xi_\nu(a) - c| \ge e^{-\alpha\nu}$ this implies that $r \le \alpha\nu \le \alpha n$. Therefore, $p(r,a) < n$ and in fact $p(r,a) \le \frac{3r}{\gamma} < n/100$. This completes the proof of Statements a) and b).

Furthermore, if $x \in I_r$, by the definition of $p = p(r,a)$, one has $|f_a^p(x) - \xi_p(a)| \le e^{-\beta p}$ and $|f_a^{p+1}(x) - \xi_{p+1}(a)| > e^{-\beta(p+1)}$. Therefore, $|f_a^p(x) - \xi_p(a)| = O(1) \cdot e^{-\beta p}$ and by (6.9),

$$e^{-\beta p} = O(1) \cdot |D_p(a)| \cdot e^{-2r}. \tag{6.10}$$

Since $|D_p(a)| \geq e^{\bar{\gamma} p}$ this implies that $p \geq C_0 \cdot r$. Moreover,

$$|Df_a^{p+1}(x)| = |Df_a(x)| \cdot |Df_a^p(f_a(x))| = O(1) \cdot e^{-r} \cdot |Df_a^p(f_a(x))|,$$

for each $x \in I_r$. Using Statement b) of this lemma,

$$|Df_a^{p+1}(x)| = O(1) \cdot e^{-r} \cdot |Df_a^p(f_a(x))| \geq C_0 \cdot e^{-r} \cdot |D_p(a)|.$$

By (6.10) and because $|D_p(a)| \geq C_0 \cdot e^{\gamma p}$ this gives $|Df_a^{p+1}(x)| \geq C_0 \cdot e^{-\beta p/2} \sqrt{|D_p(a)|} \geq e^{\gamma p/4}$ provided Δ is sufficiently large (because then $p \geq C_0 \cdot r \geq C_0 \cdot \Delta$ is large). □

From Statement a) of Proposition 6.1 the binding period following a return $\nu \leq n$ is much less than n (which will allow us to use induction) and by Statement c) the small derivative at a return is compensated by the expansion of the binding orbit which follows this return.

Proof of Theorem 6.2: By the Chain Rule

$$|D_n(a)| = |Df_a^{\nu_1 - 1}(c_1)| \times \prod_{i=1}^{s-1} |Df_a^{\nu_{i+1}-\nu_i}(\xi_{\nu_i}(a))| \times |Df_a^{n+1-\nu_s}(\xi_{\nu_s}(a))|. \tag{6.11}$$

Because of the induction assumption we can apply the previous lemma to the binding orbit. Since this lemma implies that $p_i < n$, Statement c) of Lemma 6.1 and the induction assumption gives

$$|Df_a^{p_i+1}(\xi_{\nu_i}(a))| \geq e^{\gamma p_i/4}.$$

From Proposition 6.1,

$$|Df_a^{q_i}(\xi_{\nu_i+p_i+1}(a))| \geq C_0 \cdot e^{\gamma_0 q_i}.$$

Hence,

$$|Df_a^{\nu_{i+1}-\nu_i}(\xi_{\nu_i}(a))| \geq C_0 \cdot e^{\gamma p_i/4} \cdot e^{\gamma_0 q_i}.$$

If $n > \nu_s + p_s$ then we can deal with the last factor in (6.11) by using (6.2) of Proposition 6.1:

$$|Df_a^{n+1-\nu_s}(\xi_{\nu_s})| \geq C_0 \cdot e^{\gamma_0 q_s} \cdot \inf_{\nu_s + p_s \leq j \leq n} |\xi_j(a) - c|.$$

Since for $j \leq n$, $|\xi_j(a) - c| \geq e^{-\alpha j} \geq e^{-\alpha n}$, combining all this we get in this case

$$|D_n(a)| \geq C_0^{s+1} \cdot e^{\gamma(n-F_n)/4} \cdot e^{F_n \gamma_0} \cdot e^{-\alpha n}.$$

If $n \leq \nu_s + p_s$ then we write the last factor in (6.11) as

$$|Df_a^{n+1-\nu_s}(\xi_{\nu_s}(a))| = |Df_a(\xi_{\nu_s}(a))| \times |Df_a^{n-\nu_s}(\xi_{\nu_s+1}(a))|. \tag{6.12}$$

The second factor in this expression corresponds to the first piece of a binding orbit and by Statement b) of the previous lemma is at least $C_0 \cdot |D_{n-\nu_s}(a)| \geq$

$C_0 \cdot e^{\gamma(n-\nu_s)} \geq C_0$. Since f is quadratic in c, the first factor in (6.12) is at least $C_0 \cdot |\xi_{\nu_s}(a) - c| \geq C_0 \cdot e^{-\alpha\nu_s}$ since $a \in BA_n$. Hence,

$$|D_n(a)| \geq C_0^{s+2} \cdot e^{\gamma(n-F_n)/4} \cdot e^{\gamma_0 F_n} \cdot e^{-\alpha\nu_s} \geq C_0^{s+2} \cdot e^{\gamma(n-F_n)/4} \cdot e^{\gamma_0 F_n} \cdot e^{-\alpha n}$$

if $n \leq \nu_s + p_s$. So in both cases

$$|D_n(a)| \geq C_0^{s+2} \cdot e^{\gamma(n-F_n)/4} \cdot e^{\gamma_0 F_n} \cdot e^{-\alpha n}.$$

Using (FA_n), and since $p_i \geq C_0\Delta$ we get that $s/(n-F_n) \to 0$ as $\Delta \to \infty$; hence, we have

$$|D_n(a)| \geq e^{\gamma_0(1-\tau)n-\alpha n} \geq e^{\gamma n}$$

provided $\tau, \alpha > 0$ satisfies (6.4) and Δ is sufficiently large. □

Remark. 1. In Benedicks and Carleson (1985), see also Benedicks and Young (1990), a weaker theorem is shown instead of Theorem 6.1: there exists a set of parameters a with positive Lebesgue measure for which $D_n(a)$ grows at least as fast as $e^{\sqrt{n}}$. In this case condition (FA_n) is not necessary. Indeed, as before, we certainly have $D_n(a) \geq e^{\sqrt{n}}$ for $n = N, \ldots, \nu_1(a)$, provided a is sufficiently close to a_*. Now suppose that $|D_j(a)| \geq e^{\sqrt{j}}$ for $j = N, \ldots, n-1$. Then redefine $p(r,a)$ as the largest integer so that $|f_a^j(x) - \xi_j(a)| \leq e^{-\beta\sqrt{j}}$ for $j = 0, \ldots, p(r,a)$ and all $x \in U_r$. Similarly, redefine the (BA_n) condition by $|\xi_k - c| \geq e^{-\sqrt{k}}$ for $k = N, \ldots, n$. Then, exactly as in Lemma 6.1, we get that $|Df^{p(r,a)+1}(x)| \geq e^{\sqrt{p(r,a)}/2}$ and $p(r,a) < n$. But since

$$e^{\sqrt{n}} e^{\sqrt{p(r,a)}/2} \geq e^{\sqrt{n+p(r,a)}}$$

because

$$\sqrt{n} + \sqrt{p}/2 \geq \sqrt{n+p}$$

for $p \leq n$ this gives similarly as in the proof of Theorem 6.2 that

$$D_n(a) \geq e^{\sqrt{n}}$$

when a satisfies the new (BA_n) for all n. Therefore, there is no need for the condition (FA_n) in this case.

2. In Tsujii (1992b) more parameters are excluded: a stronger condition is given which replaces conditions (BA) and (FA). This new condition implies very easily that the uniform dependence on parameters from the next part of the proof is automatically satisfied. For this reason his proof is shorter. Even so, that the exponential growth of space derivatives implies results on the parameter dependence is interesting in itself.

Step 5: Uniformity of parameter dependence over small parameter intervals

We have shown so far that parameters satisfying (FA) and (BA) give rise to an exponential expansion along the orbit of the critical value. We want to show that these conditions (FA) and (BA) are satisfied for a large set of parameters. For this we need some results which state that parameter and space derivatives are comparable and that the parameter dependence is uniform. Since this only holds for small parameter intervals we will first subdivide the parameter interval ω_0.

First of all, we say that ν is a *return* of an interval $\omega \subset \omega_0$ if $\xi_\nu(a) \in U$ for some $a \in \omega$. Moreover, since we will apply Lemma 6.1 to intervals of parameters instead of single parameters, we also define binding periods and free returns associated to parameter intervals. Firstly, let

$$p(r, \omega) = \min_{a \in \omega} p(r, a).$$

As before, the first free return $\nu_1(\omega)$ of ω is the smallest integer k for which there exists $a \in \omega$ with $\xi_k(a) \in U$. Similarly, if $\nu_i(\omega) \leq n$ is a free return, then let r be the largest integer such that $\xi_{\nu_i(\omega)}(\omega) \cap I_r \neq \emptyset$ and define $p_i(\omega) = p(r, \omega)$ to be the *binding period* associated to this free return. The next *free return of* ω is the largest integer $\nu_{i+1}(\omega)$ such that $\xi_j(a)$ is outside U for all $j = \nu_i(\omega) + p_i(\omega) + 1, \ldots, \nu_{i+1}(\omega) - 1$ and all $a \in \omega$.

Let us define inductively a partition $\mathcal{E}_n$ of $\omega_0 \cap EX_{n-1}$ for $n \geq N$. We should note that ω_0 is a neighbourhood of a_* of size $< \epsilon$ and so in this way $\mathcal{E}_n$ depends on ϵ. Let $\mathcal{E}_N = \{\omega_0\}$ and suppose that $\mathcal{E}_{n-1}$ is defined. (We should point out that this partition is in many ways a quantitative version of the partitions generated by Hofbauer's tower construction which we sketched in Section II.3.) The partition $\mathcal{E}_n$ will refine the partition $\mathcal{E}_{n-1}$ restricted to EX_{n-1}. So take an interval $\omega \subset EX_{n-1}$ in $\mathcal{E}_{n-1}$ and let us consider all the possibilities.

a) n is not a free return of ω. In this case we do not partition ω any further and we let $\omega \in \mathcal{E}_n$.

b) n is a free return of ω and $\xi_n(\omega)$ does not completely contain an interval of the form $I_{r,r'}$ with $r \geq \Delta - 1$. In this case $\xi_n(\omega)$ is strictly contained in the union of at most two intervals $I_{r,r'}$, or $\xi_n(\omega)$ is the union of a piece of at most one interval of the form $I_{\Delta,*}$ and possibly a piece outside U. Then we take again $\omega \in \mathcal{E}_n$ and we call such a return *inessential*.

c) n is a free return of ω and $\xi_n(\omega)$ contains at least one interval $I_{r,r'}$, with $r \geq \Delta$, completely. In this case we say that this return of ω is *essential* and decompose ω into a number of disjoint intervals. This is done as follows. Since $\omega \in \mathcal{E}_{n-1}$ we have by Proposition 6.2 that $\xi_n \colon \omega \to I$ is a diffeomorphism. Therefore, its image covers a number of intervals $I_{r,r'}$, and at each end possibly a piece of $I \setminus U$ or a piece of an interval $I_{r,r'}$. So one can decompose ω into pieces ω' and $\omega_{r,r'}$ with the property that ω' contains the points which are mapped by ξ_n outside U and similarly $\xi_n(\omega_{r,r'})$ is contained in $I_{r,r'}$ for $r \geq \Delta$. Now we join the two intervals at the end of ω from this partition to their neighbours if necessary: if an end interval does not cover an entire interval of the form $I_{r,r'}$ with $r \geq \Delta - 1$, $1 \leq r' \leq r^2$ then we add this interval to its neighbour in this partition, see Figure 6.1. In this way we get a partition $\tilde{\mathcal{E}}_n$ refining $\mathcal{E}_{n-1}$. We subdivide this interval ω' once more to obtain our desired partition $\mathcal{E}_n$. To do this, first notice that we have for each subinterval $\tilde{\omega} \in \tilde{\mathcal{E}}_n$ of ω (assuming this interval has an essential return at time n) one of the following possibilities.

1. $\xi_n(\tilde{\omega})$ contains an interval $I_{r,r'}$ with $r \geq \Delta$ and possibly a piece of a neighbouring interval $I_{r\pm 1,r'\pm 1}$; the interval $I_{r,r'}$ is called the *host interval* of $\xi_n(\tilde{\omega})$ or, more loosely speaking, the host interval of $\tilde{\omega}$ at time n.

2. $\xi_n(\tilde{\omega})$ is completely outside $U_{\Delta+1}$. In this case we say that $\tilde{\omega}$ is an *escape component* of ω. If $|\xi_n(\tilde{\omega})| \geq \sqrt{|U|}$ then we say that ω has a *substantial escape* at time n and we subdivide $\tilde{\omega}$ into subintervals so that the lengths of the images under ξ_n of these smaller subintervals is between $\sqrt{|U|}/2$ and $\sqrt{|U|}$. For simplicity we shall say that $I_{\Delta-1}$ is the host interval of $\xi_n(a)$ for each a in one of these subintervals of $\tilde{\omega}$.

Next we let $\mathcal{E}_n$ be the partition which we obtain after subdividing the elements from $\tilde{\mathcal{E}}_n$ as in 2) above.

Figure 6.1: The decomposition of ω.

For each $a \in \omega_0 \cap EX_{n-1}$ let $\omega_n(a)$ be the interval in $\mathcal{E}_n$ which contains a. Such an interval is generated as follows. Let $\hat{\nu}_1 = \nu_1(\omega_0)$. By definition $\xi_{\hat{\nu}_1}(\omega_0) \supset U$ and so ω_0 has an essential return at $\hat{\nu}_1$. When $n \geq \hat{\nu}_1$ let $\omega_{\hat{\nu}_1}(a)$ be the element of $\mathcal{E}_{\hat{\nu}_1}$ containing a. Then let $\hat{\nu}_2$ be the essential return of $\omega_{\hat{\nu}_1}(a)$ and if $n \geq \hat{\nu}_2$ let $\omega_{\hat{\nu}_2}(a) \subset \omega_{\hat{\nu}_1}(a)$ be the element of the partition $\mathcal{E}_{\hat{\nu}_2}$ containing a. Continuing in this way we have intervals $\omega_n(a) \subset \omega_{\hat{\nu}_s - 1}(a) \subset \cdots \subset \omega_{\hat{\nu}_1}(a) \subset \omega_0$ and essential returns $\hat{\nu}_1 < \hat{\nu}_2 < \cdots < \hat{\nu}_s \leq n$ for which $\omega_{\hat{\nu}_{i-1}}$ has an essential return at time $\hat{\nu}_i$. Moreover, $\omega_j(a)$ is subdivided exactly when $j \in \{\hat{\nu}_1, \ldots, \hat{\nu}_s\}$:

$$\omega_{\hat{\nu}_{i-1}}(a) \neq \omega_{\hat{\nu}_i}(a)$$

and

$$\omega_j(a) = \omega_{\hat{\nu}_i}(a)$$

for $j = \hat{\nu}_i, \ldots, \hat{\nu}_{i+1} - 1$. These times are called the *essential return times of* a. If $\omega_{\hat{\nu}_i}$ is an escape component then it has no return (to $U_{\Delta+1}$) at time $\hat{\nu}_i$. Escapes will play an extremely important role in the proof of Theorem 6.1.

Notice that, by definition, each parameter a in an interval $\omega \in \mathcal{E}_n$ has the same return times for the first $n-1$ iterates. In an informal sense, $\mathcal{E}_n$ defines a partition on BA_n and FA_n. To formalize this, we define BA'_n to be the set of parameters a such that $\omega_n(a) \cap BA_n \neq \emptyset$. Note that if $a \in \omega \in \mathcal{E}_n$ and $a \in BA_n$ then, by construction of ω, $|\xi_k(b) - c| \geq e^{-\alpha k}/2$ for $k = N, \ldots, n$ and each $b \in \omega$. So without altering anything else, BA'_n can be used instead of the set BA_n in the estimates of the previous step. Similarly, for a component ω of $\mathcal{E}_n$, let $\nu_1(\omega) < \cdots < \nu_s(\omega) \leq n$ be its free returns, $p_i(\omega)$ the corresponding binding periods and $q_i(\omega)$ as before. Then we say that ω is contained in FA'_n if $F_k(\omega)/k \geq (1-\tau)$ for $k = 1, \ldots, n$ and otherwise ω is outside this set FA'_n. Therefore, by definition, $\mathcal{E}_n$ defines a partition on BA'_n and FA'_n.

In the next proposition we will prove that

$$BA'_{n-1} \cap FA'_{n-1} \subset EX_{n-1}.$$

Moreover, this proposition shows that the parameter dependence of the position of the iterates $\xi_n(a)$ is uniform on the intervals of $\mathcal{E}_{n-1}$ which are contained in $BA'_{n-1} \cap EX_{n-1}$.

Proposition 6.3. [Uniform parameter dependence] *There exists a positive constant C such that for each sufficiently large Δ there exist ϵ with the following properties. For each $n \geq N$ and each $\omega \in \mathcal{E}_{n-1}$ which has a free return at n and for which $\omega \subset BA'_{n-1} \cap EX_{n-1}$ one has if $\xi_n(\omega) \subset U_{\Delta/2}$ then*

$$\frac{|\xi'_k(a)|}{|\xi'_k(b)|} \leq C \quad \text{for all } a, b \in \omega \text{ and for all } k = 0, \ldots, n.$$

Moreover,

$$BA'_{n-1} \cap FA'_{n-1} \subset EX_{n-1}.$$

In the proof of this proposition we will make use of three lemmas and the following fact. If C' is so that $|\partial_a f_a| \leq C'$ then

$$|\partial_a f_a^j(x)| \leq C' \cdot \sum_{i=0}^{j-1} |Df_a^{j-1-i}(f^i(x))|. \tag{6.13}$$

Let us prove this by induction. For $j = 1$ this expression is trivial. By the Chain Rule

$$\partial_a f_a^{j+1}(x) = \partial_a f_a(f_a^j(x)) + Df_a(f_a^j(x)) \cdot \partial_a f_a^j(x).$$

Hence, by induction,

$$\begin{aligned} |\partial_a f_a^{j+1}(x)| &\leq C' + |Df_a(f_a^j(x))| \cdot |\partial_a f_a^j(x)| \\ &\leq C' + |Df_a(f_a^j(x))| \cdot C' \cdot \sum_{i=0}^{j-1} |Df_a^{j-1-i}(f^i(x))| \\ &\leq C' + C' \cdot \sum_{i=0}^{j-1} |Df_a^{j-i}(f^i(x))| \\ &\leq C' \cdot \sum_{i=0}^{j} |Df_a^{j-i}(f^i(x))|. \end{aligned}$$

The first lemma which we will now state shows that the estimates from Lemma 6.1 still hold if we replace $p(r,a)$ by $p(r,\omega)$. So let $\omega \in \mathcal{E}_{n-1}$ and assume that $\nu < n$ is a free return of ω. Then by definition $\xi_\nu(\omega)$ is contained in the union of an interval of the form $I_{r,r'}$ with possibly pieces of its two neighbours. Let HD-dist(I,J) be the Hausdorff distance of two intervals I and J. So this number is $< \epsilon$ if I is contained in a ϵ-neighbourhood of J and vice versa.

Lemma 6.2. *For each Δ there exists $\epsilon > 0$ such that if $\omega \in \mathcal{E}_{n-1}$ is as above and is contained in $BA'_{n-1} \cap EX_{n-2}$, then for each $a, b \in \omega$, we have*

$$\text{HD-dist}(f_a^j(I_r), f_b^j(I_r)) < \frac{1}{1000}|f_a^j(I_r)|,$$

$$\text{HD-dist}(f_a^j(U_r), f_b^j(U_r)) < \frac{1}{1000}|f_a^j(U_r)|$$

and for each interval $\tilde\omega \subset \omega$,

$$\text{HD-dist}\left(\xi_{\nu+j}(\tilde\omega), f_a^j(\xi_\nu(\tilde\omega))\right) < \frac{1}{1000}|f_a^j(\xi_\nu(\tilde\omega))|$$

for all $1 \le j \le p(r,\omega)+1$.

Proof. Take $x \in I_r$. By the Mean Value Theorem $|f_a^j(x) - f_b^j(x)| \le |\partial_a f_{\tilde a}^j(x)|\,|a-b|$ for some $\tilde a \in [a,b]$. By Lemma 6.1 we have $p(r,\tilde a) < n$ and therefore $|Df_{\tilde a}^i(f_{\tilde a}(x))| = O(1)\cdot|D_i(\tilde a)|$ and that these numbers grow exponentially for $i = 1,\dots,p(r,\tilde a)$. Since (6.13) implies

$$\begin{aligned}|\partial_a f_{\tilde a}^j(x)| &\le C'\cdot|Df_{\tilde a}(x)|\cdot|Df_{\tilde a}^{j-2}(f_{\tilde a}(x))| + C'\cdot\sum_{i=1}^{j-1}|Df_{\tilde a}^{j-1-i}(f_{\tilde a}^i(x))|\\ &\le C'\cdot|Df_{\tilde a}^{j-2}(f_{\tilde a}(x))|\left(\sum_{i=0}^{j-2}\frac{1}{|Df_{\tilde a}^i(f_{\tilde a}(x))|}+1\right),\end{aligned}$$

we get $|\partial_a f_{\tilde a}^j(x)| \le C\cdot|D_{j-2}(\tilde a)| \le C\cdot|D_{j-1}(\tilde a)|$ and therefore

$$|f_a^j(x) - f_b^j(x)| \le C\cdot|D_{j-1}(\tilde a)|\cdot|a-b|.$$

By Statement 1b) of Lemma 6.1, $f_{\tilde a}^{j-1}$ has bounded distortion on $f_{\tilde a}(I_r)$ for $j = 1,2,\dots,p(r,\tilde a)+1$ and hence

$$\begin{aligned}|f_a^j(x) - f_b^j(x)| &\le C\cdot|D_{j-1}(\tilde a)|\cdot|a-b| \le C\cdot|f_{\tilde a}^j(I_r)|\cdot\frac{|\omega|}{|f_{\tilde a}(I_r)|}\\ &\le C\cdot|f_{\tilde a}^j(I_r)|\cdot\frac{|\omega|}{|\xi_\nu(\omega)|}\cdot\frac{e^{-r}/r^2}{|f_{\tilde a}(I_r)|}.\end{aligned}$$

Using Proposition 6.2 from Step 2 and using the fact that $f_{\tilde a}$ has a quadratic critical point, this gives that for some $a' \in \omega$,

$$|f_a^j(x) - f_b^j(x)| \le C\cdot|f_{\tilde a}^j(I_r)|\cdot\frac{1}{|D_{\nu-1}(a')|}\cdot\frac{e^{-r}/r^2}{e^{-2r}} \le C\cdot|f_{\tilde a}^j(I_r)|\cdot e^{-\gamma\nu+r}/r^2.$$

Since $\omega \subset BA'_{n-1}$ one has $r \le 2\alpha\nu$ and it follows that

$$\text{HD-dist}(f_a^j(I_r), f_b^j(I_r)) < \frac{1}{4000} \cdot |f_{\tilde{a}}^j(I_r)| \le \frac{1}{4000} \cdot \max_{a' \in \omega} |f_{a'}^j(I_r)|,$$

for $j = 1, \ldots, p(r, \tilde{a}) + 1$ provided $\epsilon > 0$ is small (because then ν is large). Since this last inequality holds for all $a, b \in \omega$, the first inequality follows. The other inequalities are proved similarly. □

Corollary 6.2. *The estimates from Lemma 6.1 hold for $a \in \omega$ even if one replaces $p(r, a)$ by $p(r, \omega)$.*

Proof. Statements a) and b) of Lemma 6.1 hold trivially. In order to prove Statement c), take $\tilde{a} \in \omega$ such that $p(r, \tilde{a}) = p(r, \omega)$. From the previous lemma one has that $|f_a^j(U_r)| = O(1) \cdot |f_b^j(U_r)|$ for $j = 1, \ldots, p(r, \omega) + 1$ and $a, b \in \omega$. Therefore, from the definition of $p(r, \omega)$ we get for each $a \in \omega$ that

$$\begin{aligned} |f_a^{p(r,\omega)+1}(U_r)| &= |f_a^{p(r,\tilde{a})+1}(U_r)| = O(1) \cdot |f_{\tilde{a}}^{p(r,\tilde{a})+1}(U_r)| \\ &\ge O(1)e^{-\beta p(r,\tilde{a})} = O(1)e^{-\beta p(r,\omega)}. \end{aligned}$$

From this it follows that we can replace $p(r, a)$ by $p(r, \omega)$ in the proof of Lemma 6.1. □

By Proposition 6.2 we have that for an interval $\omega \subset BA'_{n-1} \cap EX_{n-2}$, $|\omega|/|\xi_j(\omega)|$ is exponentially small in terms of j when $j \le n-1$. However, we also need that $|\xi_j(\omega)|/|\xi_\nu(\omega)|$ is exponentially small in terms of $\nu - j$ when $j < \nu \le n-1$ and ν is a free return of ω. This is proved in the next lemma.

Lemma 6.3. *There exists C_0 such that for each sufficiently large Δ there exists $\epsilon > 0$ such that for each $\omega \in \mathcal{E}_{n-1}$ which is contained in $BA'_{n-1} \cap EX_{n-1}$ and for each $\tilde{\omega} \subset \omega$ one has the following. For any consecutive free returns $\nu < \nu' \le n$ of ω,*

$$|\xi_{\nu+j}(\tilde{\omega})| \ge C_0 \cdot e^{\gamma j} \cdot |\xi_{\nu+1}(\tilde{\omega})|$$

for $j = 1, \ldots, p(r, \omega) + 1$ and $|\xi_{\nu+p(r,\omega)+1}(\tilde{\omega})| \ge e^{\gamma p(r,\omega)/4} \cdot |\xi_\nu(\tilde{\omega})|$. Moreover,

$$|\xi_{\nu'}(\tilde{\omega})| \ge C_0 \cdot e^{(\nu'-j)\gamma_0} \cdot |\xi_j(\tilde{\omega})| \tag{6.14}$$

for $j = \nu + p(r, \omega) + 1, \ldots, \nu'$ and

$$|\xi_{\nu'}(\tilde{\omega})| \ge 2 \cdot |\xi_\nu(\tilde{\omega})| \tag{6.15}$$

Proof. Let $p = p(r, \omega)$. By the second inequality from the previous lemma one has HD-dist $\left(\xi_{\nu+j}(\tilde{\omega}), f_a^j(\xi_\nu(\tilde{\omega}))\right) < \frac{1}{1000} \cdot |f_a^j(\xi_\nu(\tilde{\omega}))|$ for $j = 1, \ldots, p+1$ and so Lemma 6.1 implies $|\xi_{\nu+j}(\tilde{\omega})| \ge C_0 \cdot e^{\gamma j} \cdot |\xi_{\nu+1}(\tilde{\omega})|$ for $j = 1, \ldots, p+1$ and $|\xi_{\nu+p+1}(\tilde{\omega})| \ge e^{\gamma p/4} \cdot |\xi_{\nu+1}(\tilde{\omega})|$.

So let us prove the last two inequalities. As before, let $\bar{\gamma} = \sup_{a,x} |Df_a(x)|$. By (6.13), $|\partial_a f_a^j(x)| \le C \cdot e^{\bar{\gamma} j}$. By Proposition 6.2 and since $\omega \subset EX_{n-1}$ one has $|a-b| \le C \cdot e^{-\gamma j} \cdot |\xi_j(\tilde{\omega})|$. Hence,

$$|f_a^{k-j}(\xi_j(b)) - f_b^{k-j}(\xi_j(b))| \le C \cdot e^{\bar{\gamma}(k-j)} \cdot e^{-\gamma j} \cdot |\xi_j(\tilde{\omega})|$$

when $\nu + p + 1 \le j < k \le \nu'$. Taking $\xi_k(\tilde{\omega}) = [\xi_k(a), \xi_k(b)]$ one gets from this and from the second inequality in Proposition 6.2,

$$\begin{aligned} |\xi_k(\tilde{\omega})| &= |f_a^{k-j}(\xi_j(a)) - f_b^{k-j}(\xi_j(b))| \\ &\ge |f_a^{k-j}(\xi_j(a)) - f_a^{k-j}(\xi_j(b))| - |f_a^{k-j}(\xi_j(b)) - f_b^{k-j}(\xi_j(b))| \\ &\ge \left(|Df_a^{k-j}(\tilde{x}_j)| - C \cdot e^{\bar{\gamma}(k-j)} \cdot e^{-\gamma j}\right) \cdot |\xi_j(\tilde{\omega})| \end{aligned} \tag{6.16}$$

where $\tilde{x}_j \in \xi_j(\tilde{\omega})$. Because the factor $e^{\bar{\gamma}(k-j)}e^{-\gamma j}$ can become very large if $(k-j) >> j$ we shall have to use the previous inequality with care. Now if $k - j \le \dfrac{\gamma}{(\gamma+\bar{\gamma})} j$ then, because of the choice of β in (6.4),

$$e^{\bar{\gamma}(k-j)}e^{-\gamma j} \le e^{\gamma\left(\frac{\bar{\gamma}}{\gamma+\bar{\gamma}}-1\right)j} \le e^{-\beta j}$$

which is small if j is large. Hence, using (6.16), we get for $k - j \le \dfrac{\gamma}{(\gamma+\bar{\gamma})} j$,

$$\text{HD-dist}\left(\xi_k(\tilde{\omega}), f_a^{k-j}(\xi_j(\tilde{\omega}))\right) \le e^{-\beta j} |\xi_k(\tilde{\omega})|. \tag{6.17}$$

Now let W_i, $i = 1, \dots, 4$ be interval neighbourhoods of c as in Proposition 6.1 of size $i \cdot \delta$ where δ is some fixed positive number (independent of Δ). Choose integers $k_0, \dots, k_u$ with $\nu + p + 1 = k_0 < k_1 < \cdots < k_u = \nu'$ so that for each $i = 0, \dots, u-1$,

$$k_{i+1} - k_i \le \frac{\gamma}{(\gamma+\bar{\gamma})} k_i.$$

As before, this implies that the last term in (6.16) is very small for $k_i \le j < k \le k_{i+1}$. Moreover, we can choose these integers $k_0, k_1, \dots, k_u$ so that $\xi_{k_{i+1}}(\tilde{\omega}) \cap W_4 = \emptyset$ implies that $\dfrac{\gamma}{2(\gamma+\bar{\gamma})} k_i \le k_{i+1} - k_i$ as well as $\xi_m(\tilde{\omega}) \cap W_4 = \emptyset$ for all $m = k_i + 1, \dots, k_{i+1} - 1$. (Note that if k_0 is large then the subsequent steps $k_{i+1} - k_i$ are also large. In particular, if Δ is large and $\epsilon > 0$ is small enough then we may assume that $k_{i+1} - k_i$ is large for each $i = 0, \dots, u-1$. We shall need this remark in the next lemma.)

If $\xi_{k_{i+1}}(\tilde{\omega}) \cap W_2 \ne \emptyset$ and $\xi_{k_{i+1}}(\tilde{\omega})$ does not contain a component of $W_3 \setminus W_2$ then (6.17) implies that $f_a^{k_{i+1}-k_i}(\xi_{k_i}(\tilde{\omega}))$ is contained in W_4. Since $\xi_j(\tilde{\omega}) \cap U = \emptyset$ for $j = k_i, \dots, k_{i+1}-1$, (6.17) also implies that $f_a^{j-k_i}(\xi_{k_i}(\tilde{\omega}))$ is outside a slightly smaller interval U' for these integers j. From (6.1) of Proposition 6.1,

$$|Df_a^{k_{i+1}-j}(\tilde{x}_j)| \ge C_0 \cdot e^{\gamma_0(k_{i+1}-j)} \text{ for } j = k_i, \dots, k_{i+1}.$$

If $\xi_{k_{i+1}}(\tilde{\omega}) \cap W_2 \ne \emptyset$ and $\xi_{k_{i+1}}(\tilde{\omega})$ contains a component of $W_3 \setminus W_2$ then $\nu' - k_{i+1}$ is universally bounded from above. Hence, from (6.2) in Proposition 6.1,

$$|Df_a^{k_{i+1}-j}(\tilde{x}_j)| \ge C_0 \cdot e^{\gamma_0(\nu'-j)} \text{ for } j = k_i, \dots, \nu'$$

and the last term in (6.16) is small for $k_i \le j \le k \le \nu'$.

Finally, if $\xi_{k_{i+1}}(\tilde{\omega}) \cap W_2 = \emptyset$ we get again from (6.17) that $f_a^m(\tilde{x}_j) \notin W_1$ for $m = 0, \dots, k_{i+1} - j$ and as before we get by (6.3) from Proposition 6.1,

$$|Df_a^{k_{i+1}-j}(\tilde{x}_j)| \ge C_0 \cdot e^{\gamma_0(k_{i+1}-j)} \text{ for } j = k_i + 1, \dots, k_{i+1}.$$

From the choice of α and k_i and using (BA'_{n-1}), this also implies

$$|Df_a^{k_{i+1}-k_i}(\tilde{x}_{k_i})| \ge C_0 \cdot e^{\gamma_0(k_{i+1}-k_i)} \cdot e^{-\alpha k_i} \ge e^{\gamma(k_{i+1}-k_i)}.$$

From all this, (6.14) follows. Since $\nu' - \nu$ tends to infinity as ϵ tends to zero, the last inequality also follows. □

Lemma 6.4. **[Main Distortion Lemma]** *There exists $C < \infty$ such that for each Δ sufficiently large there exists $\epsilon > 0$ such that the following holds. For each component $\omega \in \mathcal{E}_{n-1}$ with $\omega \subset BA'_{n-1} \cap EX_{n-1}$ which has a free return at time $n \ge N$, one has*

$$\frac{|D_k(a)|}{|D_k(b)|} \le C \tag{6.18}$$

for all $k = 0, \dots, n-1$ and all $a, b \in \omega$ when $\xi_n(\omega) \subset U_{\Delta/2}$.

Proof. First we make the following remark. By the construction of the set $\mathcal{E}_n$, the iterates $\xi_i(a)$ have free returns at the same times $\nu_1 < \nu_2 < \nu_3 < \cdots \le n$ for each $a \in \tilde{\omega}$. Now choose s so that $\nu_s \le k < \nu_{s+1}$. By assumption $\xi_{\nu_j}(\omega)$ is contained inside an interval of the form I_{r_j, r'_j} (plus pieces of at most two of its neighbours) for each $j = 1, 2, \dots, s$, where $r_j \ge \Delta$ and $1 \le r'_j \le r_j^2$. By the previous lemma $|\xi_{\nu_{j+1}}(\omega)| \ge 2|\xi_{\nu_j}(\omega)|$ because $\nu_{j+1} - \nu_j$ is large when Δ is large.

Take $k \in \{0, 1, \dots, n-1\}$. By the Chain Rule

$$\frac{|D_k(a)|}{|D_k(b)|} = \prod_{i=1}^{k} \frac{|Df_a(\xi_i(a))|}{|Df_b(\xi_i(b))|}.$$

Let $k_0 \le n$ be maximal so that $|\xi_{k_0}(\omega)| \le |U|$. Let us first estimate

$$\prod_{i=1}^{k} \frac{|Df_a(\xi_i(a))|}{|Df_b(\xi_i(b))|}$$

for $k \le k_0$. Note that $\frac{|Df_a(\xi_i(a))|}{|Df_b(\xi_i(b))|} \le (1 + C \cdot |b-a|) \frac{|\xi_i(a) - c|}{|\xi_i(b) - c|}$ and, by Proposition 6.2, $|\omega| \le Ce^{-\gamma n}$ and therefore $(1 + C \cdot |b-a|)^n$ is universally bounded. So it suffices to estimate $\prod_{i=1}^{k} \frac{|\xi_i(a) - c|}{|\xi_i(b) - c|}$ or in other words

$$\sum_{i=1}^{k_0} \frac{|\xi_i(a) - \xi_i(b)|}{|\xi_i(b) - c|}.$$

Let us split this sum up in several parts: the binding pieces and the following piece of the orbit which is outside U. So let

$$S_0'' = \sum_{i=1}^{\nu_0-1} \frac{|\xi_i(a)-\xi_i(b)|}{|\xi_i(b)-c|}$$

and for $0<j<s$,

$$S_j' = \sum_{i=\nu_j}^{\nu_j+p_j} \frac{|\xi_i(a)-\xi_i(b)|}{|\xi_i(b)-c|} \text{ and } S_j'' = \sum_{i=\nu_j+p_j+1}^{\nu_{j+1}-1} \frac{|\xi_i(a)-\xi_i(b)|}{|\xi_i(b)-c|}.$$

If $\nu_s+p_s<k$ take

$$S_s' = \sum_{i=\nu_s}^{\nu_s+p_s} \frac{|\xi_i(a)-\xi_i(b)|}{|\xi_i(b)-c|}\,,\ \ S_s'' = \sum_{i=\nu_s+p_s+1}^{k-1} \frac{|\xi_i(a)-\xi_i(b)|}{|\xi_i(b)-c|}.$$

If $k\le \nu_s+p_s$ then take $S_s' = \sum_{i=\nu_s}^{k}\frac{|\xi_i(a)-\xi_i(b)|}{|\xi_i(b)-c|}$ and $S_s''=0$. So we need to estimate the sum of the terms S_j' and S_j''.

Let us first estimate the sum of the terms S_j'' for $j=1,\dots,s$. By the previous lemma, $|\xi_i(a)-\xi_i(b)|$ 'grows' exponentially fast for $i=\nu_j+p_j+1,\dots,\nu_{j+1}$. More precisely, the sum of these terms is bounded by C times the last one. Furthermore, $|\xi_i(b)-c|\ge |U|$ for $i=\nu_j+p_j+1,\dots,\nu_{j+1}$ because this part of the orbit is free. So it follows that $S_j''\le C\cdot\frac{|\xi_{\nu_{j+1}}(\omega)|}{|U|}$. Since

$$|\xi_k(\omega)|\ge 2|\xi_{\nu_{j+1}}(\omega)|\ge 4|\xi_{\nu_j}(\omega)|$$

for $j=0,1,\dots,s-1$, this implies that

$$\begin{aligned}\sum_{j=1}^{s} S_j'' &\le \sum_{j=1}^{s-1} C\cdot\frac{|\xi_{\nu_{j+1}}(\omega)|}{|U|} \quad + \quad C\cdot\frac{|\xi_k(\omega)|}{|U|}\\ &\le C\cdot\frac{|\xi_k(\omega)|}{|U|}\le C.\end{aligned}$$

Similarly, S_0'' is universally bounded.

So let us bound $\sum S_j'$. For $\nu_j<i\le\nu_j+p_j$, one has the following. By Lemmas 6.1 and 6.2, $|\xi_i(b)-\xi_i(a)|\le C\cdot|f_a^{i-\nu_j-1}(\xi_{\nu_j+1}(\omega))|$ and $Df_a^{i-\nu_j-1}$ has bounded distortion on some neighbourhood of $f_a(c)$. This gives

$$|\xi_i(b)-\xi_i(a)|\le C\cdot|D_{i-\nu_j-1}(a)|\cdot|\xi_{\nu_j+1}(\omega)|\le C\cdot|D_{i-\nu_j-1}(a)|\cdot|f_a(\omega_{\nu_j}(\omega))|.$$

Furthermore, from the definition of p_j,

$$\begin{aligned}|D_{i-\nu_j-1}(a)|\cdot|\xi_{\nu_j+1}(a)-c_1(a)| &= O(1)\cdot|f_a^{i-\nu_j}(\xi_{\nu_j}(a))-\xi_{i-\nu_j}(a)|\\ &\le C\cdot e^{-\beta(i-\nu_j)}.\end{aligned}$$

Hence,

$$\begin{aligned}|\xi_i(b)-\xi_i(a)| &\le C\cdot|D_{i-\nu_j-1}(a)|\cdot|f_a(\omega_{\nu_j}(\omega))|\le C\cdot\frac{|f_a(\omega_{\nu_j}(\omega))|\cdot e^{-\beta(i-\nu_j)}}{|\xi_{\nu_j+1}(a)-c_1(a)|}\\ &\le C\cdot\frac{|\omega_{\nu_j}(\omega)|\cdot e^{-\beta(i-\nu_j)}}{|U_{r_j}|},\end{aligned}$$

where the last step follows since $f|I_r$ has bounded distortion, $\omega_{\nu_j}(\omega) \subset I_{r_j}$ and $|I_{r_j}| = O(1)\cdot|U_{r_j}|$. Moreover, if $i \le \nu_j(a)+N$ then by definition of N, $\xi_i(b) \notin W$ and so $|\xi_i(b)-c| \ge C$. On the other hand, if $i \ge \nu_j(a)+N$, then from (BA'_{n-1}), (6.2a) and since $i \le \nu_j + p_j$,

$$\begin{aligned}|\xi_i(b) - c| &\ge |\xi_{i-\nu_j}(b) - c|/2 - |\xi_i(b) - \xi_{i-\nu_j}(b)| \\ &\ge (e^{-\alpha(i-\nu_j)}/2 - e^{-\beta(i-\nu_j)}) \ge e^{-\alpha(i-\nu_j)}/4.\end{aligned}$$

It follows that

$$S'_j \le \sum_{i=\nu_j}^{\nu_j+p_j} \frac{|\xi_i(b)-\xi_i(a)|}{|\xi_i(b)-c|} \le C\cdot \sum_{i=\nu_j}^{\nu_j+p_j} \frac{|\omega_{\nu_j}(\omega)|\cdot e^{-\beta(i-\nu_j)}}{|U_{r_j}|\cdot e^{-\alpha(i-\nu_j)}} \le C\cdot\frac{|\omega_{\nu_j}(\omega)|}{|U_{r_j}|},$$

where we have used that $0<\alpha<\beta$. Now let (r) denote the set of indices $j<s$ such that $\omega_{\nu_j}(\omega)\cap I_r \neq \emptyset$. Since $|\omega_{\nu_{j+1}}(\omega)| \ge 2|\omega_{\nu_j}(\omega)|$,

$$\sum_{j\in(r)} \frac{|\omega_{\nu_j}(\omega)|}{|U_r|} \le C \max_{j\in(r)} \frac{|\omega_{\nu_j}(\omega)|}{|U_r|}$$

and since, for $j \le s$, the largest $\omega_{\nu_j}(\omega)$ is still contained in at most three of the r^2 subintervals $I_{r,r'}$ of I_r,

$$\frac{|\omega_{\nu_j}(\omega)|}{|U_r|} \le \frac{3}{r^2}.$$

In particular,

$$\sum_{j=1}^{s} S'_j \le C\cdot\sum_{j=1}^{s}\frac{|\omega_{\nu_j}(\omega)|}{|U_{r_j}|} \le C\cdot\sum_r \max_{j\in(r)}\frac{|\omega_{\nu_j}(\omega)|}{|U_r|} \le C\cdot\sum_r\frac{3}{r^2} < 10\cdot C.$$

It remains to consider the case that $k\in\{k_0,\dots,n-1\}$. So we need to estimate

$$\prod_{i=k_0}^{k}\frac{|Df_a(\xi_k(a))|}{|Df_b(\xi_k(b))|} = \frac{Df_a^{k-k_0}(\xi_k(a))}{Df_b^{k-k_0}(\xi_k(b))}$$

from above. First notice that $|\xi_i(\omega)|$ grows exponentially and in particular,

$$|\xi_i(\omega)| \ge C_0\cdot|U| \text{ for } i=\{k_0,\dots,n\}. \tag{6.19}$$

Since $\omega\in\mathcal{E}_{n-1}$ the intervals $\xi_i(\omega)$ never contain an interval from the collection I_r with $r\ge\Delta$ for $i=k_0,\dots,n-1$. This together with (6.13) implies that the intervals $\xi_i(\omega)$ do not intersect $U_{\Delta+q}$ for $i=k_0,\dots,n-1$ where q depends on C_0 (for example, if $C_0=1$ then $q=1$ works). Now fix Δ for the moment. We claim that there exists $n_0(\Delta)<\infty$ such that $n-k_0\le n_0(\Delta)$ for each $\omega\subset\omega_0$ provided $|\omega_0|$ is sufficiently small. This holds because f_{a_*} has no homtervals and therefore (6.19) implies that there exists $n_0(\Delta)<\infty$ for which

$$\operatorname{int}(f_{a_*}^j(\xi_{k_0}(\omega))) \text{ intersects } c \text{ for some } j\in\{0,1,\dots,n_0(\Delta)\}$$

for each $\omega \subset \omega_0$. If we let $|\omega_0|$ be small enough then the previous claim follows. In particular, for each $\xi, \Delta > 0$ there exists $\epsilon > 0$ such that

$$\sup_x |f_a^j(x) - f_b^j(x)| \le \xi \text{ for all } j = 0, 1, \dots, n - k_0, \quad a, b \in \omega_0$$

provided $|\omega_0| < \epsilon$. So the Hausdorff distance between $\xi_k(\omega)$ and $f_{a_*}^{k-k_0}(\xi_{k_0}(\omega))$ is at most ξ for $k = k_0, \dots, n-1$. Since $\xi_n(\omega) \subset U_{\Delta/2}$ this implies $f_{a_*}^{n-k_0}(\xi_{k_0}(\omega))$ is contained in a small neighbourhood of c. Since f_{a_*} is a Misiurewicz map we can use the last part of Proposition 6.1 and therefore there exists a universal constant $K < \infty$ such that

$$\frac{|Df_{a_*}^j(x)|}{|Df_{a_*}^j(y)|} \le K$$

for each $j = 0, 1, \dots, k - k_0$ and each $x, y \in I$ provided $f_{a_*}^{k-k_0}[x, y]$ is contained in a sufficiently small neighbourhood of c. Combining all this shows that

$$\frac{|Df_a^j(x)|}{|Df_b^j(y)|} \le 2K$$

for each $a, b \in \omega$, $x, y \in \xi_{k_0}(\omega)$ and each $j \in \{0, 1, \dots, n - k_0\}$. Provided $\epsilon > 0$ and therefore $|\omega_0|$ is sufficiently small. □

Proof of Proposition 6.3: The proof of Proposition 6.3 follows immediately from Proposition 6.2 and the previous two lemmas.

Step 6: The exclusions of the parameters

Let us now compare the size of BA'_{n-1} with the size of BA'_n. So take an interval $\omega \in \mathcal{E}_{n-1}$ contained in $BA'_{n-1} \cap EX_{n-1}$ with a return at time n. The part of ω which fails to satisfy condition (BA'_n) corresponds precisely with the intervals $\omega_{r,r'} \subset \omega$ from the partition above with $r > \alpha n$. But we have

Lemma 6.5. *There exists a constant $C_0 > 0$ such that for every sufficiently large Δ there exists $\epsilon > 0$ such that for each $n \in \mathbb{N}$ and each element ω of $\mathcal{E}_{n-1}$ which has a return at time n and which is contained in $BA'_{n-1} \cap EX_{n-1}$,*

$$\frac{|\omega \setminus \cup_{r \ge \alpha n} \omega_{r,r'}|}{|\omega|} \ge 1 - e^{-\alpha n C_0}$$

for $n \ge N$.

Proof. Since ω is an element of $\mathcal{E}_{n-1}$, there exists a free return $\nu < n$ with binding period p such that $\xi_\nu(\omega)$ covers an interval $I_{r,r'}$ and therefore $|\xi_\nu(\omega)| \ge \frac{e^{-r}}{r^2}$ for some $r \le \alpha\nu < \alpha n$. From Lemma 6.3 in Step 5, $|\xi_n(\omega)| \ge e^{\gamma p/4}|\xi_k(\omega)|$. But by Lemma 6.1, $p \ge C_0 r$ and hence

$$|\xi_n(\omega)| \ge e^{\gamma p/4}\frac{e^{-r}}{r^2} \ge \frac{e^{(-1+C_0)r}}{r^2} \ge e^{(-1+C_0/2)\alpha n} \tag{6.20}$$

because $r \le \alpha n$ and n is large when ϵ is small. By possibly taking the largest interval $\tilde{\omega} \subset \omega$ with $\xi_n(\tilde{\omega}) \subset U_{\Delta/2}$ we get that $|\xi_\nu(\tilde{\omega})|$ also satisfied (6.20) provided $\epsilon > 0$ is sufficiently small because then n tends to infinity. As in Proposition 6.3 in Step 5, the distortion of $\xi_\nu \colon \tilde{\omega} \to I$ is bounded by C and since $\xi_\nu(\cup_{r>\alpha n}\omega_{r,i}) \subset U_{\alpha n}$ it follows by (6.20) that

$$\frac{|\cup_{r>\alpha n}\,\omega_{r,i}|}{|\omega|} \le \frac{|\cup_{r>\alpha n}\,\omega_{r,i}|}{|\tilde{\omega}|} \le C \cdot \frac{e^{-\alpha n}}{|\xi_n(\tilde{\omega})|} \le C \cdot e^{-C_0\alpha n/2} \le e^{-C_0\alpha n/4}$$

provided $\epsilon > 0$ is sufficiently small (because then n becomes very large). □

From this lemma it follows that $|BA'_n|/|BA'_{n-1} \cap EX_{n-1}| \ge 1 - e^{-\alpha n C_0}$ for $n \ge N$ and $\epsilon > 0$ sufficiently small. So the set of parameters which violates BA'_n is very small compared to BA'_{n-1}. Let us now show that a similar statement also holds for the set FA'_n. In order to do this we use a large deviation argument.

First we shall show that essential returns occur quite frequently.

Lemma 6.6. *For any sufficiently large Δ there exists $\epsilon > 0$ such that for any $\omega \in \mathcal{E}_{\hat{\nu}}$ with an essential return and host interval $I_{r,r'}$, $r \ge \Delta - 1$, at time $\hat{\nu}$ one has the following. (Here, by convention, $r = \Delta - 1$ when an escape takes place.) If $\hat{\nu}'$ is the next essential return of ω and $\omega \subset EX_{\hat{\nu}'-1} \cap BA'_{\hat{\nu}'-1}$ then*

$$\hat{\nu}' - \hat{\nu} \le 4r/\gamma \tag{6.21a}$$

and

$$\text{if } \hat{\nu}' - \hat{\nu} = 4r/\gamma \text{ then } \omega \text{ has a substantial escape at time } \hat{\nu}'. \tag{6.21b}$$

Moreover,

$$|\omega'| \le C \cdot e^{6\beta r/\gamma} \cdot e^{-\hat{r}} \cdot |\omega| \tag{6.22}$$

when $\omega' \subset \omega$ is so that $\xi_{\hat{\nu}'}(\omega') \subset I_{\hat{r},\hat{r}'}$.

Proof. If ω has an escape at time $\hat{\nu}$ then replace ω by the subset $\tilde{\omega}$ for which $\xi_{\hat{\nu}}(\tilde{\omega}) = I_{\Delta,\Delta^2}$. Let $\hat{\nu} = \nu_0 < \cdots < \nu_k = \hat{\nu}'$ be the free returns of ω between time $\hat{\nu}$ and $\hat{\nu}'$. By definition these returns are inessential and $k \ge 1$. Let p_j be the binding period of ω following the return ν_j and let $L_j = \nu_{j+1} - \nu_j - p_j - 1$. By lemma 6.2 the Hausdorff-distance between $\xi_{\nu_j+p_j}(\omega)$ and $f_a^{p_j}(\xi_{\nu_j}(\omega))$ is very small compared to the size of these intervals. Furthermore, if $\xi_{\nu_j}(\omega) \subset I_{r_j,r'_j}$ then by the choice of p_j, $|f_a^{p_j}(U_{r_j})| = O(1) \cdot e^{-\beta p_j}$ and by Lemma 6.1, $f_a^{p_j}|I_{r_j}$ has bounded distortion. Therefore, and since $p_j \le 3r_j/\gamma$,

$$\frac{|\xi_{\nu_j+p_j}(\omega)|}{|\xi_{\nu_j}(\omega)|} \ge C_0 \cdot \frac{e^{-\beta p_j}}{|I_{r_j,r'_j}|} \ge C_0 \cdot e^{-\beta p_j} r_j^2 e^{r_j} \ge e^{[1-4\beta/\gamma]r_j} \cdot e^{\beta p_j/2}.$$

because $r_j \ge \Delta - 1$ and Δ is large. Since $\omega \subset EX_{n-1}$ and $\hat{\nu}' \le n$, Lemma 6.3 gives

$$\begin{aligned} |\xi_{\nu_{j+1}}(\omega)| &\ge C_0 \cdot e^{\gamma_0 L_j} \cdot |\xi_{\nu_j+p_j}(\omega)| \\ &\ge C_0 \cdot e^{\gamma_0 L_j} \cdot e^{[1-5\beta/\gamma]r_j} \cdot e^{\beta p_j/2} \cdot |\xi_{\nu_j}(\omega)| \\ &\ge e^{\gamma L_j} \cdot e^{[1-5\beta/\gamma]r_j} \cdot |\xi_{\nu_j}(\omega)| \end{aligned}$$

for $\epsilon > 0$ sufficiently small, since $L_j + p_j \to \infty$ as ϵ tends to zero. By definition $|\xi_{\nu_0}(\omega)| = O(1)e^{-r}/r^2$ and so the last inequality gives

$$|\xi_{\nu_1}(\omega)| \geq e^{\gamma L_0} \cdot e^{[1-4\beta/\gamma]r} e^{-r}/r^2 \geq e^{\gamma L_0} \cdot e^{-[5\beta/\gamma]r}$$

since $r \geq \Delta - 1$ and Δ is large. In particular, $L_0 \leq \frac{5\beta r}{\gamma^2}$. Moreover, we have for $j = 2, \ldots, k$,

$$|\xi_{\nu_j}(\omega)| \geq |\xi_{\nu_1}(\omega)| \prod_{m=1}^{j-1} e^{\gamma L_m} e^{[1-5\beta/\gamma]r_m} \geq e^{\gamma L_0} e^{-5\beta r/\gamma} \prod_{m=1}^{j-1} e^{\gamma_0 L_m} e^{[1-5\beta/\gamma]r_m}.$$

Hence,

$$|\xi_{\nu_j}(\omega)| \geq e^{-5\beta r/\gamma} \text{ for } j = 1, \ldots, k \tag{6.23}$$

and

$$\sum_{m=1}^{k-1} \{\gamma L_m + [1 - 5\beta/\gamma]r_m\} \geq 5\beta r/\gamma \text{ implies } |\xi_{\nu_k}(\omega)| \geq 1.$$

From this and the choice of β,

$$\sum_{m=1}^{k-1} \{\gamma L_m + 3r_m\} \geq 16\beta r/\gamma \text{ implies } |\xi_{\nu_k}(\omega)| \geq 1.$$

Hence, because $p_j \leq 3r_j/\gamma$ by Lemma 6.1, either

$$\begin{aligned}
\nu_k - \nu_0 = \sum_{m=0}^{k-1} (p_m + L_m) &\leq 3r/\gamma + 5\beta r/\gamma^2 + \sum_{m=1}^{k-1} \{3r_m/\gamma + L_m\} \\
&\leq \frac{1}{\gamma} \left[3r + 5\beta r/\gamma + \sum_{m=1}^{k-1} \{\gamma L_m + 3r_m\}\right] \\
&\leq \frac{1}{\gamma} [3r + 5\beta r/\gamma + 16\beta r/\gamma] \leq \frac{4r}{\gamma}
\end{aligned}$$

or $|\xi_{\nu_k}(\omega)| \geq 1$. From this (6.21a) and (6.21b) follow. So let us prove (6.22). By Proposition 6.3, if $|\xi_{\hat\nu'}(\omega)| \leq |U|$ then the ratio of $|\omega'|$ to $|\omega|$ is at most C times the ratio of $|\xi_{\hat\nu'}(\omega')|$ and $|\xi_{\hat\nu'}(\omega)|$. This and (6.23) imply (6.22), provided Δ is sufficiently large. f $|\xi_{\hat\nu'}(\omega)| \geq |U|$ then we can prove again (6.22) simply by shrinking ω. □

Let a have an essential return at time $\hat\nu$ and let $\hat\nu = \hat\nu_0 < \hat\nu_1 < \hat\nu_2 < \cdots < \hat\nu_s \leq n$ the subsequent essential returns of a. Let $I_{r_i,r_i'}$ be the host interval of $\xi_{\hat\nu_i}(\omega_{\hat\nu_i}(a))$. By (6.20) from the previous lemma,

$$\begin{aligned}
\frac{|\omega_{\hat\nu_s}(a)|}{|\omega_{\hat\nu}(a)|} &\leq C^s \exp\left\{\sum_{i=1}^{s}[6\beta r_{i-1}/\gamma - r_i]\right\} \\
&\leq C^s \exp\left\{-(7/8)\sum_{i=1}^{s} r_i + 6\beta \cdot r/\gamma\right\}
\end{aligned} \tag{6.24}$$

where we write $r_0 = r$ and $r_i = \Delta - 1$ if an escape takes place at time ν_i.

As we shall show below, substantial escapes can be used to estimate the function F_n and therefore be used to show that the condition FA_n holds for many parameters. Therefore we define for an essential return $\hat{\nu}$ of a,

$$E(a;\hat{\nu}) = \inf\{k > 0\,;\ \omega_{\hat{\nu}+k}(a) \text{ has a substantial escape at time } \hat{\nu} + k\}.$$

Let us estimate this function.

Lemma 6.7. *For each Δ sufficiently large there exists $\epsilon > 0$ such that for each $n \in \mathbb{N}$ and each $\omega \in \mathcal{E}_{\hat{\nu}}$ which has an essential return at time $\hat{\nu} < n$, which is contained in $BA'_{n-1} \cap EX_{n-1}$ and has host interval $I_{r,r'}$ where $r \geq \Delta$ one has*

$$\int_{\{a\in\omega\,;\ 6r/\gamma \leq E(a;\hat{\nu}) \leq n-\hat{\nu}\}} e^{\gamma E(a;\hat{\nu})} da \leq e^{-r/8} \cdot |\omega|,$$

$$\int_{\{a\in\omega\,;\ E(a;\hat{\nu}) \leq 6r/\gamma\}} e^{\gamma E(a;\hat{\nu})} da \leq e^{r/6} \cdot |\omega|.$$

Here we define the integral over the empty set to be zero.

Proof. The second inequality holds trivially. Also, we may confine ourselves to parameters $a \in \omega$ for which $E(a;\hat{\nu}) \leq n - \hat{\nu}$ because the other parameters do not contribute to the first integral. So take such a parameter $a \in \omega$ and let $\hat{\nu} = \hat{\nu}_0 < \hat{\nu}_1 < \cdots < \hat{\nu}_s \leq \hat{\nu}_{s+1} \leq n$ be the essential returns of a where $\omega_{\hat{\nu}_i}(a)$ at time $\hat{\nu}_{i+1}$ has no substantial escape for $i = 0, \ldots, s-1$ and does have a substantial escape for $i = s$. Let J_i be the host interval of $\omega_{\hat{\nu}_i}(a)$ for $i = 1, \ldots, s$ and let $R = r_1 + \cdots + r_s$. Of course, $J_i = I_{r_i,r'_i}$ with $r_i \geq \Delta - 1$ when the interval does not escape and by convention $J_i = I_{\Delta-1}$ if it does escape. We want to consider all possible paths of host intervals $(J_1, \ldots, J_s)$ with given length s and fixed $R = r_1 + r_2 + \cdots + r_s$. In other words, given s and R we want to estimate the number of strings $((r_1, r'_1), (r_2, r'_2), \ldots, (r_s, r'_s))$ of non-negative integers with

$$R = r_1 + \cdots + r_s \text{ and } r'_i \leq r_i^2.$$

Since $r_i \geq \Delta - 1$, one has $s \leq R/\Delta$. Fixing s it is easy to check that there are $\binom{R+s-1}{s-1}$ solutions of $R = r_1 + \cdots + r_s$ with $r_i \geq 0$. One can see this for example by thinking of a row consisting of $R + s - 1$ holes each filled with a marble. $\binom{R+s-1}{s-1}$ configurations can be created by taking $s - 1$ of these marbles out; moreover, these empty holes partition the remaining R marbles into s groups of $r_i \geq 0$, $i = 1, \ldots, s$, marbles. Distinguishing the orbits which enter the left and right components of I_r and counting the r'_i gives

$$2^s \binom{R+s-1}{s-1} \cdot \prod_{i=1}^{s} r_i^2$$

possibilities. Let us estimate this from above. Since by Simpson's Formula, $n! \approx (n/e)^n\sqrt{2\pi n}$, and using that $s \le R/\Delta$, it follows that there are at most

$$C \cdot 2^s \frac{(s+R-1)^{s+R-1}}{R^R(s-1)^{s-1}}\sqrt{\frac{s+R-1}{(s-1)R}} \cdot e^{(1/16)\sum_{i=1}^s r_i} \le e^{R/16} \cdot (1+o(\Delta))^R$$

possibilities, where $o(\Delta)$ is a function such that $o(\Delta) \to 0$ as $\Delta \to \infty$ and where we have used that $t^2 \le e^{t/16}$ for t large. Let $A_{s,R}$ be the set of parameters $a \in \omega$ which have their first substantial escape at the $(s+1)$-th essential return following $\hat{\nu}$, which have an essential return with host interval inside I_r at time $\hat{\nu}$ and for which $\sum_{i=1}^s r_i = R$. Since $r_i \ge \Delta$, we get $A_{s,R} = \emptyset$ when $s > R/\Delta$. By the previous argument this set has at most $e^{R/16}(1+o(\Delta))^R$ components. Let $\hat{\omega}_s$ be the largest one. So we get

$$|A_{s,R}| \le e^{R/16} \cdot (1+o(\Delta))^R|\hat{\omega}_s|.$$

We claim that for each $a \in \omega \cap A_{s,R}$ for which $E(a;\hat{\nu}) \le n - \hat{\nu}$,

$$E(a;\hat{\nu}) \le (4R+4r)/\gamma.$$

Indeed, by the previous lemma, $\hat{\nu}_i - \hat{\nu}_{i-1} < 4r_{i-1}/\gamma$ for $i = 1,\dots,s$ (where $r_0 = r$) because the corresponding intervals $\omega_{\hat{\nu}_{i-1}}(a)$ do not have a substantial escape at time $\hat{\nu}_i$ for $i \le s$. Moreover, $\hat{\nu}_{s+1} - \hat{\nu}_s \le 4r_s/\gamma$ and $\xi_{\hat{\nu}_{s+1}}(\omega_{\hat{\nu}_s(a)})$ is certainly a substantial escape when equality holds. Furthermore, by definition $t \le \hat{\nu}_{s+1} - \hat{\nu}_0$. From this the claim follows. Similarly, $s > R/\Delta$ because $r_i \ge \Delta$. So if we write $B_{s,R,t} = \{a \in \omega\,;\ E(a;\hat{\nu}) = t\} \cap A_{s,R}$ then $B_{s,R,t} = \emptyset$ if $s > R/\Delta$ if $n - \hat{\nu} \le t < (4R+4r)/\gamma$. Therefore, for any $t \le n - \hat{\nu}$,

$$\begin{aligned}|\{a \in \omega\,;\ E(a;\hat{\nu}) = t\}| &= \sum_{s,R}|B_{s,R,t}| \le \sum_{s \le R/\Delta, R \ge \gamma t/4 - r}|A_{s,R}| \\ &\le \sum_{R=\gamma t/4-r}^{\infty}\sum_{s=1}^{R/\Delta} e^{R/16} \cdot (1+o(\Delta))^R \cdot |\omega_s|.\end{aligned}$$

Using (6.24) this gives

$$\begin{aligned}&|\{a \in \omega\,;\ E(a;\hat{\nu}) = t \le n - \hat{\nu}\}| \\ &\quad\le \sum_{R=\gamma t/4-r}^{\infty}\sum_{s=1}^{R/\Delta} C^s \cdot e^{R/16} \cdot (1+o(\Delta))^R \cdot \exp\{-(7/8)R + 6\beta r/\gamma\} \cdot |\omega| \\ &\quad\le \sum_{R=\gamma t/4-r}^{\infty} (1+o(\Delta))^R \cdot \exp\left\{-\frac{3}{4}R + 6\beta r/\gamma\right\} \cdot |\omega| \\ &\quad\le \sum_{R=\gamma t/4-r}^{\infty} \exp\left\{(o(\Delta) - \frac{3}{4})R\right\} \cdot \exp\{6\beta r/\gamma\} \cdot |\omega| \\ &\quad\le C \cdot \exp\left\{(o(\Delta) - \frac{3}{4})(\frac{\gamma t}{4} - r)\right\} \cdot \exp\{6\beta r/\gamma\} \cdot |\omega|.\end{aligned}$$

If $n - \hat{\nu} \le t \le 6r/\gamma$ this implies

$$|\{a \in \omega\,;\; E(a;\hat{\nu}) = t\}| \le C \cdot \exp\left\{(o(\Delta) - \frac{3}{4})\frac{t}{12}\right\} \cdot \exp\{6\beta r/\gamma\} \cdot |\omega|$$
$$\le \exp\left\{-\frac{t}{20}\right\} \cdot |\omega|,$$

provided Δ is sufficiently large. Since $\gamma < 1/40$ it follows that

$$\int_{\{a \in \omega\,;\; 6r/\gamma \le E(a;\hat{\nu}) \le k\}} e^{\gamma E(a;\hat{\nu})}\, da \le \sum_{t \ge 6r/\gamma} e^{\gamma t} e^{-t/20} |\omega| \le e^{-r/8} |\omega|,$$

for Δ large enough. The last inequality in the statement of this lemma is obvious. □

Now take $a \in \omega_0$. Let ν_1 be as in Step 3 and define $e_1 = \nu_1$. Note that $\xi_k(\omega_0) \cap U = \emptyset$ for all $k = 1, \ldots, \nu_1 - 1$ and $\xi_{\nu_1}(\omega_0) \supset W$. So ω_0 has a substantial return at this time e_1. So take $\omega_{e_1(a)}(a)$ to be the component of $\mathcal{E}_{e_1(a)}$ containing a. Suppose that $e_1(a), \ldots, e_i(a)$ are defined so that $\omega_{\bar{\nu}_{i-1}(a)}(a)$ has a substantial return at time $\bar{\nu}_i(a)$ where $\bar{\nu}_i(a) = e_1(a) + \cdots + e_i(a)$. Then let $e_{i+1}(a)$ be the smallest integer such that $\omega_{\bar{\nu}_i(a)}$ has a substantial return at time $\bar{\nu}_i(a) + e_{i+1}(a)$. Furthermore, for $i \ge 0$, let

$$E_i(a) = \begin{cases} 0 & \text{if } a \text{ escapes at time } \bar{\nu}_i(a) \\ e_{i+1} & \text{otherwise.} \end{cases}$$

Note that if a escapes at one of these times then the escape if substantial by definition. Also define

$$T_n(a) = \sum_{i=0}^{s-1} E_i(a),$$

where s is the maximal integer such that $e_1(a) + e_2(a) + \cdots + e_s(a) \le n$ and the empty sum is defined to be zero. By definition $T_n(a)$ is constant on each component of $\mathcal{E}_n$.

Lemma 6.8. *For Δ sufficiently large, we have for any $\omega \in EX_{n-1} \cap BA'_{n-1}$ the following. If a has a substantial escape at time $\bar{\nu}_i(a)$ and if $\bar{\nu}_{i+1}(a) \le n$ is the next return of a to U then this return has a substantial escape. In particular,*

$$F_n(a) \ge n - T_n(a).$$

Proof. $n - T_n(a)$ is the sum of E_i for those indices i for which $\omega_{\bar{\nu}_i(a)}(a)$ is a substantial escape component at time $\bar{\nu}_i(a)$. So for these indices, $\xi_{\bar{\nu}_i(a)}(\omega_{\bar{\nu}_i(a)})$ contains an interval of at least size $\ge \sqrt{|U|}$ and is outside U. As in the proof of Lemma 6.6, for the smallest integer $\nu'_i(a) > \bar{\nu}_i(a)$ for which $\omega_{\hat{\nu}_i(a)}(a)$ has a return one has

$$|\xi_{\nu'_i(a)}(\omega_{\bar{\nu}_i(a)})| \ge \sqrt{|U|}.$$

It follows that this return has again a substantial escape. So $\nu_i'(a) = \bar{\nu}_i(a) + E_i(a) = \bar{\nu}_{i+1}(a)$ and $\xi_k(\omega_{\bar{\nu}_i(a)})$ stays outside U for $k = \bar{\nu}_i(a), \dots, \bar{\nu}_i(a) + E_i(a)$. Moreover, this part of the orbit is not part of a bound period. So it follows that $\bar{\nu}_i(a), \dots, \bar{\nu}_i(a) + E_i(a)$ is part of a free orbit of length $E_i(a)$. It follows that $F_n(a) \geq n - T_n(a)$. □

So let us estimate T_n.

Lemma 6.9. *For Δ sufficiently large one has the following. Let $\hat{\omega}$ be the union of the components $\omega \in \mathcal{E}_{n-1}$ which are contained in $EX_{n-1} \cap BA'_{n-1}$. Then*

$$\int_{\hat{\omega}} e^{\gamma T_n(a)}\, da \leq e^{\tau n} \cdot |\hat{\omega}|$$

and

$$|\{a \in \hat{\omega}\,;\ T_n(a) > \tau n\}| \leq e^{-\tau n \gamma/2} \cdot |\hat{\omega}|.$$

Proof. Let ω^s be the part of $\hat{\omega}$ which has s substantial escapes and let $\omega_{(\bar{r}_1,\dots,\bar{r}_s)}$ be the set of parameters in ω^s which experience precisely substantial return at $\nu_1 < \nu_2 < \cdots < \nu_s$ and with host intervals J_i. Here J_i is an interval $I_{\bar{r}_i}$ as above where $\bar{r}_i$ is of the form (r_i, r_i'), $r_i \geq \Delta - 1$ and $r_i' \in \{1, \dots, r_i^2\}$. By definition $r_i = \Delta - 1$ occurs in the case of a substantial escape component. For convenience of notation let

$$\omega_{(\bar{r}_1,\dots,\bar{r}_{s-1},*)} = \bigcup_{\bar{r}_s} \omega_{(\bar{r}_1,\dots,\bar{r}_{s-1},\bar{r}_s)},$$

and, more generally,

$$\omega_{(\bar{r}_1,\dots,\bar{r}_k,*,\dots,*)} = \bigcup_{i=k+1}^{s} \bigcup_{\bar{r}_i} \omega_{(\bar{r}_1,\dots,\bar{r}_s)}.$$

Notice that $E_0, \dots, E_{s-2}$ are constant on $\omega_{(\bar{r}_1,\dots,\bar{r}_{s-1},*)}$ and therefore

$$\int_{\omega_{(\bar{r}_1,\dots,\bar{r}_{s-1},*)}} e^{\gamma T_n(a)}\, da = e^{\gamma \sum_{i=0}^{s-2} E_i(a)} \int_{\omega_{(\bar{r}_1,\dots,\bar{r}_{s-1},*)}} e^{\gamma E_{s-1}(a)}\, da.$$

Furthermore, since $E_i > 0$ only if $\bar{r}_i \geq \Delta$ and since $E_{s-1}(a) < n - \hat{\nu}_{s-1}$ for $a \in \omega_{(\bar{r}_1,\dots,\bar{r}_{s-1},*)}$,

$$\begin{aligned}
&\int_{\omega_{(\bar{r}_1,\dots,\bar{r}_{s-1},*)}} e^{\gamma E_{s-1}(a)}\, da \\
&\quad \leq \sum_{r_s=\Delta}^{\infty} \sum_{r_s'=1}^{r_s^2} \left[\int_{\omega_{(\bar{r}_1,\dots,\bar{r}_s)}} e^{\gamma E_{s-1}(a)}\, da \right] + |\omega_{(\bar{r}_1,\dots,\bar{r}_{s-1},*)}| \\
&\quad = \sum_{r_s=\Delta}^{\infty} \sum_{r_s'=1}^{r_s^2} \left[\int_{\{a \in \omega_{(\bar{r}_1,\dots,\bar{r}_s)}\,;\ 6r_s/\gamma \leq E_{s-1}(a) \leq n - \hat{\nu}_{s-1}\}} e^{\gamma E_{s-1}(a)}\, da \right. \\
&\qquad \left. + \int_{\{a \in \omega_{(\bar{r}_1,\dots,\bar{r}_s)}\,;\ E_{s-1}(a) \leq 6r_s/\gamma\}} e^{\gamma E_{s-1}(a)}\, da \right] + |\omega_{(\bar{r}_1,\dots,\bar{r}_{s-1},*)}|.
\end{aligned}$$

Hence, by Lemma 6.7,

$$\int_{\omega_{(\bar r_1,\dots,\bar r_{s-1},*)}} e^{\gamma E_{s-1}(a)}\,da \le \sum_{r_s=\Delta}^{\infty}\sum_{r_s'=1}^{r_s^2} |\omega_{(\bar r_1,\dots,\bar r_s)}| \left[e^{-r_s/8}+e^{r_s/6}\right] + |\omega_{(\bar r_1,\dots,\bar r_{s-1},*)}|.$$

Using Proposition 6.3 and that the escape is substantial, this gives

$$\begin{aligned}\int_{\omega_{(\bar r_1,\dots,\bar r_{s-1},*)}} & e^{\gamma E_{s-1}(a)}\,da \\ &\le \left\{\sum_{r_s=\Delta}^{\infty} C\cdot\frac{e^{-r_s}/r_s^2}{e^{-\Delta/2}}\left[e^{-r_s/8}+e^{r_s/6}\right] + 1\right\}|\omega_{(r_1,\dots,r_{s-1},*)}| \\ &\le C\cdot|\omega_{(\bar r_1,\dots,\bar r_{s-1},*)}|.\end{aligned}\tag{6.25}$$

Furthermore,

$$\int_{\omega_{(\bar r_1,\dots,\bar r_{s-2},*,*)}} e^{\gamma(E_{s-1}(a)+E_{s-2}(a))}\,da \le \sum_{\bar r_{s-1}=\Delta-1}^{\infty}\sum_{\bar r_s=\Delta-1}^{\infty}\left[\int_{\omega_{(\bar r_1,\dots,\bar r_s)}} e^{\gamma E_{s-2}(a)}e^{\gamma E_{s-1}(a)}\,da\right].$$

Since all the previous integrals can be written as infinite sums, we get from the last inequality and using (6.25) twice,

$$\begin{aligned}\int_{\omega_{(\bar r_1,\dots,\bar r_{s-2},*,*)}} & e^{\gamma(E_{s-1}(a)+E_{s-2}(a))}\,da \\ &\le \sum_{\bar r_{s-1}=\Delta-1}^{\infty}\left[e^{\gamma E_{s-2}(a)}\cdot C\cdot|\omega_{(\bar r_1,\dots,\bar r_{s-1},*)}|\right] \\ &= \int_{\omega_{(\bar r_1,\dots,\bar r_{s-2},*,*)}} C\cdot e^{\gamma E_{s-2}(a)}\,da \\ &\le C^2\cdot|\omega_{(\bar r_1,\dots,\bar r_{s-2},*,*)}|.\end{aligned}$$

Repeating this argument s times,

$$\int_{\omega^s} e^{\gamma T_n(a)}\,da \le C^s\cdot|\omega^s| \le e^{\tau^2 n}\cdot|\omega^s|$$

where we used that s/n tends to zero as $\Delta\to\infty$ and $\epsilon\to 0$. (This last statement holds because c is a non-periodic point for f_{a_*}. Therefore there exist for each $l\in\mathbb{N}$ constants $\kappa>0,\epsilon>0$ such that if $|a-a_*|<\epsilon$, L is an interval of length $\le\sqrt{\kappa}$ and U is a neighbourhood of c of length κ then $f_a^i(L)\cap U\neq\emptyset$ for at most one $i\in\{0,1,\dots,l\}$. It is precisely for this reason that we subdivided the substantial escape intervals in Section 5 into intervals of length $\in [\sqrt{|U|}/2, \sqrt{|U|}]$ so that they are short when a return occurs.) Since this holds for any such set ω^s, one gets

$$\int_{\hat{\omega}} e^{\gamma T_n(a)}\, da \le e^{\tau^2 n} \cdot |\hat{\omega}|.$$

Therefore,

$$|\{a \in \hat{\omega}\,;\ T_n(a) > \tau n\}| \le e^{-\tau\gamma n} \int_{\hat{\omega}} e^{\gamma T_n(a)}\, da \le e^{-\tau n(\gamma - \tau)} \cdot |\hat{\omega}|.$$

This completes the proof of this lemma. □

It follows that the proportion of ω which is not in $\mathcal{E}_{n+1}$ because it violates condition (FA_{n+1}) is exponentially small in terms of n.

Step 7: The conclusion of the proof of Theorem 6.1

It follows from the estimates from the previous step that the total length of the components from $\mathcal{E}_n$ which satisfy conditions (FA_n) and (BA_n) is at least

$$|\omega_0| \cdot \prod_{i=0}^{n} \left[1 - C \cdot e^{-iC_0}\right]$$

where these constants C, C_0 can be taken as close to 0 as one likes by choosing $\epsilon > 0$ sufficiently small. Since the infinite product of these terms is bounded away from zero, the size of the set of parameters in ω_0 which satisfy these conditions for all n, is a definite proportion of the size of the set ω_0. This proportion even tends to one as ω_0 shrinks to a_*. In particular, a_* is a density point of the set $\bigcap_{n\ge 0}(BA_n \cap FA_n)$. This completes the proof of Theorem 6.1. □

7 Some Further Remarks and Open Questions

In the first section of this chapter we have shown the ergodicity of unimodal maps satisfying the negative Schwarzian derivative condition. The ergodicity in the multimodal case has also been shown by Blokh and Lyubich (1989c) and (1990c), see also Lyubich (1991). One question which had been open for some time was whether absorbing Cantor attractors can exist. Recently, several papers have been written – using entirely different methods – showing that Cantor attractors do not exist for maps with negative Schwarzian derivative and a quadratic critical point. Jakobson and Świątek (1991a) show this for maps near the full map $f(x) = 4x(1-x)$ using an inductive inducing method, see also Guckenheimer and Johnson (1990). Milnor and Lyubich (1991) and Keller and Nowicki (1992) show that a Fibonacci map as above has no absorbing Cantor attractor by showing that it has an absolutely continuous invariant probability measure. In the paper of Milnor and Lyubich, complex methods are used similar to those developed in the last chapter which show that the summability condition from

Section 4 is satisfied. Keller and Nowicki extend the latter result: they do not use complex extensions and also show that such a measure exists provided the critical point is of order $l < 2 + \epsilon$ (in fact, if $l > 2$ the summability condition is not satisfied). Lyubich (1992a) has given the proof of the absence of absorbing Cantor attractors in the case of unimodal maps with negative Schwarzian derivative and with quadratic critical point. It is now generally believed that absorbing Cantor attractors can exist in the general case:

Conjecture 1: If $f\colon [-1,1] \to [-1,1]$ is smooth unimodal Fibonacci map with a critical point of order l then f has an absorbing Cantor attractor when $l \geq 6$.

Taking a slightly different direction we can pose questions of the following type. Suppose that f is a smooth, non-renormalizable unimodal map and has a non-flat critical point. Moreover, suppose that f has a periodic attractor which does not attract the turning point. What it the size of the basin of the attractor? If this map satisfies the Misiurewicz condition then it is shown in Van Strien (1990) that the basin has full measure. In general, it could be conceivable that the set of points which do not tend to a periodic attractor has positive Lebesgue measure.

Similarly, consider a unimodal map $f\colon [0,1] \to [0,1]$ with negative Schwarzian derivative. Martens has shown that the attractor of f has either zero Lebesgue measure or consists of intervals, see Section 1. However, there is still another way to define the attractor of a map as the union of the supports of the measures from the collection

$$\hat{\omega}_f(\delta_x) = \{\nu\,;\ \nu \text{ is an accumulation point of } \frac{1}{n}\sum_{i=0}^{n-1} \delta_{f^i(x)}\}.$$

As we saw in Section 5, Keller and Hofbauer have shown that the support of these measures can be quite unexpected. In one example, $\hat{\omega}_f(\delta_x) = \delta_p$ for almost all x where p is a repelling fixed point. So in this example, the 'physical attractor' consists of a repelling fixed point. However, this set of limit measures can also be much larger. So one may ask how large the union of the supports of these physical limit measures is. For example, is it true that

$$\cup_{x\in[0,1]} \left\{\text{supp}(\nu)\,;\ \nu \in \cup_{\hat{\omega}_f(\delta_x)}\right\}$$

either has Lebesgue measure zero or that f has an absolutely continuous invariant measure?

As we have seen in this chapter, unimodal maps for which the summability condition from Section 4 is satisfied have absolutely continuous invariant probability measures. Compared to the situation for circle diffeomorphisms this is far from satisfactory. There, some natural topological conditions (the Diophantine conditions on the rotation numbers) were sufficient to get absolutely continuous invariant probability measures. In the case of smooth unimodal maps with negative Schwarzian derivative only a few topological conditions are known to imply the existence of an absolutely continuous invariant probability measure: i) the Misiurewicz condition, ii) the Fibonacci map, see Lyubich and Milnor (1991) and also Keller and Nowicki (1992), iii) some maps satisfying some 'starting conditions', see Jakobson and Świątek (1991b) and iv) some topological conditions in the case of 'long-branched maps', see Bruin (1992b) and (1992c). Most likely, much more general topological conditions can be found which are sufficient to obtain invariant measures. One possible condition is the following:

Conjecture 2: Suppose that f is a unimodal map with $Sf < 0$. Suppose that

$$\sup\{i \geq 1\,;\ f^{ik} \text{ is a diffeomorphism on } [f^n(c), f^{n+k}(c)] \text{ for } k, n \in \mathbb{N}\} < \infty.$$

Then f has an absolutely continuous invariant measure.

This proposal is related to the one given by Nowicki and Przytycki (1989). Note that the integer i in this conjecture is like the depth we encountered in Chapter IV and which was first defined by Blokh and Lyubich (1989d). For the maps without absolutely continuous invariant probability measures which were constructed in Section I.5 and V.5 (having almost saddle-nodes and almost restrictive intervals respectively) the supremum is infinite. On the other hand, for Misiurewicz maps the supremum is finite.

Appropriate topological conditions should also imply a certain 'smoothness' of the invariant measures. The idea is that the condition stated in the conjecture above is the analogue of the condition that the rotation number of a circle diffeomorphism is of constant type. Let us make this more precise. In the circle case two conjugate circle diffeomorphisms having a sufficiently irrational rotation number are smoothly conjugate. Now we cannot expect anything like this for intervals. Indeed, if two maps are smoothly (or even Lipschitz) conjugate then their eigenvalues at corresponding periodic orbits are the same and this is rather exceptional. Even so, it is possible that some topological conditions imply that conjugacies are smooth at special points.

Another analogy with circle diffeomorphisms springs to mind. Herman's result for families states that the set of parameters for which maps from the Arnol'd family of circle maps f_a are smoothly linearizable (and therefore have an absolutely continuous invariant measure) has positive Lebesgue measure. The analogue of this result in the interval case is Jakobson's theorem from the previous section. Even though the results are similar the proofs are completely different. The proof of Herman's result is based on two facts:

1. There is a set of full Lebesgue measure such that if $\rho(f_a)$ is in this set then f_a is smoothly conjugated to a rotation;
2. if f_{a_0} is C^1 linearizable then $a \mapsto \rho(f_a)$ is Lipschitz *at* a_0.

From these facts the result follows very easily. So it would be very interesting to prove Jakobson's result in a similar way for interval maps. The analogue of 1) would be a more precise version of Conjecture 2 above. A result in the direction of 2) can be found in Guckenheimer (1980) for families of unimodal maps.

A similar question is whether there is a topological condition which allows for recurrent critical points and which still implies positive Liapounov exponents. In short, one hopes to get an analogue of Herman's theory for interval maps. Similarly, consider the quadratic family f_μ. As we remarked above, Świątek (1992b) has shown that the set of parameters μ for which $f_\mu(x) = \mu x(1-x)$ has a periodic attractor forms a dense set. From the result of Section 6 the complement of this set has positive Lebesgue measure. In fact one would expect the following

Conjecture 3: Let f_μ be the quadratic family. Consider the set of parameters μ for which f_μ has either a periodic attractor or an absolutely continuous invariant probability measure. This set has full Lebesgue measure.

Of course, this conjecture would follow from Conjecture 4 from Section III.7. Many results on ergodic properties of interval maps are not even mentioned in this chapter. For example, in Hofbauer's work and Hofbauer and Keller (1982), see also Newhouse (1991), the reader will find a discussion on invariant measures of maximal metric entropy. Those are the measures for which the metric entropy coincides with the topological entropy; in fact this is the supremum of all possible metric entropies, see for example Mañé (1987, pp. 244). Also we have not dealt at all with the thermodynamical theory, singularity spectra, decay of correlations and such matters. Some papers on these subjects are listed in the references.

Chapter VI.

Renormalization

In this chapter we will discuss the renormalization techniques which were introduced in one-dimensional dynamics independently by Feigenbaum (1978), (1979) and Coullet and Tresser (1978) to explain some quantitative and universal phenomena appearing in bifurcations of one parameter families of unimodal maps. More precisely, let f_t be a full one parameter family of unimodal maps of the interval $I = [-1, 1]$. For instance, f_t may be the quadratic family. As we saw in Section II.5, because f_t is a full family, there exists an interval $[a_1, b_1]$ in the parameter space such that for every t in this interval, f_t has a restrictive interval $I_{1,t} = [p'(t), p(t)]$ of period 2 where $p(t)$ is a fixed point of f_t and $f_t(p'(t)) = p(t)$. Furthermore, f_t^2 is a unimodal map from $I_{1,t}$ into itself and the family $[a_1, b_1] \ni t \mapsto f_t^2 \,|\, I_{1,t}$ is again full. In particular, $f_{b_1}^2 \,|\, I_{1,b_1}$ is a surjective unimodal map and there is a parameter value $\tilde{b}_1 \in (a_1, b_1)$ such that the critical point of $f_{\tilde{b}_1}^2 \,|\, I_{1,\tilde{b}_1}$ is a fixed point. Since this family of first return maps is again full, we can repeat the argument and we get, by induction, a decreasing sequence of intervals $[a_n, b_n]$ in the parameter space, and, for each $t \in [a_n, b_n]$ an interval $I_{n,t} \subset I$ such that the first return map of f_t to $I_{n,t}$ is a unimodal map which coincides with the restriction of $f_t^{2^n}$ to this interval. Furthermore, $t \mapsto f_t^{2^n} | I_{n,t}$ is a full family of unimodal maps. In particular, there exists $\tilde{b}_n \in [a_n, b_n]$ such that the critical point of $f_{\tilde{b}_n}$ is periodic of period 2^n and f_t has zero topological entropy for $t \leq \tilde{b}_n$. Let a_∞ be the limit of a_n when $n \to \infty$. As we have seen before f_{a_∞} has an attracting Cantor set and the dynamics of the restriction of f_{a_∞} to this Cantor set is conjugate to the adding machine, see Section III.4.

All these topological facts were quite well known before Coullet-Tresser and Feigenbaum made the following quantitative discoveries based on numerical experiments.

Numerical observations.

1) The parameters $\tilde{b}_n$ corresponding to the quadratic family converge to a_∞ geometrically, i.e., there exists a number $\delta > 1$ such that

$$\frac{\tilde{b}_{n+1} - \tilde{b}_n}{\tilde{b}_n - \tilde{b}_{n-1}} \to \frac{1}{\delta}.$$

Furthermore b_n also converges to a_∞ with the same rate.

2) Of course, the value a_∞ depends very much on the family f_t because one can change it by reparametrizing the family. However, numerical experiments indicate that the value δ might be universal. Indeed, for each full family of unimodal maps with a quadratic turning point one gets numerically $\delta = 4.669\ldots$.

3) There seems to be a universal metric structure related to the attracting Cantor set of f_{a_∞}. For example, if c is the critical point of f_{a_∞},

$$\frac{|f_{a_\infty}^{2^{n+1}}(c) - c|}{|f_{a_\infty}^{2^n}(c) - c|} \to \lambda$$

and again λ seems independent of the family.

In Feigenbaum (1978), (1979) and Coullet and Tresser (1978), a conjectural explanation was suggested for these numerical results. In those papers the renormalization operator (which will be defined in the next section) was introduced and it was shown that these numerical results could be explained if this operator, defined on an appropriate space of functions, would have a hyperbolic fixed point. As we will see in Section 1, these conjectures were proved using rigorous computer estimates by Lanford (1982)-(1986) and Eckmann and Wittwer (1987). The existence of this fixed point was also proved 'by hand' in Campanino et al. (1981), (1982) and Epstein (1986). For further references, see Eckmann and Epstein (1986), Eckmann (1986) and Sullivan (1992).

Recently, Sullivan (1992) has introduced many new techniques from complex analysis in order to give a conceptual proof of the above conjectures and also of some generalizations. For example he showed that the renormalization operator has a 'hyperbolic' invariant set which contains the fixed point mentioned above. From this he proves the validity of the third numerical observation. Under some additional hypothesis, evidence is given in Jiang et al. (1991) that the numerical observations 1) and 2) also hold. Most of this chapter is devoted to explaining Sullivan's ideas. In the next section we will state Sullivan's result explicitly.

1 The Renormalization Operator

In this chapter we will explain Sullivan's conceptual proof of the renormalization conjectures mentioned above. In fact, we will discuss some generalizations of these conjectures. In this section we describe these generalized conjectures in terms of the renormalization operator, we mention some of the computer-assisted proofs and we state Sullivan's theorem on the dynamics of the renormalization operator and the rigidity theorem that follows from it.

We will consider unimodal maps on a compact interval. Such a map f has a critical point $c_0 = c_0(f)$ and, disregarding some trivial dynamical situations, we may assume that the critical value $c_1 = f(c_0)$ lies to the right of the critical point c_0, that its image, $c_2 = f(c_1)$, is to the left of the critical point and that

$c_3 = f(c_2)$ belongs to the interval (c_2, c_1). Hence, the interval $[c_2, c_1]$, which is called the *dynamical interval* of f, is invariant by f and in fact it is the smallest invariant interval that contains the critical point. Conjugating f by an affine transformation we can assume that the dynamical interval of f coincides with $[0,1]$, i.e., $f^2(c_0) = 0$ and $f(c_0) = 1$. Furthermore, (in this chapter) we define

$$C(f) = \text{cl}\,\{f^n(c_0)\,;\ n \in \mathbb{N}\}.$$

We will consider smooth maps with *quadratic singularities*. More precisely, we will consider the space $\mathcal{U}^r$ of maps $f\colon [0,1] \to [0,1]$ of the form $f = \phi \circ Q \circ \psi$ where $\psi\colon [0,1] \to [\psi(0), 1]$ is an orientation reversing C^r diffeomorphism, $\psi(0) \in (-1, 0)$, $Q\colon [\psi(0), 1] \to [0,1]$ is the quadratic map $Q(x) = x^2$ and $\phi\colon [0,1] \to [0,1]$ is an orientation reversing C^r diffeomorphism. As we shall see, in the present theory it will be natural to consider the case when ϕ, ψ are analytic diffeomorphisms and also the case when ϕ, ψ are C^r diffeomorphisms with $r \geq 2$. More generally, we shall consider the case when ϕ, ψ are C^1 diffeomorphisms such that $\log D\phi$ and $\log D\psi$ satisfy the little Zygmund condition, see Section IV.2.a. The corresponding classes of maps are denoted by $\mathcal{U}^\omega$, $\mathcal{U}^r$ and $\mathcal{U}^{1+z}$. (We should note that $\mathcal{U}^{1+z} \supset \mathcal{U}^2$.) The fact that the critical point of f is of quadratic type plays an important role in the theory explained below. Note that maps that appear in generic families of sufficiently smooth unimodal maps will have this type of critical point. Take the metric on $\mathcal{U}^r$ for $1 \leq r < \infty$ defined by

$$d_r(f, g) = \sup\{D^k(f-g)(x)\,;\ 0 \leq k \leq r \text{ and } x \in [0,1]\}.$$

We say that J is a *unimodal interval* for some interval map g if J contains precisely one turning point c of g, $g(J) = J$ and if no subinterval of J has these properties. Hence $J = [g^2(c), g(c)]$. We say that J is a periodic unimodal interval for f of period m if it is a unimodal interval for f^m. As we have proved in Section II.5 for such an interval $J, \ldots, f^{m-1}(J)$ are pairwise disjoint. One of the intervals $f^k(J)$ is of the form $[f^{2m}(c_0), f^m(c_0)]$ where c_0 is the critical point of f.

Definition. First we say that $f \in \mathcal{U}^r$ is *renormalizable*, or that $f \in \mathcal{D}^r(\mathcal{R})$, if f has a periodic interval of period $m > 1$. The corresponding *renormalization operator* is the map $\mathcal{R}\colon \mathcal{D}^r(\mathcal{R}) \to \mathcal{U}^r$ defined by

$$\mathcal{R}f = A^{-1} \circ f^m \circ A$$

where $m > 1$ is the smallest possible period as above, $A\colon [0,1] \to \triangle$ is the affine map such that $A(1) = f^m(c_0)$ and where $\Delta = [f^{2m}(c_0), f^m(c_0)]$ is the unimodal periodic interval of period m that contains c_0.

Remark. 1. As before we are using the following notation: $[a, b]$ denotes the interval with endpoints a, b even when $b < a$. So we are not assuming that $f^{2m}(c_0)$ is smaller than $f^m(c_0)$.

We should note that some people prefer to use the larger restrictive intervals of period m (f^m maps the boundary of such a restrictive interval into itself) instead of the unimodal intervals from above.

3. Notice that $\mathcal{D}^r = \mathcal{D}^0 \cap \mathcal{U}^r$ and that the renormalization operator preserves all classes of differentiability. However, as we will observe later on, the renormalization operator is not smooth in the space of C^r maps endowed with a C^r metric if $r < \infty$.

3. The renormalization operator is not invertible: $\mathcal{R}f$ does not depend on f outside the orbit of the unimodal interval. However, we have

Proposition 1.1. *The restriction of the renormalization operator to the space of real analytic renormalizable maps is injective.*

Proof. Suppose $f = \mathcal{R}(g_1) = \mathcal{R}(g_2)$ where g_1 and g_2 are real analytic. Then $f = A_1^{-1} \circ g_1^{m_1} \circ A_1 = A_2^{-1} \circ g_2^{m_2} \circ A_2$. Let us first prove that $m_1 = m_2$. From this equation it follows that f extends analytically to $A_1^{-1}[0,1] \cup A_1^{-1}[0,1]$. Lemma II.5.1 implies that each unimodal interval of f in $A_i^{-1}[0,1]$ is the image under A_i^{-1} of a unimodal periodic interval of g_i of period m_i. Since g_i is a unimodal map, f maps $A_i[0,1]$ into itself. Hence f maps $A_1^{-1}[0,1] \cap A_2^{-1}[0,1]$ into itself. But then $g_i^{m_i}$ maps $A_i(A_1^{-1}[0,1] \cap A_2^{-1}[0,1])$ into itself, but since the period of the unimodal periodic interval of g_i is precisely m_i this implies that $A_1^{-1}[0,1] = A_2^{-1}[0,1]$. Therefore $m_1 = m_2$ and $A_1 = A_2$. So $g_1^m = g_2^m$ where $m = m_1 = m_2$. Let us now consider the fixed point p of g_1. Because $g_1^m = g_2^m$, p has to be also the fixed point of g_2 and the derivatives of g_1 and g_2 at p are the same. Let λ be this eigenvalue. Let $a_j(g_i)$ be the j-th coefficient of the power series expansion of g_i in p. Then one has by induction

$$a_j(g_i^m) = P_{j,m}(\lambda) \cdot a_j(g_i) + Q_{j,m}(a_1(g_i), \ldots, a_{j-1}(g_i))$$

where $P_{j,m}$ and $Q_{j,m}$ are polynomials. Hence, because $g_1^m = g_2^m$, we get by induction on j that $a_j(g_1) = a_j(g_2)$ for all $j \geq 1$. It follows that $g_1 = g_2$. □

1.1. The renormalization conjectures in the period doubling case

The domain $D^r(\mathcal{R})$ of the renormalization operator $\mathcal{R}$ has infinitely many connected components. One of these components is the set $\mathcal{U}_0^r$ of maps $f \in \mathcal{U}^r$ whose critical orbit satisfy the inequality: $0 = c_2 < c_0 < c_4 < c_3 < c_5 < c_1 = 1$. In this component the renormalization operator is just the doubling operator $\mathcal{R}(f)(x) = A^{-1} \circ f^2 \circ A$ with $A\colon [0,1] \to [f^2(c_0), f^4(c_0)]$ an orientation reversing affine map. If a map in $\mathcal{U}^r$ can be renormalized infinitely often and all the iterates of the renormalization operator belong to this component of the domain we say that f is a *Feigenbaum* map. In order to explain the quantitative numerical discovery in the bifurcation structure of one-dimensional parametrized families of unimodal maps, Feigenbaum and also Coullet and Tresser made the following conjectures on the structure of the renormalization operator

Conjectures.

1. There exists a Banach space of analytic functions $\mathcal{B}$ such that the restriction of the renormalization operator to $\mathcal{B} \cap \mathcal{U}_0^2$ is a bounded C^2 operator which has a fixed point Φ (which is often referred to as the Feigenbaum fixed point).

2. The derivative $D\mathcal{R}(\Phi)$ is a compact operator whose spectrum has a unique eigenvalue $\delta = 4.66920\ldots$ outside the unit circle and the other eigenvalues are in the interior of the unit disc.

3. Let $\Sigma_n \subset \mathcal{B} \cap \mathcal{U}_0^2$ be the set of maps in the neighbourhood of Φ having zero topological entropy and for which the critical point is periodic of period 2^n. Then Σ_n, which is a codimension one Banach submanifold, intersects the local unstable manifold of $\mathcal{R}$ transversally for n large enough.

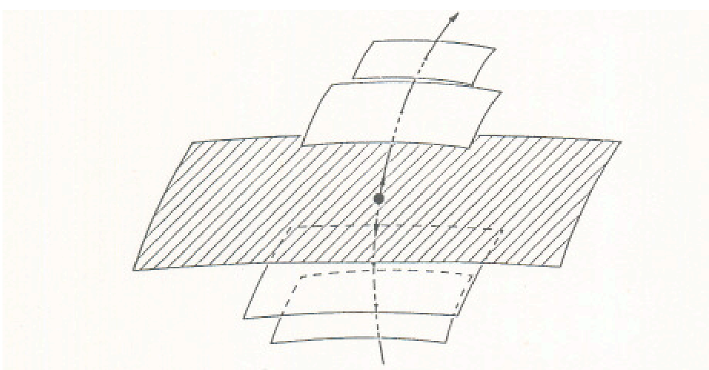

Fig. 1.1:

From the smoothness of the operator and this spectral property, the first two of these conjectures would imply, using the Stable and Unstable Manifold Theorem, see Hirsch and Pugh (1970) or for example Palis and de Melo (1982), that there exist a local unstable manifold $\mathcal{W}^u$ of dimension one, tangent to the eigenspace associated to the eigenvalue δ and a local stable manifold $\mathcal{W}^s$ intersecting $\mathcal{W}^u$ transversely at Φ. The local stable manifold of $\mathcal{R}$ is the set of maps ψ in a neighbourhood of Φ such that all iterates $\mathcal{R}^n(\psi)$ are defined, remain in this neighbourhood and converge exponentially fast to Φ. One can also define the global stable manifold of the fixed point Φ as the set of maps ψ such that $\mathcal{R}^n(\psi)$ converges to Φ. However, since the renormalization operator is not invertible, one cannot define the global unstable manifold. Even so, for each $\psi \in \mathcal{W}^u$ and for each positive integer j, there exists a unique $\psi_j \in \mathcal{W}^u$ such that $\mathcal{R}^j(\psi_j) = \psi$. Furthermore, the sequence ψ_j converges geometrically to Φ with rate δ^{-1}, i.e., $||\Phi - \psi_j|| \approx \delta^{-j}$.

As $\mathcal{R}(\Sigma_{n+1}) \subset \Sigma_n$, the third conjecture would imply that the submanifolds Σ_n accumulate at the local stable manifold geometrically with rate δ^{-1}. Similarly, many others "bifurcation submanifolds" accumulate at $\mathcal{W}^s$ with the same rate. For instance, if n is big enough, there exists a codimension one Banach submanifold $\tilde{\Sigma}_n$, transversal to $\mathcal{W}^u$, such that for all $\psi \in \tilde{\Sigma}_n$, $\mathcal{R}^n(\psi)(1) = 0$ (in which case $\mathcal{R}^n(\psi)$ is the full map). Clearly $\tilde{\Sigma}_n$ and Σ_n accumulate at $\mathcal{W}^s$ from different sides. If f_μ is a one parameter family of maps contained in the neighbourhood of the fixed point, which intersects $\mathcal{W}^s$ transversely at f_{μ_∞} then it will intersect Σ_n (resp. $\tilde{\Sigma}_n$) transversely at f_{a_n} (resp. $f_{\tilde{a}_n}$). Since the operator $\mathcal{R}$ is C^2 on $\mathcal{B} \cap \mathcal{U}_0$, it follows, from the stable manifold theory, that a_n and $\tilde{a}_n$ converge geometrically to μ_∞ with rate δ^{-1}. This explains Feigenbaum's universal constant δ. This hyperbolic picture, see Figure 1.1, is Feigenbaum's and Coullet and Tresser's explanation for the numerical discoveries 1) and 2) from the introduction of this chapter. As we will see at the end of this chapter it also explains the numerical observation 3).

Lanford (1984a) gave the first complete proof of Conjectures 1) and 2), see also Lanford (1984b) and (1986). In his case the space $\mathcal{B}$ is the set of maps of the type $x \mapsto g(x^2)$ where g is holomorphic in the complex disc containing the real interval $[-1, 1]$ and preserves the real axis. The proof combines non-rigorous computer estimates to find an approximate solution for the fixed point of $\mathcal{R}$, with rigorous computer estimates in a given neighbourhood of the approximate solution and finally a modification of Newton's method to prove the existence of a solution of the so called Cvitanović-Feigenbaum functional equation $\Phi(x) = -\frac{1}{\lambda}\Phi \circ \Phi(-\lambda x)$ in the given neighbourhood of the approximate solution.

Eckmann and Wittwer (1987) gave a different computer assisted proof of the same first two conjectures and also a proof of the third one. This proof consists of looking for fixed points of an operator, which is essentially the doubling operator, acting on a space of one-parameter families of analytic maps in a neighbourhood of an approximate solution for the unstable manifold. This approximate solution for the unstable manifold is obtained using non-rigorous computer estimates. Using rigorous computer estimates and applying Newton's method, they prove that the approximate solution has a neighbourhood where the operator is a contraction and has a unique fixed point. This is the unstable manifold. Furthermore, there are several proofs of the existence of a solution for the Cvitanović-Feigenbaum functional equation that do not rely on computer estimates: Epstein (1986), Eckmann and Epstein (1986), see also Campanino, Epstein and Ruelle (1981), (1982) and Eckmann (1986) for further references. Sullivan has given a conceptual proof of some of these conjectures. His proof gives a good conceptual understanding of the mechanism of renormalization for analytic maps. He shows that if ϕ and ψ are combinatorially equivalent infinitely renormalizable analytic maps of bounded combinatorial type (for the definition see below) in $\mathcal{U}^\omega$ then the distance between $\mathcal{R}^n(\phi)$ and $\mathcal{R}^n(\psi)$ converges to zero. This means that under iteration of the renormalization operator, any infinite renormalizable map converges to a 'hyperbolic strange attractor' of the renormalization operator. This attractor is conjectured to be expanding in the sense that the distance between the iterates of two maps which are very close together and have different kneading sequences expands exponentially to a definite size. Moreover, he shows that in the C^2 case the renormalization operator also converges. More precisely, for combinatorially equivalent infinitely renormalizable maps $\phi, \psi \in \mathcal{U}^{1+z}$ of bounded combinatorial type, the distance between $\mathcal{R}^n(\phi)$ and $\mathcal{R}^n(\psi)$ converges to zero.

Remark. 1. Let us make a remark on a technical difficulty for the understanding of renormalization theory in the space of C^r maps when $r < \infty$. Here one can formulate the same conjecture: the orbits of the renormalization operator through any two combinatorially equivalent C^r maps which are infinitely renormalizable converge exponentially to a unique orbit contained in a hyperbolic expanding attractor of the renormalization operator. To have any hope for this conjecture to hold we need to assume that r is not too small, say $r \geq 2$. However there is a difficulty: the renormalization operator is not a smooth map when r is finite. Indeed, the renormalization operator associates to a map, up to scaling, a restriction of an iterate of the map and the composition map $(h, g) \mapsto h \circ g$ is not differentiable if we consider the space of C^r maps with $r < \infty$. Note however that $\mathcal{R} \colon \mathcal{D}^{r+s} \to \mathcal{U}^r$ is C^s, see Irwin (1972). To bypass these problems most results on renormalization work in the C^∞ category. However, Davie (1992) and Lanford (1992) have some C^2 results.

2. Let us mention an interesting remark by Jakobson (1986) which uses his theorem discussed in Section V.6 and the hyperbolic structure of the renormalization operator near the fixed point from above. Let f_t be a one parameter family of unimodal maps that intersects transversely the local stable manifold of the renormalization operator at the parameter value μ_∞. Suppose that for $t \leq \mu_\infty$, the map f_t has zero topological entropy. Let $\mathcal{S} = \{t > \mu_\infty\,;\ f_t$ satisfies the Axiom A $\}$ and $\mathcal{C} = \{t > \mu_\infty\,;\ f_t$ has an absolutely continuous invariant probability measure$\}$. Then f_t is structurally stable for $t \in \mathcal{S}$ and it is 'chaotic' for $t \in \mathcal{C}$. The important consequence of Jakobson's result and of the hyperbolicity of Feigenbaum's fixed point, is that there exist a constant $k > 0$ such that

$$\frac{|\mathcal{S} \cap [\mu_\infty, \mu_\infty + \epsilon]|}{\epsilon} > k$$

and

$$\frac{|\mathcal{C} \cap [\mu_\infty, \mu_\infty + \epsilon]|}{\epsilon} > k$$

where $|\cdot|$ denotes the Lebesgue measure of a set. So, after crossing the stable manifold of the Feigenbaum's fixed point we meet, with positive probability, both stable and chaotic situations.

1.2. The domain of the renormalization operator

Let us now analyze the domain of the renormalization operator in more detail. As we have mentioned above, the domain $\mathcal{D}^r$ of the renormalization operator has infinitely many connected components. In each connected component of $\mathcal{D}^r$ the return time m of the unimodal interval $\triangle$ is the same for each map f. Since the corresponding intervals $f^i(\triangle), 0 \leq i < m$ are all disjoint it follows that they are embedded in $[0,1]$ in the same order for each map in one component of $\mathcal{D}^r$. We will code these as in Section II.5: let $X = \{x_1, \ldots, x_n\}$ be a finite set endowed with an order relation $\prec$. As before we say that a permutation $\sigma\colon X \to X$ is *unimodal* with respect to the order relation $\prec$ if it satisfies the following condition. Embed X monotonically into the real line, draw the graph of σ on $\mathbb{R}^2$ and connect the consecutive points of the graph by a line segment. If the curve so obtained is the graph of a unimodal map then we say that the permutation is unimodal. It is *renormalizable* if X is the disjoint union of p sets X_i each containing m points and such that

1. each X_i is mapped by σ onto some X_j;
2. for each $i \neq j$, either $X_i \prec X_j$ or $X_j \prec X_i$ (here $X_i \prec X_j$ means that $x_i \in X_i$, $x_j \in X_j$ implies $x_i \prec x_j$).

It follows that a unimodal renormalizable permutation σ of $(X, \prec)$ defines a unimodal permutation of $\{X_1, X_2, \ldots, X_p\}$ endowed with the order relation induced from $\prec$.

Proposition 1.2. *a) Let $f \in \mathcal{U}^r$ be renormalizable with a maximal unimodal periodic interval $\triangle$ of period m and let X be the collection of intervals*

$$\{\triangle, \triangle^1 = f(\triangle), \ldots, \triangle^{m-1} = f^{m-1}(\triangle)\}$$

with the ordering induced from that of the real line. Then the permutation $\sigma = \sigma_0(f)\colon X \to X$ defined by $\sigma(\triangle^i) = \triangle^j$ whenever $f(\triangle^i) = \triangle^j$, is unimodal and non-renormalizable.

b) Conversely, given a unimodal non-renormalizable permutation $\sigma\colon X \to X$, there exists in each full family of unimodal maps a renormalizable map f such that $\sigma_0(f) = \sigma$.

Proof. Let us prove b). Since σ is a unimodal permutation, there is an element x_0 of X that is mapped into the greatest element x_1 of X and this x_1 is mapped into the smallest element x_2. Next we represent the elements of X by disjoint subintervals of $[0,1]$ so that the order they have in X is the same as the order of the real line. So to each $x_i \in X$ there corresponds an interval $\triangle^i$. We map $\triangle$ onto $\triangle^1$ by a folding map. For each j, we map the interval $\triangle^j$, corresponding to the element $x \in X$ onto the interval $\triangle^{j+1}$ by an orientation preserving (resp. reversing) affine map g if $x \prec x_0$ (resp. $x_0 \prec x$). It is easy to see that, since σ is a unimodal permutation, we can extend g to a unimodal map of the interval $[0,1]$ (for example one can define g to be affine on each component of the complement of the intervals corresponding to elements of X). Then $\triangle$ is a unimodal interval of g. Since σ is non-renormalizable it follows that $\triangle$ is not contained in a larger unimodal interval of period ≥ 2 and therefore that $\sigma_0(g) = \sigma$. From Section II.5 it follows that there exists a map f in each full family of unimodal maps which is combinatorially equivalent to g. It follows that $\sigma_0(f) = \sigma$. $\square$

Example. The permutation σ, $\sigma(\triangle^i) = \triangle^{i+1}$, represented in Figure 1.2, is a unimodal permutation because the corresponding map is unimodal. The map depicted in Figure 1.3 is clearly renormalizable.

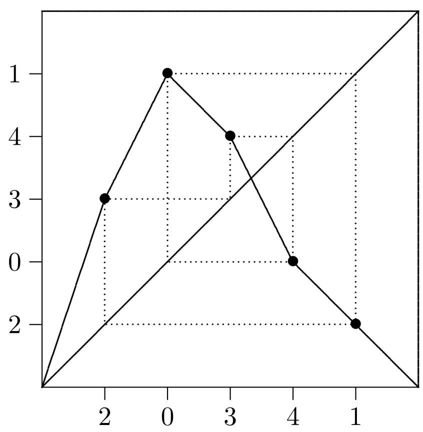

Fig. 1.2: A unimodal permutation with the corresponding unimodal map.

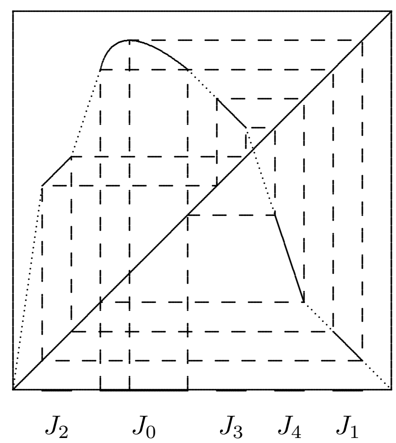

Fig. 1.3: A renormalizable map.

Let σ be a unimodal, non-renormalizable permutation and

$$\mathcal{D}^r_\sigma = \{f \in \mathcal{D}^r\,;\ \sigma_0(f) = \sigma\}.$$

From Proposition 1.2, we have that $\mathcal{D}^r$ is equal to $\bigcup_\sigma \mathcal{D}^r_\sigma$ where σ runs over all unimodal non-renormalizable permutations and each $\mathcal{D}^r_\sigma$ is non-empty. We will also consider iterates $\mathcal{R}^k = \mathcal{R} \circ \mathcal{R}^{k-1}$ of the renormalization operator. The domain $\mathcal{D}^r_k$ of $\mathcal{R}^k$ is the disjoint union of sets of the form $\mathcal{D}^r_{\sigma_0,\sigma_1,\ldots,\sigma_{k-1}}$ where σ_i is a unimodal non-renormalizable permutation and

$$\mathcal{D}^r_{\sigma_0\sigma_1,\ldots,\sigma_{k-1}} = \{f \in \mathcal{D}^r_k\,;\ \sigma_0(R^i f) = \sigma_i \text{ for } i = 0, 1, \ldots, k-1\}.$$

It follows that the set $\mathcal{D}^r_\infty = \bigcap_{k\geq 1} \mathcal{D}^r_k$ of infinitely renormalizable maps may be written as the uncountable disjoint union $\mathcal{D}^r_\infty = \bigcup \mathcal{D}^r_{\sigma_0,\sigma_1,\ldots}$. For $f \in \mathcal{D}^r_\infty$ let us write

$$\sigma(f) = (\sigma_0, \sigma_1, \ldots)$$

if $f \in \mathcal{D}^r_{\sigma_0,\sigma_1,\ldots}$. If $f \in \mathcal{D}^r_\infty$ then $\sigma(f) = \sigma(g)$ implies that these maps have the same kneading invariants and are combinatorially equivalent. Notice that $\mathcal{R}(\mathcal{D}^r_{\sigma_0,\sigma_1,\ldots}) = \mathcal{D}^r_{\sigma_1,\sigma_2,\ldots}$. This means that the renormalization operator acts as the shift map in the space of sequences of permutations. In particular, if $(\sigma_0, \sigma_1, \ldots)$ is a periodic point of period k for the shift map then $\mathcal{D}^r_{\sigma_0,\sigma_1,\ldots}$ is invariant by $\mathcal{R}^k$. Let $r \geq 1+z$ (for the definition of the class of C^{1+z} maps see Section IV.2.a) and $(\sigma_0, \sigma_1, \dots)$ as above. The main result to be proved in this chapter implies that $\mathcal{R}^k$ has a unique fixed point in $\mathcal{D}^r_{\sigma_0,\sigma_1,\dots}$, that this fixed point is a real analytic map and that it is a global attractor for $\mathcal{R}^k \mid \mathcal{D}^r_{\sigma_0,\sigma_1,\dots}$.

Let us apply these results to families of unimodal maps. Firstly, all these types occur in full families:

Proposition 1.3. *If $f_a \in \mathcal{U}^r$ is a full family of unimodal maps and $(\sigma_0, \sigma_1, \dots)$ a sequence of unimodal, non-renormalizable permutations then there exists a parameter value a_0 such that $\sigma(f_{a_0}) = (\sigma_0, \sigma_1, \dots)$.*

Proof. Follows from Theorem II.4.1 see also Theorem II.5.2. □

Let $|\sigma_i|$ denote the number of elements permuted by σ_i. The next proposition tells us that the topological type of an infinitely renormalizable map is determined by these permutations.

Proposition 1.4. *Let $f, g \in \mathcal{U}^{1+z}$ be two infinitely renormalizable maps such that $\sigma(f) = \sigma(g) = (\sigma_0, \sigma_1, \dots)$. Let $q(n) = |\sigma_0| \cdot |\sigma_1| \cdots |\sigma_n|$ and denote by $\triangle_n$ be the interval $[f^{2q(n)}(c_0), f^{q(n)}(c_0)]$. Then*

1. *$f^{q(n)}(\triangle_n) = \triangle_n$ and*
$$C(f) = \mathrm{cl}\,\{f^n(c_0)\,;\ n \in \mathbb{N}\}$$
is a Cantor set which is equal to $\bigcap F_n$ where $F_n = \cup_{i=0}^{q(n)-1} f^i(\triangle_n)$;
2. *The map*
$$h\colon \{f^n(c_0(f))\,;\ n \in \mathbb{N}\} \to \{g^n(c_0(g))\,;\ n \in \mathbb{N}\}$$
defined by $h(f^n(c_0(f))) = g^n(c_0(g))$ extends continuously to a conjugacy $h\colon C(f) \to C(g)$ between f and g;
3. *there exists $n_0 \in \mathbb{N}$ such that $\mathcal{R}^n f$ and $\mathcal{R}^n g$ are topologically conjugate for $n \geq n_0$.*

Proof. Statement 1) follows from the non-existence of wandering intervals. Indeed, since $F_{n+1} \subset F_n$, $C(f) \subset \bigcap F_n$ and each component of $\bigcap F_n$ contains a point of $C(f)$. On the other hand, since f has no wandering intervals and $q(n) \to \infty$, the set $\bigcap F_n$ contains no intervals. Thus, $C(f)$ is a Cantor set. (If f has wandering intervals then $C(f)$ could have isolated points). From this and the fact that the original map h is ordering preserving Statement 2) follows.

So let us prove Statement 3). From the finiteness of attractors, see Theorem IV.B, it follows that the turning points of f and g are not accumulated by periodic attractors when $f, g \in \mathcal{U}^r$ and $r \geq 2$. So for n large enough, $\mathcal{R}^n f$ and $\mathcal{R}^n g$ have no wandering intervals and no periodic attractors. Since $\sigma(f) = \sigma(g)$, the orbits of the turning points of f and g are ordered in the same way. Using

Theorem II.3.1 the result follows. □Next, if two infinitely

renormalizable maps are C^k conjugate with $k \geq 1$ then the renormalizations of these maps tend to each other in the C^k topology:

Proposition 1.5. *Let $1 + Z \leq r \leq \omega$ and $f, g \in \mathcal{U}^r$ be two infinitely renormalizable maps. If $h\colon [0,1] \to [0,1]$ is a C^k conjugacy between f and g with k a finite integer $\leq r$ then the C^k distance between $\mathcal{R}^n(f)$ and $\mathcal{R}^n(g)$ goes to zero as $n \to \infty$.*

If $r = \omega$ and h is analytic then there exists a neighbourhood U of $[0,1]$ in $\mathbb{C}$ such that both $\mathcal{R}^n(f)$ and $\mathcal{R}^n(g)$ have holomorphic extensions to U and $\mathcal{R}^n(f) - \mathcal{R}^n(g)$ converge uniformly to zero on U.

Proof. Let Δ_n be as before; $\mathcal{R}^n(f)$ is up to an affine change of coordinates equal to $f^{q(n)}|\Delta_n$. Because there are no wandering intervals, $|\Delta_n| \to 0$ and therefore $h_n = A_n^{-1} \circ h \circ A_n$ goes to the identity map in the C^k topology. If f, g, h are analytic then they extend to holomorphic maps F, G, H on a neighbourhood U of $[0,1]$ and H is in fact a conjugacy in this neighbourhood. Take this neighbourhood so that H is univalent on U. Hence $H_n = A_n^{-1} \circ H \circ A_n$ is univalent on $U_n = A_n^{-1}(U)$ and these sets 'get big': distance$(\mathbb{C} \setminus U_n, 0) \to \infty$. Moreover, $H_n(0) = 0$, $H_n(1) = 1$. Hence by Koebe's Distortion Lemma H_n tends uniformly to the identity on U. □In Section

2 we will improve this proposition by showing that the convergence above is exponentially fast (because we shall show that the lengths of the intervals Δ_n go to zero exponentially fast).

1.3. The results of Sullivan on renormalization

Now we state the results of Sullivan that will be proved in this chapter. First we need some definitions.

Definition. We say that $f \in \mathcal{D}^r_\infty$ has *bounded combinatorial type* if there exists an integer N such that $|\sigma_i| \leq N$ for all $i = 0, 1, 2, \ldots$, where $\sigma(f) = (\sigma_0, \sigma_1, \ldots)$ and $|\sigma_i|$ denotes the number of elements permuted by σ_i. Let $\mathcal{D}^r_{(N)}$ be the corresponding space of maps.

Definition. Let $f_n\colon [0,1] \to [0,1]$ be a sequence of real analytic maps. We say that f_n converges *strongly* to f if there exist a neighbourhood W of $[0,1]$ in the complex plane and holomorphic extensions $F_n\colon W \to \mathbb{C}$ of f_n and $F\colon W \to \mathbb{C}$ of f such that F_n converges uniformly to F on W. We say that a subset $\mathcal{C}$ of real analytic functions is strongly compact if any sequence in $\mathcal{C}$ has a subsequence that converges strongly to an element of $\mathcal{C}$.

Remark. Since the uniform convergence of a sequence of holomorphic functions implies the uniform convergence of the sequence of all of its derivatives on compact parts of the domain, the strong convergence of f_n to f implies the convergence of the sequence in the C^r norm for all $r \in \mathbb{N}$. Hence a strongly compact subset of real analytic functions is also a compact subset of the space of C^r maps for every $r \in \mathbb{N}$.

Definition. Let $a > 0$. We say that the real analytic map $f\colon [0,1] \to [0,1]$ in $\mathcal{U}$ belongs to the *Epstein class* $\mathcal{E}_a$ if $f(x) = \phi \circ Q \circ \psi$ where Q is the quadratic map $Q(z) = z^2$, ψ is an affine map and $\phi\colon [0,1] \to [0,1]$ is a diffeomorphism whose inverse ϕ^{-1} has a holomorphic extension which is univalent in the domain $(\mathbb{C} \setminus \mathbb{R}) \cup [-a, 1+a]$.

Definition. A *quadratic-like map* is a holomorphic map $F\colon U \to V$ such that U and V are simply connected domains with the closure of U contained in V and such that F is a degree two branched covering map, i.e., F has a unique critical point c and the restriction of F to $U \setminus \{c\}$ is a degree two covering map onto $V \setminus \{F(c)\}$. The quadratic-like maps we will consider here are symmetric with respect to the real line, i.e., they commute with complex conjugation. We say that the *conformal type* of a quadratic-like map $F\colon U \to V$ is bounded by B if the conformal modulus of the annulus $V \setminus \mathrm{cl}\,(U)$ is at least $\frac{1}{B}$ and the conformal modulus of the annulus $V \setminus [F^2(c), F(c)]$ is bounded by B, see the Appendix for the definition of the modulus of an annulus. Such an annulus is called a *fundamental domain* of F. An orbit through a point in U which does not belong to the filled Julia set

$$J(F) = \{z \in U\,;\ F^k(z) \in U \text{ for all } k \geq 0\}$$

has a unique point in $V \setminus \mathrm{cl}\,(U)$ or in the boundary of U.

Remark. If f is a real analytic map that has a quadratic-like extension then there exist a map g that belongs to some Epstein class and is analytically conjugate to f. Moreover, g has an extension which is quadratic-like. This fact will be proved in Corollary 3 of Theorem 4.3.

The Main Result of this chapter is the following theorem due to Sullivan (1992).

Theorem 1.1. *Let $N \in \mathbb{N}$ and let $\mathcal{D}^{1+z}_{(N)}$ be the set of infinitely renormalizable maps of combinatorial type bounded by N.*

1. *There exists a strongly compact, $\mathcal{R}$-invariant set $\mathcal{A} \subset \mathcal{D}^{\omega}_{(N)}$ such that if $f \in \mathcal{D}^{1+z}_{(N)}$ then the $C^{1+\alpha}$ distance between $\mathcal{R}^n(f)$ and $\mathcal{A}$ converges to zero as n goes to infinity for any $\alpha \in (0,1)$.*

2. *There exist $a > 0$ and $B > 0$ such that $\mathcal{A} \subset \mathcal{E}_a$ and all maps in $\mathcal{A}$ have quadratic-like extensions whose conformal type is bounded by B.*

3. *The restriction of $\mathcal{R}$ to $\mathcal{A}$ is a homeomorphism which is topologically conjugate to the full shift on a finite number of symbols.*

4. *If $f \in \mathcal{A}$ then the stable set of f,*

$$W^s(f) = \{g\,;\ \mathcal{R}^n(g) - \mathcal{R}^n(f) \text{ converges to zero}\},$$

is the set of maps such that $\mathcal{R}^m(g)$ and $\mathcal{R}^m(f)$ have the same combinatorial type for some $m \geq 0$.

5. *There exists a strongly compact subset $\mathcal{C} \supset \mathcal{A}$ with the following properties: i) for any real analytic map g of combinatorial type bounded by N and belonging to some Epstein class, there exists $n_0(g)$ such that $\mathcal{R}^n(g) \in \mathcal{C}$ for $n \geq n_0(g)$; ii) if $f, g \in \mathcal{C}$ have the same combinatorial type then $\mathcal{R}^n(f) - \mathcal{R}^n(g)$ converges strongly to zero.*

In Section 9 we prove that if two infinite renormalizable maps $f, g \in \mathcal{D}^{1+z}$ have the same bounded combinatorial type and the distance between $\mathcal{R}^n(f)$ and $\mathcal{R}^n(g)$ converges to zero exponentially fast, then, for each $0 \leq \alpha < 1$, there exists a $C^{1+\alpha}$ diffeomorphism h of the real line that maps the attracting Cantor set $C(f)$ of f onto the attracting Cantor set $C(g)$ conjugating the two maps in their attracting Cantor sets. Using Theorem 1.1 we also prove a weaker rigidity result: if two maps $f, g \in \mathcal{D}^{1+z}$ have the same bounded combinatorial type then the asymptotic geometry of their attracting Cantor set is the same, i.e., they have the same scaling function.

1.4. An outline of the proof of Theorem 1.1

The complete proof of Theorem 1.1 will occupy Sections 2 to 8. Here we will give an outline of the proof and discuss the tools we will use.

Step 1. The real bounds

If f is an infinitely renormalizable map of combinatorial type bounded by N, there exists a decreasing sequence $\triangle_n$ of intervals containing the critical point and a sequence $q(n)$ of return times to these intervals such that $\mathcal{R}^n(f)$ is, up to affine conjugacy, equal to the restriction of $f^{q(n)}$ to $\triangle_n$. The combinatorial type is bounded by N if and only if $\frac{q(n+1)}{q(n)} \leq N$. The orbit of the interval $\triangle_n$ is a closed set F_n that has $q(n)$ connected components. In Section 2 we prove the following fundamental result: the ratio of the length of a connected component of F_n to the length of either a component of F_{n+1} or of $F_n \setminus F_{n+1}$ contained in the first component, is bounded away from zero and from one by a constant that depends only on f. Furthermore, if n is large enough, this bound becomes universal, i.e., we get a bound that does not even depend on f. This holds for maps in $\mathcal{U}^{1+z}$.

The ingredients of the proof of the real bounds are the control of the distortion of the cross-ratio under iteration and Koebe's Distortion Principle we have discussed in Chapter 4.

From the real bounds we prove in Section 2 that any limit of renormalization belongs to an Epstein class.

Step 2. The complex bounds

The main result of Section 5 is the existence of a universal constant B, depending only on the bound for the combinatorial type N with the following property. Let f be a real analytic map which is infinitely renormalizable of combinatorial type bounded by N and either has a quadratic-like extension or belongs to some Epstein class. Then, provided n is sufficiently large, $\mathcal{R}^n(f)$ has a holomorphic extension to a neighbourhood of the dynamical interval in the complex plane which is quadratic-like and has a conformal type bounded by B.

The main ingredient of the proof of the complex bounds are the real bounds. Here we use some hyperbolic geometry, the Schwarz Theorem on the contraction of the hyperbolic metric by holomorphic maps and Koebe's Distortion Theorem.

Koebe's Distortion Theorem implies that the set $\mathcal{C}$, defined as the set of real analytic unimodal maps that have a quadratic-like extension of conformal type bounded by B, is strongly compact. So take a real analytic function f of bounded combinatorial type which is either in some Epstein class or has a quadratic-like extension. The complex bounds give that for n sufficiently large $\mathcal{R}^n(f)$ is in the above strongly compact set. In particular, any limit of renormalization has a quadratic-like extension of bounded combinatorial type.

Step 3. The pullback argument

In Section 4 we prove that two quadratic-like maps with the same bounded combinatorial type are quasiconformally conjugate. Furthermore, the conformal distortion of this quasiconformal conjugacy depends only on the behaviour of the maps near the closure of a fundamental domain in the complex plane. In particular, we get that if the maps are C^0 close to each other near the closure of a fundamental domain, the conformal distortion of the quasiconformal conjugacy is close to one.

The proof uses the real bounds and the compactness of the set of quasiconformal homeomorphisms with uniformly bounded conformal distortion. The conjugacy is obtained by a pullback construction starting with a map that conjugates the two maps at the critical orbit and at the boundary of the domain.

As a consequence of the pullback argument we get that given two quadratic-like maps F and G of the same bounded combinatorial type, there exists a non-negative number $d_{JT}(F,G)$ such that if $K > \exp(d_{JT}(F,G))$ then there exists a

K-quasiconformal conjugacy between F and G defined in some neighbourhood of the *filled Julia set* $J(F) = \{z \in U; F^n(z) \in U \text{ for all } n \in \mathbb{N}\}$. We also prove in Section 4 that $d_{JT}(F,G) = 0$ if and only F and G are holomorphically conjugate in some neighbourhood of their filled Julia sets. This clearly defines an equivalence relation and d_{JT} defines a metric, called the Julia-Teichmüller metric, in the space $\mathcal{G}_{\underline{\sigma}}$ of holomorphic equivalence classes of quadratic-like maps of the same bounded combinatorial type $\underline{\sigma}$.

The renormalization operator extends to a map $\mathcal{R}\colon \mathcal{G}_{\sigma_0,\sigma_1,\sigma_2,\ldots} \to \mathcal{G}_{\sigma_1,\sigma_2,\ldots}$. In fact, the renormalized map of the restriction of a quadratic-like map to the dynamical interval is, up to affine conjugacy, the restriction of an iterate of this map to an interval around the critical point. As we will see, if the map is quadratic-like, the restriction of this iterate has an extension to a neighbourhood of the interval in the complex plane which is quadratic-like. Since the renormalization operator is just the restriction of the iterate to some neighbourhood of the critical point it does not increase the Julia-Teichmüller distance.

Let $\mathcal{C}$ be the strongly compact set defined in Step 2). If $f, g \in \mathcal{C}$ have the same combinatorial type we may define $d_{JT}(f,g)$ as the Julia-Teichmüller distance between their quadratic-like extension. This is again a metric on the set of conformal equivalence classes. From Koebe's Distortion Principle and the property of the pullback argument discussed above we get that there exists a constant l_0 such that the Julia-Teichmüller distance between any two maps in $\mathcal{C}$ with the same combinatorics is bounded by l_0.

Step 4. The contraction of the Julia-Teichmüller metric

This is the main part of the proof of Theorem 1.1: if $f, g \in \mathcal{C}$ have the same combinatorial type then $d_{JT}(\mathcal{R}^n(f), \mathcal{R}^n(g))$ converges to zero as $n \to \infty$.

The proof of the above statement relies on the complex bounds and on extension of some ideas from Teichmüller theory. Using the Measurable Riemann Mapping Theorem with parameters due to Ahlfors and Bers, see the Appendix, we define a special class of one parameter deformations of holomorphic equivalence classes of quadratic-like maps, called Beltrami paths, that have the following properties.

a) The image of a Beltrami path under the renormalization operator is again a Beltrami path.

b) Roughly speaking, we say that a Beltrami path is efficient if the distance between its end points is not much smaller then the sum of the distance of consecutive points on the path. A Beltrami path satisfies the Almost Geodesic Principle: if an infinitesimal beginning piece of the Beltrami path is efficient then it is efficient on a definite piece of the path.

The second property is the main ingredient in the proof that the renormalization operator contracts the Julia-Teichmüller distance. We start with two maps f, g of the same bounded combinatorial type. Take a quasiconformal conjugacy between quadratic-like extensions of f and g whose conformal distortion is very close to the exponential of their Julia-Teichmüller distance. We consider the Beltrami coefficient of this quasiconformal conjugacy and we deform it along a Beltrami path up to a distance (in the parametrization) much larger than the constant l_0 in Step 3. Notice that since the Beltrami path is very efficient between f and g, by the above property b), it remains quite efficient up to the other endpoint. Hence the Teichmüller distance between the endpoints is much larger than l_0. On the other hand, by the complex bounds, there exists an n such that the endpoints of the image of this Beltrami path by $\mathcal{R}^n$ belongs to $\mathcal{C}$. Using part b) of the Almost Geodesic Principle we conclude that the Teichmüller distance between $\mathcal{R}^n(f)$ and $\mathcal{R}^n(g)$ must be smaller than the time needed to go from one to the other along the Beltrami path by a definite factor which is essentially equal to the Teichmüller distance between f and g. To get a definite contraction of the Julia-Teichmüller distance under this iteration we use a converse of the Almost Geodesic Principle that is proved in Section 8.

In order to prove this Almost Geodesic Principle we associate in Section 6 to each germ of a quadratic-like map a compact space that is laminated by Riemann surfaces so that two maps are holomorphically equivalent if and only if the corresponding Riemann surface laminations are equivalent in the sense of Teichmüller. In this way we embed the space of holomorphic equivalence classes of quadratic-like maps in a Teichmüller space of Riemann surface laminations. In Section 6 we reproduce the main ingredients of the classical Teichmüller theory for compact Riemann surfaces and extend them to Riemann surface laminations. As we will show in Section 7, the Almost Geodesic Principle will then follow from an extension of the so-called Grötsch inequality.

Step 5. The structure of the attractor $\mathcal{A}$ of the renormalization operator

The set $\mathcal{A}$ of limit points of the renormalization operator is clearly non-empty, Namely, if f is real analytic infinitely renormalizable of bounded combinatorial type then $\mathcal{R}^n(f) \in \mathcal{C}$ for n large enough and $\mathcal{C}$ is strongly compact. Also $\mathcal{A}$ is a subset of $\mathcal{C}$.

Let $f_0 \in \mathcal{A}$. Since $f_0 = \lim \mathcal{R}^{n_i}(f)$ and $\mathcal{C}$ is strongly compact we can, by taking convergent subsequences of the sequences $\mathcal{R}^{n_i - m}(f)$, construct a bisequence $\ldots, f_{-m}, \ldots, f_{-1}, f_0, f_1, \ldots$ in $\mathcal{A}$ such that $\mathcal{R}(f_m) = f_{m-1}$ for all $m \in \mathbb{Z}$. If $\ldots, g_{-m}, \ldots, g_{-1}, g_0, g_1, \ldots$ is another such sequence and f_m has the same combinatorial type as g_m for all $m \in \mathbb{Z}$, then we prove in Section 8 that $g_0 = f_0$. The proof of this uniqueness theorem follows from the contraction of the Julia-Teichmüller metric (which implies that $d_{JT}(f_m, g_m) = 0$), and the pullback argument.

Next one proves that given any bisequence $\ldots, \sigma_{-m}, \ldots, \sigma_0, \sigma_1, \ldots$ of unimodal permutations of bounded type there exists a unique map $f_0 \in \mathcal{A}$ and a bisequence f_m as above such that $f_{-m} \in \mathcal{D}_{\sigma_{-m},\ldots}$. The proof of the existence of f_0 in the case of periodic combinatorial type follows from the compactness of $\mathcal{C}$

and the uniqueness. In the general situation it follows from the existence in the periodic case, the uniqueness and the compactness of $\mathcal{C}$. From this result we get the third statement of Theorem 1.1 since the restriction of the renormalization operator to the set of renormalizable real analytic maps is injective and the attractor $\mathcal{A}$ is contained in this set.

From the real bounds we prove in Section 2, it follows that any limit point of the renormalization operator in the C^0 topology is real analytic. This, together with the previous step proves the first statement of Theorem 1.1. The other statements follow from the compactness property given by the complex bounds and the contraction of the Julia-Teichmüller distance.

2 The Real Bounds

In this section we will prove that the attracting Cantor set of a C^2 (or C^{1+z}) infinitely renormalizable unimodal map of bounded combinatorial type has bounded geometry. As we will see, this shows that any map in the ω-limit set of the orbit of an infinite renormalizable C^{1+z} map of bounded combinatorial type under the renormalization operator is real analytic. In the next section we will analyze the bounded geometry property and conclude that given two such maps, with the same combinatorics, there exists a quasiconformal homeomorphism which maps the critical orbit of one map onto the critical orbit of the other map conjugating them along the attracting Cantor sets.

To state the next result we introduce some notation. For $f \in \mathcal{D}_\infty$, with $\sigma(f) = (\sigma_0, \sigma_1, \dots)$, we denote by $\triangle_n$ the interval $[f^{2q(n)}(c_0), f^{q(n)}(c_0)]$ (which contains the turning point) where $q(n) = |\sigma_0| \cdot |\sigma_1| \cdots |\sigma_n|$. Note that $f^{q(n)}$ maps $\triangle_n$ onto itself. Let

$$F_n = \triangle_n \cup \triangle_n^1 \cup \cdots \cup \triangle_n^{q(n)-1}$$

where $\triangle_n^j = f^j(\triangle_n)$. Hence each $\triangle_n^j$ is a connected component of F_n and it is called an *interval of generation n*. Each connected component of $F_{n-1} \setminus F_n$ is called a *gap of generation n*. An interval or a gap of generation n is called an element of generation n.

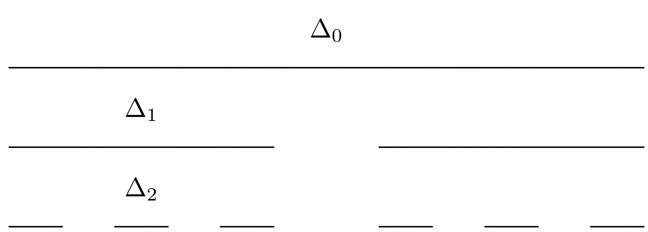

Fig. 2.1: The intervals and gaps of generation 2 and 3.

In the next theorem we shall prove that if we take an infinitely renormalizable C^{1+z} unimodal map (not necessarily of bounded combinatorial type), with non-flat critical point, then its renormalizations are all contained in a compact space of unimodal maps. First we give a definition.

Definition. We say that a bound $L(f) < \infty$ depending on an infinitely renormalizable map f is *beau*, i.e., bounded and eventually universally (bounded), if the following statement holds. $L(\mathcal{R}^n f)$ is finite for each infinitely renormalizable maps f within the class considered and furthermore there exists n_0 such that $L(\mathcal{R}^n f)$ is bounded from above by some constant L for all $n \geq n_0$ and every infinitely renormalizable map f. Similarly, if the estimate is from below or away from one. In other words, an estimate is beau if the bound appearing in it holds for each infinitely renormalizable map after a sufficient number of renormalizations.

Definition. The *Hausdorff dimension* of a subset $C \subset \mathbb{R}$ is defined as follows. If $\mathcal{O}$ is a finite cover of C by intervals we denote by $|\mathcal{O}|$ the maximal length of elements of $\mathcal{O}$. Let s be a positive real number. The s-Hausdorff measure of C is defined as

$$HM_s(C) = \lim_{\epsilon \to 0} \inf_{|\mathcal{O}| \leq \epsilon} \{\sum_{I \in \mathcal{O}} |I|^s\}.$$

The Hausdorff dimension of C is the unique number $\mathrm{HD}(C)$ such that $HM_s(C)$ is zero for $s > \mathrm{HD}(C)$ and is equal to ∞ if $s < \mathrm{HD}(C)$.

Theorem 2.1. **[Real C^1 Bounds]** *Let $f \in \mathcal{U}^{1+z}$ be an infinitely renormalizable map. Then there exists a constant $L(f)$ such that the following statements hold.*

1. *For each $n \in \mathbb{N}$, the distortion of*

$$f^k \colon \triangle_n^j \to \triangle_n^{j+k}$$

is bounded by $L(f)$ if $0 < j \leq j + k \leq q(n)$ where $\triangle_n^{q(n)} = \triangle_n$. The Cantor set $C(f) = \cap_{n=0}^{\infty} F_n$ is the closure of the critical orbit and has zero Lebesgue measure.

1'. *The previous bound is beau: there exists a constant L such that for each f as above, $L(\mathcal{R}^n f) \leq L$ for n sufficiently large.*

2. *If the combinatorial type of f is bounded then there exist constants $0 < \lambda(f) < \mu(f) < 1$ depending only on f, such that if I is an interval of generation n and $J \subset I$ is an element of generation $n+1$ then*

$$\lambda(f) < \frac{|J|}{|I|} < \mu(f).$$

In particular, the Hausdorff dimension of the attracting Cantor set is positive and bounded away from one in this case.

2'. *The previous bounds are beau: given $N \in \mathbb{N}$, there exist constants $0 < \lambda < \mu < 1$ such that for each f as above of combinatorial type $\leq N$, $\lambda \leq \lambda(\mathcal{R}^n f) < \mu(\mathcal{R}^n f) \leq \mu$ for n sufficiently large.*

The proof of this theorem is based on two Koebe Principles – Theorems IV.1.2 and IV.3.3 – which can be stated in our setting as follows. We say that V contains a *δ-scaled neighbourhood of U* if each component of $V \setminus U$ has at least length $\delta|U|$. Furthermore we say that a collections of intervals T_i has *intersection multiplicity κ* if each point is contained in at most κ of these intervals. These Koebe Principles state that for each $\kappa \in \mathbb{N}$ there exist positive functions $B_0, K \colon \mathbb{R}^+ \to \mathbb{R}^+$ with the following property. Let $f \in \mathcal{U}^{1+Z}$ and suppose that for some pair of intervals $M \subset T$, and some $i \in \mathbb{N}$ $f^i|T$ is a diffeomorphism and let κ be the intersection multiplicity of the collection of intervals $T, \dots, f^{i-1}(T)$. If τ is so that $f^i(T)$ contains a τ-scaled neighbourhood of $f^i(M)$ then the Macroscopic Koebe Principle claims that

$$T \text{ contains a } B_0(\tau)\text{-scaled neighbourhood of } M$$

and the (Infinitessimal) Koebe Principle gives that

$$\frac{Df^i(x)}{Df^i(y)} \le K(\tau) \text{ for all } x, y \in M.$$

To prove Theorem 2.1 will need the following lemma. Before stating this lemma we should note that $q(2) \ge 4$ and that the interval Δ_n^3 has two neighbours from the collection $\Delta_n^0, \dots, \Delta_n^{q(n)-1}$ if $n \ge 2$. Since Δ_n^1 and Δ_n^2 have only one neighbour we shall therefore assume $n \ge 2$ and consider the interval Δ_n^3.

Lemma 2.1. *For each infinitely renormalizable map $f \in \mathcal{U}^{1+z}$ there exist $\tau(f) > 0$ and $L(f) < \infty$ such that for each integer $n \ge 2$ there is an interval $S_n \supset \Delta_n^3$ with the following properties:*

1. *The restriction of $f^{q(n)-3}$ to S_n is a diffeomorphism and $f^{i-3}(S_n)$ is a τ-scaled neighbourhood of Δ_n^i for each $3 \le i < q(n)$.*
2. *The distortion of $f^{q(n)-i}$ on $f^{i-3}(S_n) \supset \Delta_n^i$ is bounded by $L(f)$.*
3. *The first return map to $\triangle_n$ is the composition of a quadratic map with a diffeomorphism that has bounded distortion, more precisely, for each $n \ge 0$ $\mathcal{R}^n f$ is of the form $f = \phi_n \circ Q \circ \psi_n$ where ψ_n is a C^{1+z} diffeomorphism whose distortion tends to zero as $n \to \infty$, Q is the quadratic map $Q(x) = x^2$ and ϕ_n is a C^{1+z} diffeomorphism whose distortion is bounded by $L(f)$.*

Proof. We shall prove that S_n is not too small, by finding some space on either side of some interval $\triangle_n^i$ which can be pulled back to some space near $\triangle_n^3$ by the Koebe Principle. Let $I_n = f^{-1}(\triangle_n^1) \supset \triangle_n$. Each component of $I_n \setminus \{c\}$ is mapped homeomorphically onto $\triangle_n^1$ and the intersection of I_n with F_n is equal to $\triangle_n$. For each $3 < i < q(n)$ let $M_n^i \supset \triangle_n^3 = f^3(\triangle_n) = f^3(I_n)$ be the maximal interval such that f^{i-3} is monotone on M_n^i. We claim that then each connected component of $f^{i-3}(M_n^i) \setminus \triangle_n^i$ contains a component of F_n.

Indeed, let $M_n^{i,\beta}$, $\beta \in \{+,-\}$, be the connected components of $M_n^i \setminus \triangle_n^1$. We want to prove that $f^{i-3}(M_n^{i,+})$ and $f^{i-3}(M_n^{i,-})$ both contain a component of F_n. Indeed, let $\beta \in \{-,+\}$; by the maximality of M_n^i there exists $k < i-3$ such that $f^k(M_n^{i,\beta})$ contains the critical point in its boundary or endpoints of the dynamical interval $[f^2(c), f(c)]$. The other boundary point of $f^k(M_n^{i,\beta})$ is an endpoint of $f^k(\triangle_n^1) = \triangle_n^{k+1}$. Since $\triangle_n^{k+1} \cap (I_n \cup \Delta_n^1 \cup \Delta_n^2) = \emptyset$ and since $k < q(n) - 3$, the interval $f^k(M_n^{i,\beta})$ contains either one of the components I_n' of $I_n \setminus \{c\}$, Δ_n^1 or Δ_n^2. As $f(I_n') = \triangle_n^1$ we get that $f^{i-3}(M_n^{i,\beta})$ contains either $\triangle_n^{i-3-k}$, or $\triangle_n^{i-2-k}$ or $\triangle_n^{i-1-k}$. This proves the claim. Now we can define $T^i \supset \triangle_n^3$ to be the smallest interval such that $f^{i-3}|T^i$ is monotone and such that each component of $f^{i-3}(T^i) \setminus \triangle_n^i$ contains a component of F_n. From the same easy arguments we have used before in Chapter IV (see, for instance, Lemma 10.3) we get that the intersection multiplicity of the intervals

$$\{T^i, f(T^i), \dots, f^{i-3}(T^i)\}$$

is at most equal to three.

Let K_n be the smallest interval containing $\triangle_n$ such that each component $K_n \setminus \triangle_n$ contains a component of F_n. We claim that there exists a constant d_0, independent of n, such that K_n is a d_0-scaled neighbourhood of $\triangle_n$. Let us prove this. For $3 \le i < q(n)$ there are components of F_n on either side of $\triangle_n^i$. So for $3 \le i < q(n)$ we can take K_n^i to be the smallest interval containing $\triangle_n^i$ such that each component of $K_n^i \setminus \triangle_n^i$ contains a component of F_n. Let $3 \le l < q(n)$ be so that $|\triangle_n^l| \le |\triangle_n^i|$ for all $3 \le i < q(n)$. Of course, since $\triangle_n^3 = f^2(\triangle_n^1)$, there exists a universal constant d_1 such that $|\triangle_n^1|, |\triangle_n^2| \ge d_1 |\triangle_n^3|$. Hence from the choice of l we get that K_n^l is a d_1-scaled neighbourhood of $\triangle_n^l$. The mapping f^{l-3} maps $T^l \supset \Delta_n^3$ diffeomorphically onto K_n^l and, because of the disjointness statement above we can use the Macroscopic Koebe Principle and we get that T^l is a d_2-scaled neighbourhood of $\triangle_n^3$, where d_2 depends only on d_1 (and so is universal). Since $f^{-3}(T^l)$ is contained in K_n the claim follows immediately.

Now we can complete the proof of the lemma. $f^{q(n)-3}$ maps $T^{q(n)-3}$ diffeomorphically onto K_n and K_n contains a d_0-scaled neighbourhood of Δ_n. So by the Koebe Principle there exists an interval S_n with $\Delta_n^3 \subset S_n \subset T^{q(n)-3}$ such that the restriction of $f^{q(n)-3}$ to S_n has bounded distortion (simply take S_n such that $f^{q(n)-3}(S_n)$ is a $d_0/2$-scaled neighbourhood of Δ_n). In particular, $f^{i-3}(S_n)$ is a $\tilde{d}_0$-scaled neighbourhood of Δ_n^i and, applying Koebe again, $f^{q(n)-i}\colon f^{i-3}(S_n) \to f^{q(n)-3}(S_n)$ has also bounded distortion for each $3 \le i < q(n)$. The last statement of the lemma follows since the restriction of f to $\triangle_n$ is quadratic-like and the first return map to $\triangle_n$ is

$$f^{q(n)} = (f^{q(n)-3}|\triangle_n^3) \circ (f^3|\triangle_n).$$

Hence $\mathcal{R}^n f$ is of the form $\phi_n \circ Q \circ \psi_n$, where $Q \circ \psi_n$ is just a rescaled version of f and ϕ_n is a diffeomorphism whose distortion is universally bounded. □

Proof of Theorem 2.1: Lemma 2.1 gives already the bounded distortion part of the Statement 1) of Theorem 2.1. Let us prove that the attracting Cantor set has zero Lebesgue measure. Indeed, by Lemma 2.1 there exists an interval $S_n \supset \triangle_n^3$ such that $f^{q(n)-3} \mid S_n$ is monotone, has bounded distortion and $f^{q(n)-3}(S_n)$ is a τ-scaled neighbourhood of $\triangle_n = f^{q(n)-3}(\triangle_n^3)$ as in Figure 2.2. In particular, S_n is a τ'-scaled neighbourhood of $\triangle_n^3$.

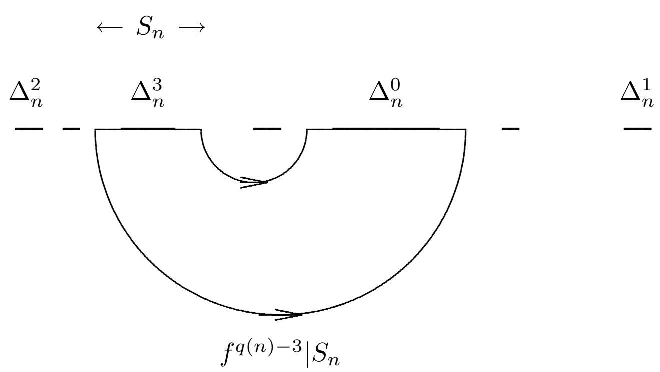

Fig. 2.2:

If we let T_n be the maximal interval containing $\triangle_n$ such that $f^3(T_n) \subset S_n$ then we obviously have that T_n contains a τ''-scaled neighbourhood of $\triangle_n$. Since $f^{q(n)}$ restricted to T_n is unimodal, T_n does not intersect any component of F_n except $\triangle_n$. Hence, the components of $T_n \setminus \triangle_n$ are contained in the complement of F_n as indicated in Figure 2.2 and at least one of them belongs to F_{n-1}. From all this it follows that the gap of generation n around $\triangle_n$ is not too small. Now take $3 \leq 3 < q(n)$. The image of $W_n^i = f^{i-3}(S_n)$ under $f^{q(n)-i}$ certainly contains T_n and this map has bounded distortion. Since F_n is forward invariant, one can pull this gap back to a gap around Δ_n^i by this map. So Δ_n^i is contained in a $\tilde{\tau}$-scaled neighbourhood S_n^i for which $S_n^i \cap F_n \subset \Delta_n^i$. Therefore, the size of each component of $F_{n-1} \setminus F_n$ is at least a constant times the size of the neighbouring components of F_n. This implies that there exists a constant $\lambda < 1$ such that $|F_n| < \lambda |F_{n-1}|$ and thus the Lebesgue measure of $C(f)$ is zero.

Let us show that the Hausdorff dimension of the attracting Cantor set is less than one. In order to show this, we claim that given $\mu > 0$ and $N < \infty$ there exists $s_0 < 1$ with the following property. If $s \in (s_0, 1)$, $q \leq N$, V is an interval and $U_1, \dots, U_q$ are subintervals such that the size of U_i is at most $(1/\mu)$ times the size of the components of $V \setminus (U_1 \cup \dots \cup U_q)$ neighbouring U_i then

$$|V|^s \geq (|U_1|^s + \dots + |U_q|^s).$$

Indeed, from this assumption $|V|^s \geq ((1 + \frac{\mu}{2})|U_1| + \dots + (1 + \frac{\mu}{2})|U_q|)^s$. Hence $|V|^s \geq (q \cdot (1 + \frac{\mu}{2}))^s \times (\frac{1}{q}|U_1| + \dots + \frac{1}{q}|U_q|)^s$. Since $s < 1$, the function $t \to t^s$ is concave and we get that $(\frac{1}{q}|U_1| + \dots + \frac{1}{q}|U_q|)^s \geq \frac{1}{q}|U_1|^s + \dots + \frac{1}{q}|U_q|^s$. Therefore, $|V|^s \geq (q \cdot (1 + \frac{\mu}{2}))^s \cdot \frac{1}{q} \cdot (|U_1|^s + \dots + |U_q|^s) \geq (|U_1|^s + \dots + |U_q|^s)$ if s is sufficiently close to one. This proves the claim.

Since the combinatorics is bounded by N we get from this

$$(*) \qquad \sum_{i=0}^{q(n)-1} |\triangle_n^i|^s \geq \sum_{i=0}^{q(n+1)-1} |\triangle_{n+1}^i|^s$$

when s is close to one. Since the maximal length of components of F_n goes to zero exponentially with n, we get that the s-Hausdorff measure of C is finite. Hence the Hausdorff dimension of C is smaller or equal to s.

To prove 2) let $\tilde{L} = L(f)$ and $\mathcal{C}_{\tilde{L}}$ be the class from Lemma 2.1. This set is closed and therefore compact in the C^0 topology. Let $\mathcal{C}_{\tilde{L},N}$ be the maps in $\mathcal{C}_{\tilde{L}}$ of type σ with $|\sigma| \leq N$. Again this set is compact. The functions which associate to a map in $\mathcal{C}_{\tilde{L},N}$ the length of the smallest interval, the length of the smallest gap, the length of the largest interval and the length of the biggest gap, are continuous and strictly positive. Hence, by compactness they are bounded and bounded away from zero. Since $\mathcal{R}^n f(x) \in \mathcal{C}_{\tilde{L},N}$ for all $n \in \mathbb{N}$ this proves 2).

The proof of 1') and 2') is based on the following fact. By Theorem IV.2.1, provided $f \in \mathcal{U}^{1+z}$ the lower bound for the cross-ratio distortion of $\mathcal{R}^n f$ goes to one if the total length of the intervals $\Delta_n^0, \ldots, \Delta_n^{q(n)-1}$ goes to zero as $n \to \infty$ and by Statement 1) this assumption is satisfied. From Theorem IV.1.2, it follows that the constants appearing in Lemma 2.1 can be chosen uniformly. □

In the next theorem we shall improve on the previous result and show that if we take an infinitely renormalizable C^{1+z} unimodal map then its renormalizations are even bounded in the $C^{1+\alpha}$ sense.

Theorem 2.2. [**Real** $C^{1+\alpha}$ **Bounds**] *Let $f \in \mathcal{U}^{1+z}$ be an infinitely renormalizable map of bounded combinatorial type and let $\alpha < 1$. Then*

$$f^k \colon \triangle_n^j \to \triangle_n^{j+k}$$

is uniformly $C^{1+\alpha}$ for all $0 < j < j+k \leq q(n)$ and all $n \in \mathbb{N}$. More precisely, there exists some constant $L(f)$ such that

$$\left| \frac{Df^k(x)}{Df^k(y)} - 1 \right| \leq L(f) \left[\frac{|x-y|}{|\triangle_n^j|} \right]^\alpha$$

for all $x, y \in \triangle_n^j$. Moreover, this estimate is beau: there exists a constant L such that for each f as above, $L(\mathcal{R}^n f) \leq L$ for n sufficiently large.

Proof. Since the Hausdorff dimension of C is less than one, there exists $s < 1$ for which $\sum_{i=0}^{q(n)-1} |\triangle_n^i|^s$ is universally bounded (in fact this is inequality $(*)$ in the proof of Theorem 2.1). So the theorem follows immediately from the $C^{1+\alpha}$ Koebe Principle, see Theorem IV.3.2. □

Remark. Notice that for Statements 1) and 2) in Theorem 2.1 we did not need to require the map f to be in $\mathcal{U}^{1+z}$. Indeed, as we have seen in Sections IV.2

and IV.3 the natural differentiability hypothesis in the Koebe Principle is that the map should belong to the Zygmund class C^{1+Z} for Statements 1) and 2) and should belong to the class C^{1+z} for Statements 1') and 2'). (For these statements it is necessary that the distortion tends to one.)

Next we want to show that any sequence of renormalizations is precompact in the following sense. From Lemma 2.1, $f_n = \mathcal{R}^n(f)$ can be written in the form $\phi_n \circ Q \circ \psi_n$ where ϕ_n is a C^{1+z} diffeomorphism whose distortion is uniformly bounded, Q is the quadratic map and ψ_n is a C^{1+z} diffeomorphism whose distortion tends to zero. It follows that each subsequence of ϕ_n has a subsequence which converges in the C^0 topology to a homeomorphism ϕ. Moreover, ψ_n converges to the space of affine maps. In particular, each subsequence of f_n has a convergent subsequence in the C^0 topology. In the next theorem it is shown that this limit is not just C^0 but even real analytic.

Theorem 2.3. **[Infinitely renormalizable C^{1+z} maps are asymptotically in the Epstein class]**

Let f be a C^{1+z} infinitely renormalizable unimodal map of bounded type. Let $f_n\colon [0,1] \to [0,1]$ be the sequence $f_n = \mathcal{R}^n(f)$ of renormalized maps. Then there exists $a = a(f) > 0$ such that any map $\tilde{f}$ that is a C^0 limit of a convergent subsequence of f_n belongs to the Epstein class $\mathcal{E}_a$. Writing $f_n = \phi_n \circ Q \circ \psi_n$ and $\tilde{f} = \tilde{\phi} \circ Q \circ \tilde{\psi}$ we get $\phi_n \to \tilde{\phi}$ in the $C^{1+\alpha}$ topology and $\tilde{\phi}^{-1}$ has a univalent holomorphic extension to the domain $(\mathbb{C}\backslash\mathbb{R}) \cup [-a, 1+a]$. Moreover, ψ_n converges to ψ in the $C^{1+\alpha}$ topology and ψ is an affine map. The number $a(f)$ is beau.

Proof. Because of Proposition 1.5 we may take a C^{1+z} coordinate change and we can assume that f is of the form $h \circ Q \circ A$ where $Q(x) = x^2$, h is C^{1+z} and A is an affine map. From the proof of Lemma 2.1 there exists $a > 0$ and a-scaled intervals S_n^i of $\triangle_n^i$ such that $f(S_n^i) = S_n^{i+1}$ for $i = 1, \dots, q(n) - 1$. Let us write, up to affine conjugacy,

$$f_n = (\, f \circ f \circ \cdots \circ f \,) \circ f$$

and split off the first term f. So consider the chain $g_{q(n)-1} \circ g_{q(n)-2} \circ \cdots \circ g_1$ where

$$g_i = f|S_n^i = h \circ Q \circ A|S_n^i.$$

Furthermore, let $\tilde{h}_i$ be the affine map from $QA(S_n^i)$ onto $h \circ f(S_n^i)$ and let

$$\tilde{g}_i = \tilde{h}_i \circ Q \circ A|S_n^i$$

and

$$\tilde{f}_n = \tilde{g}_{q(n)-1} \circ \tilde{g}_{q(n)-2} \circ \cdots \circ \tilde{g}_1.$$

We want to show that

$$(*) \qquad \begin{aligned} &|D(g_{q(n)-1} \circ \cdots \circ g_1) \,-\, D(\tilde{g}_{q(n)-1} \circ \cdots \circ \tilde{g}_1)| \\ &\le K'' \cdot \left(\frac{|x-y|}{|S_n^1|} \right)^{\alpha} \cdot \sum_{i=1}^{q(n)-1} |S_n^i|^{\alpha}. \end{aligned}$$

So let us show that replacing g_i by $\tilde{g}_i$ causes only an error of order $|S_n^i|^\alpha$ in the $C^{1+\alpha}$ metric. So write $x_i = g_{i-1} \circ \cdots \circ g_0(x)$ and $y_i = g_{i-1} \circ \cdots \circ g_0(y)$. Since h is a $C^{1+\alpha}$ diffeomorphism for each $\alpha < 1$, and from the $C^{1+\alpha}$ bounds in the previous theorem this gives

$$|Dh_i(Q(A(x_i))) \; - \; Dh_i(Q(A(y_i)))| \leq K' \left(\frac{|x-y|}{|S_n^1|} \right)^\alpha \cdot |S_n^i|^\alpha.$$

Similarly, using again the $C^{1+\alpha}$ bounds from Theorem 2.2, we can estimate both

$$\left| D(g_{q(n)-1} \circ \cdots \circ g_{i+1})(g_i(x_i)) \; - \; D(g_{q(n)-1} \circ \cdots \circ g_{i+1})(g_i(y_i)) \right|$$

and

$$\left| D(g_{q(n)-1} \circ \cdots \circ g_{i+1})(\tilde{g}_i(x_i)) \; - \; D(g_{q(n)-1} \circ \cdots \circ g_{i+1})(\tilde{g}_i(y_i)) \right|$$

from above by

$$\leq K' \left(\frac{|x-y|}{|S_n^1|} \right)^\alpha \cdot |S_n^i|^\alpha.$$

Combining this, it follows by the chain-rule that the term

$$D(g_{q(n)-1} \circ \cdots \circ g_1)(x) - D(g_{q(n)-1} \circ \cdots \circ g_1)(y)$$

and the corresponding term with g_i replaced by $\tilde{g}_i$ differ in absolute value at most by

$$K' \left(\frac{|x-y|}{|S_n^1|} \right)^\alpha \cdot |S_n^i|^\alpha.$$

Using this repeatedly gives $(*)$. After scaling it follows from $(*)$ that

$$\left| \; \left[Df_n(x) - Df_n(y) \right] \; - \; \left[D\tilde{f}_n(x) - D\tilde{f}_n(y) \right] \; \right| \leq K'' \cdot |x-y|^\alpha \cdot \sum_{i=1}^{q(n)-1} |S_n^i|^\alpha.$$

By Theorem 2.1 this is exponentially small in n and so the $C^{1+\alpha}$ distance between f_n and $\tilde{f}_n$ goes exponentially fast to zero. Note that $\tilde{f}_n$ is a composition of quadratic and linear maps, that $\tilde{g}_n \colon S_n^1 \to S_n^{q(n)}$ is a diffeomorphism and that S_n^i is a a-scaled neighbourhood of $\triangle_n^i$ for each $i = 1, \ldots, q(n)$ and each $n \geq 0$. Hence $\tilde{f}_n$ is in $\mathcal{E}_a$. It therefore follows that f_n is exponentially close in the $C^{1+\alpha}$ sense to a map from the Epstein class $\mathcal{E}_a$. Let us show that any convergent subsequence of f_n even tends to a map from the Epstein class $\mathcal{E}_a$ by showing that the set $\mathcal{E}_a$ (endowed with the topology of uniform convergence) is compact. Write $\tilde{f}_n = \tilde{\phi}_n \circ Q \circ A_n$ where $\tilde{\phi}_n^{-1}$ has a holomorphic univalent extension to $(\mathbb{C} \setminus \mathbb{R}) \cup [-a, 1+a]$. By Montel's Theorem there exists a subsequence of $\tilde{\phi}_n$ such that $\tilde{\phi}_n^{-1}$ converges uniformly on compact subsets of $(\mathbb{C} \setminus \mathbb{R}) \cup [-a, 1+a]$ to a holomorphic map $\tilde{\phi}^{-1}$ and by Lemma 2.1 $\tilde{\phi}^{-1}$ is not constant. It follows that $\tilde{f}_n$ is also in $\mathcal{E}_a$. From this compactness it follows that the sequence f_n tends to a map from the Epstein class $\mathcal{E}_a$. □

3 Bounded Geometry

Let $C \subset \mathbb{R}$ be a Cantor set. Because the Cantor sets we will consider are dynamically defined they are naturally given as an intersection of a nested sequence of finite unions of intervals. In this section we will show that these defining intervals satisfy some very rigid structure when they correspond to an infinitely renormalizable map of bounded combinatorial type. Let us start with some definitions.

Definition. A *presentation* of a Cantor set C is a decreasing collection $F_n \supset F_{n+1}$ of closed sets such that each F_n is a finite union of closed intervals whose boundary points are in C; each connected component of F_n contains the same number a_n of connected components of F_{n+1} and $\cap_{n=0}^{\infty} F_n = C$. Each component of F_n is called an *interval* of generation n and each component of $F_n \setminus F_{n+1}$ is called a *gap* of generation $n+1$. Of course, there are many presentation of a Cantor set. However, if a Cantor set is the forward orbit of the turning point of an infinitely renormalizable unimodal map then it is natural to take F_n to be the union of the intervals $\triangle_n^i$ from the previous section.

Definition. We say that the presentation $\{F_n\,;\ n = 0, 1, 2, \dots\}$ of C has *bounded geometry* if there exist $0 < \lambda < \mu < 1$ such that if I is an interval of generation n and $J \subset I$ is either an interval of generation $n+1$ or a gap of generation $n+1$ then

$$0 < \lambda < \frac{|J|}{|I|} < \mu < 1.$$

It follows from the above definition that if the presentation $\{F_n\,;\ n = 0, 1, \dots\}$ of C has bounded geometry then it has *bounded combinatorics*, namely, the number a_n of components of F_{n+1} in each component of F_n is bounded independently of n.

The main result of this section is the following

Theorem 3.1. *Let $\{F_n^{(i)}\,;\ n = 0, 1, 2, \dots\}$ be presentations of bounded geometry of the Cantor sets $C^{(i)} \subset \mathbb{R}$, $i = 1, 2$. Suppose that these presentations have the same combinatorics, i.e., the number of components of $F_{n+1}^{(i)}$ in each component of $F_n^{(i)}$ does not depend on i. Then there exists a quasiconformal homeomorphism $h\colon \mathbb{C} \to \mathbb{C}$ which is symmetric and maps $C^{(1)}$ onto $C^{(2)}$.*

Corollary 3.1. *Let $f, \tilde{f}$ be C^2 infinitely renormalizable with the same bounded combinatorial type. Then there exists a quasiconformal homeomorphism $h\colon \mathbb{C} \to \mathbb{C}$ such that $h(f^n(c)) = \tilde{f}^n(c)$.*

Proof. This follows immediately from the real bounds given in Theorem 2.1 and the previous theorem. □

In order to prove this theorem we will consider a large disc D containing the Cantor set C and we will construct a decomposition of the Riemann surface $S = D \setminus C$ into countable many pairs of pants whose boundaries are geodesics in the hyperbolic metric of S. The main step is to prove that the lengths of these geodesics are bounded from above and from below.

Definition. We say that a set is a *pair of pants* if it is homeomorphic to an open disc with two closed discs taken out. Let F_n be a presentation of the Cantor set C and $S = D \setminus C$ as above. We say that a countable family $\mathcal{F}C$ of disjoint simple closed curves in S determines *a standard decomposition of S in pairs of pants* if the following properties hold:

1. each connected component of $D \setminus \cup\{\alpha\,;\ \alpha \in \mathcal{F}C\}$ not containing points of C is a pair of pants.
2. for each interval I of level n there exists a unique curve $\alpha \in \mathcal{F}C$ such that the disc $D_\alpha \subset D$ bounded by α contains I and contains no other interval of level n;
3. let $\alpha, \beta, \gamma \in \mathcal{F}C$ be the boundary of a pair of pants; assume that the disc D_γ bounded by γ contains α and β and that $C \cap D_\alpha$ is to the left of $C \cap D_\beta$; then α is as in 2), namely, D_α contains one and only one interval of level n for some n.

Notice that the above properties define a unique decomposition of S into pairs of pants up to homotopy. In the special case when each component of F_n contains exactly two components of F_{n+1} then all the boundary curves of the pants are as in 2), see Figure 3.1 on the left. In Figure 3.2 a decomposition is drawn in a surface with a hyperbolic structure.

In the proof of Theorem 3.1 we will need the following lemma.

Lemma 3.1. *Let F_n be a presentation of bounded geometry of a Cantor set C. Let $\mathcal{F}C$ be the set of boundary curves of a standard decomposition into pairs of pants of $S = D \setminus C$. For $\alpha \in \mathcal{F}C$, let α' be the unique hyperbolic geodesic in the homotopy class of α. Then $\mathcal{F}C' = \{\alpha'\,;\ \alpha \in \mathcal{F}C\}$ defines a standard decomposition into pairs of pants of S and there exist positive constants $b > a > 0$ such that the hyperbolic length of each $\alpha' \in \mathcal{F}C'$ lies between a and b.*

Proof. Let $l_S(\alpha)$ denotes the length of the curve α in the hyperbolic metric of S. Let $\alpha \in \mathcal{F}C$ and let n be such that the disc D_α bounded by α contains an interval I of level n and does not contain any interval of smaller level. Hence all intervals of level n contained in D_α are contained in the same interval of level $n-1$. Let $A \subset S$ be the annulus defined as the union of the upper half plane,

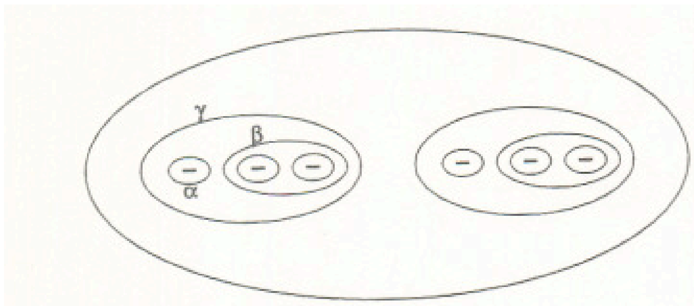

Fig. 3.1: Two standard decompositions in pairs of pants. If each component of F_n contains exactly two respectively three components of F_{n+1} then all the boundary curves of the pairs of pants are as shown above.

the lower half plane and the two gaps of level $n-1$ adjacent to the convex hull of the union of intervals of level n contained in D_α. Now the size of these gaps is of the same order as the length of this convex hull and therefore we get that the modulus of A is bounded from below by a universal constant (which does not depend on α). Hence, if $\gamma \subset A$ is the simple closed geodesic in the hyperbolic metric of A, then $l_A(\gamma)$ is bounded from above by a universal constant. On the other hand, by construction of A, we have that γ is homotopic to α. Therefore, $l_S(\gamma) \geq l_S(\alpha')$. On the other hand, since $A \subset S$, we get, from the Schwarz Lemma (see the Appendix), that $l_A(\gamma) > l_S(\gamma)$. Therefore $l_S(\alpha')$ is bounded from above by a universal constant b.

Let us prove now that the hyperbolic length $l_S(\alpha')$ is also bounded from below by a constant a which depends only on the bounds of the geometry of C. So let us show that $l_S(\alpha)$ cannot be too small. For this we will use the following fact about Riemann surfaces: if γ is a (small) homotopically non-trivial simple closed geodesic in a Riemann surface S then γ has an annulus neighbourhood $A \subset S$ which is homotopically equivalent to γ and whose modulus is, up to universal constants, inversely proportional to the hyperbolic length of γ, see for example Theorem 6.3 from Douady and Hubbard (1985b). So if $l_S(\alpha)$ is very small we get from this that α is contained in an annulus A homotopic to α with very large modulus. So, by trimming the annulus and using Koebe's Lemma, see Lemma 4.2 in Section 4 of this chapter, there is a univalent map G from a very thick annulus $A_R = \{z \in \mathbb{C}\,;\ 1 < |z| < R\}$ onto an annulus A homotopic to α whose distortion is bounded by a universal constant. Hence, there is a curve γ connecting the two boundary components of A_R whose image is contained in a gap J of order n and the image of a circle $\{z\,;\ |z| = 2\}$ goes around an interval I of level n such that I and J are both contained in the same interval of level $n-1$. Thus the Euclidean lengths of I and J are comparable, by bounded geometry. This is not possible for R big since the distortion of G is universally

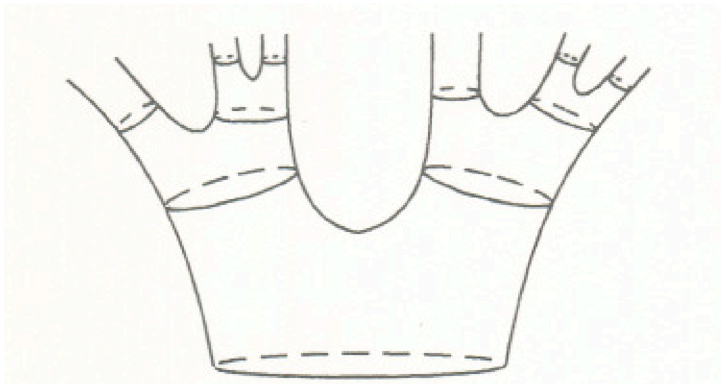

Fig. 3.2: The decomposition in pairs of pants in a surface with a hyperbolic structure.

bounded. This proves that the length of the curves in $\mathcal{F}C'$ is bounded from below and the lemma is proved. $\square$

Proof of Theorem 3.1: Let D be a very large disc centered at the origin that contains both Cantor sets $C^{(i)}$, $i = 1, 2$. Let $\mathcal{F}C^i$ be the family of simple closed geodesics of $S^{(i)} = D \setminus C^{(i)}$ which are the boundary curves of the standard decomposition into pairs of pants. Notice that, since complex conjugation is a conformal involution of $S^{(i)}$, all gaps of the Cantor sets are geodesics and all pairs of pants are symmetric (with respect to complex conjugation). Each boundary of a pair of pants intersects the real axis at two points which we call special points. The intersection of the real axis with each pair of pants consists of three geodesic segments connecting special points of the different boundary components. Since the two Cantor sets have the same combinatorics, we can start by defining a monotone map $\phi\colon C^{(1)} \to C^{(2)}$ and extend ϕ to a monotone map of the set of special points so that special points in the same pair of pants are mapped into special points in the same pair of pants. This gives a correspondence between pairs of pants. Next we extend ϕ to a quasiconformal map between each corresponding pair of pants in the following way. We start by extending ϕ to the boundary of each pair of pants by interpolating 'linearly' between two special points. Next we extend it symmetrically to the interior of each pair of pants. Clearly we can do this so that the quasiconformal constant of ϕ on each pair of pants depends only on the lengths of the boundary components of the domain and of the range. From Lemma 3.1, there is a common bound for the complex dilatation of ϕ in all pairs of pants. As the definition of ϕ in the boundary of each pair of pants depends only on the special points and the lengths of the corresponding boundary components, ϕ extends continuously to a quasiconformal map of $S^{(1)}$ onto $S^{(2)}$ (ϕ is K-quasiconformal in each pair of pants, is continuous and the boundary of the pairs of pants have zero

Lebesgue measure). Since the restriction of ϕ to the real axis is monotone and the Euclidean diameter of the pairs of pants tends to zero, we get that ϕ extends continuously to the closure of $S^{(1)}$. It follows that ϕ is K-quasiconformal because the Cantor sets C^i have zero Lebesgue measure. This completes the proof of the theorem. □

Remark. It is possible to give a more elementary proof of Theorem 3.1 using circles which intersect the real axis in special points. However, the notation becomes more cumbersome in that case.

4 The Pullback Argument

In the previous two sections it was shown that two infinitely renormalizable unimodal maps in $\mathcal{U}^2$ (or even in $\mathcal{U}^{1+Z}$) of the same bounded combinatorial type are quasi-symmetrically conjugate on the attracting Cantor set. In particular, there exists a quasiconformal homeomorphism on $\mathbb{C}$ which maps the first of these Cantor sets on the second one and acts as a conjugacy on the Cantor sets.

Now we will show that if these maps have holomorphic extension to a neighbourhood of the interval and these holomorphic extensions are quadratic-like, as defined below, then we can even find a quasiconformal homeomorphism which is a conjugacy between their holomorphic extensions to neighbourhoods of the attracting Cantor sets. In Section 6 we will show that if one sufficiently often renormalizes a real analytic map, which belongs to some Epstein class and has bounded combinatorial type, then the resulting map has a holomorphic extension which is quadratic-like.

4.1. Quadratic-like maps and the pullback argument

The maps we shall consider are quadratic-like in the sense defined by Douady and Hubbard (1985a):

Definition. Assume that U and V are simply connected domains in $\mathbb{C}$. Then a holomorphic map $F: U \to V$ is called *quadratic-like* if the closure of U is contained in V and if there exists a unique critical point c of F such that F restricted to $U \setminus \{c\}$ is a covering map of degree two onto $V \setminus \{F(c)\}$. The subset

$$J(F) = \{z \in U\,;\ F^n(z) \in U \text{ for all } n \geq 0\}$$

is called the *filled Julia set* of F.

Let f be a real analytic unimodal map whose critical point is non-degenerate, i.e., the second derivative of f at the critical point c is non-zero. Then f has a holomorphic extension F to a simply connected neighbourhood U of the dynamical interval in the complex plane. By taking U small enough, it is clear that F maps U onto an open set V as a branched covering, i.e. the restriction of F to $U \setminus \{c\}$ is a holomorphic covering map of degree two onto $V \setminus \{f(c)\}$, where c is the critical point of f. In Figure 4.1 we indicate by a dotted line the segment L in the real axis, between the critical value $v = f(c)$ and the boundary of V, which is not in the image of f. The pre-image of this segment by F is a curve transversal to the real axis that splits U in two connected components U_- and U_+, as indicated in Figure 4.1, and F maps each of them homeomorphically onto $V \setminus L$. We denote the corresponding two branches of the inverse of F by $F_\pm^{-1}\colon V \setminus L \to U_\pm$. Being the holomorphic extension of a real analytic map of an interval, F is symmetric with respect to the real axis, i.e., $\overline{F(z)} = F(\bar{z})$ where $\bar{z}$ denotes the complex conjugate of the complex number z. If the closure of U is contained in V then $F\colon U \to V$ is quadratic-like map as defined above. These are special type of quadratic-like maps because they are symmetric with respect to the real axis. So we will call them *symmetric quadratic-like*. Clearly $J(F)$ contains the dynamical interval and all its negative iterates under F. So it is a rather complicated subset of the plane (except when the second iterate of the critical point is a fixed point; in this case the filled Julia set is just the dynamical interval).

Definition. Two symmetric quadratic-like maps $F\colon U \to V$ and $\tilde{F}\colon \tilde{U} \to \tilde{V}$ are said to be *holomorphically equivalent* if there exists a symmetric holomorphic diffeomorphism $\phi\colon U_1 \to \tilde{U}_1$ such that $\phi \circ F = \tilde{F} \circ \phi$, where $U_1 \subset U$ (resp. $\tilde{U}_1 \subset \tilde{U}$) is a symmetric neighbourhood of $J(F)$ (resp. $J(\tilde{F})$).

The space of equivalence classes of quadratic-like maps is denoted by $\mathcal{G}$ and $[F]$ is the equivalence class of F. The elements of $\mathcal{G}$ will also be called germs of quadratic-like maps.

Definition. The *conformal type* of a symmetric quadratic-like map $F\colon U \to V$ are the following two positive real numbers $m_1 < m_2$. Here m_1 is the modulus of the annulus $V \setminus \overline{U}$ and m_2 is the modulus of $V \setminus [F(v), v]$, where v is the critical value of F. We say that the conformal type of F is bounded by B if $\frac{1}{B} \le m_1 < m_2 < B$.

Remark. A. Douady has observed that the first inequality of the above inequalities is the essential one since once we have this inequality we can get the other one by changing the domain of the map. However this requires a little argument and we will not discuss it here.

Any map of the form $f(x) = 1 - ax^2$ is quadratic-like (just take U to be some very large disc). Moreover, by the so-called Straightening Theorem which is due to Douady and Hubbard (1985a), any quadratic-like map is quasiconformally conjugate to one of the above maps (with real valued a if it is symmetric). This result is stated as Theorem 5.10 of the Appendix.

Definition. Let $F\colon U \to V$ be a quadratic-like map. We say that F is *infinitely renormalizable of combinatorial type* $(\sigma_0, \sigma_1, \dots)$ if the restriction of F to the interval $[F(v), v]$ of the real line, if v is the critical value of F, is an infinitely renormalizable unimodal map of combinatorial type $\sigma_0, \sigma_1, \dots$. We denote by $\mathcal{G}_{\sigma_0, \sigma_1, \dots}$ the set of equivalence classes of quadratic-like maps of combinatorial type $(\sigma_0, \sigma_1, \dots)$. Similarly we say that a quadratic-like map $F\colon U \to V$ is *of bounded combinatorial type* if it is infinitely renormalizable of combinatorial type $\sigma_0, \sigma_1, \dots$ and if $|\sigma_i|$ is bounded.

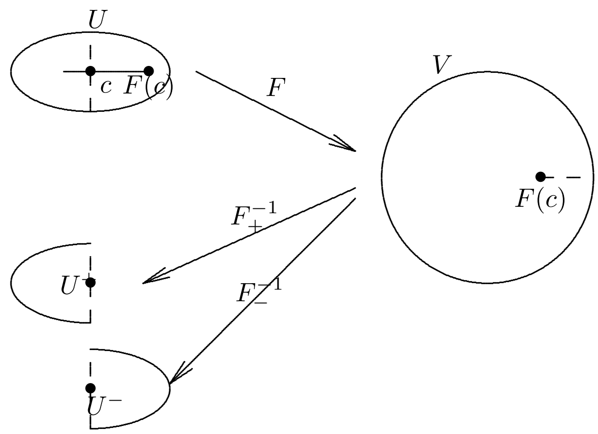

Fig. 4.1:

From the Straightening Theorem it follows that if F is infinitely renormalizable then $J(F)$ is connected and, by Sullivan's Theorem on the non-existence of wandering domains for rational maps, $J(F)$ has empty interior (and hence coincides with the Julia set of F).

In the next theorem we shall show that two symmetric quadratic-like maps of the same bounded combinatorial type are quasiconformally conjugate.

Theorem 4.1. **(Pullback)** *Let $F\colon\ U \to V$ and $\tilde{F}\colon\ \tilde{U} \to \tilde{V}$ be two infinitely renormalizable symmetric quadratic-like maps of the same bounded combinatorial type. Then there exists a symmetric quasiconformal map $h\colon\ F(U_1) \to \tilde{F}(\tilde{U}_1)$ such that $h \circ F = \tilde{F} \circ h$ on the neighbourhood U_1 of $J(F)$.*

Proof. Let $V_1 \supset U$ be a symmetric neighbourhood of U whose boundary is a real analytic curve in V near the boundary of V. Then the closure of $U_1 = F^{-1}(V_1)$ is contained in V_1. By conjugating with the Riemann mapping of V_1 we may assume that $V_1 = \mathbb{D}$. Similarly we may assume that $\tilde{V}_1 = \mathbb{D}$.

First we show that there exists a quasiconformal homeomorphism $h_0\colon\ \mathbb{D} \to \mathbb{D}$ such that

$$h_0(\overline{z}) = \overline{h_0(z)}$$
$$\tilde{F} \circ h_0 = h_0 \circ F \text{ on } \partial U_1$$
$$h_0(F^n(c)) = \tilde{F}^n(c) \text{ for all } n \in \mathbb{N}.$$

In fact, since F and $\tilde{F}$ have the same bounded combinatorial type, by Theorem 3.1 there exists a quasiconformal homeomorphism $\phi\colon\ \mathbb{C} \to \mathbb{C}$ that is symmetric with respect to the real axis and maps the critical orbit of F onto the critical orbit of $\tilde{F}$ conjugating F and $\tilde{F}$ on the critical orbit. This quasiconformal homeomorphism may be taken to be a smooth diffeomorphism in the complement of a neighbourhood of the critical orbit. Next, since the boundary of U_1 and $\tilde{U}_1$ are smooth curves, we can construct a smooth diffeomorphism $\psi\colon\ \mathbb{C} \to \mathbb{C}$ such that ψ is symmetric, is the identity on the boundary of $\mathbb{D}$, and conjugates F with

$\tilde{F}$ at points of the boundary of U_1. We can now construct h_0 by gluing these two quasiconformal homeomorphisms: let h_0 coincide with ψ on a neighbourhood of the fundamental domain $\mathbb{D} \setminus U_1$ and coincide with ϕ on a neighbourhood of the critical orbit.

Next we will construct a sequence $h_n \colon \mathbb{D} \to \mathbb{D}$ of quasiconformal homeomorphism satisfying the following properties:

$$\begin{aligned}
&\tilde{F} h_{n+1} = h_n F,\\
&h_{n+1} = h_n \text{ on } \mathbb{D} \setminus F^{-n}(U_1),\\
&h_{n+1}(\overline{z}) = \overline{h_{n+1}(z)}.\\
&h_{n+1} \text{ conjugates } F \text{ and } \tilde{F} \text{ along the critical orbit}
\end{aligned}$$

This is done by induction. We start with h_0 and assume by induction that we have constructed h_n. Since h_n maps the critical value v of F in the critical value $\tilde{v}$ of $\tilde{F}$, there exists a unique lift of h_n to a map $h_{n+1} \colon U_1 \to \tilde{U}_1$ that maps the upper half plane in the upper half plane. In fact, $h_{n+1} = \tilde{F}_{\pm}^{-1} \circ h_n \circ F$ on $U_{1,\pm}$, where $\tilde{F}_{\pm}^{-1}$ are the holomorphic inverse branches of $\tilde{F}$ as above. Since h_n is quasiconformal and the other maps are conformal it follows that h_{n+1} is quasiconformal with the same conformal distortion as h_n. On the other hand, since h_n coincides with h_0 on $\mathbb{D} \setminus U_1$ and h_0 conjugates F with $\tilde{F}$ on the boundary of U_1 we see that h_{n+1} coincides with h_0 in the boundary of U_1. Hence it can be extended continuously to $\mathbb{D}$ by setting it equal to h_0 on $\mathbb{D} \setminus U_1$. This extension is quasiconformal and has the same quasiconformal distortion as h_n because the boundary of U_1 is smooth (hence has zero Lebesgue measure). Now we claim that h_{n+1} maps the critical orbit of F in the critical orbit of $\tilde{F}$. This follows because, by induction, h_n has this property and F has the same combinatorial type as $\tilde{F}$. This last condition implies that if $p \in U_{1,\pm}$ is in the critical orbit of F then the point in $\tilde{F}^{-1}(h_n(F(p))$ that belongs to the critical orbit of $\tilde{F}$ lies in $\tilde{U}_{1,\pm}$. That is, they have the same kneading sequence.

We claim that the sequence h_n converges uniformly to a quasiconformal homeomorphism h which is a conjugacy between F and $\tilde{F}$. Indeed, let K be the quasiconformal distortion of h_0. Since all maps h_n are K-quasiconformal, and the set of K-quasiconformal homeomorphisms is compact, we see that there are subsequences that converges uniformly. On the other hand, since h_{n+1} is equal to h_n outside of $F^{-n}(\mathbb{D})$ we see that any two limits of convergent subsequences must coincide in the complement of the Julia set of F. The claim follows because the filled Julia set of F has empty interior in this setting. $\square$

Remark. Suppose there exists a quasiconformal homeomorphism $\phi \colon \mathbb{D} \to \mathbb{D}$ that conjugates F with $\tilde{F}$ in the boundary of U_1 and that the quasiconformal distortion of ϕ on the fundamental neighbourhood $\mathbb{D} \setminus U_1$ is bounded by k. The conformal distortion K of the map h_0 in the above proof may be much larger than k. However, the conformal distortion of the conjugacy h is precisely equal to k in the complement of the Julia set of F. This is not sufficient to prove that the conjugacy is k-quasiconformal and not just K-quasiconformal because the Julia set might have non-zero Lebesgue measure (this is an open question). We will prove in the next subsection that the conformal distortion of the conjugacy has to be zero on the Julia set and, therefore, the conjugacy is k-quasiconformal.

4.2. The monotonicity of the kneading invariant for the quadratic family

Let us first deduce two theorems from this pullback argument which were already announced in Chapter II. These theorems are based on the following lemma.

Lemma 4.1. *Let $h\colon \overline{\mathbb{C}} \to \overline{\mathbb{C}}$ be some quasiconformal homeomorphism and $f\colon U \to V$ be a conformal map. Then $h \circ f \circ h^{-1}$ is conformal if the Beltrami coefficient μ associated to h satisfies*

$$\mu(z) = \mu(f(z)) \cdot \frac{\overline{f'(z)}}{f'(z)} \tag{4.1}$$

almost everywhere.

Proof. It follows from an elementary calculation that (4.1) implies that $\bar{\partial}(h \circ f \circ h^{-1}) = 0$. Therefore, by Weyl's Lemma, $h \circ f \circ h^{-1}$ is almost everywhere equal to a conformal map. Since this map is continuous it is conformal, see also Theorem 5.1 on page 28 of Lehto and Virtanen (1973). □

The pullback argument gives Sullivan's unpublished proof of the theorem below that was also proved by Douady and Hubbard (1982) and Thurston, see Milnor (1983).

Theorem 4.2. *Let $f_a\colon [-1,1] \to [-1,1]$ be the quadratic family $f_a(x) = 1 - ax^2$. Then the mapping $a \to \nu(a)$, where $\nu(a)$ is the kneading invariant of f_a, is monotone. So $b > a$ implies that $\nu(b) \succeq \nu(a)$ where $\succeq$ is the Milnor-Thurston's order relation on the set of itineraries.*

Proof. Because of results from Section II.10 it is enough to prove that if a_1, a_2 are such that f_{a_1} and f_{a_2} are critically finite maps (this means that the critical orbit is a finite set, say with cardinality N) with the same kneading invariant then $a_1 = a_2$. So suppose, by contradiction, that $a_1 \neq a_2$ and f_{a_1}, f_{a_2} are combinatorially equivalent maps as above. Since $\nu(a_1) = \nu(a_2)$, the orbits of the critical points of f_{a_1} and f_{a_2} are finite and have the same ordering in $\mathbb{R}$. Therefore there exists a C^∞-diffeomorphisms $h_0\colon \overline{\mathbb{C}} \to \overline{\mathbb{C}}$ and a big neighbourhood U of $[-1,1]$ in $\mathbb{C}$ such that $f_{a_1}(\mathbb{C} \setminus U) \subset \mathbb{C} \setminus U$ and

$$\begin{aligned} &h_0(\bar{z}) = \overline{h_0(z)}, \\ &h_0(f_{a_1}^n(c)) = f_{a_2}^n(c) \quad \text{for all } n = 0, 1, \ldots, N, \\ &f_{a_2} h_0(z) = h_0 f_{a_1}(z) \quad \text{for } z \in \mathbb{C} \setminus U. \end{aligned}$$

Here $c = 0$ is the critical point of f_a. Since h_0 is K-quasiconformal for some K, as in the proof of Theorem 4.1, we can construct a sequence $h_n\colon \overline{\mathbb{C}} \to \overline{\mathbb{C}}$ of

K-quasiconformal homeomorphisms so that

$$f_{a_2}h_{n+1} = h_n f_{a_1},$$
$$h_{n+1} = h_n \text{ on } \overline{\mathbb{C}} \setminus f^{-n}(U),$$
$$h_{n+1}(\overline{z}) = \overline{h_{n+1}(z)}.$$

As before, h_n converges uniformly to a K-quasiconformal map h_∞ which is a conjugacy between f_{a_1} and f_{a_2}. Let $\beta(z)$ be the Beltrami coefficient of h_∞. Since h_∞ is a conjugacy between f_{a_1} and f_{a_2} it follows that the field of ellipses defined by β is invariant under f_{a_1} and also by complex conjugation. In other words, β satisfies the invariance condition (4.1). Then $\beta_u = u \cdot \beta$ with $u \in \mathbb{C}$, $|u| \leq 1$ also satisfies the invariance condition (4.1). From the Measurable Riemann Mapping Theorem there exists a unique quasiconformal homeomorphism $H_u \colon \overline{\mathbb{C}} \to \overline{\mathbb{C}}$ whose Beltrami coefficient is β_u and such that $H_u(\infty) = \infty$, $H_u(0) = 0, H_u(1) = 1$. From this theorem, see Ahlfors and Bers (1960), it also follows that $u \mapsto H_u(z)$ is analytic for each $z \in \overline{\mathbb{C}}$. As f_{a_1} preserves the field of ellipses defined by β_1 and f_{a_1} is conformal, f_{a_1} also preserves the field ellipses defined by β_u. From Lemma 4.1 it follows that

$$g_u = H_u \circ f_{a_1} \circ H_u^{-1}$$

is a conformal map. Since f_{a_1} is quadratic, g_u is also a quadratic polynomial. Therefore, because g_u also has its branch point in 0 and $g_0(0) = 1$, one gets

$$g_u(z) = 1 - \varphi(u)z^2$$

where $u \mapsto \varphi(u)$ is a holomorphic function. Since $\varphi(0) = a_1$ and $\varphi(1) = a_2$ it follows that the image of φ contains a neighbourhood of a_1 in the complex plane. In particular, every quadratic polynomial $f_w(z) = 1 - wz^2$ for w in this neighbourhood is critically finite because it is conjugate to f_{a_1}. But since $f_w(z)$ depends analytically on w it follows that for every w the map $f_w(z) = 1 - wz^2$ is critically finite. This is clearly not the case. □

Theorem 4.3. *Let $f_a \colon [-1, 1] \to [-1, 1]$ be the family of quadratic polynomials $f_a(x) = 1 - ax^2$. If f_{a_1} is infinitely renormalizable of bounded combinatorial type (or if f_{a_1} is a Misiurewicz map) then there is no $a \neq a_1$ such that f_a is conjugate to f_{a_1}.*

Proof. From the monotonicity proved in Theorem 4.2a, the set of parameters a such that f_a is conjugate to f_{a_1} is an interval. If it is not a single point then it is a closed interval $[b, c]$. Let us consider f_b and f_c as polynomial maps of the Riemann sphere. If f_{a_1} is infinitely renormalizable then, as in the proof of Theorem 4.1, we can construct a quasiconformal homeomorphism $h_0 \colon \overline{\mathbb{C}} \to \overline{\mathbb{C}}$ and a big neighbourhood U of $[-1, 1]$ in $\mathbb{C}$ such that $f_a(\mathbb{C} \setminus U) \subset \mathbb{C} \setminus U$ and

$$h_0(\overline{z}) = \overline{h_0(z)},$$
$$h_0(f_b^n(0)) = f_c^n(0) \text{ for all } n \in \mathbb{N},$$
$$h_0 f_b = f_c h_0 \text{ on } \mathbb{C} \setminus U.$$

If f_{a_1} is a Misiurewicz map then the same holds using Exercise III.6.1. So starting from h_0 we construct, as before, a sequence of quasiconformal maps h_n converging to a quasiconformal conjugacy $h\colon \overline{\mathbb{C}} \to \overline{\mathbb{C}}$ between f_b and f_c. As before h satisfies (4.1). As in the previous theorem we consider the family $H_u\colon \overline{\mathbb{C}} \to \overline{\mathbb{C}}$ of quasiconformal homeomorphisms, fixing ∞, 0 and 1, whose Beltrami coefficient is $u \cdot \mu(z)$ where $\mu(z)$ is the Beltrami coefficient of h. That these maps exist follows from the Measurable Riemann Mapping Theorem and this theorem also implies that $z \mapsto H_u(z)$ is analytic. By Lemma 4.1, $H_u \circ f_b \circ H_u^{-1}(z)$ is conformal and therefore is of the form

$$H_u \circ f_b \circ H_u^{-1}(z) = 1 - \varphi(u)z^2$$

where $\varphi(u)$ is a holomorphic function with $\varphi(0) = b$ and $\varphi(1) = c$. In particular, the image of φ contains a neighbourhood of b in $\mathbb{C}$. This implies that all maps f_a for $a \in [b-\varepsilon,\ c]$, for $\varepsilon > 0$ small, are conjugate to f_{a_1}. This contradicts the definition of b. □

Corollary 4.1. *The maximal distortion of the quasiconformal map h from Theorem 4.1 is almost everywhere zero on $J(F)$.*

Proof. Let h be the quasiconformal map from Theorem 4.1 and suppose by contradiction that the maximal conformal distortion of h is positive on a set of positive Lebesgue measure. Then the Beltrami coefficient $\mu(z)$ of h is non-zero in an invariant set of positive Lebesgue measure $S \subset J(F)$ and hence defines an F-invariant line-field on S. Now it is merely a conjecture that $J(F)$ has Lebesgue measure zero if it has non-empty interior. So in order to get a contradiction we again use the so-called Straightening Theorem, see Theorem 5.10 of the Appendix. This theorem states that every quadratic-like map F is quasiconformally conjugate by a map $\psi\colon\ V \to W$ to a quadratic polynomial $f_a(z) = 1 - az^2$, see the Appendix. Since F is symmetric the parameter a is real. Here W is a neighbourhood of the Julia set of f_a. But since $J(F)$ has positive Lebesgue measure and ψ is a quasiconformal conjugacy, the Julia set $J(f_a)$ of f_a also has positive measure and has an f_a-invariant line field defined on an invariant subset $E \subset J(f_a)$ of positive Lebesgue measure. Let $\mu(z) = 0$ for $z \notin E$ and for $z \in E$ let μ be so that $|\mu(z)| = 1/2$ and $\mu(z)$ points in the direction of the linefield through z. Since the quasiconformal structure defined by $u\mu(z)$ is invariant by f_a we have as before that

$$H_u \circ f_a \circ H_u^{-1}(z) = 1 - \varphi(u)z^2,$$

where $H_u\colon\ \overline{\mathbb{C}} \to \overline{\mathbb{C}}$ is the quasiconformal homeomorphism fixing $0, 1, \infty$ whose Beltrami-coefficient is $u\mu(z)$ and $\varphi(u)$ is a holomorphic function. We claim that φ is not constant. Indeed, otherwise H_u would commute with f_a for all u since H_0 is the identity map. Therefore H_u is equal to the identity in the Julia set of f_a. Consider the set of points in the Julia set which are Lebesgue density points of $J(f_a)$ as well as in the set of points where H_u is differentiable. This set has full Lebesgue measure (as a subset of the Julia set $J(f_a)$). Let us prove that the derivative of H_u is the identity at each point x of this set. Take two disjoint cones based at x. Since x is a density point we can find sequences $z_n \to x$ $w_n \to x$ which belong to the Julia set and such that the line through z_n and x converges to a line in one of the cones while the line through w_n and x converges to a line in the other cone. Hence the derivative of H_u at x, which by assumption exists, is equal to the identity in two linearly independent directions. Thus it is equal to the identity and we have shown that the derivative of H_u is equal to the identity map in almost every point of $J(f_a)$. On the other hand, at almost all points of $J(f_a)$ the derivative of H_u maps an ellipse defined by μ (and therefore with eccentricity away from 1) into a circle. This is clearly impossible because as we have seen in almost all points of $J(f_a)$ the derivative of H_u is the identity map. So the claim is proved. It follows that the image of φ contains a neighbourhood of $\varphi(0) = 0$ and this contradicts Theorem 4.2b. □

Remark. 1. It is not known whether $J(F)$ has Lebesgue measure zero. But if not then at least we know, by the proof above, there is no measurable invariant line field on $J(f)$.

2. Let f_a be the quadratic family. Świątek (1992a) shows that if f_a and f_b are conjugate then they are quasisymmetrically conjugate. This and the previous result implies the following famous conjecture: the set of real parameters for which f_a has a periodic attractor is dense.

Theorem 4.4. *Let $F\colon U \to V$ and $\tilde{F}\colon \tilde{U} \to \tilde{V}$ be infinitely renormalizable symmetric quadratic-like maps with the same bounded combinatorial type. Let $k > 1$. Suppose there exists a quasiconformal homeomorphism $\phi\colon U_1 \to \tilde{U}_1$ with the following properties:*

1. *U_1 is a neighbourhood of the Julia set $J(F)$ and $F(U_1)$ contains the closure of U_1, i.e., $F|U_1$ is quadratic-like and similarly for $\tilde{U}_1$ and $\tilde{F}$.*
2. *ϕ conjugates F and $\tilde{F}$ in the boundary of U_1;*
3. *the conformal distortion of the restriction of ϕ to the fundamental neighbourhood $F(U_1) \setminus U_1$ is bounded by k.*

Then there exists a quasiconformal conjugacy between $F|U_1$ and $\tilde{F}|\tilde{U}_1$ whose conformal distortion is bounded by k.

Proof. Let h be the quasiconformal conjugacy constructed in Theorem 4.1, starting with a quasiconformal homeomorphism h_0 that coincides with ϕ in the fundamental neighbourhood $F(U_1) \setminus U_1$. As we have remarked after the proof of Theorem 4.1, the conformal distortion of h restricted to the complement of the Julia set is bounded by k. On the other hand, by the Corollary of Theorem 4.2b, the conformal distortion of h on the Julia set of F is almost everywhere equal to zero. This proves the theorem. □

4.3. The definition of the renormalization operator on the space of germs of quadratic-like maps

Of course, if $F\colon U \to V$ is a quadratic-like map then its iterate is not even defined. So in order to define the renormalization operator on the space $\mathcal{G}$ of equivalence classes of symmetric quadratic-like maps we need the result below.

Proposition 4.1. *Let $F\colon\ U \to V$ be a symmetric quadratic-like map which is k times renormalizable of type $\sigma = (\sigma_0, \dots, \sigma_{k-1})$, with $|\sigma_0| \cdot \ldots |\sigma_{k-1}| = m$. Then there exist unique (topological) discs $U_\sigma \subset U$ and V_σ such that $F^m\colon\ U_\sigma \to V_\sigma$ is a symmetric quadratic-like map where U_σ contains the critical point of F and such that V_σ and V coincide outside the real axis.*

Remark. As we will see in the proof of this proposition the set V_σ is equal to V minus two intervals in the real axis if $m > 3$. So V_σ is a disc with two slits. It follows that the boundary of U_σ is the union of eight arcs. Four of these arcs are mapped injectively into $\partial V \cap \mathbb{D}_+$ or $\partial V \cap \mathbb{D}_-$ and the other four are mapped into the real axis. The dotted arcs are mapped into the slits in the real axis; the solid and dashed arcs are mapped into the boundary of the disc as indicated in Figure 4.2 (for $m = 2, 3$ there is just one slit).

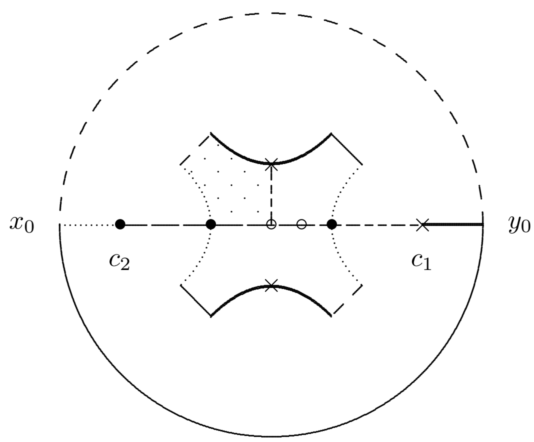

Fig. 4.2: The shaded region (and its symmetric with respect to the origin) is mapped to the upperhalf of the disc.

Proof. By conjugating with the Riemann mapping of V we may assume that $V = \mathbb{D}$ and the critical point of F is 0. Let us denote by F_s^{-1}, $s = \pm$, the two inverse branches of F defined on the disc with a slit $\mathbb{D} \setminus [F(0), 1)$ as in Figure 4.2. So F_- maps $\mathbb{D} \setminus [F(0), 1)$ onto the 'left-half' U_- of U and F_+ maps $\mathbb{D} \setminus [F(0), 1)$ onto the 'right half' U_+ of U. Let $\triangle^0 \ni 0$ be the interval bounded by $F^m(0)$ and $F^{2m}(0)$ and $\triangle^i = F^i(\triangle^0)$ for $0 < i \le m = |\sigma|$. The intervals $\triangle^i$, $i = 0, \dots, m-1$ are disjoint. For each $i = 1, \dots, m-1$ there exists a unique choice of $s(i) = \pm$ such that $F_{s(i)}^{-1}(\triangle^{m-i}) = \triangle^{m-i-1}$. The mappings

$$G_+ = F_+^{-1} \circ F_{s(m-2)}^{-1} \circ \cdots \circ F_{s(0)}^{-1}$$

$$G_- = F_-^{-1} \circ F_{s(m-2)}^{-1} \circ \cdots \circ F_{s(0)}^{-1}$$

restricted to $\mathbb{D}_+ = \mathbb{D} \cap$ (upper half space) and to $\mathbb{D}_- = \mathbb{D} \cap$ (lower half space) are univalent maps. Let $U_{++} = G_+(\mathbb{D}_+)$, $U_{+-} = G_+(\mathbb{D}_-)$, $U_{-+} = G_-(\mathbb{D}_+)$ and $U_{--} = G_-(\mathbb{D}_-)$. The restriction of F^m to U_{++} is a conformal isomorphism between U_{++} and $\mathbb{D}_+$. Since the mapping $z \to z' = F_+^{-1}F(z)$ is a conformal isomorphism between U_- and U_+, which is the identity on the common boundary of U_- and U_+ and since $F^m(\overline{z}) = \overline{F^m(z)}$ it follows that

$$U_{+-} = \{z; \overline{z'} \in U_{++}\}, U_{-+} = \{z; z' \in U_{++}\}$$

and $U_{--} = \{z; \overline{z} \in U_{++}\}$. Hence F^m is a conformal isomorphism between

$$U_{+-} \text{ and } \mathbb{D}_- \,,\ U_{-+} \text{ and } \mathbb{D}_+, \text{ and also between } U_{--} \text{ and } \mathbb{D}_-.$$

Now we take U_σ to be the interior of the closure of the union of U_{++}, U_{+-}, U_{-+}, U_{--} and we define $V_\sigma = F^m(U_\sigma)$. Then $F^m \colon U_\sigma \to V_\sigma$ is a symmetric quadratic-like map and the proposition is proved. □

The shape of the regions U_σ and V_σ will play an important role in the next section. Therefore we will now study these regions in an example.

Example. Take $m = 5$ and the combinatorial type of the map $f\colon\ [-1,1] \to [-1,1]$ associated to F as in Figure 4.3. It is easy to check that

$$G_+ = F_+^{-1} \circ F_-^{-1} \circ F_+^{-1} \circ F_+^{-1} \circ F_-^{-1},$$

$$G_- = F_-^{-1} \circ F_-^{-1} \circ F_+^{-1} \circ F_+^{-1} \circ F_-^{-1}.$$

The image of $\mathbb{D}_+$ under these compositions of the univalent maps is described in Figure 4.4 where we use the following notation: x_{-n} denotes one point such that $F^n(x_{-n}) = x$. Hence $F^5 \mid \tilde{U}_\sigma$ is as in Figure 4.5.

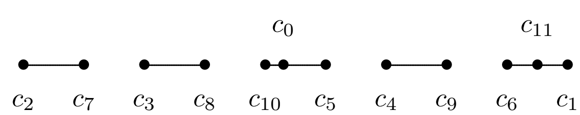

Fig. 4.3: The relative position of the first 11 iterates of c.

Definition. If $[F] \in \mathcal{G}$ is renormalizable of combinatorial type σ, with $|\sigma| = m$, then the renormalization of $[F]$, which we denote by $\mathcal{R}([F])$, is the equivalence class of

$$F^m \colon\ U_\sigma \to V_\sigma$$

where U_σ and V_σ are as in the above proposition.

The renormalization operator $\mathcal{R}$ maps $\mathcal{G}_{\sigma_0,\sigma_1,\ldots}$ into $\mathcal{G}_{\sigma_1,\sigma_2,\ldots}$. Now we want to define a positive function on pairs of elements of $\mathcal{G}$ that will play a crucial role in the theory. In the next subsection we will prove that this function is a distance function and at the end of the chapter we will prove that it is contracted under iteration of the renormalization operator.

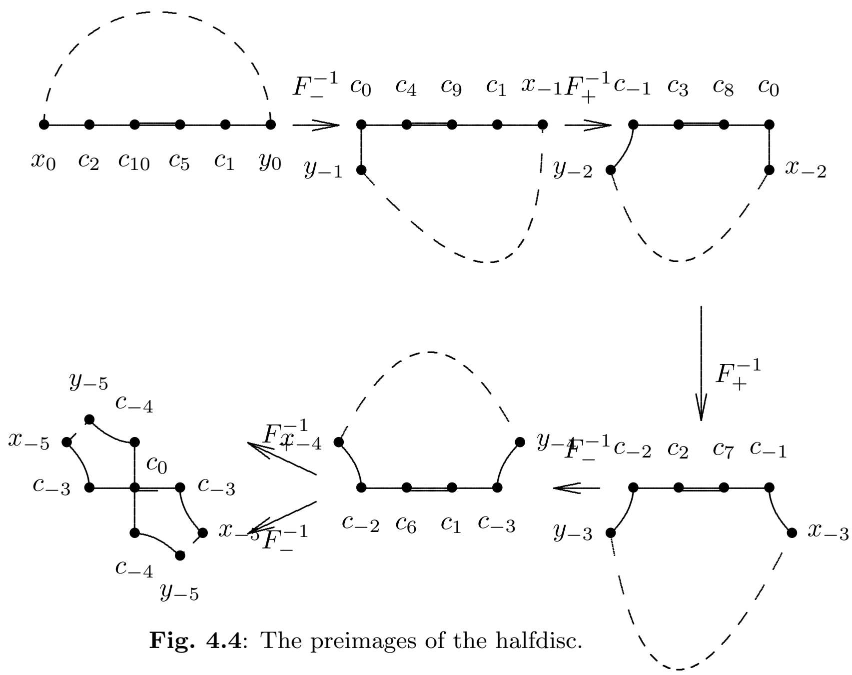

Fig. 4.4: The preimages of the halfdisc.

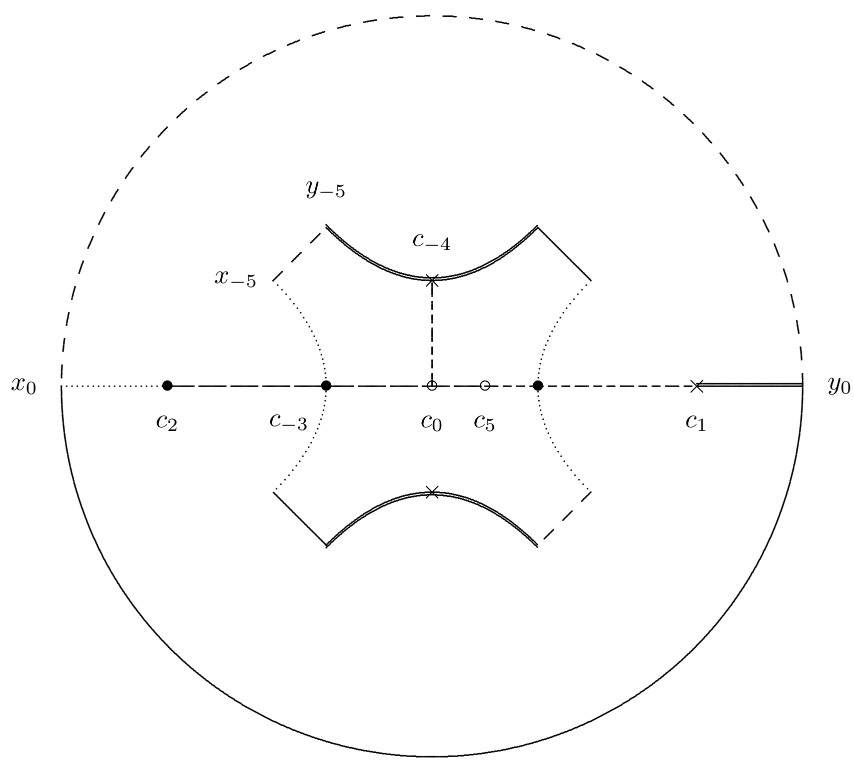

Fig. 4.5: The disc is slit in the arc connecting x_0 and c_2 and in the arc connecting y_0 and c_1. Each curve is marked in the same way as its image.

Definition. Let $[F]$ and $[\tilde{F}]$ be germs of maps as above and let $d_{JT}([F],[\tilde{F}]) = \log K$ where K is the smallest number for which there exists a $(K+\epsilon)$-quasiconformal conjugacy between two representatives F and $\tilde{F}$ of the germs $[F]$ and $[\tilde{F}]$ for every positive ϵ. This number is called the *Julia-Teichmüller* distance between the two elements of $\mathcal{G}$. It is clear that this function satisfies the triangle inequality. It is not completely clear however that it defines a metric on $\mathcal{G}$. Indeed, if the Julia-Teichmüller distance between the germs of F and $\tilde{F}$ is zero then for every ϵ we can find a neighbourhood U_ϵ of the Julia set of F and a conjugacy between F and $\tilde{F}$ on this neighbourhood that has conformal distortion smaller than $1+\epsilon$. However, as ϵ goes to zero, the domain of the conjugacy may shrink down to the Julia set and we may not get a conformal conjugacy in some neighbourhood. In the next subsection we will prove that d_{JT} is indeed a distance function.

Proposition 4.2. *If F and $\tilde{F}$ are infinitely renormalizable quadratic-like maps of the same bounded combinatorial type then*

$$d_{JT}(\mathcal{R}([F]),\mathcal{R}([\tilde{F}])) \le d_{JT}([F],[\tilde{F}]).$$

Proof. This is clear because the renormalization operator is just the restriction of some iterate to a smaller neighbourhood and the restriction of any conjugacy to this smaller neighbourhood is a conjugacy between the corresponding iterates. □

The next theorem is a very important consequence of the pullback argument.

Theorem 4.5. *Let $B > 0$. Let $\mathcal{G}^B$ be the set of all germs in $\mathcal{G}$ that have a representative whose conformal type is bounded by B. Then there exists $D > 0$, that depends only on B, such that if F and $\tilde{F}$ are infinitely renormalizable quadratic-like map of bounded combinatorial type whose germs belong to $\mathcal{G}^B$ then*

$$d_{JT}([F],[\tilde{F}]) \le D.$$

Proof. The ingredients of the proof are Theorem 4.2c and Koebe's Distortion Theorem for univalent functions which states that given $\epsilon > 0$, there exists a constant $K = K(\epsilon)$ such that if $H\colon \mathbb{D} \to \mathbb{C}$ is a univalent holomorphic function then the restriction of H to the disc $\mathbb{D}_{1-\epsilon}$, with centre at the origin and radius $1-\epsilon$, is bounded by K, see the Appendix.
We split the proof into several lemmas.

Lemma 4.2. *Given $a > 0$ there exists $b = b(a)$ such that if $H\colon A_a \to \mathbb{C}$ is a univalent function on the annulus $A_a = \{z \in \mathbb{C}; 1 < |z| < 1+a\}$ then the distortion of the restriction of H to the annulus $\frac{1}{3}A_a = \{z \in \mathbb{C};\ 1+\frac{1}{3}a < |z| < 1+\frac{2}{3}a\}$ is bounded by b.*

Proof. We can cover the inner annulus by balls of radius $\frac{2}{5}a$ such that the concentric ball of radius $\frac{1}{2}a$ is contained in A_a and the number of such balls depends only on a. Using the Koebe's distortion lemma in each of these balls we finish the proof. □

Lemma 4.3. *Given $m > 0$ there exists $\epsilon = \epsilon(m)$ with the following property. If $A \subset \mathbb{D}$ is an annulus of modulus at least m such that one of the boundaries of A is equal to the unit circle and such that A does not contain the origin then A contains the boundary of the disc $\mathbb{D}_{1-\epsilon}$.*

Proof. Suppose this is not the case. This means that there exists one such annulus such that the other boundary has a point p very close to the unit circle. On the other hand, there exists a holomorphic homeomorphism $H\colon A_a \to A$ where $a = a(m)$. It follows that there exists a curve connecting the boundaries of A_a whose image has a very small length. Hence, there exists a point in the annulus $\frac{1}{3}A_a$ such that the derivative of H at this point is very small in norm. Then the derivative of H is very small on all points of $\frac{1}{3}A$ because H has bounded distortion in this annulus. This is not possible because the image of the circle of radius $1 + \frac{1}{2}a$ and centre at the origin is a curve that encloses the origin and passes near a point in the boundary of $\mathbb{D}$ and, therefore, cannot have a small length. $\square$

Let $F\colon U \to V$ be a quadratic-like map. Then there exist univalent functions $\phi\colon \mathbb{D} \to U$, $\psi\colon \mathbb{D} \to V$ such that $F = \psi \circ Q \circ \phi^{-1}$ where $Q(z) = z^2$.

Lemma 4.4. *Given $B > 0$ there exist $\delta = \delta(B)$ and $M = M(B)$ satisfying the following property. Let $\mathcal{S}$ be the set of pairs (ϕ, ψ) of univalent functions on the unit disc $\mathbb{D}$ such that $F = \psi \circ Q \circ \phi^{-1}$ is a quadratic-like map of conformal modulus bounded by B and the dynamical interval of F is $[0, 1]$. Then the following holds for all $(\phi, \psi) \in \mathcal{S}$:*

1. *$\phi(\mathbb{D}_{1-\delta})$ contains the Julia set of $F = \psi \circ Q \circ \phi^{-1}$;*
2. *the restriction of F to $\phi(\mathbb{D}_{1-\delta})$ is quadratic-like;*
3. *$\psi(\mathbb{D}_{1-\delta})$ contains the closure of $F(\phi(\mathbb{D}_{1-\delta}))$ and is contained in the disc $\mathbb{D}_M$.*

Proof. Let $(\phi, \psi) \in \mathcal{S}$ and let $U = \phi(\mathbb{D})$, $V = \psi(\mathbb{D})$. By hypothesis, V contains the closure of U and the annulus $V \setminus \bar{U}$ has at least modulus $\frac{1}{B}$. Hence, the modulus of the annulus $U \setminus F^{-1}(U)$, which is equal to one half of the modulus of $V \setminus U$, is at least $\frac{1}{2B}$. If $\delta = \epsilon(\frac{1}{2B})$, where ϵ is as in Lemma 4.3, Statements 1 and 2 are satisfied. The first inclusion in Statement 3 is obvious because the closure of $Q(\mathbb{D}_{1-\epsilon})$ is contained in $\mathbb{D}_{1-\epsilon}$. To prove the second inclusion we will use Koebe's Distortion Theorem together with Koebe's $\frac{1}{4}$-Theorem. This last theorem states that the image of a univalent function on the unit disc whose derivative at the origin has norm one contains a disc of radius $\frac{1}{4}$ centered at the image of the origin, see the Appendix. Let $R > 0$ be such that the modulus of the annulus $\mathbb{D}_R \setminus [0, 1]$ is at least B. Let $m > 0$ be such that $\frac{m(1-\epsilon)}{4} > R + 1$. From Koebe's $\frac{1}{4}$-Theorem, it follows that if ψ is univalent on the disc $\mathbb{D}_{1-\epsilon}$, maps the origin to a point in the interval $(0, 1)$ and its derivative at the origin has norm larger than m then $\psi(\mathbb{D}_{1-\epsilon})$ contains the disc $\mathbb{D}_R$. Since ψ is univalent in the unit disc, it follows from Koebe's Distortion Theorem that its distortion on $\mathbb{D}_{1-\epsilon}$ is bounded by $b = b(\epsilon)$. Let $M = mb$. From the Mean Value Theorem we have that if $\psi(\mathbb{D}_{1-\epsilon})$ is not contained in $\mathbb{D}_M$ then the derivative of ψ at some point of $\mathbb{D}_{1-\epsilon}$ has norm at least M. Hence, since the distortion is bounded by b, the norm of the derivative of ψ at zero is at least m. This is a contradiction and proves the lemma. ☐

Lemma 4.5. *Let $\mathcal{S} = \mathcal{S}(N, B)$ be the set of pairs (ϕ, ψ) of univalent functions on the unit disc with the following properties:*

i) *the closure of $\phi(\mathbb{D})$ is contained in $\psi(\mathbb{D})$;*

ii) *the mapping $\psi \circ Q \circ \phi^{-1}$ (where $Q(z) = z^2$) is a symmetric quadratic-like map with dynamical interval $[0, 1]$, which is infinitely renormalizable of combinatorial type bounded by N and conformal type bounded by B.*

Then, any sequence $(\phi_n, \psi_n) \in \mathcal{S}$ has a subsequence that converges uniformly in compact subsets and any limit point of a convergent subsequence belongs to $\mathcal{S}(N, B - \beta)$ for any $\beta > 0$.

Proof. The lemma follows from Lemma 4.4 because any uniformly bounded sequence of holomorphic maps has a subsequence that converges uniformly on compact subsets; the uniform limit of a sequence of univalent mappings is either univalent or constant. Hence it must be univalent. ☐

Proof of Theorem 4.2d: The theorem follows easily from the above lemmas and from Theorem 4.2c. ☐

4.4. Quadratic-like maps and expanding maps of the circle

Let us finish this section by quoting a result due to Douady and Hubbard which relates quadratic-like maps to degree two maps of the circle. Using this result we will prove that the Julia-Teichmüller metric defined above is really a metric in each set of $\mathcal{G}_{\sigma_0,\sigma_1,\ldots}$ of germs of infinitely renormalizable quadratic-like maps of a given bounded combinatorial type. We will also prove that the Julia-Teichmüller distance between two infinitely renormalizable quadratic-like maps of the same bounded combinatorial type is small if the corresponding degree two expanding circle maps satisfy certain conditions. This will play an important role in the proof that $d_{JT}(\mathcal{R}^n([F]), \mathcal{R}^n([\tilde{F}]))$ converges to zero as $n \to \infty$ in Section 8.

As we have observed before, if $F \colon U \to F(U)$ is an infinitely renormalizable quadratic-like map, then its Julia set $J(F)$ is connected (and also has empty interior). Therefore, $F(U) \setminus J(F)$ is homeomorphic to an annulus. By the Riemann Mapping Theorem there exist a positive number $a > 0$ and a holomorphic diffeomorphism $\phi\colon\ F(U) \setminus J(F) \to A_a$, where A_a is the annulus $\{z \in \mathbb{C};\ 1 < |z| < 1 + a\}$. Let $A^+ = \phi(U \setminus J(F))$ and $G\colon\ A^+ \to A_a$ be the holomorphic map $G(z) = \phi(F(\phi^{-1}(z)))$. By the Schwarz Reflection Principle, see the Appendix, we can extend G to a holomorphic map of a cylindrical neighbourhood A of the unit circle by symmetry, i.e., $G(\frac{1}{\bar{z}}) = 1/\overline{G(z)}$ if $z \in A^+$. Since points near the unit circle are mapped by G near the unit circle, the Schwarz Reflection Principle implies that G extends continuously to the circle and hence to a holomorphic map in the neighbourhood A of the unit circle. Thus the unit circle is invariant under G. Since $G^{-1}(A)$ is contained in A it follows from the Lemma of Schwarz that $G\colon G^{-1}(A) \to A$ expands the Poincaré metric on A. Since the restriction of the Poincaré metric on A to the unit circle and the standard metric on the circle are equivalent, the restriction of G to the unit circle is an expanding degree two map f which is called the *external map* of F. Therefore we have proved the following result.

Theorem 4.6. **(Douady and Hubbard)** *Let $F\colon U \to F(U)$ be a quadratic-like map whose Julia set $J(F)$ is connected. Then there exists an expanding degree two analytic circle map $f\colon S^1 \to S^1$ and a holomorphic conjugacy between $F\colon U \setminus J(F) \to F(U) \setminus J(F)$ and a holomorphic extension $f\colon A^+ \to f(A^+)$ of f. Here A^+ is a component of $A \setminus S^1$ and A is a neighbourhood of $S^1 = \{z \in \mathbb{C};\ |z| = 1\}$.*

Let us first state the main application of this theorem.

Corollary 4.2. *Let $F, \tilde{F}$ be symmetric quadratic-like maps of the same bounded combinatorial type and let $f, \tilde{f}$ be the corresponding expanding analytic circle maps. Then the following three statements are equivalent:*

a) $d_{JT}([F], [\tilde{F}]) = 0$*;*

b) *f and $\tilde{f}$ are analytically conjugate;*

c) *$F, \tilde{F}$ are equivalent, i.e., $[F] = [\tilde{F}]$.*

Proof of the Corollary: Notice that if there exists a conjugacy h between F and $\tilde{F}$ with conformal distortion $1+\epsilon$ defined in some neighbourhood of the Julia set, then there exists a quasiconformal conjugacy between the holomorphic extensions G and $\tilde{G}$ of the corresponding external maps with the same conformal distortion. Indeed, let ϕ (resp. $\tilde{\phi}$) be the holomorphic conjugacy between F and G (resp. $\tilde{F}$ and $\tilde{G}$) in the complement of the Julia set. Since the external maps f and g are expanding maps of the circle they are quasisymmetrically conjugate (this can be proved using the Naive Distortion Lemma for $C^{1+\alpha}$ expanding maps, see Exercise II.2.3), and this quasisymmetric conjugacy can be extended, by a theorem of Ahlfors and Beurling, to a quasiconformal homeomorphism of the plane. By gluing this homeomorphism with $\tilde{\phi} \circ h \circ \phi^{-1}$ we get a quasiconformal homeomorphism h_0 that has quasiconformal distortion $1+\epsilon$ at a fundamental domain of G, conjugates G and $\tilde{G}$ at the boundary of this fundamental domain (because it coincides with $\tilde{\phi} \circ h \circ \phi^{-1}$ in a neighbourhood of this fundamental domain) and conjugates G and $\tilde{G}$ in the circle. Hence, from the pullback argument we get a conjugacy between G and $\tilde{G}$,on a neighbourhood of the circle, with the same conformal distortion. By the same argument the converse is true: if there is a quasiconformal conjugacy h between G and $\tilde{G}$ on some neighbourhood of the circle with conformal distortion $1+\epsilon$, the same happens between F and $\tilde{F}$. If a) holds, then for every ϵ there exists a quasiconformal conjugacy h between G and $\tilde{G}$ with conformal distortion $1+\epsilon$. Since a quasiconformal homeomorphism with conformal distortion close to one is Hölder continuous with exponent close to one, the eigenvalues of the external maps f and $\tilde{f}$ at corresponding periodic points are equal. Since these maps are real analytic, if follows from Shub and Sullivan (1986) that they are analytically conjugate. This proves that a) implies b). If there exists an analytic conjugacy between f and $\tilde{f}$, the holomorphic extension of this map is a holomorphic conjugacy between G and $\tilde{G}$. Hence there exists a conformal conjugacy between F and $\tilde{F}$ in some neighbourhood of the Julia set (the above argument with $\epsilon = 0$). □

Corollary 4.3. *Let $F\colon U \to V$ be a quadratic-like map such that $F(\overline{z}) = \overline{F(z)}$ and $F(-z) = F(z)$. Then there exists a neighbourhood V_1 of $J(F)$ such that V_1 is invariant by the reflections in the real and imaginary axis and such that $F\colon F^{-1}(V_1) \to V_1$ is a quadratic-like map.*

Proof. Let R_1 and R_2 be defined by $R_1(z) = \overline{z}$ and $R_2(z) = -z$. Then $F \circ R_1 = R_1 \circ F$ and $F \circ R_2 = F$. Hence $R_i(J(F)) = J(F)$. Take a neighbourhood $W \subset U$ of $J(F)$ such that $R_i(W) = W$. Let $W_1 \subset W$ be a neighbourhood of $J(F)$ such that $F(W_1) \subset W$. As before let ϕ be the Riemann mapping of $W \setminus J(F)$ to the annulus $A_a = \{z \in \mathbb{C};\ 1 < |z| < 1+z\}$. Let $G\colon \phi(W_1) \to A_a$ be defined by $G(z) = \phi \circ F \circ \phi^{-1}(z)$ and let $T_i = \phi \circ R_i \circ \phi^{-1}$. Since W is symmetric under R_i this defines a diffeomorphism of A_a. By the Schwartz Reflection Principle, one can extend T_i to

$$A = \{z \in \mathbb{C};\ z \in A_a\ ,\ \frac{1}{z} \in A_a \text{ or } |z| = 1\}.$$

Similarly we can extend G holomorphically to the symmetric

$$\{z \in \mathbb{C};\ z \in \phi(W_1)\ ,\ \frac{1}{z} \in \phi(W_1) \text{ or } |z| = 1\}$$

of $\phi(W_1)$. The maps T_i are conformal diffeomorphisms of A. Moreover, T_1, $T_2 \circ T_1$ are conformal reflections of A and $T_i \circ G = G \circ T_i$. We claim that T_1 and $T_2 \circ T_1$ extend to conformal reflections of the plane (with respect to lines through the origin). Indeed, let p be a fixed point of T_1 and let S be the reflection with respect to the line through p and the origin. Then $S(A) = A$ and $T_1 \circ S$ is a holomorphic automorphism of A with a fixed point. Hence $T_1 \circ S = id$ and $T_1 = S$. Similarly, $T_2 \circ T_1$ is a reflection with respect to some line through the origin. In particular, T_i are Euclidean isometries. Now take N so large that $F^{-N}(W_1) \subset W_1$. Then $G^N(\phi(F^{-N}(W_1))) \subset \phi(W_1)$ and therefore from the Lemma of Schwartz, G^N expands the Poincaré metric on $\phi(W_1)$. So let us define a new metric in a neighbourhood of the circle which is invariant under T_i such that this metric is expanded by the first iterate of G. Hence some iterate N' of G expands the Euclidean metric on a small neighbourhood of the circle. Let v be a vector in z, let $|v|_z$ be the Euclidean norm of v and let $||v||_z = \sum_{i=0}^{N'} |DG^i(z)v|_{G^i(z)}$. Then $||DG(z)v||_{G(z)} > ||v||_z$ for z sufficiently near the unit circle. Let d be the resulting metric. Since G preserves the unit circle the annulus $\hat{V}_1 = \{z\,;\ d(z, S^1) < \epsilon\}$ is invariant by T_i and the closure of $G^{-1}(\hat{V}_1)$ is contained in $\hat{V}_1$. Hence $\phi^{-1}(\hat{V}_1)$ has the required properties (by construction it is invariant under R_i). $\square$

Corollary 4.4. *If f is a real analytic map that has a quadratic-like extension F then there exists a map g that belongs to some Epstein class and is analytically conjugate to F. Moreover, g has an extension which is quadratic-like.*

Proof. Let $F\colon U \to V$ be a quadratic-like extension of f. Then one has $F(\overline{z}) = \overline{F(z)}$ and there exist holomorphic univalent symmetric maps Φ and Ψ such that $F = \Phi \circ Q \circ \Psi$. So let $U_1 = \Psi(F^{-1}(U))$ and $V_1 = \Psi(U)$ then $F_1\colon U_1 \to V_1$ defined by $F_1 = \Psi \circ F \circ \Psi^{-1}$ is a quadratic-like map conjugate to F and $F_1 = \Psi_1 \circ Q$ where $\Psi_1(z) = \Psi \circ \Phi$. Hence $F_1(-z) = F(z)$ and $F_1(\overline{z}) = \overline{F_1(z)}$. As we will see in Corollary 2 to Theorem 4.3, it follows that one can find a neighbourhood of the dynamical interval U_2 such that $V_2 = F_1(U_2)$ contains the closure of U_2 and such that V_2 is invariant under the involutions $z \mapsto -z$ and $z \mapsto \overline{z}$. By the Riemann Mapping Theorem, there exists a holomorphic diffeomorphism H from V_2 onto $\mathbb{C}_J = (\mathbb{C} \setminus \mathbb{R}) \cup J$ where J is an open interval that contains the dynamical interval of f (we may choose H so that it fixes the critical point and such that H' is positive in the fixed point). Since V_2 is invariant under the involutions, H is also invariant under these involutions. Let us show that the restriction of $G = H \circ F \circ H^{-1}$ to the dynamical interval of f belongs to an Epstein class. The closure of $U = G^{-1}(\mathbb{C}_J)$ is contained in $\mathbb{C}_J$. Let Q be the real quadratic map which sends c_0 to $G(c_0)$. Let $U_{\pm,\pm}$ be the components of U minus the horizontal and vertical lines through c_0. Since both G and Q are symmetric with respect to both involutions, it follows that G maps the upper-right component $U_{+,+}$ diffeomorphically onto the intersection of $\mathbb{C}_J$ with the lower halfplane. Using the fact that this map is invariant by the involutions, it follows that there exists a unique holomorphic diffeomorphism Φ from $Q(U)$ to $\mathbb{C}_J$ that maps the lower halfplane to the lower halfplane and satisfies $G = \Phi \circ Q$. Hence the restriction of G to the dynamical interval belongs to an Epstein class. $\square$

5 The Complex Bounds

In this section we will establish the basic connection between real and complex dynamics by proving the following theorem which tells us that an important collection of real analytic infinitely renormalizable maps, after a number of renormalizations, can be extended to quadratic-like maps of bounded conformal type. Of course this result is especially useful because it enables us to use the pullback technique of the previous section and we will conclude that these renormalized maps are contained in some compact set of maps. So the Main Result of this section is the following theorem.

Theorem 5.1. *For each $N > 0$ there exists $L = L(N)$ such that if $f \in \mathcal{U}^\omega$ is an infinitely renormalizable map of combinatorial type bounded by N that either belongs to some Epstein class or has a quadratic-like extension, then there exists*

$n_0 = n_0(f)$ such that for $n \geq n_0$, the mapping $\mathcal{R}^n(f)$ extends to a symmetric quadratic-like map whose conformal type is bounded by L.

If f is quadratic-like then we know already by Proposition 4.1 that $\mathcal{R}^n(f)$ is also quadratic-like. However, now we shall also give bounds for the conformal moduli. One of the main ingredients of the proof are the real bounds from Sections 2 and 3. We should emphasize that it is enough to prove Theorem 5.1 for maps in some Epstein class, see Corollary 3 of Theorem 4.3.

Since Theorem 5.1 might also be true for real analytic maps, we first extend a real analytic map $f\colon [0,1] \to [0,1]$ to a small open set in the complex plane. Then we get the following:

Lemma 5.1. *For each $f \in \mathcal{U}^\omega$ there are neighbourhoods U and V of the dynamical interval $[0,1]$ in the complex plane such that f extends to a holomorphic map $F\colon U \to V$ with the following properties:*

1. *F is a proper degree two covering map of $U \setminus \{c\}$ onto $V \setminus \{F(c)\}$ and it is real, i.e., commutes with complex conjugation.*

2. *F factors as $F = \Phi \circ Q \circ \Psi$ where Q is the real quadratic map and Φ and Ψ are holomorphic univalent functions defined on neighbourhoods of $[0,1]$ and $Q(\Psi[0,1])$.*

Proof. Because f is real analytic one can extend f uniquely to a complex analytic map on a neighbourhood of $[0,1]$ in the complex plane. By taking this neighbourhood small enough we get the above properties since f has only one critical point in the origin and this is a quadratic one. □

The holomorphic map F of Lemma 5.1 is in general not a quadratic-like map because V need not contain the closure of U. Nevertheless, we will use the notation of Section 4, namely, we slit V at the intersection of $[F(0), +\infty) \cap V$ and denote by U_- and U_+ the components of $U \setminus F^{-1}([F(0), +\infty))$ as in Figure 5.1. We let $F_s^{-1}\colon V \setminus [F(0), \infty) \to U_s$, for $s = \pm$, be the holomorphic inverse branches of F.

Let us assume that f is infinitely renormalizable and, as in the previous sections, let $q(n)$ denote the renormalizing return times and let $\triangle_n$ be the interval $[f^{2q(n)}(c), f^{q(n)}(c)]$. This interval is $f^{q(n)}$-invariant and contains the critical point c. Moreover, let Ξ_n be the disjoint collection of intervals $\triangle_n^j = f^j(\triangle_n)$, $j = 0, 1, \ldots, q(n) - 1$. Hence the collection Ξ_n is invariant by f and the intersection over all n of the union of the intervals in Ξ_n is the attracting Cantor set Λ_f of f. The real bounds give that the sum of the lengths of the intervals in Ξ_n goes exponentially to zero as n goes to infinity if f has bounded combinatorics.

In order to show that some renormalized version of f is quadratic-like, we want to analyze the backward orbit of F along these intervals. The main aim is to find a small neighbourhood V of $\triangle_n$ whose $q(n)$-preimage U is mapped strictly into itself. In order to make this precise, we want to define which preimages we should take.

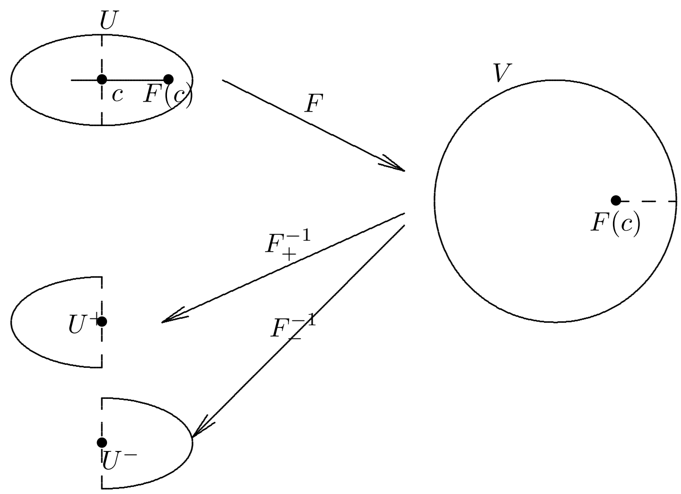

Fig. 5.1:

Definition. Let f be infinitely renormalizable and x_{-j} be a sequence defined by $x_0 = c$ and by $x_{-(j+1)}$ as the only point in the attracting Cantor set such that $f(x_{-(j+1)}) = x_{-j}$, i.e., x_{-j} is the only backward branch of the critical point that remains in the attracting Cantor set. (After all, f is injective on the attracting Cantor set.) The *sign map* is the map $s\colon \mathbb{N} \to \{\pm\}$ satisfying $F^{-1}_{s(j)}(x_{-j}) = x_{-(j+1)}$ for all $j \in \mathbb{N}$. The *basic backward compositions* are the maps

$$B_n = F^{-1}_{s(q(n)-2)} \circ \cdots \circ F^{-1}_{s(j)} \circ \cdots \circ F^{-1}_{s(0)}.$$

Let S_n be the maximal neighbourhood of $\triangle_n^1$ on which $f^{q(n)-1}$ is monotone and let $W_n = f^{q(n)-1}(S_n)$. Observe that the basic backward composition B_n maps $\triangle_n$ diffeomorphically onto $\triangle_n^1$ and, as we have already seen in Section 2, it is also defined and monotone on W_n. From Section 2 we also get that the size of each component of $W_n \setminus \triangle_n$ is bigger than a universal constant times the size of $\triangle_n$. In order to prove the Main Theorem of this section, we will have to understand the distortion properties of the basic backward compositions in neighbourhoods of the renormalizing interval $\triangle_n$ in the complex plane. For that we notice that each factor in a basic composition is a square root composed with a map which is univalent in a region which is very big compared to the sizes of the intervals under consideration. So we will analyze the distortion of these maps with several different arguments, using the information coming from the real bounds on the exponential decay in the length of the intervals in Ξ_n.

First we need some background on the hyperbolic metric on a slit region in the complex plane.

Definition. Let J be an interval of the real line and $\mathbb{C}_J$ be the set of all complex numbers except for the real numbers that are not in J, i.e.,

$$C_J = (\mathbb{C} \setminus \mathbb{R}) \bigcup J.$$

Let $\mathbb{C}_J$ be endowed with the hyperbolic metric and let ρ_J be the corresponding distance function. It is clear that J is a geodesic of $\mathbb{C}_J$ since complex conjugation, being conformal, is an isometry of the hyperbolic metric and fixes J. If $J = [-1, 1]$, the composition of the maps $z \mapsto z^2$ and $z \mapsto \frac{z-1}{z+1}$ is a holomorphic isomorphism between the half-plane $\{z \in \mathbb{C};\ Re(z) > 0\}$ and $\mathbb{C}_J$. Hence the set of points in $\mathbb{C}_J$ whose hyperbolic distance to J is at most k is the union of two Euclidean discs, symmetric to each other with respect to the real axis, whose boundaries intersect at the boundary of J (in the next exercise it is shown that these discs intersect the positive real axis at an angle θ such that $k = \log \tan(\frac{\pi}{2} - \frac{\theta}{4})$). We shall denote by

$$D_k(J) \text{ or } D(J; \theta)$$

any of these so-called Poincaré neighbourhoods of J.

Exercise 5.1. Prove these statements. Moreover, show that the Euclidean radius of each of the circles which contain $\partial D(J; \theta)$ is equal to $|J|/(2\sin(\theta))$. (Hint: $z \mapsto z^2$ sends $\mathbb{C}_+ = \{z \in \mathbb{C};\ Re(z) > 0\}$ homeomorphically onto $\mathbb{C} \setminus \{\text{positive real numbers}\}$ preserving the positive real axis and $z \mapsto \frac{z-1}{z+1}$ maps the set $\mathbb{C}_+$ homeomorphically onto $\mathbb{C}_J$ while mapping the positive real axis onto $J = [-1, 1]$. So $D_k(J)$ is the image under this composition of the set with distance k to the positive real axis in the hyperbolic metric of $\mathbb{C}_+$. Of course $\mathbb{C}_+$ is a rotated version of the usual upper half-plane and therefore the metric on this set is given by $(1/\mathrm{Re}(z))|dz|$. So in this right half-plane the geodesics are lines and circles perpendicular to the imaginary axis. So the boundary of the set with distance $\leq k$ to the positive real axis consists of two straight lines going through 0. Moreover, if the angle of such a line with the positive real axis is $\theta/2$ then the distance of this line with the real axis is equal to

$$\int_0^{\theta/2} \frac{1}{\cos(t)}\, dt = \int_0^{\theta} \frac{1}{2\cos(t/2)}\, dt = \log \tan(\frac{\pi}{2} - \frac{\theta}{4}).$$

Since the corresponding set in $\mathbb{C}_J$ is the disc $D(J; \theta)$ the first part follows. From Figure 5.2 the second statement also follows.)

This hyperbolic metric on $\mathbb{C}_J$ is very useful because 'square roots contract this metric'. More precisely, let T be a *square root*, i.e., $T = AF_1 \circ S \circ AF_2$ where

$$S \colon \mathbb{C} \setminus \{\text{negative real numbers}\} \to \mathbb{C}$$

is the standard square root map and AF_i are real affine maps. Suppose T maps J onto J diffeomorphically (in particular, its singularity is in $\mathbb{R} \setminus J$) then T is a holomorphic map of $\mathbb{C}_J$ into itself. Therefore, by the Lemma of Schwarz (which can be found in the Appendix), it contracts the Poincaré metric. Since T preserves J this implies that

$$T(D_k(J)) \subset D_k(J) \text{ for all } k > 0.$$

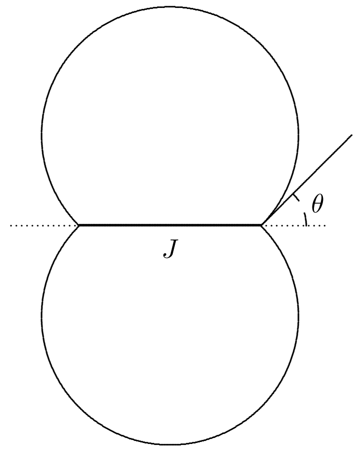

Fig. 5.2: The set of points whose hyperbolic distance to J is equal to k.

Similarly, if T is as above but maps J diffeomorphically to $T(J)$ then

$$T(D_k(J)) \subset D_k(T(J)) \text{ for all } k > 0.$$

In the next lemma we will show that something similar holds for univalent holomorphic maps which are just defined near J and not on all of $\mathbb{C}_J$. This lemma will only be used in Lemma 5.3 below. We should point out that Lemma 5.3 will only be used for maps in some Epstein class in which case the proof is immediate. However, because Theorem 5.1 might also be true for any real analytic map we have included a proof of Lemma 5.3 in this more general setting. Given an interval $J \subset \mathbb{R}$, let D_J be the Poincaré neighbourhood of J and ρ_J the Poincaré metric on $\mathbb{C}_J$ as above.

Lemma 5.2. *Given $a > 0$ and $r_0 > 0$ there exist $K = K(r_0, a)$ and $l_0 > 0$ with the following property. If Φ satisfies the conditions:*

1. *Φ is holomorphic and univalent on a Euclidean disc of radius a centered at a point of an interval $J \subset \mathbb{R}$ and $D_{r_0}(J)$ is contained in this Euclidean disc;*
2. *Φ maps the real axis into the real axis;*
3. $|J| \leq l_0$.

Then, provided $k \leq r_0$,

$$\rho_{\Phi(J)}(\Phi(x), \Phi(y)) \leq (1 + K|J|)\rho_J(x, y)$$

for all $x, y \in D_k(J)$; so in particular $\Phi(D_k(J)) \subset D_{(1+K|J|)k}(\Phi(J))$.

Proof. Without loss of generality we may assume that $\Phi(J) = J$. Let $B(x_0; a)$ be the Euclidean disc of radius a centered at a point $x_0 \in J$ which contains $D_{r_0}(J)$ as above. Let us take x_0 to be the middle point of J. Furthermore, let

ρ_J be the Poincaré metric of $\mathbb{C}_J$ and let $\tilde{\rho}_J$ be the Poincaré metric of this set intersected with $B(x_0;a)$, i.e., on $(B(x_0;a)\setminus\mathbb{R})\cup J$.

We claim that there exist $l_0>0$ and $K>0$ such that if $x,y\in D_{r_0}(J)$ and $|J|<l_0$ then $\tilde{\rho}_J(x,y)<(1+K|J|)\rho_J(x,y)$.

Before proving this claim let us show that it implies the lemma. Let $\overline{\rho}$ be the Poincaré metric on $(\Phi(B(x_0;a))\setminus\mathbb{R})\cup\Phi(J)$. Since Φ is univalent we have that $\overline{\rho}(\Phi(x),\Phi(y))=\tilde{\rho}_J(x,y)$. By the Lemma of Schwarz, $\rho_{\Phi(J)}(\Phi(x),\Phi(y))\leq\overline{\rho}(\Phi(x),\Phi(y))$. So the claim implies

$$\rho_{\Phi(J)}(\Phi(x),\Phi(y))\leq\overline{\rho}(\Phi(x),\Phi(y))=\tilde{\rho}_J(x,y)<(1+K|J|)\rho_J(x,y)$$

if $x,y\in D_{r_0}(J)$.

So it remains to prove the claim. Let $\alpha_0=\alpha_0(J)$ be the smallest angle such that $D(J;\alpha_0)\subset B(x_0;a)$. Notice that $\alpha_0(J)\to 0$ as $|J|\to 0$ and, for $|J|\leq l_0$ we have $\alpha_0(J)\leq K_0|J|$ for some constant $K_0>0$ depending on l_0. This is true because, if $|J|$ is small enough, $D(J;\alpha)\subset B(x_0;a)$ whenever $(2|J|)/(2\sin\alpha)<a$ (here $\dfrac{|J|/2}{\sin\alpha}$ is the radius of the Euclidean circle which contain $\partial D(J;\alpha)$, see the exercise above). Because α_0 is the smallest α for which this is true this gives $\frac{1}{2}\alpha_0<\sin(\alpha_0)\leq\frac{|J|}{a}$. Now, there exists $K_1>0$, independent of J, such that the hyperbolic distance $\rho_J(x,\partial B(x_0;a))>\log\frac{K_1}{|J|}$ for all $x\in D_{r_0}(J)$. This is so because

$$\rho_J(x,\partial B(x_0;a))\geq\rho_J(x,\partial D(J;\alpha_0))\geq\rho_J(\partial D_{r_0}(J),\partial D(J,\alpha_0))$$

and since both these curves $\partial D_{r_0}(J)$ and $\partial D(J,\alpha_0)$ have a constant distance to J. This last distance is greater or equal to

$$\log\tan\left(\frac{\pi}{2}-\frac{\alpha_0}{4}\right)-r_0\geq\log\frac{K_1}{|J|}$$

for $|J|$ sufficiently small. Here we used $\frac{1}{2}\alpha_0<\sin(\alpha_0)\leq\frac{|J|}{a}$ in the last inequality. Let $\phi_J(z)|dz|$, $\tilde{\phi}_J(z)|dz|$ and $\psi(z)|dz|$ be the Poincaré metrics of $\mathbb{C}_J$, $(B(x_0;a)\setminus\mathbb{R})\cup J$ and of $D_{x,\frac{K_1}{|J|}}=\{z\,;\ \rho_J(z,x)<\frac{K_1}{|J|}\}$. If $x\in D_{r_0}(J)$, $D_{x,\frac{K_1}{|J|}}\subset B(x_0;a)$ and therefore, by the Lemma of Schwarz, $\psi(x)\geq\tilde{\phi}_J(x)$. On the other hand if $\Psi\colon\mathbb{C}_J\to\mathbb{D}$ is a hyperbolic isometry such that $\Psi(x)=0$ then $\Psi(D_{x,\frac{K_1}{|J|}})$ is the Euclidean disc $\mathbb{D}_s$ with radius s, where $\log\frac{K_1}{|J|}=\log\frac{1+s}{1-s}$. But the scalar map $z\mapsto z/s$ is an isometry between the hyperbolic metric on $\mathbb{D}_s$ and that of $\mathbb{D}$. So from this one gets immediately $\frac{\psi(x)}{\phi_J(x)}=\frac{1}{s}$. Thus $\tilde{\phi}_J(x)\leq(1+K|J|)\phi_J(x)$. Hence, for every $x,y\in D_{r_0}(J)$ we have that

$$\tilde{\rho}_J(x,y)\leq(1+K|J|)\rho_J(x,y)$$

and this proves the claim and, therefore, the lemma. □

Remark. We can use the same argument with the inverse of ϕ and we get

$$(1-K|J|)\rho_J(x,y)\leq\rho_{\phi(J)}(\phi(x),\phi(y))\leq(1+K|J|)\rho_J(x,y)$$

Lemma 5.3. *Let $R > 0$ be given. Let S_n be the maximal neighbourhood of $\triangle_n^1$ on which $f^{q(n)-1}$ is monotone and let $W_n = f^{q(n)-1}(S_n)$. Then there exists n_0 such that if $n \geq n_0$ then the basic composition B_n is defined and it is univalent on the Poincaré neighbourhood $D_R(W_n)$.*

Proof. In fact we shall use Lemma 5.3 only for maps which are in some Epstein class. In that case the proof is obvious. In order to be complete we give the proof in the general case. Let W_n^i be the interval neighbourhood of $\triangle_n^i$ which is mapped monotonically onto W_n by $f^{q(n)-i}$. Since W_n^i is contained in an interval of the form $\triangle_{n-3}^j$, one gets from the bounded geometry that $\sum_{j=1}^{q(n)-1} |W_n^j|$ goes to zero exponentially with n. Let $F = \Phi \circ Q$ and let $a > 0$ be such that Φ is univalent on the disc of radius a centered at any point of $Q(\triangle_n)$. Let $K = K(2R, a)$ be as in Lemma 5.2 and take n_0 so large that for $n \geq n_0$, we have $\prod_{j=1}^{q(n)-1}(1 + K|W_n^j|) < 2$ and $|Q(W_n^j)| < l_0$. Because $|W_n^i| \to 0$ as $n \to \infty$ we can also assume that n_0 is so large that $D_{2R}(W_n^i)$ is contained in a disc with radius a. From Lemma 5.2 one gets

$$F^{-1}(D_R(W_n)) = Q^{-1}\Phi^{-1}(D_R(W_n)) \subset Q^{-1}D_{(1+K|W_n|)R}(\Phi^{-1}W_n).$$

Moreover,

$$Q^{-1}D_{(1+K|W_n|)R}(\Phi^{-1}W_n) \subset D_{(1+K|W_n|)R}(F^{-1}W_n) = D_{(1+K|W_n|)R}(W_n^{q(n)-1})$$

because, from the Lemma of Schwarz, $Q^{-1} \colon \mathbb{C} \setminus \Phi^{-1}W_n \to \mathbb{C} \setminus F^{-1}W_n$ contracts the Poincaré metric on these spaces. Hence

$$F^{-1}(D_R(W_n)) \subset D_{(1+K|W_n|)R}(W_n^{q(n)-1}).$$

By assumption the set $D_{(1+K|W_n|)R}(W_n^{q(n)-1})$ is contained in a disc of radius a and so Φ is univalent on the Q image of this set. So we can repeat all this and we get that the basic composition B_n maps $D_R(W_n)$ univalently into $D_{2R}(B_n(W_n^1))$. $\square$

To prove Theorem 5.1, we will show that for n big enough, the map $F_+^{-1} \circ B_n$ maps some neighbourhood of the dynamical interval $\triangle_n$ well inside itself. In view of Lemma 5.2, we will start by analyzing the situation in which all factors of the composed map B_n are in fact square root maps. Later, in Remark 2 below Lemma 5.4, we will extend the proof to the situation where some of the factors are square root maps but the others may be more general holomorphic maps that are univalent on the upper half-plane. We will select some of these factors and will treat them as square root maps and we will group the others into maps with some distortion properties. We state this abstractly as follows. As before we say that T is a *square root* if T is of the form $T = AF_1 \circ S \circ AF_2$ where

$$S \colon \mathbb{C} \setminus \{\text{negative real numbers}\} \to \mathbb{C}$$

is the standard square root map and AF_i are real affine maps. Next let Σ be the set of square roots maps of the upper half-plane which induce orientation preserving homeomorphisms from some fixed interval $I = [a, b]$ onto itself (and therefore the singularity of T is not contained in I). If a map $T \in \Sigma$ is such that its singularity t to the left of a we say that it is a *left root*; if its singularity is to the right of b we say that it is a *right root*. We consider compositions of the form

$$A_m \circ C_m \circ \cdots \circ A_2 \circ C_2 \circ A_1 \circ C_1$$

satisfying the following properties depending on parameters (k, K, λ) with $1 < k < K$ and $\lambda > 0$.

1. A_1 is a left root with singularity a and A_m is a right root with singularity b;

2. If $i = 2, 3, \ldots, m-1$, A_i is a left root whose singularity $a_i < a$ moves exponentially fast to the left: $1 < k < \frac{|a-a_{i+1}|}{|a-a_i|} < K$ (so $a_2 < a - 1$);

3. For $i = 2, 3, \ldots, m-1$, C_i is a univalent map of the upper half plane which is a self homeomorphism of I. This composition has a precise kind of bounded distortion near the interval $[a_i, a]$. Namely, if $I_i \supset I$ is the maximal subinterval of the real axis which is mapped homeomorphically into the real axis by C_i then the image of I_i contains $[a_i, b] \supset [a, b]$ and the preimage J_i of $[a_i, a]$ under C_i is well within I_i (by this we mean that I_i contains a λ-scaled neighbourhood of J_i).

These abstract compositions will later be related to the composition B_n by looking carefully at the dynamics of f. In Figure 5.3 we describe the image of the upper half-plane by the factors of the above composition.

Let N be the Poincaré neighbourhood of I whose Euclidean radius is the maximum of the numbers $|a_k - a|$.

Lemma 5.4. (The Sector Lemma) *For each (k, K, λ) as above, there exists $\theta > 0$ such that the image of the neighbourhood N under any composition*

$$A_m \circ C_m \circ \cdots \circ A_2 \circ C_2 \circ A_1 \circ C_1$$

which satisfies Properties 1, 2 and 3 from above with parameters (k, K, λ), is contained in the sector bounded by the vertical line at b, the interval I and the line through a with an angle θ with the negative real axis as in Figure 5.4.

Proof. Let the angle of a point z be the angle of the line through a and z with the negative real axis. We start by analyzing the distortion properties of the map C_i. By definition, $C_i = T_{n(i)} \circ \cdots \circ T_2 \circ T_1$ where $T_i \in \Sigma$. Since C_i is a diffeomorphism on the interior of I_i, any factor T_j of the above composition is a diffeomorphism on $I_i^j = T_{j-1} \circ \cdots \circ T_1(I_i)$. In particular, T_i is holomorphic and univalent in the Poincaré neighbourhood $D_1(I_i^j)$ and, by the Lemma of Schwarz, $T_j(D_1(I_i^j)) \subset D_1(I_i^{j+1})$ and C_i is holomorphic and univalent on $D_1(I_i)$. By Koebe's Distortion Theorem (see the Appendix), C_i has bounded distortion on $U_i = D_1(J_i')$ where $J_i' \supset J_i$ is well inside I_i and J_i is well inside J_i'. By Property 3 the interval $J_i \cup I$ is well inside U_i and therefore the image of U_i by $A_i \circ C_i$ contains a rectangle R_i in the upper half-plane whose base is $[A_i(a_i), a]$ and

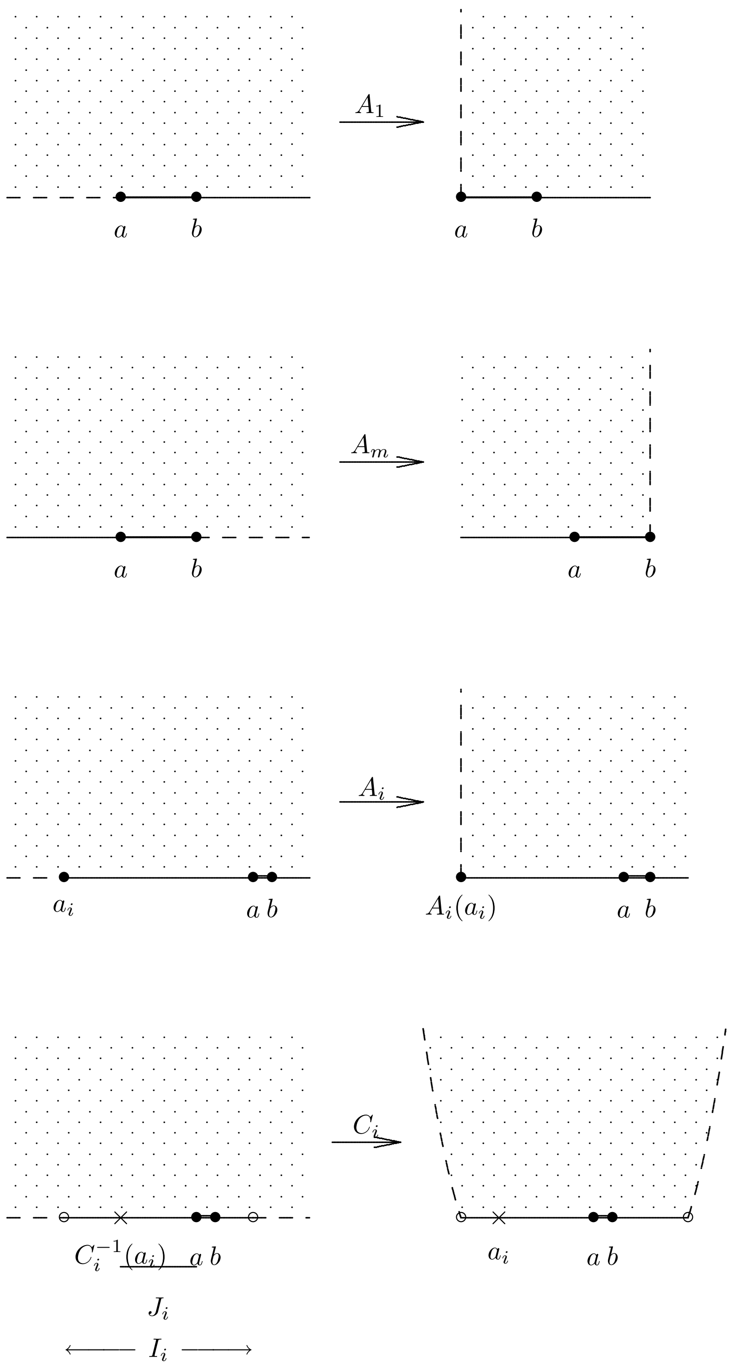

Fig. 5.3: The action of the maps A_i, C_i and the intervals J_i and I_i.

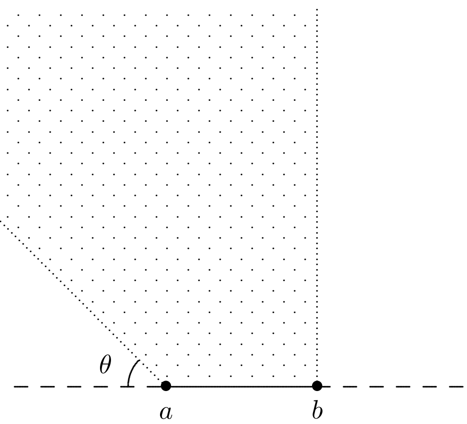

Fig. 5.4:

whose height is a definite proportion of the length of the base as indicated in Figure 5.5.

Now we prove the Sector Lemma. Start with any point p_1 in the upper half-plane and define $p_2 = A_1 \circ C_1(p_1)$, ..., $p_{i+1} = A_i \circ C_i(p_i)$ for $i > 1$. First note that p_2 lies to the right or on a vertical line through a. We will complete the lemma in three steps now.

Step 1: If p_i is not in U_i then $p_{i+1} \notin R_i \subset A_i \circ C_i(U_i)$ as in Figure 5.5. Since the height of R_i is a definite proportion of its base $[a_i, a]$, the angle θ_{i+1} of p_{i+1}, is bounded away from 0. This is so because $p_{i+1} = A_i \circ C_i(p_i)$ lies to the right of the vertical line through $A_i(a_i)$ and above the rectangle R_i as in Figure 5.6. So if $p_i \notin U_i$ for all i then we are finished.

Step 2: If p_i is in U_i but the angle θ_i is very close to π, then, by the bounded distortion property, θ_{i+i} is not too small.

Step 3: So assume there is a first i so that $p_i \in U_i$ and the angle θ_i is not too near π. Since i is minimal we may assume that the angle θ_i of p_i with the negative real axis is not too small because of Steps 1 and 2 and because p_2 is to the right or on the vertical line through a. Then we apply a fixed number l of the factors $A_i \circ C_i$, $A_{i+1} \circ C_{i+1}$, ..., until a_{i+l} is much further away from a than p_{i+l}. This happens because of Property 2 and the remark below. During these l iterations the angle as viewed from a is only distorted by a bounded amount. The point p_i is contained in a Poincaré neighbourhood whose Euclidean size depends on the scale (i.e., on i) and on the angle of p_i which is not too small. The next l iterates remain in this Poincaré neighbourhood since this neighbourhood is invariant whereas the root moves away. Hence, the subsequent factors $A_j \circ C_j$, for $m > j > i + l$, only cause a sequence of distortions decaying geometrically by Property 2 and the remark below. Thus the angles of $p_i, \ldots, p_{m-1}$ stay away from zero. Moreover, as we are assuming that $p_i \in U_i$, this point also belongs to the Poincaré neighbourhood N. Since each $T \in \Sigma$ preserves I, the lemma of Schwarz implies that C_j and A_j map N

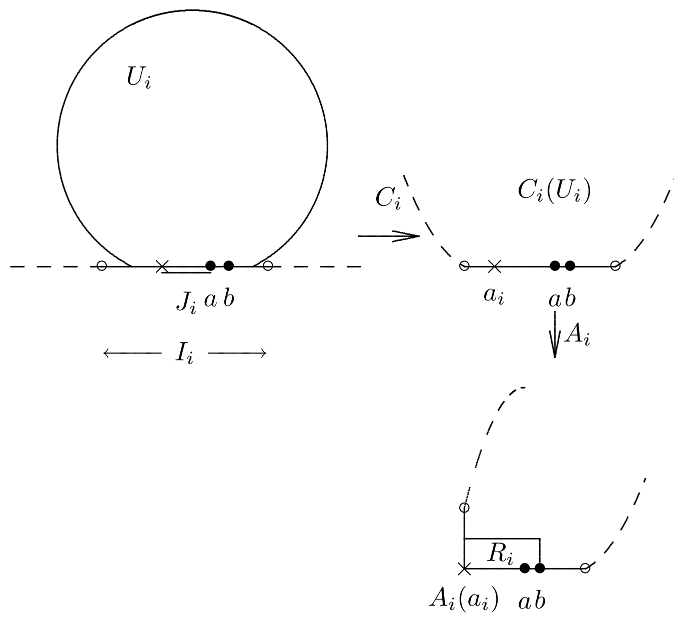

Fig. 5.5:

into itself and so $p_{i+1}, \ldots, p_{m-1}$ are also in N. But C_m has bounded distortion on N. This can be seen as follows. Since I_m contains a scaled neighbourhood of (a_{m-1}, a) and N has radius $|a_{m-1} - a|$ (and is 'based in I) it follows that N has bounded diameter in terms of the Poincaré metric on $\mathbb{C}_{I_m}$. But as before, this and Koebe's Distortion Theorem implies that C_m has bounded distortion on N. Therefore C_m only reduces the angle by a bounded factor. Finally A_m, the right root, reduces the angle by at most a factor two. This completes the proof of the Sector Lemma. □

Remark. If a holomorphic map is univalent on $\mathbb{C}_J$ then Koebe's Distortion Theorem implies that it has bounded distortion on any region R_0 as in Figure 5.7. The constants depend on the shape of R_0. Furthermore, it has exponentially small non linearity on a region R_n which is exponentially small. Hence the maps have bounded distortion and this yields the bounded distortion of the composition of the first l iterates (l fixed) we needed in the proof above. The fact that the regions R_n go down exponentially fast to zero, yields that the distortion of the subsequent maps goes exponentially fast to zero.

Next we will prove that once we get control on the angle of the image of N as in the Sector Lemma, we can bound the height of the image of a smaller region. Let $\theta \in (0, \frac{\pi}{2})$. Let $a < b < d$, $a < a' < b' < d$ and let d' be a point in the upper half space with the same real part as b'. Let S be the cone in the upper half space whose boundary is the union of the interval $[a'b']$, the vertical line through b', d' and the ray in the positive half plane that starts in a' making an angle θ with the negative real axis as in Figure 5.8. For each $R > 0$, D_R denotes the

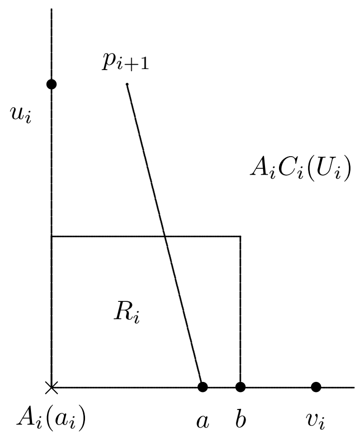

Fig. 5.6: The region $A_iC_i(U_i)$ is bounded by the line segments $[A_i(a_i), u_i]$, $[A_i(a_i), v_i]$ and a curve connecting u_i and v_i in the positive quadrant. p_{i+1} lies above R_i (and outside $A_iC_i(U_i)$).

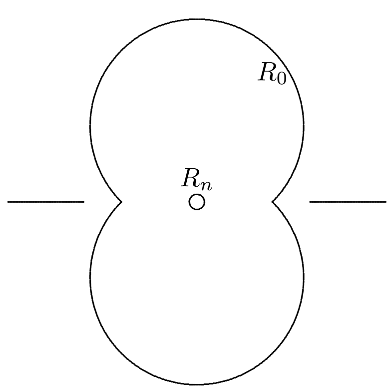

Fig. 5.7: The regions R_n and R_0.

intersection with the upper half space of the Poincaré neighbourhood of radius R of the geodesic (a,d) of the Riemann surface $\mathbb{C}_{(a,d)} = (\mathbb{C} \setminus \mathbb{R}) \cup (c,d)$.

Lemma 5.5. *Given $\theta > 0$ and $C > 0$, there exists $r' > 0$ such that if the distance between any two of the points a, b, d, a', b', d' is bigger than $\frac{1}{C}$ and smaller than C, then there exists for each $L > r'$, a constant $R > 0$ satisfying the property below. If $F\colon D_R \to S$ is a univalent map that maps $[a,d]$ homeomorphically onto the interval of the boundary of S with endpoints a', d', then*

$$F(D_L) \subset D_{L/2}.$$

Proof. Let $\Psi\colon \mathbb{H} \to S$ be the Riemann mapping normalized so that $\Psi(\infty) = \infty$, $\Psi(a) = a'$, $\Psi(d) = d'$. Let $\Psi_R\colon D_R \to S$ be the holomorphic diffeomorphism that maps a into a', d into d' and the highest point in the boundary of D_R into infinity. Hence $F \circ \Psi_R^{-1}$ is a univalent map from S into S that maps the arc a', d' of the boundary of S homeomorphically onto itself. We split the proof in some steps.

Step 1: If $\Phi\colon S \to \Phi(S) \subset S$ is a univalent function that maps homeomorphically the arc $[a', d']$ of the boundary of S homeomorphically onto itself, then

$$\Phi(\Psi(D_r)) \subset \Psi(D_r) \quad \text{for all } r > 0.$$

Proof. $\Psi^{-1} \circ \Phi \circ \Psi$ extends, by symmetry, to a univalent map of $\mathbb{C}_{(a,d)}$ into itself. Hence, by Schwarz, it maps D_r inside itself for all $r > 0$. □

Step 2: There exists $r' > 0$ such that, if $L > r'$ then

$$\Psi(D_{2L}) \subset D_{\frac{L}{2}}.$$

Proof. Let $\tilde{S}$ be the sector bounded by the vertical ray from b' and the horizontal ray from b' in the negative direction. Let $G\colon \mathbb{H} \to \tilde{S}$ be the holomorphic diffeomorphism that maps ∞ into ∞, a into a' and d into d'. From the hypothesis, we get that G is the composition of the square root map with an affine map of bounded distortion. Hence, $G(D_{2L}) \subset D_{\frac{L}{2}}$ if L is big enough. On the other hand, since $S \subset \tilde{S}$ we have that $G^{-1} \circ \Psi$ is a univalent map of $\mathbb{C}_{(a,d)}$ into itself. Therefore, by Schwarz, $G^{-1}(\Psi(D_{2L})) \subset D_{2L}$. Hence, $\Psi(D_{2L}) \subset G(D_{2L}) \subset D_{\frac{L}{2}}$. □

Step 3: Given L_0, there exists R_0 such that if $L \le L_0$ and $R \ge R_0$, then

$$\Psi_R(D_L) \subset \Psi(D_{2L}.$$

Step 4: Let Φ be as in 1) and let $\bar{R} \ge r'$ as in 2). If $R \ge R_0(\bar{R})$, R_0 as in 3), we have

$$\Phi\Psi_R(D_{\bar{R}}) \subset \Phi\Psi(D_{2\bar{R}}) \subset \Psi(D_{2\bar{R}}) \subset D_{\frac{\bar{R}}{2}}.$$

□

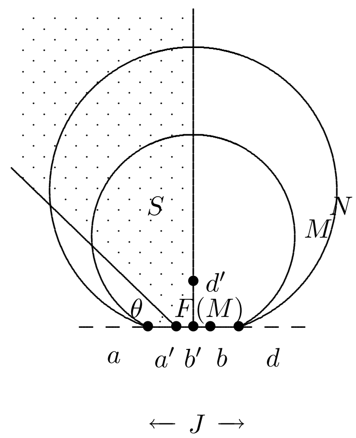

Fig. 5.8: Fig VI.5.8

In the remainder of this section we will show that if f is Epstein and it is infinitely renormalizable of combinatorial type bounded by N then one can define a factoring of all the basic backward compositions with the properties needed in the Sector Lemma. In order to define this factorization, let us recall the notation we have already introduced in the beginning of this section. Let $F\colon U \to V$ be the holomorphic extension of f to a degree two branched covering map from Proposition 5.1; F_s^{-1}, $s = \pm$ are the two inverse branches of the inverse of f; c_{-k} is the inverse branch of the backward orbit of the critical point staying in the attracting Cantor set and $s\colon \mathbb{N} \to \{\pm\}$ is the sign map: $F_{s(k)}^{-1}(c_{-k}) = c_{-(k+1)}$. Also Ξ_n is the collection of intervals

$$\{\triangle_n, \triangle_n^1 = f(\triangle_n), \ldots, f^{q(n)-1}(\triangle_n) = \triangle_n^{q(n)-1}\}.$$

Definition. The *scale map* $sc\colon \mathbb{N} \to \mathbb{N}$ is defined by $sc(k) = \max\{i\,;\ c_{-k} \in \triangle_i^1\}$. Since $c_{-k} \in \Delta_0^1$ for any k this definition makes sense. Note that sc maps $\{0, \ldots, q(n) - 2\}$ onto $\{0, \ldots, n - 1\}$.

We also consider a partition of the integers $\{0, 1, \ldots, q(n) - 2\}$ in 'epochs' as follows. Let $e_n(j)$ be the smallest integer in $\{0, 1, \ldots, q(n) - 2\}$ such that $sc(k) < j$ for all $k > e_n(j)$. Clearly $q(n)-2 \geq e_n(1) > e_n(2) > \cdots > e_n(n-1) \geq 0$. So define

$$E_n(1) := \{e_n(2) + 1, \ldots, q(n) - 2\},$$

$$E_n(j) := \{e_n(j + 1) + 1, \ldots, e_n(j)\}$$

for $j = 2, \ldots, n - 2$ and

$$E_n(n - 1) := \{1, \ldots, e_n(n - 1)\}.$$

These collections of sets form a partition of $\{0, 1, \ldots, q(n) - 2\}$. Note that for $j > 1$, by definition, the maximal value of the scale function restricted to $E_n(j)$ is j and that this maximal value is assumed at the right boundary point of $E_n(j)$. Moreover, the number of points where the scale map assumes the maximal value in an epoch is universally bounded because of the bounded combinatorics. One of these maxima corresponds to the closest approach of c_{-k} to the critical value for k in the corresponding epoch.

Example. Let us consider the periodic doubling case $q(k) = 2^k$, $k \le n = 6$. In Figure 5.9 we have drawn the graph of the scale map from 0 to $q(6) = 64-2 = 62$ and the corresponding epochs.

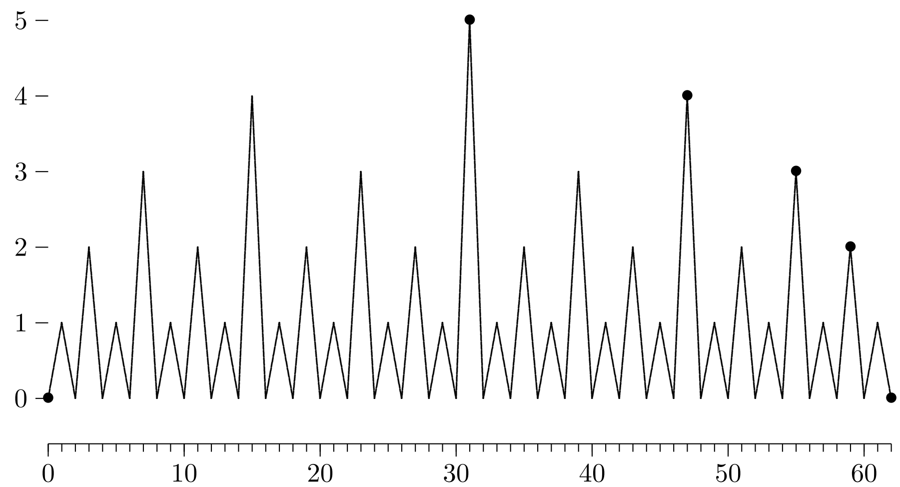

Fig. 5.9: The scale map in the period doubling case for $n = 6$. The end of each epoch is marked in the graph.

Definition. A *root at level n* is an integer $k \in \{0, \ldots, q(n) - 1\}$ such that $k \in E_n(sc(k))$, i.e., k is an integer where the scale map assumes its maximum in its epoch. (The maximal value of the scale map on $E_n(j)$ is by definition j and so there certainly exists an integer $k \in E_n(j)$ such that $sc(k) = j$.) A root k is a *left* (resp. *right*) root if the interval of Ξ_n that contains c_{-k} has the opposite (resp. the same) orientation as the critical value interval $\triangle_n^1$, i.e., when the iterate of f (or $F|\mathbb{R}$) mapping the interval Δ_n^1 to the interval Δ_n^i containing c_{-k} is orientation preserving (respectively orientation reversing). Below we shall show that this terminology corresponds to the one introduced before.

Let us show that the largest element in $E_n(j)$ is always a left root. Indeed, let k be the largest element in $E_n(j)$. By definition we have that $sc(k) = j$. Let $\triangle_n^i \in \Xi_n$ be the interval that contains c_{-k}. Since $sc(k) = j$ we have that $\triangle_n^i \subset \triangle_j^1$ and $\triangle_n^i$ is not contained in $\triangle_{j+1}^1$. In fact, $\triangle_n^i$ is precisely the element of Ξ_n contained in $\triangle_j^1 \setminus \triangle_{j+1}^1$ which is further (in terms of backward iteration) from $\triangle_n$. Of course this is the first iterate of $\triangle_n^1$ which is contained in $\triangle_j^1 \setminus \triangle_{j+1}^1$. Since $f^{q(j)}$ is the first return map to $\triangle_j^1$ one has $J = f^{q(j)}(\triangle_n^1)$. Since $f^{q(j)} \colon \triangle_j^1 \to \triangle_j^1$ is a unimodal map – it has only one turning point (but does not map the boundary into itself) – and $\triangle_j^1$ and $\triangle_n^1$ contain the critical value of f, we have that $f^{q(j)}|\triangle_n^1$ is orientation reversing. This shows that the largest element in $E_n(j)$ is always a left root.

We can now define the factoring of the n-th basic backward composition

$$B_n = F^{-1}_{s(q(n)-2)} \circ \cdots \circ F^{-1}_{s(j)} \circ \cdots \circ F^{-1}_{s(0)}.$$

We partition this composition in blocks of maps containing no left roots and maps which are left roots. More precisely, let $C_1 = F^{-1}_{s(i)-1} \circ \cdots \circ F^{-1}_{s(0)}$ where $s(i)$ is maximal so that the block $\{0, \ldots, s(i)-1\}$ contains no left roots. Then $s(i)$ is the first left root and so we set $A_1 = F^{-1}_{s(i)}$. For $m \geq 2$ let $C_m = F^{-1}_{s(j+l)-1} \circ \cdots \circ F^{-1}_{s(j)+1}$ where j is the $(m-1)$-th left root and $j+l$ is the next left root and we set $A_m = F^{-1}_{s(j+l)}$. In this way we get

$$B_n = A_m \circ C_m \circ \ldots A_1 \circ C_1.$$

Since the largest elements of $E_n(j)$ are left roots, the indices $s(j)+1, \ldots, s(j+l)$ of the maps appearing in

$$A_j \circ C_j = F^{-1}_{s(j+l)} \circ \cdots \circ F^{-1}_{s(j)+1}$$

are all contained in one epoch.

Let us now explain how all this relates to the abstract situation of compositions in Σ we studied above. So let $I = [a, b]$ be some fixed interval and let $\triangle_n^{j(k)}$ be the interval in Ξ_n containing c_{-k}. We identify $\triangle_n^{j(k)} \subset \mathbb{R}$ with $I \subset \mathbb{R}$ via an affine map U_k which is orientation preserving if $\triangle_n^{j(k)}$ and Δ_n^1 have the same orientation and orientation reversing otherwise. Then we define

$$T_{s(k)} = U_k \circ F^{-1}_{s(k)} \circ U^{-1}_{k-1}.$$

Clearly, if F is a quadratic map then $T_{s(k)} \in \Sigma$ and so $T_{s(k)}$ maps the upper-half plane into itself. Let us analyze when $T_{s(k)}$ is a left root. $F^{-1}_{s(k)}$ is orientation preserving on $\mathbb{R}$ if $s = -$ and reversing in $s = +$ (this is because f is increasing to the left of c and decreasing to its right; see also Figure 5.1). In particular, U_k is orientation preserving if and only if $s(k) = -$ and U_{k-1} is orientation preserving or if $s(k) = +$ and U_{k-1} is orientation reversing. Furthermore, since F_-^{-1} maps the upper-half plane to the left of some line and F_+^{-1} maps it to the right of some line it follows that $U \circ F_s^{-1} \circ U'$ is a left root if and only if $s = +$ and U is orientation preserving or $s = -$ and U is orientation reversing (the orientation of U' is irrelevant for this). But this implies that $F^{-1}_{s(k)}$ is a left root if U_{k-1} is orientation preserving and a right root otherwise. Hence the above terminology.

Lemma 5.6. *If $f = \Phi \circ Q$ is an infinite renormalizable map of bounded combinatorial type that belongs to some Epstein class then the above factoring of the basic backward compositions satisfies the conditions of the Sector Lemma.*

Proof. The topology is right for property iii). Hence the statement follows from the real bounds. □

Proof of Theorem 5.1: The main tool for the proof of Theorem 5.1 is the Sector Lemma. Let θ be as in the Sector Lemma, where the parameters k, K, λ are given by the real bounds. Let R and $\overline{R}$ be as in Lemma 5.5.

Let $W_n = [a, c]$ be the maximal interval containing $\triangle_n$ such that the basic composition B_n is defined and monotone on W_n. Let $W_{n_i} = f^{i-1}(B_n(W_n))$ for $i = 1, \cdots, q(n)$. Let $[a', a'']$ be the maximal symmetric interval containing the critical point such that $f^{q(n)}([a', a'']) \subset W_n$. We may assume that $f^{q(n)}(a') = a$. Let b' be the critical point of f. Of course $b = f^{q(n)}(b') \in [a', b]$ is the extremal value of $f^{q(n)}|\triangle_n$.

The basic composition is well defined and univalent on the Poincaré neighbourhood $D_R(W_n)$. From the choice of W_n, the root A_1 has a left singularity at a and A_m is a right root with singularity $b = F^{q(n)}(b') = F^{q(n)}(c)$. By Lemma 5.7, the factorization of the basic composition satisfy the hypothesis of the Sector Lemma. From the Sector Lemma we conclude that the image of $D_R(W_n)$ by $F^{-1} \circ B_n$ is contained in a sector S based on the interval $[a', b']$ and with angles θ and $\frac{\pi}{2}$. Because of the real bounds the interval $[a', a'']$ is contained well inside W_n as indicated in Figure 5.10. Now we can use Lemma 5.5 to prove that $F_+^{-1} \circ B_n$ maps $D_{\overline{R}}(W_n)$ well inside itself and the theorem is proved. This implies that the conformal moduli are bounded from above and below. The statement of Theorem 5.1 holds also for maps that have quadratic-like extensions because, as we have already observed in Section 1, such maps are analytically conjugated to maps that belong to some Epstein class. $\square$

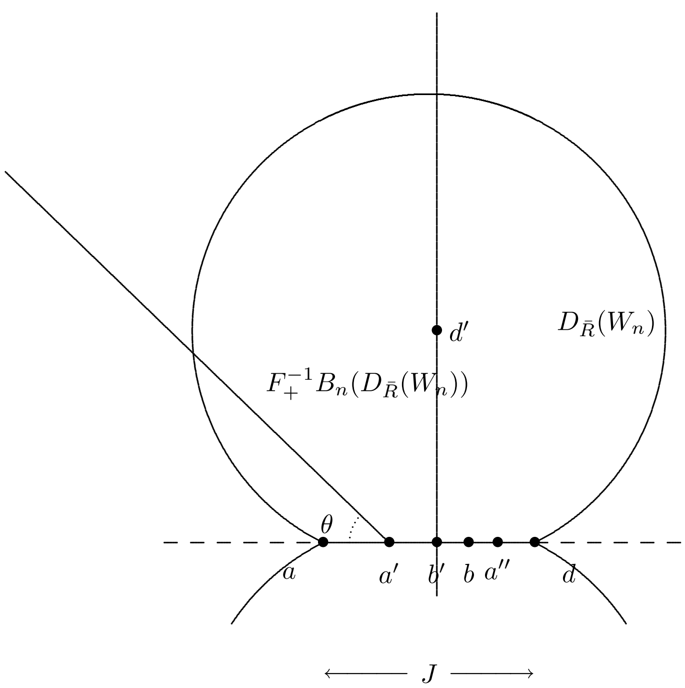

Fig. 5.10: The region $F_+^{-1}B_n(D_{\bar{R}}(W_n))$ is bounded by the line segments $[a', b']$, $[b', d']$ and a curve connecting a' and d'.

As we saw before if F has a quadratic-like extension then it is holomorphically equivalent to a map in some Epstein class. This is the reason we were able to assume in the proof of Theorem 5.1 above that the maps are Epstein. Using the compactness properties of the set of quadratic-like maps of bounded conformal type, see the proof of Theorem 4.2d, we get the following uniform version of Theorem 5.1.

Theorem 5.2. *For each $N > 0$ there exists $L = L(N)$ such that for each $B_0 > 0$ there exists $n_0(N, B_0)$ such that for any quadratic-like map F of conformal type bounded by B_0 one has the following properties.*

1. *If F is an infinitely renormalizable of combinatorial type bounded by N then for $n \geq n_0$, the mapping $\mathcal{R}^n(f)$ extends to a symmetric quadratic-like map whose conformal type is bounded by L.*

2. *If F can be renormalized $10n_0$ times all with type bounded by N then $\mathcal{R}^{n_0}(f)$ extends to a symmetric quadratic-like map whose conformal type is bounded by L.*

6 Riemann Surface Laminations

In the previous section we have seen that, if a map f belongs to some Epstein class and is infinitely renormalizable of bounded combinatorial type, then $\mathcal{R}^n(f)$ extends to a quadratic-like map for n large. Furthermore, the renormalization operator extends to an operator in the space of germs of such quadratic-like maps. In order to prove the contraction of this operator we will embed the space of germs of quadratic-like maps in a sort of Teichmüller space of laminations by Riemann surfaces. In this section we will develop some aspects of a Teichmüller theory for Riemann surface laminations.

6.1. Riemann surface laminations

Let us start this section by describing Sullivan's notion of a Riemann surface lamination and his Teichmüller theory on these laminations. As usual, a Riemann surface is a topological surface with an open cover U_i and homeomorphisms z_i (called coordinate systems or charts) from U_i to an open subset of the complex plane $\mathbb{C}$ with the property that when $U = U_i \cap U_j \neq \emptyset$,

$$z_{j,i} = z_i \circ z_j^{-1} \colon z_j(U) \to z_i(U)$$

is holomorphic.

Definition. Let $\mathcal{L}$ be a Hausdorff topological space. A *Riemann surface lamination atlas* on $\mathcal{L}$ is a collection of open homeomorphisms $Z_i \colon U_i \to \mathbb{D} \times \Lambda$ (each one of which we shall call a *flow box*) where $\mathbb{D}$ is the unit disc in $\mathbb{C}$ Λ

is a topological space and $\{U_i\}$ is an open cover of $\mathcal{L}$, satisfying the following property: the overlap maps $Z_i \circ Z_j^{-1}$ are of the form

$$(z, \lambda) \mapsto (Z_{j,i}(z, \lambda), \Lambda_{j,i}(\lambda))$$

where $z \mapsto Z_{j,i}(z, \lambda)$ is holomorphic for each λ. A Riemann surface lamination structure on $\mathcal{L}$ is an equivalence class of Riemann surface lamination atlases on $\mathcal{L}$, where two atlases are equivalent if their union is again an atlas. We call $Z_i^{-1}(\mathbb{D} \times \{\lambda\})$ a *plaque*. Furthermore, we say x and y in $\mathcal{L}$ are equivalent if there exists a chain $x = x_0, x_1, \ldots, x_n = y$ such that x_i, x_{i+1} belong to the same plaque. An equivalence class L of this equivalence relation in $\mathcal{L}$ is called a *leaf*. By definition each leaf has a Riemann surface structure.

Definition. A Riemann surface lamination $\mathcal{L}$ is *hyperbolic* if: i) each leaf L of $\mathcal{L}$ is a hyperbolic Riemann surface, i.e., there exists a holomorphic covering map $\pi_L \colon \mathbb{D} \to L$, where $\mathbb{D}$ is the Poincaré disc; ii) the Poincaré metric on the leaves of $\mathcal{L}$ is continuous on $\mathcal{L}$, i.e., if V is a continuous vector field tangent to the leaves of $\mathcal{L}$ then the hyperbolic norm of V, $p \mapsto |V(p)|$, is a continuous function on the lamination.

Remark. Candel (1991) has characterized the Riemann surface laminations which satisfy condition ii) above. In particular, he proves that if there is no invariant transversal measure then i) implies ii).

Let us give two examples of such laminations. The second example, which is a special case of the first example as we will show below, will play a fundamental role in this Chapter.

Example. Let $f \colon S^1 \to S^1$ be an expanding $C^{1+\alpha}$ map of degree two. The reason we consider such a map is because any quadratic-like infinitely renormalizable map corresponds to a degree two expanding map of the circle, see the last result in Section VI.4. To such a circle map we associate a hyperbolic Riemann surface lamination $\mathcal{L}_f$ as follows. First we need to associate a solenoid to f. Let $\mathcal{S}_f$ be the collection of all preorbits of f, i.e.,

$$\mathcal{S}_f = \{\underline{x} = (\ldots, x_n, \ldots, x_1, x_0)\,;\ f(x_n) = x_{n-1}\}$$

endowed with the product topology and let $\pi_f \colon \mathcal{S}_f \to S^1$ be defined by $\pi_f(\ldots, x_n, \ldots, x_1, x_0) = x_0$. Furthermore, let $\bar{f} \colon \mathcal{S}_f \to \mathcal{S}_f$ be the homeomorphism defined by

$$\bar{f}((\ldots, x_n, \ldots, x_1, x_0)) = (\ldots, x_n, \ldots, x_1, x_0, f(x_0)).$$

This is a homeomorphism because its inverse is the shift map, and $\bar{f}$ covers f: $\pi_f \circ \bar{f} = f \circ \pi_f$. It is called the *natural extension* of f. Let us show that $\pi_f \colon \mathcal{S}_f \to S^1$ is a locally trivial fibration whose fibers are Cantor sets. Indeed, let p be a fixed point of f and $p' \neq p$ so that $f(p') = f(p)$ and denote the arcs of the circle connecting p to p' by I_0 and I_1. If $x \neq p$ then one of the preimages of x is in I_0 and the other in I_1. In this way each point $\underline{x}$ in $\pi_f^{-1}(x_0)$ corresponds in a unique way to a sequence $\rho(\underline{x}) = (\dots, \rho_n, \dots, \rho_1) \in \{0,1\}^{\mathbb{N}}$, $\rho_n = i$ if $x_n \in I_i$, which represents its history. Thus $\pi_f^{-1}(S^1 \setminus \{p\})$ is homeomorphic to $(S^1 \setminus \{p\}) \times \{0,1\}^{\mathbb{N}}$. Moreover, define $\text{add}\colon \{0,1\}^{\mathbb{N}} \to \{0,1\}^{\mathbb{N}}$ as the 'adding one map', i.e.,

$$\text{add}(\underline{a}) = \text{add}(\dots, a_n, \dots, a_1, a_0) = (\dots, a_n, \dots, a_1, a_0 + 1).$$

Here we use the cash-register convention: if $a_0 + 1 = 2$ we should read zero for this term and add one term to its left (and repeat this again and again if necessary). If we define the equivalence relation $\{p^+\} \times (\underline{a}) \sim \{p^-\} \times \text{add}(\underline{a})$ on $(S^1 \setminus \{p\}) \times \{0,1\}^{\mathbb{N}}$ (where we write $S^1 \setminus \{p\} = (p^-, p^+)$ and $f(p^+) = f(p^-)$) then $\mathcal{S}_f$ becomes homeomorphic to $(S^1 \setminus \{p\}) \times \{0,1\}^{\mathbb{N}})/\sim$.

A connected component of this lamination is called a *leaf*. Each leaf is homeomorphic to $\mathbb{R}$ and through the local homeomorphism π_f it inherits naturally a smooth structure. Because the orbits of $\text{add}\colon \{0,1\}^{\mathbb{N}} \to \{0,1\}^{\mathbb{N}}$ are dense, each leaf is dense in $\mathcal{S}_f$. Moreover, to each periodic point of f corresponds precisely one leaf of $\mathcal{S}_f$ which is periodic under $\bar{f}$: if p is a periodic point of f of period n then $\underline{p} \in \mathcal{S}_f$ satisfying $p_{kn} = p \forall k \geq 0$ is the periodic point of $\bar{f}$. Using this lamination and the distortion theorem for expanding maps we get a global linearization of f. Indeed, since f is $C^{1+\alpha}$ each leaf L of $\mathcal{S}_f$ even carries a natural affine structure. More precisely, one can define a $C^{1+\alpha}$ diffeomorphism $h_L\colon L \to \mathbb{R}$ depending continuously (up to postcomposition with affine transformations) on the leaf L such that $\bar{f}$ becomes affine in terms of these homeomorphisms restricted to each leaf. Indeed, if we take three consecutive points a, b, c on a leaf L of $\mathcal{S}_f$ we define

$$\frac{h_L(c) - h_L(a)}{h_L(b) - h_L(a)} = \lim_{n\to\infty} \frac{d\left(\bar{f}^{-n}(c)\,,\, \bar{f}^{-n}(a)\right)}{d\left(\bar{f}^{-n}(b)\,,\, \bar{f}^{-n}(a)\right)}.$$

Here d is the metric on the leaves such that π_f is a local isometry. Since f is $C^{1+\alpha}$, because of the distortion lemma this defines an affine structure which is compatible with the smooth structure (i.e., h_L is a $C^{1+\alpha}$ diffeomorphism) and for which $\bar{f}$ becomes affine on each leaf. A periodic leaf L of period k contains a unique periodic point $\hat{q}$ of $\hat{f}$ that is mapped by π_f into a periodic point q of f of the same period k. The multiplication factor of $\hat{f}^k$, which is equal to $\frac{h_L(f^k(q_1)) - h_L(f^k(q))}{h_L(q_1) - h_L(q))}$ for any $q_1 \in L$ different from q, is equal to the eigenvalue of f at the periodic point q.

Now define a Riemann surface lamination $\tilde{\mathcal{L}}_f$ as follows. Let $\hat{\mathcal{L}}_f = \{(\underline{x}, v)\,;\ \underline{x} \in \mathcal{S}_f, v \in T_{\underline{x}} l_{\underline{x}}\}$. Here $l_{\underline{x}}$ denotes the leaf of $\mathcal{S}_f$ through $\underline{x}$ and $T_{\underline{x}} l_{\underline{x}}$ its tangent line at $\underline{x}$. We will give a Riemann surface structure on $\hat{\mathcal{L}}_f$ by covering it with two flow boxes. Let p be the fixed point of f and $\underline{x} \in \mathcal{S}_f \setminus \pi_f^{-1}(p)$. Let $\underline{x}_-$ and $\underline{x}_+$ be the only points on the leaf $l_{\underline{x}}$ such that the interval $(\underline{x}_-, \underline{x}_+)$ of $l_{\underline{x}}$ contains $\underline{x}$ and is mapped homeomorphically onto $S^1 \setminus \{p\}$ by π_f. Let $h_{\underline{x}}\colon l_{\underline{x}} \to \mathbb{R}$ be the unique $C^{1+\alpha}$ diffeomorphism that maps $\underline{x}_-$ to zero, $\underline{x}_+$ to one and is affine. We define $\Phi\colon \hat{\mathcal{L}}_f \setminus \pi_f^{-1}(p) \to \{0,1\}^{\mathbb{N}} \times (0,1) \times \mathbb{R}$ to be

$$\Phi(\underline{x}, v) = (\rho(\underline{x}), h_{\underline{x}}(\underline{x}), \pm|Dh_{\underline{x}}(\underline{x}) \cdot v|),$$

where we take the $+$ sign if $D\pi_f(\underline{x}) \cdot v$ is a positively orientated vector in S^1 and the $-$ sign otherwise. We get the other flow box by the same construction, using p' instead of p. The lamination $\tilde{\mathcal{L}}_f$ is defined as

$$\tilde{\mathcal{L}}_f = \{(\underline{x}, v) \in \hat{\mathcal{L}}_f\,;\ D\pi_f(\underline{x}) \cdot v > 0\}.$$

The restrictions of the above trivializations to $\tilde{\mathcal{L}}_f$ are the trivializations of $\tilde{\mathcal{L}}_f$. Since the Poincaré metric at the plaque that contains $\underline{x}$ is the pullback of the Poincaré metric of the upper half space by the map $Dh_{\underline{x}}$ we get the continuity of this metric.

Let $\tilde{F}\colon \tilde{\mathcal{L}}_f \to \tilde{\mathcal{L}}_f$ be the map

$$\tilde{F}(\underline{x}, v) = (\bar{f}(\underline{x}), D\bar{f}(\underline{x}) \cdot v),$$

where $D\bar{f}(\underline{x})$ is the derivative of the restriction of $\bar{f}$ to the leaf through $\underline{x}$. In terms of the local chart Φ, the expression of $\tilde{F}$ is an affine map on each leaf. Clearly $\tilde{F}$ is an extension of the lift $\bar{f}$. Let $\mathcal{L}_f = \tilde{\mathcal{L}}_f/\sim$ be the orbit space of $\tilde{F}$, i.e., the quotient space of $\tilde{\mathcal{L}}_f$ by the equivalence relation $\sim$ that identify two points if and only if they are in the same orbit of $\tilde{F}$. It is easy to see that $\mathcal{L}_f$ is a compact Riemann surface lamination. Each leaf of $\mathcal{L}_f$ is the image under the quotient map, of a leaf of $\tilde{\mathcal{L}}_f$. A leaf associated to leaf of $\tilde{L}_f$ which is periodic under $\tilde{F}$ is a cylinder. Furthermore, the conformal modulus of each cylindrical leaf determines the eigenvalues at the corresponding periodic point of f and vice versa. Clearly, all leaves are dense in $\mathcal{L}_f$.

Example. We will construct the same example as above from a different point of view. Let U be a simply connected domain and $F\colon U \to F(U) \supset U$ be a quadratic-like holomorphic map whose filled Julia set $J(F)$ is connected. As before, we consider the inverse limit

$$\tilde{\mathcal{L}}_F = \{\underline{z} = (\ldots, z_n, \ldots, z_0)\,;\ z_0 \in F(U) \setminus J(F),\ F(z_n) = z_{n-1}\},$$

the fibration $\pi_F\colon \tilde{\mathcal{L}}_F \to F(U) \setminus J(F)$ defined by $\pi_F(\underline{z}) = z_0$ and the natural extension $\tilde{F}\colon \pi_F^{-1}(U \setminus J(F)) \to \tilde{\mathcal{L}}_F$. The Riemann surface lamination we will actually consider is the space $\mathcal{L}_F$ of orbits of $\tilde{F}$. Before proving that this is a Riemann surface lamination and that it is holomorphically equivalent to the one of the previous example, we should note that $\mathcal{L}_F$ does not depend of F but only on the germ of F near the Julia set. Indeed, if $W \subset U$ is a smaller neighbourhood of $J(F)$ with $F(W) \supset W$ then $\mathcal{L}_{F|W}$ is the orbit space of $\tilde{F}$ restricted to $\pi_F^{-1}(F(W) \setminus J(F))$ and this is clearly equal to the orbit space of $\tilde{F}$. It is also easy to see that if two maps are holomorphically conjugate then the corresponding Riemann surface laminations are holomorphically equivalent, i.e., there exists a homeomorphism between them which is holomorphic on each leaf.

Let us prove that the new example is holomorphically equivalent to the previous one. Let $g\colon S^1 \to S^1$ be the analytic degree two expanding map associated to F as in Theorem 4.3, let A be an annular neighbourhood of S^1 in the complex plane and $G\colon A \to G(A) \supset A$ be the holomorphic extension of g. By Theorem 4.3, there exists a holomorphic conjugacy between $F|(U \setminus J(F))$ and $G|B$ where B is a component of $A \setminus S^1$. Let

$$\tilde{\mathcal{L}}_A = \{\underline{z} = (\ldots, z_n, \ldots, z_0)\,;\; z_0 \in G(A) \text{ and } G(z_n) = z_{n-1}\}.$$

Let π_G be the projection onto $G(A)$ and let $\tilde{G}$ be the natural extension of G. Notice that π_G is a local trivial fibration whose fibre is the diadic Cantor set. Indeed, consider the complex structure on the leaves so that π_G restricted to each leaf is a conformal map. Let us prove that the leaves of $\tilde{\mathcal{L}}_G$ are dense and that each leaf is conformally equivalent to a disc. To do this, we consider the following trivialization. Let $J \subset G(A)$ be a real analytic arc through the fixed point p of G such that J is transversal to the circle, $G(A) \setminus J$ is topologically a square and $G(J \cap A) = J$. This curve J can be obtained by taking a holomorphic linearization of G at the fixed point p and by taking a line which is invariant by the corresponding linear map. Let J_1 be the component of $G^{-1}(J)$ through the other preimage p_1 of p and denote by A_0 and A_1 the components of $A \setminus (J \cup J_1)$. In order to be definite, let A_0 be the component one meets first after moving from p in the positive orientation of the circle. The trivialization of $\tilde{\mathcal{L}}_A \setminus \pi_G^{-1}(J)$,

$$\Phi\colon \tilde{\mathcal{L}}_A \setminus \pi_G^{-1}(J) \to (G(A) \setminus J) \times \{0,1\}^{\mathbb{N}},$$

is given by $\Phi(\underline{z}) = (z_0, i(\underline{z}))$, where $i(\underline{z}) = (\ldots, i_n, \ldots, i_1)$ and $i_j = k \in \{0,1\}$ if and only if $z_j \in A_k$. Notice that if we take a curve $\underline{z}(t)$ in $\tilde{\mathcal{L}}_A$ such that $z_0(t)$ is a closed curve which starts in A_0 near J and goes once around the annulus A then $i(\underline{z}(1)) = \mathrm{add}(i(\underline{z}(0)))$ where $\mathrm{add}\colon \{0,1\}^{\mathbb{N}} \to \{0,1\}^{\mathbb{N}}$ is the adding machine. Since all orbits of the adding machine are dense, it follows that the leaves of $\tilde{\mathcal{L}}_A$ are dense. Also, each leaf of $\tilde{\mathcal{L}}_A$ is obtained from a countable number of copies of the square $G(A) \setminus J$ by gluing the boundaries via the adding machine. Hence, each leaf is homeomorphic to an open disc. Therefore each leaf is conformally equivalent to either a disc or the plane. As the restriction of π_F to a leaf is a holomorphic covering map of the annulus $G(A)$ which has finite modulus, the second alternative cannot occur. Thus, each leaf is conformally equivalent to a disc. So we may picture a leaf as an infinite strip, the preimage of J divides this strip in a countable number of squares and π_G maps each square holomorphically and bijectively onto the square $G(A) \setminus J$.

If we consider the Riemann surface lamination $\mathcal{L}_G$, i.e., the orbit space of $\tilde{G}$ restricted to $\pi_G^{-1}(B)$, then we get that $\mathcal{L}_F$ and $\mathcal{L}_G$ are conformally equivalent. So, we have to prove that $\mathcal{L}_G$ is conformally equivalent to the Riemann surface lamination $\mathcal{L}_g$ from Example 1. We will construct a global linearization of G via a direct limit construction on the inverse $\tilde{H}$ of $\tilde{G}$. To do this, we first note that $\tilde{H}\colon \tilde{\mathcal{L}}_G \to \tilde{\mathcal{L}}_G$ is just the shift map. It is continuous, one to one, maps leaves into leaves holomorphically but it is not onto since any point in the image of $\tilde{H}$ is mapped into A by π_G and not into $G(A)$. Next we consider the following direct limit system. For each integer i, let $\mathcal{L}_i$ be equal to a copy of $\tilde{\mathcal{L}}_A \times \{i\}$ and for $j \geq i$ let $H_{i,j} \colon \mathcal{L}_i \to \mathcal{L}_j$ be equal to $\tilde{H}^{j-i}$. Hence for $i \leq j \leq k$, $G_{j,k} \circ G_{i,j} = G_{i,k}$ and so we can define the direct limit $\hat{\mathcal{L}}$ of this system. In other words, $\hat{\mathcal{L}}$ is the quotient space of $\tilde{\mathcal{L}}_A \times \mathbb{N}$ by the equivalence relation that identifies (z,i) with (w,j) if and only if there exists $k \geq i,j$ such that $\tilde{H}^{k-i}(z) = \tilde{H}^{k-j}(w)$. Next define $\hat{H} \colon \hat{\mathcal{L}} \to \hat{\mathcal{L}}$ by $\hat{H}(z,i) = (\tilde{H}(z),i)$ and let $\pi_i \colon \mathcal{L}_i \to \hat{\mathcal{L}}$ be the inclusion map which sends a pair (z,i) to its equivalence class. We should note that $\pi_i(\mathcal{L}_i) \subset \pi_{i+1}(\mathcal{L}_{i+1})$. We have that $\hat{H}$ is well defined and onto. Moreover, since $\tilde{H}$ is injective, one also has that both π_i and $\hat{H}$ are injective. Hence $\hat{\mathcal{L}}$ has a unique Riemann surface lamination structure such that the $\pi_i \colon \mathcal{L}_i \to \hat{\mathcal{L}}$ are continuous embeddings and map leaves of $\mathcal{L}_i$ holomorphically into leaves of $\hat{\mathcal{L}}$. With this structure, $\hat{H}$ is a homeomorphism which maps leaves holomorphically onto leaves. It is easy to see that the leaves of $\hat{\mathcal{L}}$ are simply-connected (a non-null closed curve is a compact set and hence belongs to the image of π_i for i big enough; since the leaves of $\mathcal{L}_i$ are simply connected, this closed curve is homotopic to a point in this leaf and π_i pushes-forward this homotopy to a homotopy in the leaf of $\hat{\mathcal{L}}$).

Next we prove that the leaves of $\hat{\mathcal{L}}$ are holomorphically homeomorphic to the plane, and not to the disc. Let L be a leaf of $\hat{\mathcal{L}}$. Then $\pi_i(\mathcal{L}_i) \cap L$ is a strip L_i. One has $L_{i+1} \supset L_i$, $L = \cup L_i$ and the map $\phi_i = \pi_G \circ \pi_i^{-1} | L_i$ is a holomorphic covering map from L_i onto the annulus $G(A)$. Notice that $\phi_i^{-1}(J)$ is a sequence of lines that divides the strip L_i into squares that are mapped holomorphically, injective and onto $G(A) \setminus J$. Notice that the restriction of ϕ_{i+2} to each square of L_{i+1} maps this square injectively onto a component of $A \setminus J \cup J_1$. The restriction of ϕ_{i+2} to each square of L_i maps this square injectively onto a component of $G^{-1}(A) \setminus G^{-2}(J)$. Let Q_i and Q_i' be two consecutive square of L_i whose common boundary is mapped by ϕ_{i+2} onto $J_1 \cap G^{-1}(A)$. Let $Q_{i+1} \supset Q_i$ and $Q_{i+1}' \supset Q_i'$ denote squares of L_{i+1} and Q_{i+2} be the square of L_{i+2} that contains $Q_{i+1} \cup Q_{i+1}'$ as in Figure 6.1. Let A_i be the annulus $Q_{i+2} \setminus \operatorname{cl}(Q_i \cup Q_i')$. We have that the restriction of ϕ_{i+2} to each of these annuli A_i is a holomorphic bijection between A_i and a fixed annulus contained in $G(A) \setminus J$, namely, $(G(A) \setminus J) \setminus (B_1 \cup B_2)$, where B_i are the components of $G^{-1}(A) \setminus G^{-2}(J)$ whose boundaries intersect J_1. Hence the annuli A_i all have the same modulus. It is easy to see that we can construct a sequence of such annuli, $A_i, A_{i+3}, A_{i+6}, \ldots$ which gives a decomposition of L as a union of a disc and a sequence of annuli going around this disc. This shows that L cannot be holomorphic to a disc. Hence the leaves of $\hat{\mathcal{L}}$ are holomorphically homeomorphic to the plane and, therefore, have the induced affine structure. Since the restriction of $\hat{G}$ to each leaf is a holomorphic homeomorphism onto another leaf, it follows that this restriction is an affine map. Notice that if $\mathcal{S}_f \subset \tilde{\mathcal{L}}_A$ is the solenoid, then $\pi_i(\mathcal{S}_f) = \pi_j(\mathcal{S}_f)$, and therefore $\mathcal{S}_f$ is a subset of $\hat{\mathcal{L}}$. As the annulus A is symmetric, each leaf of $\mathcal{S}_f$ is an affine one-dimensional subspace of a leaf of $\hat{\mathcal{L}}$. This gives a new affine structure on the leaves of the solenoid that is invariant under the natural extension. By uniqueness it must coincide with the affine structure of the first example: the space of orbits of the restriction of $\hat{G}$ to the lamination $\hat{\mathcal{L}}_G = \cup_i \pi_i(\pi_G^{-1}(B) \times \{i\})$

is holomorphically equivalent to the orbit space of $\tilde{G}$ on $\pi_G^{-1}(B)$.

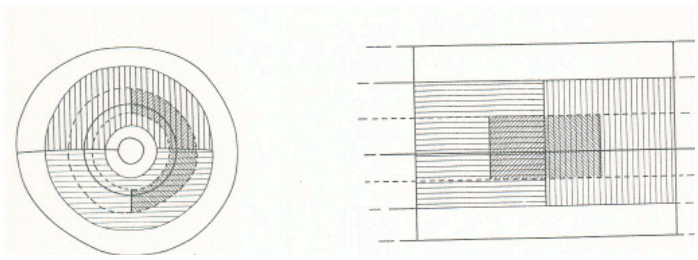

Fig. 6.1: The annuli A_i.

Definition. Let $\mathcal{L}$ be a Riemann surface lamination. We say that $\mathcal{L}$ has a *transversal measure* if there exist a covering U_i of $\mathcal{L}$ by flow boxes, $Z_i \colon U_i \to \mathbb{D} \times \Lambda_i$, and for each i a finite measure m_i in the transversal Λ_i such that the following holds. If $(z, \lambda) \mapsto (Z_{j,i}(z, \lambda), \Lambda_{j,i}(\lambda))$ is the overlapping map $Z_i \circ Z_j^{-1}$ then $\Lambda_{j,i}$ is a measure preserving local homeomorphism, i.e., if $A \subset \Lambda_j$ is in the domain of $\Lambda_{j,j}$ then $m_i(\Lambda_{j,i}(A)) = m_j(A)$. Similarly, a transversal measure class in the lamination is a choice of measures on the transversals of a finite cover by flow boxes such that the overlapping maps are absolutely continuous, i.e., they preserve the sets of zero measure in the transversal.

Remark. We claim that the laminations of the above examples do not have a transversal measure. We will just indicate how to prove this statement. A local transversal through a point of the lamination is a subset which contains this point and is mapped homeomorphically onto an open set of Λ by the composition of a flow box $Z \colon U \to \mathbb{D} \times \Lambda$ with the projection on the second factor. Let T be a local transversal of the lamination containing a point z and let γ be a closed curve in the leaf of the lamination through z. By covering γ by a finite number of flow boxes and using the overlapping maps between two consecutive intersecting flow boxes we can construct a local homeomorphism f of T satisfying the property that x and $f(x)$ are always in the same leaf and z is a fixed point of f. One can show that the germ of this map at z depends only on the homotopy class of the closed curve and all such germs form a group of transformation called the holonomy of the leaf. Now, if the lamination has a transversal measure, the corresponding measure at a local transversal is preserved by the holonomy group. In our examples a local transversal is homeomorphic to a diadic Cantor set and we can show that the group of holonomy of a cylindrical leaf is generated by the shift map. Since the shift map is expanding at a fixed point, the only invariant measure by this map is the Dirac measure at the fixed point. But this is not possible because each leaf is dense and the measure would not be finite.

In later sections we will consider transversal measure classes instead of transversal measures in our laminations.

Let us show that these laminations $\mathcal{L}_f$ can be used to describe the space of all expanding analytic maps f of the circle up to analytic conjugacy. We want to define a metric on this space via the space of Riemann surface laminations up to conformal equivalence (this notion will be defined presently). In order to show that something like this is possible fix an expanding circle map f_0 as above. Any other such map f is quasiconformally conjugate to f_0 (on a neighbourhood of the circle) and so there exists a quasiconformal homeomorphism $h_{f_0,f}$ which sends $\mathcal{L}_{f_0}$ to $\mathcal{L}_f$. As usual in Teichmüller theory we say that two Riemann surface laminations $\mathcal{L}_{f_1}$ and $\mathcal{L}_{f_2}$ are *conformally equivalent* if $h_{f_0,f_1} \circ h^{-1}_{f_0,f_2}$ is isotopic to a homeomorphism between $\mathcal{L}_{f_1}$ and $\mathcal{L}_{f_2}$ which sends leaves to leaves and which is conformal on each leaf. In the next theorem it is shown that the space of all expanding analytic maps f of the circle up to analytic conjugacy embeds in a one to one fashion to the space of Riemann surface laminations of the form $\mathcal{L}_f$ up in conformal equivalence.

Theorem 6.1. *Two analytic expanding maps $f_1, f_2\colon S^1 \to S^1$ of degree two are analytically conjugate if and only if the corresponding Riemann surface laminations $\mathcal{L}_{f_1}, \mathcal{L}_{f_2}$ are conformally equivalent.*

Proof. If f_1 and f_2 are analytically conjugate then this conjugacy can be lifted to a homeomorphism between $\mathcal{L}_{f_1}$ and $\mathcal{L}_{f_2}$ which is conformal on each leaf. On the other hand if these two Riemann surface laminations are conformally equivalent then this equivalence will send a leaf associated to a periodic orbit of f_1 to a leaf associated to the corresponding periodic orbit of f_2. As we remarked, above these leaves are cylinders and since the equivalence is conformal the conformal moduli of these cylinders coincide. It follows that the eigenvalues of corresponding periodic orbits of f_1 and f_2 are equal. By Shub and Sullivan (1986) this implies that f_1 and f_2 are analytically conjugate. □

Remark. If the f_i are merely $C^{1+\alpha}$, $0 < \alpha < 1$, then one can show that f_1 and f_2 are $C^{1+\alpha}$ conjugate if and only if the corresponding Riemann surface laminations $\mathcal{L}_{f_1}, \mathcal{L}_{f_2}$ are conformally equivalent.

Next we define a kind of Teichmüller theory on these Riemann surface laminations. To explain this carefully we define the objects in this theory simultaneously with the corresponding objects in the classical Teichmüller theory of Riemann surfaces.

6.2. Vector fields and Beltrami vectors on Riemann surfaces and on Riemann surface laminations

A continuous vector field V on a Riemann surface is given by specifying in each local chart $z_i\colon U_i \to \mathbb{C}$ a continuous function $V_i\colon z_i(U_i) \to \mathbb{C}$, called the expression of V in the chart z_i, such that, when $z_i\colon U_i \to \mathbb{C}$ and $z_j\colon U_j \to \mathbb{C}$ are two overlapping coordinate systems, then

$$V_j(z_{i,j}(z)) = \frac{dz_j}{dz_i}(z) \cdot V_i(z) \tag{6.1}$$

for all $z \in z_i(U_i)$ where $z_{i,j} = z_j \circ z_i^{-1}$ and $\frac{dz_j}{dz_i}$ denotes the derivative of $z_{i,j}$. This transformation rule is often indicated by writing $V(z)\frac{d}{dz}$.

A *vector field on a Riemann surface lamination* $\mathcal{L}$ associates to each flow box $Z_i\colon U_i \to \mathbb{D} \times \Lambda$ a function $V_i : \mathbb{D} \times \Lambda \to \mathbb{C}$ such that in overlapping flow boxes Z_i and Z_j we have

$$V_j(Z_{i,j}(z, \lambda)) = \frac{\partial Z_{i,j}}{\partial z} \cdot V_i(z, \lambda)$$

for all $z \in Z_i(U_i \circ U_j)$. In particular, restricted to each leaf it is a vector field. We say that it is *continuous* if it is continuous in each local chart. (This notion makes sense because the overlapping maps are holomorphic along leaves and continuous in the transversal direction.)

A *Beltrami vector* μ on a Riemann surface is represented, in each coordinate system $z_i\colon U_i \to \mathbb{C}$, by a measurable function $\mu_i\colon z_i(U_i) \to \mathbb{C}$ with the transformation rule

$$\mu_j(z) = \mu_i(z_{j,i}(z))\frac{\overline{dz_i/dz_j}\,(z)}{dz_i/dz_j\,(z)}$$

for each $z \in z_j(U_j)$ on overlapping domains of the charts $z_i\colon U_i \to \mathbb{C}$ and $z_j\colon U_j \to \mathbb{C}$. Here $z_{j,i} = z_i \circ z_j^{-1}$ and $dz_i/dz_j = \dfrac{d}{dz}(z_{j,i})$ where it is defined. In other words, given a point $p \in U$ we have

$$\mu_j(z_j(p)) = \rho_{i,j}(p) \cdot \mu_i(z_i(p)), \tag{6.2}$$

where

$$\rho_{i,j}(p) = \overline{\left[\frac{dz_i}{dz_j}\right]} \cdot \left[\frac{dz_i}{dz_j}\right]^{-1}$$

has norm one. This transformation rule can be abbreviated by writing the expression $\mu\dfrac{d\bar{z}}{dz}$. From (6.2) it follows that $|\mu(p)| = |\mu_i(z_i(p))|$ is well defined because in each chart this number is the same. In particular,

$$||\mu||_\infty = \text{ess sup}|\mu(p)|$$

is well defined (here ess sup stands for the essential supremum and is defined to be the infimum of all $k > 0$ such that the Lebesgue measure of $\{p\,;\ |\mu(p)| > k\}$ is zero). If such a differential μ is in L^∞ (i.e., if $||\mu||_\infty < \infty$) then μ is called a *Beltrami vector* and if $||\mu||_\infty < 1$ then it is called a *Beltrami coefficient*. The corresponding spaces are denoted by $L^\infty(S)$ respectively $L^\infty_1(S)$. We shall endow $L^\infty(S)$ with the norm $||.||_\infty$ is a Banach space.

If $g\colon \tilde{S} \to S$ is a holomorphic covering map we can define the pullback of Beltrami vectors on S in the following way. If μ is a Beltrami vector on S

then $g^*\mu$ is defined to be the Beltrami vector on $\tilde{S}$ whose local expression in a local chart $\tilde{z}\colon \tilde{U} \to \mathbb{D}$ is the same as the local expression of μ in the local chart $\tilde{z} \circ (g|\tilde{U})^{-1}$, where $\tilde{U}$ is such that the restriction of g to $\tilde{U}$ is a conformal homeomorphism. In particular, if $\pi\colon \mathbb{D} \to S$ is a holomorphic covering map then $\pi^*\mu$ is an essentially bounded measurable function ν on the unit disc that satisfy the invariance condition $\nu(x) = \nu(A(x)) \cdot \frac{\overline{A'(x)}}{A'(x)}$ for all A in the group of deck transformations Γ of the covering map π. Conversely, any essentially bounded function with the above invariance condition defines a Beltrami vector on S. Hence, if F is a fundamental domain for the action of the group of deck transformations on $\mathbb{D}$ we can identify $L^\infty(S)$ with $L^\infty(F)$ since any essentially bounded measurable function on F can be extended to a function on $\mathbb{D}$ by using the above invariance condition.

Given two Beltrami coefficients μ and ν we can define their hyperbolic distance as follows: because of (6.2), the hyperbolic distance between μ and ν at a point $p \in S$ is independent of the chart. So let $h(\mu(p), \nu(p))$ be this hyperbolic distance and define the hyperbolic distance, $h(\mu, \nu)$, between μ and ν, to be the essential supremum of $h(\mu(p), \nu(p))$. To each Beltrami coefficient μ and each Beltrami vector ν we associate a *Beltrami path in the direction of* ν as follows. In each coordinate system z_i the Beltrami coefficient $\mu_i(z_i)$ is almost everywhere a well defined point in the Poincaré disc and $\nu_i(z_i)$ is a well defined vector in $\mathbb{C}$. Now let $\mu_{i,t}$ be such that $\mu_{i,0} = \mu_i$, so that for Lebesgue almost all p the arc $t \mapsto \mu_{i,t}(z_i(p)) \in \mathbb{D}$ has constant speed and goes along a geodesic in the Poincaré metric on $\mathbb{D}$ and such that $\dfrac{d}{dt}\mu_{i,t}(z_i(p))$ at $t = 0$ is equal to $\nu_i(z_i(p))$ (for almost all p). Because of (6.2) this path does not depend on the choice of the coordinate system and so we get a well defined deformation of μ. For example, if $\mu = 0$ then

$$\mu_{i,t}(z_i(p)) = \frac{e^{|\nu_i(z_i(p))|t} - 1}{e^{|\nu_i(z_i(p))|t} + 1} \cdot \frac{\nu_i(z_i(p))}{|\nu_i(z_i(p))|}$$

defines a Beltrami coefficient for each $t \geq 0$. Next we show that if $g\colon S \to S_1$

is a quasiconformal homeomorphism, then it induces a pullback map g^* from the Beltrami coefficients of S_1 to the Beltrami coefficients of S that preserves the hyperbolic distance and maps Beltrami paths into Beltrami paths. This construction will follow from the a simple formula concerning quasiconformal maps between open sets of the complex plane. Let f, g be such maps. The Beltrami coefficient of f is defined to be

$$\mu_f(z) = \frac{\bar{\partial} f(z)}{\partial f(z)}.$$

From the chain rule we get

$$\mu_{f\circ g}(z) = \frac{\mu_g(z) + \mu_f(g(z)) \cdot p_g(z)}{1 + \mu_f(g(z)) \cdot \overline{\mu_g(z)} \cdot p_g(z)} \qquad (6.3)$$

where $p_g(z) = \frac{\overline{\partial g(z)}}{\partial g(z)}$. Notice that $p_g(z)$ has absolute value one. If $g\colon U \to V$ is a quasiconformal homeomorphism between open sets of the plane and μ is a Beltrami coefficient on V then, by the Measurable Riemann Mapping Theorem, there exists a quasiconformal homeomorphism $f\colon V \to W$, whose Beltrami coefficient is μ. We then define $g^*(\mu)$ as the Beltrami coefficient of $f \circ g$. From (6.3) it follows that $g^*\colon L_1^\infty(W) \to L_1^\infty(V)$ satisfies the required conditions since the mapping

$$u \in \mathbb{D} \mapsto \frac{\mu_g(z) + u \cdot p_g(z)}{1 + u \cdot \overline{\mu_g(z)} \cdot p_g(z)}$$

is a hyperbolic isometry. To define g^* for Riemann surfaces we just take charts.

One of the essential ideas in Teichmüller theory is that each Beltrami coefficient μ defines, via the Measurable Riemann Mapping Theorem, a complex structure S_μ on a Riemann surface S. A deformation of the Beltrami coefficient as we defined above, corresponds to deformations of the initial complex structure. (This deformation might be constant as is the case for paths of trivial Beltrami coefficients defined below.) Let us be more specific about this. As a topological space, S_μ is equal to S. The charts of the new conformal structure are obtained from the charts of S as follows. If $z_i\colon U_i \to \mathbb{C}$ is a chart in S and $\mu_i\colon z_i(U_i) \to \mathbb{C}$ is the expression of the Beltrami coefficient μ then, by the Measurable Riemann Mapping Theorem, there exists a quasiconformal homeomorphism $h_i\colon z_i(U_i) \to h_i(z_i(U_i)) \subset \mathbb{C}$ whose Beltrami coefficient is μ_i. The chart of S_μ is $h_i \circ z_i$. It is easy to see that the overlapping maps of the new charts are holomorphic and hence define a new complex structure. The identity map on S, considered as a map $g\colon S \to S_\mu$, is quasiconformal since its local expressions with respect to the above charts are precisely the h_i. Furthermore, from the above definitions we see that $\mu = g^*(0)$, i.e., μ is the pullback of the Beltrami coefficient zero of S_μ. From the above discussion we see that a complex structure on S is a new Riemann surface S_1 together with a conformal map $F_1\colon S \to S_1$ or the Beltrami coefficient $\mu = F_1^*(0)$. We say that the complex structures $F_i\colon S \to S_i$, $i = 1, 2$, are *equivalent in the sense of Teichmüller*, if there exists a conformal map $F\colon S_1 \to S_2$ such that $F \circ F_1$ is homotopic to F_2. The Teichmüller space of S, $T(S)$, is the set of equivalence classes of conformal structures on S. Alternatively, we say that two Beltrami coefficients of S, μ_1 and μ_2 are *equivalent* if there exists a quasiconformal homeomorphism $g\colon S \to S$, homotopic to the identity, such that $g^*(\mu_2) = \mu_1$. From the above discussion it is easy to conclude that the Teichmüller space of S is again the quotient space of $L_1^\infty(S)$ by the above equivalence relation. The quotient map $Q\colon L_1^\infty(S) \to T(S)$ projects Beltrami paths into curves of the Teichmüller space that we also call Beltrami paths. The Beltrami coefficient that are equivalent to 0 are called the *trivial Beltrami coefficient*. So, a Beltrami path does not deform a complex structure, in the sense of Teichmüller, if its Beltrami coefficients are trivial. If $g\colon S \to S$ is a quasiconformal homeomorphism, the conformal distortion of g

with respect to the complex structures μ_1 and μ_2 is defined to be

$$K_g = \text{ ess sup}_z K_g(z),$$

where

$$K_g(z) = \exp(h(\mu_1(z), g^*(\mu_2(z)))$$

and h is the Poincaré metric defined above. The Teichmüller distance between the equivalence class of μ_1 and the equivalence class of μ_2 is defined as the infimum of $\log K_g$ over all quasiconformal homeomorphisms $g\colon S \to S$ that are isotopic to the identity. This clearly defines a metric on $T(S)$. One of the main theorems we will prove in section 7 states that if a Beltrami path is an almost geodesic between two points with respect to the above metric it remains almost a geodesic up to a definite distance. In particular we recover a classical result in Teichmüller theory: if the Beltrami path is a geodesic between two points it remains a geodesic between any other two points.

A Beltrami vector is *infinitesimally trivial* if it can be written as $\bar\partial V$ where V is a continuous vector field. Here $\bar\partial V \in L^\infty$ is defined in the distribution sense. In other words, in local coordinates $z_i\colon U_i \to z_i(U)$, $\bar\partial V$ is the function $\bar\partial V_i \in L^\infty$ which satisfies

$$\int (\bar\partial V_i)\phi = \int V_i \cdot \bar\partial \phi$$

for all smooth functions ϕ with compact support in $z_i(U_i)$. Here $\bar\partial$ is defined as in Section III.1, see also the appendix. In particular, it follows that in this case $\bar\partial V$ is indeed a Beltrami vector. In order to motivate these notions let us state the following proposition (which will not be needed).

Proposition 6.1. *On a compact Riemann surface a Beltrami vector is infinitesimally trivial if and only if it can be written as* $\nu = \frac{d}{dt}\mu_t$ *where* μ_t *is a path of trivial Beltrami coefficients depending smoothly on* t.

Proof. Because this proposition is only meant as a motivation for the definition of the notion of infinitesimally trivial Beltrami vectors and is not actually used we shall not give a complete proof here, see, for example, Gardiner (1987, p. 107). We can also prove the proposition by observing that a continuous vector field, whose $\bar\partial$ derivative in the distribution sense is in L^∞, has a modulus of continuity $\delta \log \delta$. This is enough to guarantee the uniqueness of solutions for the corresponding ordinary differential equation. Thus it generates a flow which, in the case of a compact surface, is defined for all time. By a careful approximation by smooth flows one can show that for small values of t, the flow of this smooth approximation is quasiconformal (with constants bounded by the L^∞ norm of the $\bar\partial$ derivative of the smooth approximation). Hence one can use the compactness of the space of quasiconformal homeomorphisms with a given constant to conclude that the flow of the original continuous vector field is quasiconformal. From this the proposition follows. □A *Beltrami vector on a*

Riemann surface lamination $\mathcal{L}$ is represented, in each *flow box* $Z_i\colon U_i \to \mathbb{D}\times\Lambda$, by a measurable function $\mu_i\colon Z_i(U_i) \to \mathbb{C}$ with the following transformation rule. If $Z_i \circ Z_j^{-1}(z,\lambda) = (Z_{j,i}(z,\lambda), \Lambda_{j,i}(\lambda))$ then

$$\mu_j(z,\lambda) = \frac{\overline{\partial Z_{j,i}/\partial z\,(z,\lambda)}}{\partial Z_{j,i}/\partial z\,(z,\lambda)} \cdot \mu_i(Z_i \circ Z_j^{-1}(z,\lambda)) \tag{6.4}$$

for all $(z,\lambda) \in Z_j(U_i)$ and for each pair of overlapping flow boxes $Z_i\colon U_i \to \mathbb{D}\times\Lambda$ and $Z_j\colon U_j \to \mathbb{D}\times\Lambda$. Moreover, in each such flow box it is weakly continuous in the leaf direction: if $\lambda_n \to \lambda_\infty$ then

$$\lim_{n\to\infty} \int \mu_i(z,\lambda_n)\phi\,dz\,d\bar z = \int \mu_i(z,\lambda_\infty)\phi\,dz\,d\bar z$$

for each $\phi \in L^1(\mathbb{D})$. The space of such vectors endowed with the $||\quad||_\infty$ norm is denoted by $L^\infty(\mathcal{L})$. If V is a continuous vector field for which $\bar\partial V$ (where the derivatives are taken in the sense of distributions) is a well defined L^∞ function in each local coordinate system, then it is easy to show that $\bar\partial V$ is a Beltrami vector. As before Beltrami vectors which can be written in this form will be called *infinitesimally trivial*.

If we have a Beltrami coefficient μ whose local expression $\mu_i\colon \mathbb{D}\times\Lambda \to \mathbb{C}$, in each flow box $Z_i\colon U_i \to \mathbb{D}\times\Lambda$ is such that $\lambda \mapsto \mu_i(\cdot,\lambda) \in L^\infty(D)$ is continuous in the L^∞ topology, we can use the Ahlfors-Bers Measurable Riemann Mapping Theorem, and construct a homeomorphisms $H_i\colon \mathbb{D}\times\Lambda \to \mathbb{D}\times\Lambda$ by integrating each Beltrami coefficient on $\mathbb{D}\times\{\lambda\}$ (this theorem guarantees the continuity with respect to λ of the quasiconformal homeomorphisms of the disc). Hence, the flow boxes $H_i \circ Z_i$ define a new Riemann surface lamination. However, if the Beltrami coefficients are only weakly continuous with respect to the transversal parameter we cannot guarantee the continuity of H_i in λ. So it seems that we have more Beltrami coefficients than complex structures on a Riemann surface lamination: some Beltrami coefficients can be integrated and give rise to new complex structures on a lamination while some others cannot! However, in our context we shall be able to forget about the complex structure of the lamination and just work with Beltrami coefficients and vectors.

6.3. Quadratic differentials on Riemann surfaces and Riemann surface laminations

A *quadratic differential on a Riemann surface* is given in local coordinates $z_i\colon U_i \to z_i(U_i)$ by integrable complex valued functions $\phi_i\colon z_i(U_i) \to \mathbb{C}$ such that

$$\phi_i(z) = \phi_j(z_{i,j}(z))(\frac{dz_j}{dz_i}(z))^2 \tag{6.5}$$

for each $z \in z_i(U_i)$, where $z_{i,j} = z_j \circ z_i^{-1}$ and $\dfrac{dz_j}{dz_i} = \dfrac{dz_{i,j}}{dz}$. In other words, these differentials are of the form $\phi\,dz^2$. To such a quadratic differential one can associate a measure $|\phi|$. Indeed, since $|\phi_i(z_i)| = |\phi_j(z_j)| \cdot |dz_j/dz_i|^2$ and $|dz_j/dz_i|^2$ is the Jacobian of this overlap map, $|\phi|$ defines a surface element and therefore a measure. Similarly, $\sqrt{|\phi|}$ defines a metric (a line element). A quadratic differential is holomorphic (meromorphic) if all its local expressions are holomorphic (meromorphic) functions. If the quadratic differential ϕ is holomorphic and non-zero in a neighbourhood of a point, we can find a chart in this neighbourhood such that $\phi = dz^2$, i.e., its local expression is identically equal to one in this chart. In such a coordinate system the metric $\sqrt{|\phi|}$ is just the euclidean metric and its geodesics are the straight lines. Some of the geodesics of ϕ play a special role: a parametrized curve γ is called a *horizontal (vertical) trajectory* of ϕ if $\phi(\gamma(t))\gamma'(t)^2 > 0$ (respectively $\phi(\gamma(t))\gamma'(t)^2 < 0$) in local coordinates. These horizontal and vertical trajectories define two transverse foliations in the complement of the zeros of ϕ. If ϕ has a zero of order $m \geq -1$ in p then $m+2$ horizontal trajectories and also $m+2$ vertical trajectories emanate from p. Trajectories which enter a singular point are called *separatrices*.

Since $|\phi|$ defines a measure on S we can define

$$||\phi|| = \int_S |\phi|.$$

Quadratic differentials for which $||\phi|| < \infty$ are called *integrable* and the corresponding space is denoted by $L^1(S)$. As in the case of Beltrami differentials, we can define the pullback of the quadratic differentials via a holomorphic covering map $\pi\colon \mathbb{D} \to S$. Associated to each quadratic differential ϕ on S is a locally integrable function $\pi^*\phi = \tilde{\phi}\colon \mathbb{D} \to \mathbb{C}$ which satisfies the invariance condition: $\tilde{\phi}(A(z))(A'(z))^2 = A(z)$ for all elements A of the group of deck transformations Γ. However, $\tilde{\phi}$ is not integrable over $\mathbb{D}$: the integral (Lebesgue) of $|\tilde{\phi}|$ over a fundamental domain F of Γ coincides with the integral of $|\phi|$ over S. Conversely, any integrable function on the fundamental domain F generates a function with the above invariance condition and, therefore, a quadratic differential on S. Thus, $L^1(S)$ can be identified with $L^1(F)$.

One of the main features of the integrable quadratic differentials is the natural duality between Beltrami vectors and quadratic differentials defined below. If μ is a Beltrami vector and ϕ is a quadratic differential then $\mu\phi$, which in local charts is just the product of the expression of μ by the expression of ϕ, is a measure on S. Therefore, the following bilinear map is well defined:

$$L^\infty(S) \times L^1(S) \to \mathbb{C}, \quad (\mu, \phi) \mapsto \int_S \mu\phi$$

If $\pi\colon \mathbb{D} \to S$ is a holomorphic covering map, $\tilde{\mu} = \pi^*\mu$ and $\tilde{\phi} = \pi^*\phi$ we have that $\int_S \mu\phi = \int_F \tilde{\mu}\tilde{\phi}$, where F is a fundamental domain of the group of deck transformations. Hence the above duality is precisely the classical duality between L^∞ and L^1. This duality will play a fundamental role in the characterization of the infinitesimal trivial Beltrami differentials in subsection 6.5.

The notion of *quadratic differential on a Riemann surface lamination* $\mathcal{L}$ is somewhat more tricky. The definition is such that a sort of duality will be established similar to the classical case explained above. Up to now, all the objects defined on Riemann surfaces and on Riemann surface laminations are specified by their local expressions which are functions, defined on open sets, satisfying some invariance condition with respect to the overlapping maps. A quadratic differential on a Riemann surface lamination is defined by specifying two objects: a transversal invariant measure class and the local expressions that are functions depending on a choice of a measure in the measure class. Let us describe this notion more carefully. Firstly, a quadratic differential associates a measure class to each local transversal in such a way that this measure class is invariant under the change of coordinate systems. More precisely, let $Z_i\colon U_i \to \mathbb{D}\times\Lambda$, $Z_j\colon U_j \to \mathbb{D}\times\Lambda$ be two flow boxes and denote by $(\Lambda, [m])$. the space Λ with measure class $[m]$. Then there are measure classes $[m_i]$ and $[m_j]$ defined on each of these transversals Λ such that the function $\Lambda_{i,j}\colon (\Lambda, [m_i]) \to (\Lambda, [m_j])$ defined by

$$Z_j \circ Z_i^{-1}(z,\lambda) = (Z_{i,j}(z,\lambda), \Lambda_{i,j}(\lambda))$$

is absolutely continuous. Secondly, given a measure m_i in the measure class $[m_i]$ the quadratic differential is represented by *a measurable function $\phi_i\colon \mathbb{D}\times\Lambda \to \mathbb{C}$, (which depends on the choice of the measure) such that the following holds. In overlapping coordinate systems Z_i and Z_j one has

$$\phi_i(z,\lambda) = \phi_j\left(Z_{i,j}(z,\lambda), \Lambda_{i,j}(\lambda)\right) \cdot \left[\frac{\partial Z_{i,j}(z,\lambda)}{\partial z}\right]^2 \cdot Jac(\Lambda_{i,j})(\lambda), \tag{6.6}$$

where $Jac(\Lambda_{i,j})$ is the Jacobian of $\Lambda_{i,j}$ with respect to the measures m_i and m_j. Moreover, the functions ϕ_i are integrable with respect to the product measure of the Lebesgue measure on the disc $\mathbb{D}$ and the measure m_i in the transversal. If we pick two measures in this measure class, the local expression of the quadratic differential is multiplied by the Radon-Nikodym derivative of the first measure with respect to the second one. Because of this, integration in local charts can be defined as the usual Lebesgue integral of the local expression $|\phi_i|$ of ϕ with respect to the measure $dzd\bar{z}dm_i$. We shall denote the measure associated to ϕ by $|\phi|$. If $\int_{\mathcal{L}} d|\phi|$ is finite then we say that ϕ is an *integrable quadratic differential*. A quadratic differential is called *holomorphic* if it is holomorphic on almost all leaves (w.r.t. to the measure class). We will prove the existence of holomorphic quadratic differentials on Riemann surface laminations in the next subsection.

Note that a quadratic differential on a Riemann surface lamination does not satisfy any continuity requirements in the transversal direction. On each leaf which is in the support of the transverse measure class of a holomorphic quadratic differential, we have horizontal and vertical trajectories as before. Because the local expression ϕ_j depends on the choice of the measure in the measure class, a holomorphic quadratic differential on a Riemann surface lamination does not define a line element on the leaves. This will create some fun in the next section.

As in the case of Riemann surfaces, there is a pairing $L^\infty(\mathcal{L}) \times L^1(\mathcal{L}) \to \mathbb{C}$ between Beltrami vectors and quadratic differentials. Indeed, if $\mu \in L^\infty(\mathcal{L})$

and $\phi \in L^1(\mathcal{L})$ then we obtain a (complex) measure $\mu\phi$ by defining the measure of a subset A of the domain of a flow box $Z_i : U_i \to \mathbb{D} \times \Gamma$ to be equal to $\int_{Z_i(A)} \mu_i \phi_i dz d\bar{z} dm_i$. It follows that this number does not depend on the choice of flow boxes and of the measure m_i in the measure class defined by ϕ. Also, $(\mu, \phi) \mapsto \int_{\mathcal{L}} d(\mu\phi)$ is a well defined bilinear map. We do not claim that $L^\infty(\mathcal{L})$ is the dual space to $L^1(\mathcal{L})$ as is the case for Riemann surfaces. However, the examples discussed in the next subsection show that the functionals of $L^\infty(\mathcal{L})$ defined by this pairing separate points in $\mathcal{L}^\infty(\mathcal{L})$. This is enough for the characterization of the infinitesimally trivial Beltrami differentials in subsection 6.5.

Let us finish this subsection by discussing some differences between quadratic differentials on Riemann surfaces and on Riemann surface laminations. Suppose that we are in the case of a Riemann surface S. Then a quadratic differential ϕ defines a Beltrami vector $\dfrac{|\phi|}{\phi}$. Now consider the Beltrami path $t \mapsto \mu_t$ starting at μ with Beltrami vector $\dfrac{|\phi|}{\phi}$. Therefore, given a quasiconformal homeomorphism Ψ of S and a quadratic differential ϕ we can consider the *conformal distortion* of Ψ between the original complex structure of S and the complex structure obtained by stretching the original structure by a factor e^s in the direction of ϕ. This is defined to be

$$K(z) = \exp(h(\mu_0(z), \Psi^*(\mu_s)(z))), \tag{6.7}$$

where μ_t is the Beltrami path determined by ϕ and h is the hyperbolic metric on $\mathbb{D}$.

In the Riemann surface lamination case a quadratic differential ϕ does not define such a Beltrami vector. Indeed, $\dfrac{|\phi(z)|}{\phi(z)}$ need not be a Beltrami vector; indeed, it need not even be defined on every leaf and also it is not continuous with respect to the transversal in the weak topology. However, it does transform like a Beltrami vector when we change coordinates and hence it defines a Beltrami vector on almost every leaf of the lamination (with respect to the transversal measure class). So a quadratic differential ϕ on a Riemann surface lamination $\mathcal{L}$ still defines a Beltrami path $t \mapsto \mu_t$ starting at μ with Beltrami vector $\frac{|\phi|}{\phi}$ not on the whole lamination but on almost every leaf (with respect to the measure class corresponding to ϕ). Therefore, the distortion $K(z)$ of a quasiconformal map Ψ between the structure μ and the stretched structure μ_s is defined as above for almost every z (with respect to the measure $|\phi|$). It follows that the expression $\int K(z) d|\phi|$ is well defined. We shall need this expression in the next section.

6.4. Examples of quadratic differentials

We have seen how to define the pullback Beltrami vectors by holomorphic covering maps. In general, if a map is not invertible, we cannot push-forward differentials: some compatibility condition must be satisfied in order to obtain a differential on the range of the map that corresponds to a differential on the domain. However, we can always push-forward measures: the measure of a subset of the range is defined to be equal to the measure of its pre-image. In this subsection we will show that we can define a similar operation for quadratic differentials and that this operation respects the pairing between quadratic differentials and Beltrami differentials.

Let us show that each quadratic differential on $\mathbb{D}$ defines a quadratic differential on a hyperbolic Riemann surface or on a hyperbolic Riemann surface lamination.

Let us first do this for a Riemann surface. Let $\pi\colon \mathbb{D} \to S$ be a holomorphic covering map. To this map we can associate a map $\pi_*\colon L^1(\mathbb{D}) \to L^1(S)$ associating to each quadratic differential ϕ on $\mathbb{D}$ a quadratic differential $\pi_*(\phi)$ on S satisfying

$$\int_{\mathbb{D}} (\pi^*\mu) \cdot \phi = \int_S \mu \cdot (\pi_*\phi) \tag{6.8}$$

for all Beltrami vectors $\mu \in L^\infty(S)$, where $\pi^*\mu$ is the pullback of μ. Indeed, let $\tilde{\phi}(z) = \sum_{A\in\Gamma} \phi(A(z)) \cdot (A'(z))^2$ where Γ is the group of deck transformations (this is called the Poincaré theta series). Let F be a fundamental domain of the action of Γ. Since for every $A \in \Gamma$, $\int_F |\phi(A(z))||A'(z)|^2 dz d\bar{z} = \int_{A(F)} |\phi(w)| dw d\bar{w}$ and $\sum_{A\in\Gamma} \int_{A(F)} |\phi(w)| dw d\bar{w} = \int_{\mathbb{D}} |\phi| < \infty$, it follows that the above series converges absolutely almost everywhere and $\int_F |\tilde{\phi}| dz d\bar{z} \le \int_{\mathbb{D}} |\phi| dz d\bar{z}$. Since $\tilde{\phi}$ obviously satisfy the required invariance condition, it defines an integrable quadratic differential $\pi_*\phi$ on S. Let us prove that $\pi_*\phi$ satisfy (6.8). Indeed, the function $\tilde{\mu} = \pi^*\mu$ satisfy the invariance condition $\tilde{\mu}(A(z) = \dfrac{\overline{A'(z)}}{A'(z)}\mu(z)$ for all $A \in \Gamma$. Hence,

$$\begin{aligned}\int_{\mathbb{D}} \phi\tilde{\mu} &= \sum_{A\in\Gamma} \int_{A(F)} \phi(w)\tilde{\mu}(w) dw d\bar{w} = \\ &= \sum_{A\in\Gamma} \int_A \phi(A(z))\tilde{\mu}(z)(A'(z))^2 dz d\bar{z} = \\ &= \int_F (\tilde{\phi}(z) \cdot \tilde{\mu}(z) dz d\bar{z} = \int_S (\pi_*\phi)\mu.\end{aligned}$$

One can prove that if ϕ is holomorphic then $\pi_*\phi$ is a holomorphic quadratic differential, see Lehto (1987, p. 223).

Similarly, we can construct quadratic differentials with support on a leaf L of a Riemann surface lamination $\mathcal{L}$. First, take as transversal measure class the class of atomic measures on the transversals with support on a leaf L. Next take a classical quadratic differential ϕ_L on the leaf L. Using this differential we want to construct a quadratic differential ϕ on $\mathcal{L}$. Let us describe its expression in a flow box $Z_i\colon U_i \to \mathbb{D} \times \Lambda$. For each λ such that $Z_i^{-1}(\mathbb{D} \times \{\lambda\})$ is a plaque of the leaf L, let $\phi_{L,i}(z, \lambda)$ be the local expression of ϕ_L in the chart Z_i evaluated at (z,λ). Let Λ_L be the collection of $\lambda \in \Lambda$ for which $h(\mathbb{D}\times\{\lambda\}) \subset L$. Since L is a single leaf, Λ_L is countable. So write $\Lambda_L = \{\lambda_i\}_{i\in\mathbb{N}}$ and define a measure m_i on Λ by choosing a sequence $a_i > 0$ with $\sum_{i\in\mathbb{N}} a_i = 1$ and take $m_i = \sum_{i\in\mathbb{N}} \delta_{\lambda_i}$, where δ_λ denotes the Dirac measure on Λ with support $\{\lambda\}$. After choosing this measure, the local expression of ϕ is given by

$$\phi_i(z,\lambda) = \begin{cases} \dfrac{1}{a_j}\phi_{L,i}(z,\lambda) & \text{if } \lambda = \lambda_j \in \Lambda_L \\ 0 & \text{otherwise.} \end{cases}$$

It is clear that this defines a quadratic differential ϕ on $\mathcal{L}$ and that $\int_{\mathcal{L}} |\phi|$ is equal to the integral of $|\phi_L|$ over L. Thus we have constructed a push-forward map which we shall denote by

$$(i_L)_*\colon L^1(L) \to L^1(\mathcal{L})$$

such that

$$\int_{\mathcal{L}} \mu((i_L)_*\phi) = \int_L ((i_L)^*\mu)\phi \tag{6.9}$$

for each $\mu \in L^\infty(\mathcal{L})$. Here $i_L\colon L \to \mathcal{L}$ is the inclusion of the leaf L and $(i_L)^*\colon L^\infty(\mathcal{L}) \to L^\infty(L)$ is the map that restricts Beltrami vectors of the lamination to the leaf L.

6.5 Infinitesimally trivial Beltrami coefficients on Riemann surfaces and on Riemann surface laminations

The next theorem holds for both Riemann surfaces and Riemann surface laminations and is the main result of this section.

Theorem 6.2. *Let S be a compact Riemann surface or a compact hyperbolic Riemann surface lamination. Then one has the following properties.*

1. *The set of infinitesimally trivial Beltrami vectors is closed in the weak topology on $L^\infty(S)$ (this means that $\mu_n \to \mu$ when $\int \mu_n\phi \to \int \mu\phi$ for all quadratic differentials in $L^1(S)$).*

2. *Two Beltrami vectors μ,ν differ by a trivial deformation, i.e., $\mu-\nu = \bar\partial V$ for some continuous vector field V on S if and only if $\int \mu\phi = \int \nu\phi$ for all holomorphic quadratic differentials $\phi \in L^1(S)$.*

Proof. Let us first show that 1) implies 2). Consider the space of all infinitesimally trivial Beltrami vectors and the pairing map

$$L^\infty \times L^1 \to \mathbb{C}.$$

Since we have assumed that Statement 1) holds, the trivial vectors form a closed set with respect to the weak topology on L^∞ and therefore we get by Hahn-Banach that there exists a closed subspace $Z \subset L^1$ such that $\mu - \nu$ is infinitesimally trivial if and only if $\int \mu\phi = \int \nu\phi$ for all $\phi \in Z$. So it remains to show that all forms in Z are holomorphic. But this is clear because if $\phi \in Z$ then $\int \bar{\partial} V \phi = 0$ for all V. In particular, if we take a smooth vector field V with compact support contained in some flow box one has $\int \bar{\partial} V \phi = \int V \bar{\partial}\phi$ and since this holds for every such V one gets $\bar{\partial}\phi = 0$. Hence, by Weyl's Lemma, which states that if the $\bar{\partial}$ derivative of a function is zero in the distribution sense then the function is equal almost everywhere to a holomorphic function, one has that ϕ is holomorphic.

So it remains to be shown that the space of all trivial Beltrami vectors forms a closed set. Take a sequence μ_n of trivial Beltrami vectors converging weakly to a vector μ. Then μ_n can be written as $\mu_n = \bar{\partial} V_n$ where V_n is a continuous vector field on S. If we are working on a Riemann surface then we lift everything to the universal cover. Thus we get a sequence $\tilde{\mu}_n, \tilde{V}_n \colon \mathbb{D} \to \mathbb{C}$ on the universal cover such that $\tilde{\mu}_n = \bar{\partial}\tilde{V}_n$. By Green-Stokes' Theorem

$$\tilde{V}_n(z) = \int_{\mathbb{D}} \frac{\bar{\partial}\tilde{V}_n(\omega)}{w - z}\, dw\, d\bar{w} + \int_{\partial\mathbb{D}} \frac{\tilde{V}_n(\omega)}{w - z}\, dw\,.$$

But since $\tilde{V}_n$ is the lift of a continuous vector field V_n on S, $\tilde{V}_n$ is bounded in the Poincaré metric of $\mathbb{D}$ and therefore, its Euclidean norm goes to zero at the boundary of $\mathbb{D}$. Consequently,

$$\tilde{V}_n = \int_{\mathbb{D}} \frac{\tilde{\mu}_n(\omega)}{w - z}\, dw\, d\bar{w}.$$

Now we need the construction from (6.4): there exists a push-forward map $\pi_* \colon L^1(\mathbb{D}) \to L^1(S)$ such that for each quadratic differential ϕ on $\mathbb{D}$ the quadratic differential $\pi_*(\phi)$ on S satisfies

$$\int_{\mathbb{D}} \phi \pi^* \mu = \int_S (\pi_*\phi)\mu \tag{6.13}$$

for all Beltrami vectors $\mu \in L^\infty(S)$. Next choose $\phi_z \in L^\infty(\mathbb{D})$ to be $\phi_z(\omega) = \dfrac{1}{\omega - z} d\omega^2$. So from (6.13) one gets

$$\tilde{V}_n(z) = \int_{\mathbb{D}} \frac{\tilde{\mu}_n(\omega)}{w - z}\, dw\, d\bar{w} = \int_{\mathbb{D}} \tilde{\mu}_n(\omega)\phi_z dw d\bar{w} = \int_S (\pi_*\phi_z)\mu_n.$$

Since μ_n converges in the weak topology it follows that $V_n(z)$ converges for each z to

$$\tilde{V}(z) = \int_{\mathbb{D}} \frac{\tilde{\mu}(\omega)}{w - z}\, dw\, d\bar{w}.$$

From this expression it follows that V is continuous and that $\tilde{V}_n$ converges uniformly on compacta. In particular, the limit V is also a vector field on S, i.e., $\tilde{V}(g(z)) = g'(z)\tilde{V}(z)$ for every element g of the group of deck transformations Γ.

The above expression also implies that $\bar{\partial} V = \mu$ proving that μ is infinitesimally trivial. Thus the proof is completed in the case of Riemann surfaces.

If we are on a Riemann surface lamination we go about it in the following way. Take a sequence of V_n such that $\mu_n = \bar{\partial} V_n$ and μ_n converges weakly to μ in L^∞. Take a leaf L of the lamination and let $\pi_L\colon \mathbb{D} \to L$ be the universal cover. As before let $\tilde{\mu}_{n,L}$ and $\tilde{V}_{n,L}$ be the lifts of μ_n and V_n (through $\pi_L\colon \mathbb{D} \to L$). Since V_n is bounded, $\tilde{V}_{n,L}$ is zero at the boundary of $\mathbb{D}$ (here we use the fact that the Poincaré metric on L extends to a continuous metric on the lamination and, since the lamination is compact, the Poincaré norm of the vector field is again bounded). Therefore, as before,

$$\tilde{V}_{n,L}(z) = \int_{\mathbb{D}} \frac{\tilde{\mu}_n(\omega)}{w - z}\, dw\, d\bar{w}.$$

The map $(\pi_L)_*\colon L^1(\mathbb{D}) \to L^1(\mathcal{L})$ constructed in subsection 6.4 has the property (6.13), i.e.,

$$\int_{\mathbb{D}} \phi(\pi^* \mu) = \int_{\mathcal{L}} ((\pi_L)_* \phi)\, \mu$$

for each Beltrami differential μ on $\mathcal{L}$ and each quadratic differential ϕ. So if we take the quadratic differential $\phi(\omega) = \dfrac{dw^2}{\omega - z}$ with transverse measure class supported on L we get as in the Riemann surface case that

$$\tilde{V}_{n,L}(z) = \int_{\mathbb{D}} \frac{\tilde{\mu}_{n,L}(w)}{w - z}\, dw\, d\bar{w}$$

converges pointwise to

$$\tilde{V}_L(z) = \int_{\mathbb{D}} \frac{\tilde{\mu}_L(w)}{w - z}\, dw\, d\bar{w}.$$

This convergence holds for all leaves L. To prove the continuity notice that one can approximate $\tilde{V}_L(z)$ for z near the center of the circle by integrating the previous expression on a disc of (Euclidean) radius close enough to one. Fixing this closeness and using that the restriction of the universal covering to this compact disc depends continuously on the leaf (because the Poincaré metric is continuous) one gets the continuity. Therefore these vector fields $\tilde{V}_L$ push forward to a continuous vector field on the lamination. □

6.6. Dynamical Beltrami coefficients and the push-forward of quadratic differentials

In this subsection we will construct some Beltrami vectors and corresponding Beltrami paths in the Riemann surface laminations of Example 2. These Beltrami vectors will be related to the dynamics which defines the lamination and will define Beltrami paths not only of laminations but also paths of dynamical systems.

Let $F\colon U \to V$ be a quadratic-like map and let $\mathcal{L}$ be the Riemann surface lamination associated to its germ: $\mathcal{L}$ is the orbit space of the natural extension

$\tilde{F}\colon \tilde{F}^{-1}(\tilde{\mathcal{L}}) \to \tilde{\mathcal{L}}$ of F to the inverse limit $\tilde{\mathcal{L}}$ of

$$F|(U \setminus J(F))\colon U \setminus J(F) \to V \setminus J(F).$$

Recall that a Beltrami path in $\mathcal{L}$ in the direction of a Beltrami vector ν is a one parameter family $\mu_t \in L_1^\infty(\mathcal{L})$ such that in each flow box, the curve $t \mapsto \mu_t(z,\lambda) \in \mathbb{D}$ goes along a hyperbolic geodesic and is tangent at $t = 0$ to the vector $\nu(z,\lambda)$. Given such a Beltrami path, we can pull it back to a Beltrami path $\tilde{\mu}_t$ in $\tilde{\mathcal{L}}$. For each t, the Beltrami coefficient $\tilde{\mu}_t$ is invariant, as a Beltrami differential, by the natural extension $\tilde{F}$, i.e., $\tilde{F}^*\tilde{\nu} = \tilde{\nu}$. Conversely, any $\tilde{F}$-invariant Beltrami path on $\tilde{\mathcal{L}}$ gives rise to a Beltrami path in $\mathcal{L}$. Similarly, the Beltrami vectors in $\mathcal{L}$ correspond to $\tilde{F}$-invariant Beltrami vectors in $\tilde{\mathcal{L}}$. Here we will construct a special class of Beltrami vectors and Beltrami paths on $\mathcal{L}$ which we shall call *dynamical*. We start by taking a representative F of the germ that defines $\mathcal{L}$. Next, we take an L^∞ function $\hat{\nu}\colon V \setminus J(F) \to \mathbb{C}$ satisfying the following invariance condition:

$$\hat{\nu}(z) = \hat{\nu}(F(z))\frac{\overline{F'(z)}}{F'(z)} \tag{6.10}$$

for every $z \in U \setminus J(F)$. Such a function can be obtained by starting with any measurable and essentially bounded function on the fundamental domain $V \setminus U = F(U) \setminus U$ and extending it to $V \setminus J(F)$ by using the above equation. It follows that, if we take the pullback $\tilde{\mathcal{L}}$ of $\hat{\nu}$ via the natural projection, we get a Beltrami vector $\tilde{\nu}$ which is invariant under $\tilde{F}$ and hence, a Beltrami vector ν on $\mathcal{L}$. Such a ν is called a dynamical Beltrami vector. Notice that we can cover $\tilde{\mathcal{L}}$ by flow boxes such that the expression of $\tilde{\nu}$ in each of these boxes is independent of the transversal coordinate. Let us mention some special properties these Beltrami vectors have:

Property 1: In contrast to general Beltrami vectors on a Riemann surface lamination, a Beltrami path generated by such a dynamical Beltrami vector can be 'integrated' and gives rise to a one parameter family of Riemann surfaces laminations $\mathcal{L}_t$ which are quasiconformally equivalent to $\mathcal{L}$.

Proof. If μ_t is this Beltrami path and $\tilde{\mu}_t$ is the corresponding $\tilde{F}$-invariant Beltrami coefficients on $\tilde{\mathcal{L}}$, then $\tilde{\mu}_t$ is the pullback of $\hat{\mu}_t$ where $\hat{\mu}_t$ is the Beltrami path on V tangent to $\hat{\nu}$ at $t = 0$. By the invariance of $\hat{\nu}$ under F as a Beltrami differential, it follows that $\hat{\mu}_t$ is also invariant under F. We extend $\hat{\mu}_t$ to V by setting it equal to zero on the Julia set. By the Measurable Riemann Mapping Theorem, there exists a quasiconformal homeomorphism $H_t\colon V \to V$ whose Beltrami coefficient is $\hat{\mu}_t$. From the F-invariance of $\hat{\mu}_t$ it follows that $F_t = H_t \circ F \circ H_t^{-1}$ is conformal and therefore a quadratic-like map quasiconformally conjugated to F. The Riemann surface lamination $\mathcal{L}_t$ associated to the map F_t is the one parameter family of Riemann surface laminations which we associate to the Beltrami path μ_t.

Property 2: Given two dynamical Beltrami coefficients μ_0 and μ_1 there exists a dynamical Beltrami path μ_t between them.

Proof. Proof This can be obtained using the Beltrami coefficient of a quasi-conformal conjugacy between the corresponding quadratic like maps.

Property 3: The integral of the product of a dynamical Beltrami vector and a quadratic differential can also be interpreted in the dynamical plane. Indeed, if ϕ is an integrable quadratic differential on $\mathcal{L}$ we can pull it back to $\tilde{\mathcal{L}}$ by the quotient map and we get a quadratic differential $\tilde{\phi}$ on $\tilde{\mathcal{L}}$ which is invariant by $\tilde{F}$ as a quadratic differential. Notice that $\tilde{\phi}$ is not integrable over all $\tilde{\mathcal{L}}$. However, if ν is a Beltrami vector on $\mathcal{L}$ and $\tilde{\nu}$ its pullback to $\tilde{\mathcal{L}}$, then $\int_{\mathcal{L}} \nu\phi$ is equal to the integral of $\tilde{\nu}\tilde{\phi}$ over the fundamental domain

$$\tilde{\mathcal{L}} \setminus \tilde{F}^{-1}(\tilde{\mathcal{L}}).$$

It follows that we can define the push-forward map

$$\pi_*\colon L^1(\mathcal{L}) \to L^1(V \setminus U) \tag{6.11}$$

by: $\pi_*\phi(z)$ is the average of $\tilde{\phi}$ on the fibre $\pi^{-1}(z)$ with respect to the transversal measure of the quadratic differential. (Compare this also with the push-forward map from (6.9).) More precisely, let W be a small disc around $z_0 \in V \setminus U$ such that $F^{-n}(W) \cap W = \emptyset$ for all $n \in \mathbb{N}$. Then there exists a flow box $Z\colon \pi_F^{-1}(W) \to W \times \{0,1\}^{\mathbb{N}}$ that maps each fiber $\pi_F^{-1}(z)$ onto $z \times \{0,1\}^{\mathbb{N}}$. The local expression of $\tilde{\nu}$ in this flow box is defined to be $(z,\lambda) \mapsto \bar{\nu}(z)$ (it does not depend on the second coordinate). It is easy to see that by covering $\tilde{\mathcal{L}}$ by such flow boxes and taking these local expressions we get a Beltrami vector which is $\tilde{F}$-invariant. Therefore, we get a special type of Beltrami vectors on $\mathcal{L}$ that we call dynamical Beltrami vectors. If we start with an F-invariant function whose L^∞ norm is smaller than one we get a Beltrami coefficient. Since these Beltrami coefficient are continuous with respect to the transversal parameter in the L^∞ topology, we get a new Riemann surface lamination structures on $\mathcal{L}$ as we have seen in subsection 6.2. Here we get even more: such a dynamical Beltrami coefficient defines a new dynamical system; the Riemann surface lamination associated to it by the construction of Example 2 is precisely the lamination we are discussing. To see this, we just extend $\bar{\nu}$ to an F-invariant Beltrami vector. The arc F_t from 1) above defines an arc of dynamical systems.

If m is a measure on $\{0,1\}^{\mathbb{N}}$ in the measure class of the quadratic differential, then, with respect to this measure, the quadratic differential $\tilde{\phi}$ is represented by a function $\phi(z,\lambda)$. Then we define $\hat{\phi}(z) = \int_\Lambda \phi(z,\lambda)dm(\lambda)$. The push-forward map $\pi_*\colon \phi \mapsto \hat{\phi}$ is natural in the sense that the pairing between a dynamical Beltrami vector and any quadratic differential in $L^1(\mathcal{L})$ can also be interpreted in the dynamical plane. In fact, given any fundamental domain $N \subset B$ of F,

$$\int_{\mathcal{L}} \nu\phi = \int_N \bar{\nu}\pi_*\phi \tag{6.12}$$

for all quadratic differentials ϕ.

Notice that we can push forward quadratic differentials on $\mathcal{L}$ to integrable quadratic differentials on any fundamental domain of any representative of the germ. In general we cannot push forward Beltrami vectors: only the dynamical Beltrami vectors are pullbacks of Beltrami vectors in the dynamical plane.

Remark. If ν is a dynamical Beltrami vector on the lamination $\mathcal{L}$ then we can cover $\mathcal{L}$ by a finite number of flow boxes such that the expression of ν in these flow boxes does not depend on the transversal direction. Conversely, any Beltrami vector of $\mathcal{L}$ that is locally constant in the transversal direction is a dynamical Beltrami vector that 'lives' in a small enough neighbourhood of the Julia-set. Using a partition of unit – see Moore and Schochet (1988, pp. 44) – we can prove that the set of dynamical Beltrami vectors is dense in $L^\infty(\mathcal{L})$ with respect to the weak topology.

Theorem 6.3.

1. *Suppose S is a Riemann surface. Then given a Beltrami vector μ, there exists a Beltrami vector ν such that $\mu - \nu$ is infinitesimally trivial and such that*
$$||\nu||_\infty = \sup_{|\phi|=1} |\int \mu\phi|$$
where the supremum runs over all holomorphic quadratic differentials with respect to a given complex structure.

2. *If $\mathcal{L}$ is the Riemann surface lamination from the examples above, then given a Beltrami vector μ and any $\epsilon > 0$, there exists a Beltrami vector ν such that $\mu - \nu$ is infinitesimally trivial and such that*
$$(1-\epsilon) \cdot ||\nu||_\infty \leq \sup_{|\phi|=1} |\int_{\mathcal{L}} \mu\phi|$$
where the supremum runs over all holomorphic quadratic differentials with respect to a given complex structure.

3. *If in 2) the Beltrami vector μ is dynamical then there exists a dynamical Beltrami vector ν such that $\mu - \nu = \bar{\partial}V$ with V dynamical and*
$$(1-\epsilon) \cdot ||\nu||_\infty \leq \sup_{|\phi|=1} |\int_{\mathcal{L}} \mu\phi|.$$

Proof. To prove 1), let us consider the functional $M\colon H \to \mathbb{C}$, defined by $M(\phi) = \int_S \mu\phi$, where $H \subset L^1(S)$ is the space of holomorphic quadratic differentials. By Hahn-Banach, M extends to a functional of $L^1(S)$ of the same norm. Since $L^\infty(S)$ is the dual of $L^1(S)$, there exists $\nu \in L^\infty(S)$ whose L^∞ norm is the norm of M as a functional and such that $M(\phi) = \int_S \nu\phi$ for all $\phi \in L^1(S)$. This proves 1. For 2) and 3) we use a slightly extended version of Hahn-Banach.
To state this version of Hahn-Banach, suppose that B_0 and B_1 are the following Banach spaces related by a non degenerate pairing. For the proof of Statement 2) we take B_0 to be the Beltrami vectors on $\mathcal{L}$, B_1 to be the L^1 quadratic differentials on $\mathcal{L}$, and the pairing to be the one defined above: $(\mu, \phi) = \int_{\mathcal{L}} \phi \cdot \mu$.

For the proof of Statement 3), we take B_0 to be the dynamical Beltrami vectors and B_1 and the pairing as before.

Fact 1: The L^1 norm on B_1 induces by the pairing the L^∞ norm on B_0 in each case above. That is, if a norm $|.|$ on B_0 is defined by $\forall x \in B_0$, $|x| \leq 1$ if and only if $|(x, y)| \leq 1$ for all $y \in B_1$ with $|y| \leq 1$, then this is the L^∞ norm on B_0.

Fact 2: The L^1 norm on B_1 is induced from the pairing by the L^∞ norm on B_0 (in either case).

Proof. Fact 1 can be proved by using Hahn-Banach and using local charts. Fact 2 follows from Fact 1 and the symmetry of the above definition.

Now in the proof of Statement 2), let $T_0 \subset B_0$ be the trivial Beltrami vectors and A_0 the quotient B_0/T_0 with the induced L^∞ norm on B_0 and let $H^1 \subset B_1$ be the holomorphic quadratic differentials in B_1. By Theorem 6.2, T_0 is closed in B_0 with respect to the norm topology. Now, A_0 is dually paired to H^1 by Theorem 6.2. The norm on H^1 induced by the norm on A_0 is clearly the restriction of the norm on B_1 induced by B_0 because the unit ball in A_0 is the image of the unit ball in B_0. Thus the L^1 norm on H^1 and the L^∞ norm on $A_0 = B_0/T_0$ are related by the pairing. Thus the statement of 2) follows by definition.

Let us now prove Statement 3). As in the proof of Statement 2), one has a Beltrami vector ν_1 such that the inequality from Statement 3) holds; furthermore, $\mu - \nu_1 = \bar{\partial} V_1$ but V_1 is not necessarily dynamical. However, locally, $\bar{\partial} V_1$ is constant in the transversal direction and V_1 is continuous. Thus locally we can approximate V_1 in the $|V|_0 + |\bar{\partial} V|_\infty$ topology by a dynamical V. A standard partition of unit argument approximates V_1 globally in the $|V|_0 + |\bar{\partial} V|_\infty$ topology by a transversally locally constant V. Now let ν be defined by $\mu - \nu = \bar{\partial} V$. $\square$

Definition. A Beltrami vector for which

$$||\nu||_\infty = \sup_{|\phi|=1} \int |\nu \phi|$$

is called *extremal*.

Next define the *infinitesimal Teichmüller norm* of a Beltrami vector by

$$|\nu|_T = \sup_{|\phi|=1} | \int_{\mathcal{L}} \nu \phi|$$

where the supremum is taken over all holomorphic quadratic differentials of L^1 norm one. Notice that the infinitesimal Teichmüller norm of a Beltrami vector is bounded by its L^∞ norm. Furthermore, from Theorem 6.3 it follows that the Teichmüller norm of a Beltrami vector ν is equal to the infimum of the L^∞ norms of all Beltrami vectors that differ from ν by an infinitesimal trivial Beltrami vector.

7 The Almost Geodesic Principle

In this section we shall prove that a Beltrami path which is almost efficient near one endpoint remains almost efficient up to a given size. The precise statement of this 'Almost Geodesic Principle' will be given below. This principle and the complex bounds from Section 5 will imply that renormalization contracts the Teichmüller metric.

1. A Geometric Inequality

Let Ψ be a quasiconformal homeomorphism of a Riemann surface or a Riemann surface lamination S which is quasiconformally isotopic to the identity map and let ϕ be a holomorphic quadratic differential on S. Let $|\phi|$ be the measure corresponding to ϕ.

Proposition 7.1. *Let S be a Riemann surface or a Riemann surface laminations, Ψ and ϕ as above with $||\phi|| = 1$ and let*

$$D(z) = \frac{|D\Psi(z)v|_\phi}{|v|_\phi}$$

where v is tangent to the horizontal foliation of ϕ. (Since Ψ is quasiconformal $D\Psi$ and therefore D is defined almost everywhere.) Then

$$\int_S D\, d|\phi| \geq 1.$$

Proof in the case of a Riemann surface: First note that there exists a constant M such that for any arc α in a horizontal trajectory of ϕ the following holds. The length of $\Psi(\alpha)$ (with respect to the metric corresponding to ϕ) is at least the length of α minus $2M$. The proof of this fact is elementary, see Lehto (1987, p. 174) and also the proof below of Proposition 7.1 in the case of Riemann surface laminations. Let us show that

$$1 = ||\phi|| = \int_S d|\phi| \leq \int_S D\, d|\phi|.$$

By the first argument, there exists a constant M such that if $\alpha\colon [0,T] \to S$ is a parametrization of an horizontal arc by arc length (with respect to the metric $\sqrt{\phi}$), then $\int_0^T D(\alpha(t))dt \geq T - 2M$. Hence $\frac{1}{T}\int_0^T D(\alpha(t))dt \geq 1 - \frac{2M}{T}$. Suppose first the horizontal foliation is orientable. The set A of all separatrices of the horizontal foliation has ϕ-measure zero. In $S \setminus A$ we can consider the flow X_t such that for each z, $t \mapsto X_t(z)$ is a parametrization of the horizontal leaf by arc length. This flow preserves the measure $|\phi|$. By the Birkhoff Ergodic Theorem, see the Appendix, we see that the time average $\frac{1}{T}\int_0^T D(X_t(z))dt$ converges almost everywhere and its integral with respect to $|\phi|$ is the space average of D. By the above condition, the time average is greater or equal to one. This proves the inequality. If the horizontal foliation is not orientable then lift everything to a double cover $\tilde{S}$, such that the lift of the horizontal foliation is orientable. Then one can use again the Ergodic Theorem "upstairs". Notice that the lift of the quadratic differential has total mass 2, and so, in order to use the Ergodic Theorem, one has to multiply this quadratic differential upstairs by a factor 1/2 in order to get a probability measure. □

In the remaining of this subsection we shall prove the analogous result in the case of a hyperbolic Riemann surface lamination. This proof will also work for Riemann surfaces.

Let Ψ be a quasiconformal homeomorphism of a hyperbolic Riemann surface lamination $\mathcal{L}$ and let Ψ_t be a quasiconformal isotopy from the identity map to $\Psi_1 = \Psi$ preserving the leaves. Lift the map Ψ and the isotopy to the universal cover to an arbitrary leaf L and denote this lifts by $\Psi_L \colon \mathbb{D} \to \mathbb{D}$ and $\Psi_{L,t}$; notice that the lift $\Psi_{L,t}$ of the isotopy Ψ_t is uniquely defined if we start with $\Psi_{L,0}$ equal to the identity. Let us fix a finite cover of $\mathcal{L}$ by flow boxes. By compactness, there exists an integer N such that for every $p \in \mathcal{L}$, the curve $t \mapsto \Psi_t(p)$ is covered by at most N plaques of the lamination coming from the chosen finite collection of flow boxes. Since the diameter of a plaque in the Poincaré metric is uniformly bounded, it follows that, the Poincaré distance between z and $\Psi_L(z)$ is bounded and that these bounds do not dependent on the leaf L. In particular, Ψ_L extends continuously to the identity map on the boundary of the disc $\mathbb{D}$.

Let ϕ be a holomorphic quadratic differential on the lamination. By definition, the expression of a quadratic differential differs by a multiplicative constant on each leaf if we choose two different measures in the measure class. Therefore, ϕ defines in almost every plaque, with respect to the transversal measure class, a holomorphic quadratic differential which is well defined up to multiplication by a positive real number. Hence there exists a quadratic differential ϕ_L in the universal cover $\mathbb{D}$ of L, for almost every leaf L (with respect to the transversal measure class), such that the restriction of $(i_L)_*(\phi_L)$ to each plaque, coincides with a positive multiple of the local expression of ϕ ($(i_L)_*$ is the push-forward map of section 6.4). This quadratic differential ϕ_L is obtained by analytic continuation. Since the universal cover $\mathbb{D}$ of L is simply connected, ϕ_L is everywhere well-defined but it is only unique up to multiplication by a positive constant. Of course, ϕ_L is not necessarily invariant, as a quadratic differential, by the group of deck transformations because if $A \colon \mathbb{D} \to \mathbb{D}$ is an element of the group of deck transformation then $A_*\phi_L$ is a positive multiple of ϕ_L and this multiple is not necessarily equal to one (this means that the corresponding quadratic differential on the leaf is multivalued). Even so, ϕ defines horizontal and vertical foliations on almost every leaf; these are simply the push-forward of the horizontal lamination of ϕ_L, since the horizontal foliation does not change if we multiply the quadratic differential by a positive number. The Teichmüller metric $\sqrt{|\phi|}$ is not well-defined since ϕ_L is only defined up to multiplication by a positive constant. Moreover, the ϕ_L-length of the lift $\tilde{C}$ of a curve C to L depends on the lift of the curve. In particular, the ϕ_L distance between the endpoints of the lift of the curve $t \mapsto \Psi_t(p)$, i.e., the ϕ_L-distance between z and $\Psi_L(z)$, may not be bounded. However, ratios of lengths are well defined because the metric is defined up to a multiplicative constant. So ϕ_L defines an affine structure on each leaf: ratios of lengths of vectors are well defined.

Another difference between the present situation and the case of a compact Riemann surface discussed before, is that here, the horizontal trajectories of ϕ may have finite ϕ_L-length. In fact, this is the case for the special quadratic differentials discussed before, obtained by pushing forward, via $i_L : \mathbb{D} \to L \subset \mathcal{L}$, a quadratic differential on the disc $\mathbb{D}$ whose trajectories have finite length (dz^2 is an example). If this is the case for most trajectories, then, the lift each such trajectory to the universal covering of the corresponding leaf is a curve connecting two boundary points of the disc. As the lift Ψ_L of the map Ψ is the identity at the boundary of the disc, it follows that the geometric inequality we need in the proof is automatically satisfied: the ϕ_L-length of the Ψ_L image of a full horizontal trajectory divided by the length of the trajectory is bigger or equal to one because the endpoints do not move. There is another possibility, which is similar to the case of the compact situation where the horizontal trajectories of the quadratic differentials are recurrent. In fact, in the general case we have a decomposition of $\mathcal{L}$ into four subsets X, C_1, C_2, Y_1, Y_2 which are invariant by the horizontal trajectories as follows: i) the trajectories in X are homeomorphic to the real line and have infinite Teichmüller length; ii) the trajectories in C_1 are compact and the affine structure the quadratic differential induces in each trajectory in C_1 is equivalent to $\mathbb{R}/\mathbb{Z}$; iii) the affine structure of each trajectory in C_2 is equivalent to the Hopf circle $\mathbb{R}^+/\lambda\mathbb{R}^+$, $\lambda > 0$; iv) the trajectories in Y_1 are homeomorphic to the real line but have finite Teichmüller length; the trajectories in Y_2 are affinely equivalent to $\mathbb{R}^+$, i.e they are half infinite. This statement follows because there are only two affine structures on S^1; the first is the one coming from the additive structure on $\mathbb{R}$ and the other from the multiplicative structure on $\mathbb{R}^+$.

The first and the second case also occur in compact Riemann surfaces. The third situation may occur in a cylindrical leaf if the quadratic differential is multivalued. In this case the compact trajectory must be necessarily isolated in the leaf because the nearby trajectories are asymptotic to it. (This can be seen by going to the universal cover of the leaf and noting that the quadratic differential is multivalued). Hence C_2 has $|\phi|$-measure zero. The dynamics of the horizontal trajectories is dissipative in Y_1. Indeed, we can define a Borel subset of Y_1 by choosing in each trajectory the middle point (with respect to the Teichmüller length). This proves the dissipativeness of Y_1. We will prove below that Y_2 has $|\phi|$-measure zero. Since the quadratic differential is in L^1, it follows from the Poincaré recurrence theorem that almost all trajectories in X are recurrent.

Let us formulate the geometric inequality we have to prove more precisely. Notice first that if z is in the universal cover of a typical (with respect to the measure $|\phi|$) leaf L then the ratio $D_L(z)$ between the ϕ_L-length of the $D\Psi_L(z)$ image of a vector tangent to the horizontal foliation by the ϕ_L-length of this vector is well defined (i.e., does not depend on the choice of ϕ_L) and is invariant by the group of deck transformations. Hence it defines a function D at almost all points of $\mathcal{L}$. We have to prove that the integral of D with respect to the measure $|\phi|$ is at least one. To prove the geometric inequality we may disregard the subset C_2 since it has $|\phi|$-measure zero.

Definition. Let ξ be a partition of a subset T of $\mathcal{L}$ and let $d\xi$ be the push-forward measure of $|\phi|$ by the quotient map from T to T/ξ. Then the partition ξ is called *measurable* if the measure $|\phi|$ on T can be disintegrated on the atoms of ξ, i.e., for almost every $C \in \xi$ (with respect to $d\xi$) there exists a measure μ_C on a σ-algebra $\mathcal{M}_C$ of subsets of C satisfying the properties: i) a subset $A \subset T$ is measurable if and only if $A \cap C \in \mathcal{M}_C$ for almost all C (with respect to $d\xi$) and ii) $|\phi|(A) = \int_{T/\xi}(\mu_C(A \cap C))d\xi$. A *tube* in $\mathcal{L}$ is a set T of positive measure having a measurable partition ξ whose atoms are arcs of horizontal trajectories of the quadratic differential ϕ. We refer to Rohlin (1962) and (1966) for properties of measurable partitions.

Lemma 7.1. *Let U be a flow box of the lamination $\mathcal{L}$. Let T be the complement in U of the set of zeros of the quadratic differential. Then T is a tube and the $|\phi|$-measure of T is equal to the $|\phi|$-measure of U.*

Proof. Each connected component of the intersection of a trajectory with T is an interval that have finite $|\phi|$-length. Hence the set S of middle points of the above components is a measurable set that intersects each component of a trajectory in T in a unique point. The partition of T by these components is clearly a measurable partition. □

Lemma 7.2. *Given any compact arc of horizontal trajectories, there exists a flow box that contains this arc.*

Proof. First we cover the arc by a finite number of flow boxes. Using a partition of unity, we can glue these boxes together and construct a neighbourhood of the arc and a quasiconformal homeomorphism of this neighbourhood to a product of the disc by a neighbourhood of the transversal (i.e., a quasiconformal trivialization). Next we push forward, via this quasiconformal homeomorphism, the conformal structure on the leaves and we get a quasiconformal structure on the plaques of the product. Finally, by the Measurable Riemann Mapping Theorem with parameters (the Ahlfors-Bers Theorem, see the Appendix), there is a quasiconformal homeomorphism of $\mathbb{D} \times \Lambda$ that maps this quasiconformal structure on $\mathbb{D} \times \{\lambda\}$ in the usual conformal structure of $\mathbb{D} \times \{\lambda\}$. The composition of the two quasiconformal homeomorphisms above gives the required trivialization. □

Lemma 7.3. *The $|\phi|$-measure of the set Y_2 of half infinite horizontal trajectories is equal to zero.*

Proof. First we note that we may assume that the foliation of Y_2 is ergodic: this means that each invariant subset either has measure zero or its complement has zero. Indeed, if the foliation of Y_2 by trajectories is not ergodic then we take the ergodic decomposition and use the arguments from below for each ergodic component. So let us assume that the foliation by Y_2 is ergodic and assume by contradiction that Y_2 has positive measure. To reach a contradiction let us call a map $\tau\colon A \to B$ is admissible if: i) τ is one-to-one and measure preserving; ii) for each $x \in A$, $\tau(x)$ belongs to the same horizontal trajectory as x and is closer to the finite end than x; iii) the restriction of τ to each horizontal trajectory preserves any measure in the trajectory that is compatible with the affine structure (notice that any two measures on a trajectory compatible with the affine structure differ from each other by a positive constant factor; thus if a map preserves one of these measures it preserves any other such a measure).

Claim: Given two disjoint sets A and B in Y_2, with positive measure, there exist subsets $A_0 \subset A$, $B_0 \subset B$ of positive measure and a measure preserving bijection $\tau\colon A_0 \to B_0$ such that either τ or τ^{-1} is admissible.

Before proving the claim let us show that it gives a contradiction. Indeed, let $\mathcal{C}$ be the collection of admissible maps $\tau\colon A \to B$. In $\mathcal{C}$ we consider the partial order relation $\tau_1 \preceq \tau_2$, $\tau_i\colon A_i \to B_i$, if there exists a subset $Z_1 \subset A_1$ of measure zero such that $A_1 \setminus Z_1 \subset A_2$ and $\tau_1(z) = \tau_2(z)$ for every $z \in A_1 \setminus Z_1$. Let $\mathcal{F} \subset \mathcal{C}$ be a totally ordered subfamily. Let $\tau_n\colon A_n \to B_n$ be a sequence of elements of $\mathcal{C}$ such that the measure of A_n converges to the supremum of the measures of the domains of maps in $\mathcal{F}$. Clearly, there exists $\tau\colon A \to B$ in $\mathcal{C}$ such that $\tau_n \preceq \tau$ for all n. Here $A = \cup A_n \setminus Z$ where Z has measure zero, if $x \in A$ then $x \in A_n$ for all $b \geq n_0$ and $\tau_n(x) = \tau_{n_0}(x)$ for all $n \geq n_0$. Let $\tilde{\tau}\colon \tilde{A} \to \tilde{B}$ be an element of $\mathcal{F}$. If $\tilde{\tau} \preceq \tau_n$ for some n then $\tilde{\tau} \preceq \tau$. If $\tau_n \preceq \tilde{\tau}$ for all n, it follows that the measure of $\tilde{A}$ is equal to the measure of A and $\tilde{\tau} \preceq \tau$ (as well as $\tau \preceq \tilde{\tau}$). Therefore, $\mathcal{F}$ has an upper bound. Thus, by Zorn's Lemma, there exists $\tau\colon A \to B$ which is a maximal element of $\mathcal{C}$. By the claim, A must have full measure in Y_2. The intersection of a typical trajectory with A has full measure in the trajectory with respect to any measure compatible with the affine structure. The restriction of τ to a typical trajectory gives a one to one measure preserving mapping $\tau\colon \mathbb{R}^* \to \mathbb{R}^*$ such that $\tau(x) < x$ for every x. This contradicts the Poincaré's Recurrence Theorem.

Let us now prove the claim. Since A has positive measure, we can choose a tube T, with cross section S, such that any plaque T_x, $x \in S$, intersects A in a set of positive measure with respect to the measure in T_x obtained by disintegration of the measure $|\phi|$. For each $n \in \mathbb{Z}$. let $T_x(n)$ be the interval in the trajectory through x that contains n consecutive interval of equal length (with respect to the affine structure) in the direction of the infinite end if $n > 0$ and, if $n < 0$, either $|n|$ intervals in the direction of the finite end or the whole interval between the finite end and T_x if these intervals do not exist (for each x the last situation will always occur if $-n$ is big enough). Let $T(n) = \cup_{x \in S} T_x(n)$. The union of $T(n)$ for $n \in \mathbb{Z}$ is a union of trajectories and have positive measure. By ergodicity it intersects B in a set of positive measure. Hence, there exists $n \in \mathbb{Z}$

such that $T(n)$ intersects B in a set of positive measure. Suppose $n > 0$. Let $\tilde{T} = T(n) \setminus S_-$, where S_- is the set of end points of the plaques T_x that is closer to the finite end of the corresponding trajectory. We claim that $\tilde{T}$ is a tube such that a positive set of plaques of $\tilde{T}$ intersects both A and B. In fact, $S_- \cap B$ has zero measure and each $T_x(n)$ is divided into a finite number of intervals by S_-. So we can construct a cross section $\tilde{S}$ of $\tilde{T}$ by choosing the middle point of each of these intervals and the conclusion follows. Let $\bar{T}$ be the union of all plaques $\tilde{T}_x$ of $\tilde{T}$ that intersects both A and B in a set of positive measure in the plaque. Then $\bar{T}$ is again a tube. Using this tube it is easy to construct the local map τ. For that we choose sets of positive measure $A_1 \subset A \cap \bar{T}$, $B_1 \subset B \cap \bar{T}$ such that: i) the points of $A_1 \cap \bar{T}_x$, resp. $B_1 \cap \bar{T}_x$, are density points of $A \cap \bar{T}_x$, resp. $B \cap T_x$ for almost all $x \in \bar{S}$: each plaque $\bar{T}_x$ contains an interval J_x such that $A_1 \cap \bar{T}_x$ and $B_1 \cap \bar{T}_x$ are contained in different components of $\bar{T}_x - J_s$ and they have the same measure. Therefore, for each x we take the measure preserving bijection τ_x from $A_1 \cap \bar{T}_x$ onto $B_1 \cap \bar{T}_x$. If the set of $x \in \bar{S}$ such that $B_1 \cap \bar{T}_x$ lies between the finite end and $A_1 \cap \bar{T}_x$ has positive measure, we can construct a map τ which is admissible by taking A_0 to be the intersection of A_1 with the corresponding plaques of $\bar{T}$. If this set has measure zero, we construct a map τ from A_1 to B_1 such that the inverse of τ is admissible. □

Let $\epsilon > 0$. For each integer n, let X_n be the set of points $x \in X$ such that the ϕ_L distance between $\hat{x}$ and $\Psi_L(\hat{x})$ is less than ϵ times the ϕ_L-length of each arc of horizontal trajectory with endpoint $\hat{x}$ and Poincaré length n.

Lemma 7.4. $X = \cup_n X_n$.

Proof. The horizontal trajectory through each point $x \in X$ is not bounded. Hence $x \in X_n$ if n is large enough. □

Definition. A tube T is ϵ-good if for almost all arcs C (with respect to $d\xi$) of the partition of T by horizontal arcs satisfies: the sum of the ϕ_L distance between each endpoint of C and its Ψ_L image is less than ϵ times the ϕ_L-length of the arc.

Lemma 7.5. *Let $A \subset X$ be a set of positive measure. Then there exists an ϵ-good tube T such that $T \cap A$ has positive measure.*

Proof. Since A has positive measure and the sets X_n grows to X, it follows that $B = A \cap X_n$ has positive measure if n is big enough. Since the horizontal foliation is recurrent in X, almost every point $x \in X$ is an endpoint of an arc of the horizontal foliation having Poincaré length bigger than n and the other endpoint y is also in B. Now take a flow box containing one of those arcs and let T be the union of all of the above arcs that are contained in the flow box. It follows from Lemma 7.1 that T is a tube and we can choose T so that $T \cap B$ has positive measure. □

Lemma 7.6. *Given $\delta > 0$ there exists a finite number of disjoint ϵ-good tubes $T_1, \dots, T_n$ in X such that $\mu(\cup T_i) \geq (1-\delta)\mu(X)$.*

Proof. First we notice that if T_1 and T_2 are ϵ good tubes then $T_1 \cup T_2$ is covered by the disjoint union of three ϵ-good tubes. In fact, let S_1 be the union of all horizontal arcs in the partition of T_1 that do not intersect T_2, let S_2 be the set of horizontal arcs in the partition of T_2 that do not intersect T_1 and let S_3 be the union of each arc C which is the union of an arc of T_1 that intersects T_2 with the corresponding arc of T_2. We have that S_i, $i = 1, 2, 3$, are disjoint ϵ-good tubes whose union is $T_1 \cup T_2$. Now, from Lemma 7.5 it follows that there exists a countable family of ϵ-good tubes whose union contains almost all points in X. Hence, by taking a finite number of them we cover a subset of X of measure $(1-\delta) \times \mu(X)$. By the above argument, we can make them disjoint. □

Lemma 7.7. *Let C be an atom of the measurable partition of a tube T by horizontal arcs and μ_C the measure on C given by the disintegration of μ. Then if an endpoint x of C belongs to a typical leaf L and $\tilde{C}$ is a lift of C to the universal cover then $\int_C D d\mu_C$ is equal to the ratio between the ϕ_L-length of $\Psi_L(\tilde{C})$ and the ϕ_L-length of $\tilde{C}$. In particular, if T is an ϵ-good tube then*

$$\int_C D d\mu_C \geq 1 - \epsilon.$$

Proof. This is simply the chain rule. □

Proof of Proposition 7.1 for Riemann surface laminations: Let $T_1, \dots, T_N$ be a collection of disjoint ϵ-good tubes whose union has measure at least $(1-\epsilon)$ times the measure of X. Then

$$\begin{aligned}\int_X D d\mu \geq \sum_{i=1}^{N} \int_{T_i} D d\mu = \sum \int (\int_C D d\mu_C) d\xi \\ \geq \sum_{i=1}^{N} (1-\epsilon)\mu(T_i) = (1-\epsilon)(1-\epsilon)\mu(X).\end{aligned}$$

On the other hand, let us consider tubes whose atoms are full horizontal trajectories. Let us call a tube like that a special tube. Given any subset A of positive measure in the dissipative part, there exist a special tube T that intersects A in a set of positive measure. As above, we can then construct a finite number of disjoint ϵ-good tubes in the conservative part such that the measure of the union of these tubes is almost equal to the measure of the conservative part Y. Since Ψ_L does not move the endpoints of the trajectories (they are in the boundary of $\mathbb{D}$) we have that the ϕ_L-length of the Ψ_L image of such a trajectory is at least the ϕ_L-length of the trajectory. Therefore, $\int_Y D d\mu$ is bigger than the measure of the union of these special tubes which is almost the measure of Y. This concluded the proof of the proposition. The subset C_2 is treated as in the classical case; in fact the closed trajectories can only occur in cylindrical leaves and in this case all the trajectories in the same leaf are also closed. Since the number of cylindrical leaves is at most countable, we get that either C_2 has measure zero or the measure gives positive mass to some cylindrical leaves and the situation the same as in the classical case. □

2. The Almost Geodesic Principle

Now we come to the main result of this section. Assume that ν is a Beltrami vector at some Beltrami coefficient μ which is ϵ-extremal. This means that there exists a holomorphic quadratic differential ϕ with $||\phi|| = 1$ with respect to the μ structure such that

$$|\int \phi\nu| \geq (1-\epsilon)||\nu||_\infty.$$

So the Beltrami vector ν is 0-extremal if

$$\sup|\int \phi\nu| = ||\nu||_\infty.$$

As we have seen in the corollary to Theorem 6.1 such Beltrami coefficients exist. In the next theorem it is shown that the Beltrami path corresponding to an almost extremal vector does not coil: if the tangent Beltrami vector is almost extremal then the Beltrami path remains almost a geodesic for a long (but a priori fixed) time.

Theorem 7.1. **('Almost Geodesic Principle')**

Given $\epsilon, l > 0$ there exists $\delta > 0$ such that the following holds. Suppose that ν is a δ-extremal Beltrami vector at a Beltrami coefficient μ_0 on a Riemann surface or on a compact Riemann surface lamination S. Let Ψ be quasiconformally isotopic to the identity and let μ_l be the Beltrami coefficient which is obtained from $\mu_0 = \mu$ by stretching a distance l in the direction of ν. If K is the maximal distortion of Ψ between μ_0 and μ_l then

$$l \leq K(1+\epsilon).$$

Proof. We may assume that $||\nu||_\infty = 1$ and that μ_t is the Beltrami path passing through μ_0 and tangent to ν. Since ν is δ-extremal, there exists a holomorphic quadratic differential ϕ (holomorphic with respect to the conformal structure of μ_0), such that

$$|\int \phi\nu| \geq (1-\delta)||\nu||_\infty = (1-\delta).$$

Let $\tilde{\mu}_l$ be the conformal structure we get by deforming μ_0 in the direction $|\phi|/\phi$ a distance l. Let

$$\tilde{K}(z) = \rho\left(\left(\Psi^{-1}\right)^*(\mu_0)(z), \tilde{\mu}_l(z)\right)$$

and

$$K(z) = \rho\left(\left(\Psi^{-1}\right)^*(\mu_0)(z), \mu_l(z)\right),$$

where ρ is the Poincaré distance in $\mathbb{D}$. Since μ_0 is obtained from $\tilde{\mu}_l$ by contracting by $\frac{1}{l}$ in the direction of the horizontal trajectories of ϕ, we have the *Grötzsch inequality*,

$$\int \tilde{K}(z)\, d|\phi| \geq l.$$

Indeed, if T is the Jacobian of Ψ^{-1} with respect to the measures $|\phi|$ and the measure obtained by contracting the horizontal trajectories of ϕ by $\frac{1}{l}$ we have that $D(z) \leq \tilde{K}(z) \cdot T(z)$ almost everywhere with respect to $|\phi|$. Here $D(z)$ is the ratio of the Teichmüller length of the image by the derivative of Ψ^{-1} at z of a vector tangent to the horizontal trajectory of ϕ having Teichmüller length equal to one. By the geometric inequality of section 6.1, we have that $\int_{\mathcal{L}} D(z) d|\phi| \geq 1$. Therefore, since $\int_{\mathcal{L}} T(z) d|\phi| = \frac{1}{l}$, we get:

$$\begin{aligned}
1 &\leq (\int_{\mathcal{L}} D d|\phi|)^2 \leq (\int_{\mathcal{L}} \sqrt{\tilde{K}(z)} \cdot \sqrt{T(z)} d|\phi|)^2 \\
&\leq \int_{\mathcal{L}} \tilde{K}(z) d|\phi| \cdot \int_{\mathcal{L}} T(z) d|\phi| \\
&\leq \int_{\mathcal{L}} \tilde{K}(z) d|\phi| \cdot \frac{1}{l}.
\end{aligned}$$

The third inequality above is the Cauchy-Schwarz inequality. This proves the Grötzsch inequality.

Let us compare the Beltrami vector ν with the Beltrami vector $|\phi|/\phi$. By considering local charts we see that there exist globally defined functions $\theta\colon S \to \mathbb{R}^+$ and $\lambda\colon S \to \mathbb{R}^+$ such that

$$\nu_i(z_i(p)) = e^{i\theta(z)} e^{-\lambda(z)} \frac{|\phi_i(z_i(p))|}{\phi_i(z_i(p))}$$

for any local chart z_i on S. Hence,

$$\int_S \nu\phi = \int_S e^{i\theta(z)} e^{-\lambda(z)}\, d|\phi|.$$

By the choice of ϕ, it follows that the last integral is real and $\geq (1-\epsilon)$. Thus,

$$\int_S \cos(\theta(z))\, e^{-\lambda(z)}\, d|\phi| \geq 1 - \epsilon.$$

From the last inequality it follows that the set

$$A_\nu = \{z \in S\,;\ \theta(z) > \sqrt{\nu} \text{ or } \lambda(z) > \sqrt{\nu}\}$$

has ϕ-measure of the order $O(\sqrt{\rho})$. Moreover, if $\nu > 0$ is sufficiently small then $\rho(\mu_l(z), \tilde{\mu}_l(z)) \leq \epsilon l$, for $z \notin A_\nu$. Therefore, for $z \notin A_\nu$, $|\tilde{K}(z) - K(z)| \leq \epsilon l$. Thus,

$$\begin{aligned}
\int K(z)\, d|\phi| &= \int \tilde{K}(z)\, d|\phi| \quad - \quad \int (K(z) - \tilde{K}(z))\, d|\phi| \\
&\geq l - \left(\int_{A_\nu} |K(z) - \tilde{K}(z)|\, d|\phi| \quad + \quad \int_{A_\nu^c} |K(z) - \tilde{K}(z)|\, d|\phi| \right) \\
&\geq l - O(\sqrt{\nu}) - l\epsilon \geq \frac{l}{1+\epsilon/2}.
\end{aligned}$$

This proves the theorem. □

8 Renormalization is Contracting

In this section we will combine the results of the previous sections to prove the contraction of the renormalization operator. We will split the proof into two steps.

Step 1: The Coiling Lemma

In this step we will deduce a result for dynamical Beltrami paths which gives a kind of converse to the Almost Geodesic Principle. By the Almost Geodesic Principle, if we start with a dynamical Beltrami vector which is δ-efficient and consider the deformation of the initial germ by a hyperbolic distance d along the tangent Beltrami vector, then the Julia-Teichmüller distance between the endpoints is at least $(1-\epsilon)d$. Now we want the converse: given that this distance is at least $(1-\delta)d$ we want to conclude that the tangent Beltrami vector at the origin is ϵ-efficient, where δ depends only on d and ϵ. This means that if we deform a conformal structure along a Beltrami path tangent to a Beltrami vector which is not efficient we cannot go too far in terms of the Julia-Teichmüller distance, i.e., the Beltrami path necessarily 'coils'.

Let us formulate this statement more precisely. We fix a germ of a quadratic-like map and let $F\colon U \to V$ be a representative of this germ. A *dynamical Beltrami vector* ν in the lamination $\mathcal{L}_{[F]}$ is given by a measurable function ν which is defined in some neighbourhood W of the filled Julia set of F, vanishes at the Julia set, satisfies the invariance condition $F^*\nu = \nu$, where $F^*\nu(z) = \nu(F(z)) \cdot \frac{\overline{\partial F(z)}}{\partial F(z)}$ and is essentially bounded. As we have seen in Section 6, the Teichmüller norm of ν coincides with the infimum of the L^∞ norms of all Beltrami vectors that are equivalent to ν in the sense that they have the same period, i.e., they give the same value when paired with L^1-holomorphic quadratic differentials of the lamination $\mathcal{L}_{[F]}$. A dynamical element of the Teichmüller space of $\mathcal{L}_{[F]}$ may be represented by a measurable function μ, defined in some neighbourhood of the filled Julia set of F such that: i) μ vanishes at the filled Julia set of F; ii) $F^*\mu = \mu$ and iii) $|\mu|_\infty < 1$. This function μ is called a dynamical Beltrami coefficient and the original structure, i.e., the element of the Teichmüller space represented by $\mathcal{L}_F$ is represented by the Beltrami coefficient identically zero. Conversely, any Beltrami coefficient μ satisfying the above conditions defines an element $[\mu]$ of the Teichmüller space $\mathcal{T}(\mathcal{L})$. Indeed, if ϕ is a quasiconformal homeomorphism whose Beltrami coefficient is μ then, by the invariance condition ii) it follows that $G = \phi \circ F \circ \phi^{-1}$ is a quadratic-like map and the Teichmüller equivalence class of the associated Riemann surface lamination is an element of $\mathcal{T}$.

Two dynamical Beltrami coefficients μ_1 and μ_2 are equivalent as Beltrami coefficients if $\phi_1 \circ F \circ \phi_1^{-1}$ is holomorphically conjugate to $\phi_2 \circ F \circ \phi_2^{-1}$ in some neighbourhood of the Julia set, where ϕ_i is a quasiconformal homeomorphism with Beltrami coefficient μ_i defined on some neighbourhood of the filled Julia set of F.

Definition. We denote by $|\nu|_T$ the Teichmüller norm of a Beltrami vector ν and by $|\mu|_{JT}$ the infimum of the L^∞ norm of Beltrami coefficients that are equivalent to μ. Notice the Julia-Teichmüller distance between $[F]$ and the germ associated to μ is equal to $\log \frac{1+|\mu|_{JT}}{1-|\mu|_{JT}}$. We say that ν is ϵ-efficient as a Beltrami vector if $|\nu|_T \geq (1-\epsilon)|\nu|_\infty$. Similarly, a Beltrami coefficient is ϵ-extremal if $|\mu|_{JT} \geq (1-\epsilon)|\mu|_\infty$. We can restate the Almost Geodesic Principle in the above

language as:

Theorem 8.1. **(The Almost Geodesic Principle)** *Given $\epsilon > 0$ and $1 > d > 0$, there exists $\delta = \delta(d, \epsilon)$ such that if ν is a δ-efficient Beltrami vector of F then the Beltrami coefficients $\mu_t = t\nu$, $0 \leq t \leq \frac{d}{|\nu|_\infty}$ are ϵ-extremal.*

Now we will prove the converse.

Theorem 8.2. **(The Coiling Lemma)** *Given $\epsilon > 0$ and $d > 0$, there exists $\delta = \delta(\epsilon, d)$ such that if ν is a Beltrami vector of the germ of F and the Beltrami vector $\mu_t = t\nu$ is δ-extremal for some $t < \frac{d}{|\nu|_\infty}$ then ν is ϵ-efficient.*

Firstly we will prove the statement of Theorem 8.2 for Teichmüller spaces of Riemann surfaces. We start with the classical situation. Let Γ be a Fuchsian group, i.e., a group of Moebius transformations that acts discontinuously in $\mathbb{D}$. A *Beltrami vector of* Γ is an essentially bounded measurable map $\nu \in L^\infty(\mathbb{D})$ which is invariant under Γ as a Beltrami differential:

$$A^*\nu(z) = \nu(A(z))\frac{\overline{\partial A(z)}}{\partial A(z)} = \nu(z)$$

for all $z \in \mathbb{D}$ and for all $A \in \Gamma$. A *holomorphic quadratic differential of* Γ is a holomorphic map $\phi\colon \mathbb{D} \to \mathbb{C}$ such that $|\phi|$ is integrable on each fundamental domain of Γ and ϕ is invariant under Γ as a quadratic differential, i.e.,

$$A_*\phi(z) = \phi(A^{-1}(z)) \cdot (\partial(A^{-1}(z))^2 = \phi(z)$$

for all $z \in \mathbb{D}$ and for all $A \in \Gamma$. The Teichmüller norm $|\nu|_T$ of the Beltrami vector ν is defined as the infimum of $|\eta|_\infty$ over all Beltrami vector η of Γ satisfying the property that $\int_N \eta \cdot \phi dz d\bar{z} = \int_N \nu \cdot \phi dz d\bar{z}$ for all holomorphic quadratic differentials ϕ of Γ, where N is a fundamental domain of Γ. A Beltrami coefficient of Γ is a Beltrami vector of L^∞ norm smaller than one. The Beltrami coefficients μ_1 and μ_2 are equivalent if the quasiconformal homeomorphisms $h^{\mu_i}\colon \mathbb{D} \to \mathbb{D}$

coincide in the boundary of $\mathbb{D}$, where h^{μ_i} denotes the quasiconformal homeomorphism with Beltrami coefficient μ_i normalized so that the points $1, i, -1$ are fixed points. The Beltrami coefficient μ is ϵ-extremal if

$$|\mu|_\infty \leq (1+\epsilon)\inf\{|\tilde{\mu}|_\infty\,;\ \tilde{\mu} \text{ is a Beltrami coefficient of } \Gamma \text{ equivalent to } \mu\}.$$

Lemma 8.1. *Given $\epsilon > 0$ and $0 < d < 1$, there exists $\delta > 0$ with the following property. If Γ is any Fuchsian group, ν is a Beltrami vector of Γ and the Beltrami coefficient $\mu_t = t\nu$ is δ-extremal for some $t \leq d/|\nu|_\infty$ then ν is ϵ-efficient.*

Proof. We will only sketch the main steps of the proof and refer to Section 7, Chapter V from Lehto (1987) and Chapters 5 and 6, Gardiner (1987) for the details.

Step 1: Let ν be a Beltrami vector of Γ and take $w \in \mathbb{C}$ with $|w| \leq 1$. Let $h_{w\cdot\nu}$ denote the quasiconformal homeomorphism of the Riemann sphere whose Beltrami coefficient coincides with $w \cdot \nu$ on $\mathbb{D}$ and vanishes on $\bar{\mathbb{C}} \setminus \mathbb{D}$. Then

$$\lim_{w\to 0} \frac{S(h_{w\cdot\nu})(z)}{w} = 0$$

for all $z \in \mathbb{C} \setminus \mathbb{D}$, where S denotes the Schwarzian derivative, if and only if ν is infinitesimally trivial.

Proof of Step 1: We can normalize $h_{w\nu}$ so that it fixes infinity and its derivative at infinity is equal to one. With this normalization, we can use the representation of quasiconformal mappings, see Lehto (1986, pp. 27). This gives

$$h_{w\nu}(z) = z + w \cdot X(z) + O(w^2)$$

where X is the vector field in the Riemann sphere that vanishes at infinity and satisfies $\bar{\partial}X = \nu$, i.e., $X(z) = T\nu(z) = -\frac{1}{\pi}\int_{\mathbb{C}} \frac{\nu(u+iv)}{u+iv-z}dudv$. Since ν vanishes in $\mathbb{C} \setminus \mathbb{D}$ we also have that $X(z) = \int_{\mathbb{D}} \frac{\nu(u+iv)}{u+iv-z}dudv$. On the other hand, since ν is infinitesimally trivial, there exists a vector field V on $\mathbb{D}$ that vanishes at the boundary of $\mathbb{D}$ and is such that $\bar{\partial}V = \nu$ on $\mathbb{D}$, see Theorem 6.2. Hence V is equal to X on $\mathbb{D}$. This implies that X vanishes at the boundary of $\mathbb{D}$ and since X is holomorphic on $\mathbb{C} \setminus \mathbb{D}$ we have that X vanishes at $\mathbb{C} \setminus \mathbb{D}$. Since $h_{w\nu} = z + w \cdot X(z) + O(w^2)$, the limit in question at a point $z \in \mathbb{C} \setminus \mathbb{D}$ is equal to the third derivative of X at z which vanishes. This proves step 1.

Step 2: If ν is an infinitesimally trivial Beltrami vector then, for each $0 \leq t \leq \frac{1}{2|\nu|_\infty}$, there exists a Beltrami coefficient σ_t, which is equivalent to $t\nu$ as a Beltrami coefficient and which satisfies

$$|\sigma_t|_\infty \leq 3(|\nu|_\infty)^2 t^2.$$

Proof. Proof of Step 2 The proof of this statement follows from Step 1) and some universal estimates on the Schwarzian derivatives of univalent functions. In fact, the proof of this step requires many arguments from classical Teichmuller theory and is contained in Lemma 7.1 from Lehto (1987, p.227), see also Chapters 5 and 6 from Gardiner (1987). Let us briefly sketch the proof of Step 2 and refer to these monographs for the details. (Notice that there is a difference in the formulas in Lehto and Gardiner because Lehto uses the Poincaré disc $\mathbb{D}$ as the universal cover of the Riemann surfaces whereas Gardiner uses the upper half-plane. In our book we are following the notation of Lehto. Hence we consider the Poincaré disc $\mathbb{D}$ and also the disc $\mathbf{E}$ which is the complement of $\mathbb{D}$ in the Riemann sphere. The group Γ of deck transformations acts both in $\mathbb{D}$ and in $\mathbf{E}$. The mapping $I\colon \mathbf{E} \to \mathbb{D}$, $I(z) = \frac{1}{\bar{z}}$ is an anti-holomorphic diffeomorphism that commutes with the elements of Γ and can be used to transform Γ-invariant objects on $\mathbf{E}$ into Γ invariant objects on $\mathbb{D}$. In particular, if ϕ is a Γ-invariant quadratic differential of $\mathbf{E}$ then $z \mapsto z^{-4}\overline{\phi(\frac{1}{\bar{z}})}$.

Let us summarize the arguments that are used for the proof of Lemma 7.1 of Lehto (1987).

The argument goes as follows. Define $\beta\colon w \mapsto S(h_{w\cdot\nu})$. This is a map from the space of Beltrami coefficients into the space of holomorphic quadratic differentials. Now let the norm in the space of holomorphic Γ-invariant quadratic differentials be the hyperbolic supremum norm, i.e., the sup of the absolute value of ϕ multiplied by the Poincaré density to the power -2 (this defines a function on the quotient space because this product is Γ-invariant). Note that β is complex analytic and goes into the ball of radius 6. This follows from Nehari's inequality, see Lehto (1987, Thm.1.3, p.60) and Gardiner (1987, Lemma 6, p.99).

This map β from the space of Beltrami coefficients into the space of holomorphic quadratic differentials has a holomorphic cross section $\varphi \to \mu_\varphi$ where $\mu_\varphi = \overline{\varphi}\cdot$ (Poincaré metric)$^{-2}$ over the ball of radius 2. This follows from a result of Ahlfors-Weill, see Gardiner (1987, Lemma 7, p.100) and Lehto (1987, Thm 5.1, p. 87).

Now β exactly collapses Teichmüller equivalence classes to points because the quadratic differential precisely determines the quasicircle boundary (up to Moebius transformations) which is the image of the boundary of the upper half plane (or disk), see Gardiner (1987, Lemma 3, p.98).

Since the Teichmüller metric is induced from the metric on Beltrami coefficients and β is holomorphic, Schwarz's lemma implies that the Teichmüller metric is bigger than the Poincaré metric $P(6)$ on the ball of radius 6. Since the cross section is holomorphic on the ball of radius 2 we see that the Teichmüller metric on β^{-1} (ball of radius 2) is smaller than the Poincaré metric $P(2)$ on the ball of radius 2. Since $P(6)$ and $P(2)$ are equivalent to the Banach space supremum metric on the ball of radius 1 we deduce the map β is an equivalence of metrics between the Teichmüller metric and the Banach space metric on a small ball of universal size (say 1). (Notice we are close but have not had to introduce a complex manifold structure on the Teichmüller space for this argument.) See Gardiner (1987, Thm 3, p.104) for this.

From Step 1 when ν is infinitesimally trivial, the map β on the ν disk $\{w \cdot \nu\}$ vanishes to first order at the origin. By Cauchy's Integral Formula the 2^{nd} derivative is bounded (between domain β and the Banach space) on a fixed size ball. By the above the Teichmüller length is $O(|w|^2)$ with universal constants.

Step 3: Next we use a small modification of the proof of Lemma 7.2 from Lehto (1987, p. 227). Suppose that μ is a δ-extremal Beltrami coefficient satisfying $|\mu|_\infty \leq d$ which is not ϵ-efficient as a Beltrami vector. So by definition there exists a Beltrami vector μ_1 such that $|\mu_1|_\infty \leq (1-\epsilon)|\mu|_\infty$ and $\nu = \mu - \mu_1$ is infinitesimally trivial. For $t < \frac{|\mu|_\infty}{10}$, let λ_t be the Beltrami coefficient of the normalized quasiconformal homeomorphism

$$h^{\lambda_t} = h^\mu \circ (h^{t\nu})^{-1}$$

and hence

$$\lambda_t(w) = \frac{\mu(z) - t\nu(z)}{1 - t\mu(z)\overline{\nu(z)}},$$

where $w = h^{t\nu}(z)$. We claim that there exist positive numbers δ_0 and t_0 that depend only on ϵ and on d such that

$$|\lambda_t|_\infty \leq |\mu|_\infty - \delta_0 t \quad \text{for} \quad t \leq t_0 \cdot |\mu|_\infty.$$

Proof of Step 3: Let $E_1 = \{z \in \mathbb{D};\ |\mu(z)| \leq (1 - \frac{\epsilon}{2})|\mu|_\infty\}$ and $E_2 = \mathbb{C} \setminus E_1$. Since $|\nu|_\infty \leq |\mu|_\infty + |\mu_1|_\infty \leq 2|\mu|_\infty \leq 2$, we have that, for $z \in E_1$,

$$|\lambda_t(w)| \leq \frac{|\mu(z)| + 2t|\mu|_\infty}{1-2t} \leq |\mu|_\infty - \delta_1 \cdot t$$

if $t \leq t_1|\mu|_\infty$ where $t_1 = \frac{\epsilon}{100}$ and $\delta_1 = 1$. For $u \in \mathbb{C} \setminus \{0\}$ and $v \in \mathbb{C}$, $|v| \leq 2$, the derivative at $t = 0$ of the mapping $t \mapsto |\frac{u - tv}{1 - tu\overline{v}}|$ is equal to $-\frac{1-|u|^2}{|u|} Re(u \cdot \overline{v})$. Therefore, the supremum C over the set

$$\left\{ \frac{1}{t^2} \left| \frac{u - tv}{1 - tu\overline{v}} - u + \frac{1 - |u|^2}{|u|} Re(u \cdot \ \overline{v}) \cdot t \right|;\ |u| \geq (1-\epsilon)|\mu|_\infty,\ t \leq \frac{|\mu|_\infty|}{10} \right\}$$

is positive and depends only on ϵ and on $|\mu|_\infty$. Therefore, if $z \in E_2$, we have that

$$|\lambda_t(w)| \leq |\mu(z)| - \frac{1 - |\mu(z)|^2}{|\mu(z)|} Re(\mu(z)\overline{\nu(z)}) \cdot t + Ct^2$$

for all $t \leq \frac{|\mu|_\infty}{10}$. On the other hand, for $z \in E_2$,

$$\begin{aligned} \frac{1 - |\mu(z)|^2}{|\mu(z)|} Re(\mu(z)\overline{\nu(z)}) &\geq (1 - |\mu(z)|^2)(|\mu(z)| - |\mu_1(z)|) \\ &\geq (1 - |\mu|_\infty^2) \left((1 - \frac{\epsilon}{2})|\mu|_\infty - |\mu_1|_\infty \right) \\ &\geq (1 - |\mu|_\infty) \frac{\epsilon}{2} |\mu|_\infty \geq (1-d) \frac{\epsilon}{2} |\mu|_\infty. \end{aligned}$$

Therefore, there exist positive numbers t_2 and δ_2 that depends only on ϵ and on d such that $|\lambda_t(w)| \leq |\mu|_\infty - \delta_2 t$ for $t \leq t_2|\mu|_\infty$. Taking $\delta_0 = \min\{\delta_1, \delta_2\}$ and $t_0 = \min\{t_1, t_2\}$, we complete the proof of the Step 3.

Step 4: Let us now complete the proof of the lemma. Given $\epsilon > 0$ and $d > 0$, let $t_0 = t_0(\epsilon, d)$ and $\delta_0 = \delta_0(\epsilon, d)$ be as in the previous step. Let $\delta \in (0, \delta_0 \bar{t}/2)$ where $\bar{t} \leq t_0$ satisfies $\left(1 - \delta_0 \bar{t} + 12\bar{t}^2\right) / \left(1 - 12\bar{t}^2\right) \leq 1 - \delta_0 \bar{t}/2$. Because of Step 2 and since $|\nu| \leq 2$, there exists σ_t equivalent to $t\nu$ as a Beltrami coefficient, such that $|\sigma_t|_\infty \leq 12t^2$. Therefore, if τ_t is the Beltrami coefficient of the quasiconformal homeomorphism $h^{\tau_t} = h^{\lambda_t} \circ h^{\sigma_t}$ then τ_t is equivalent to μ as a Beltrami coefficient and

$$|\tau_t|_\infty \leq \frac{|\lambda_t|_\infty + |\sigma_t|_\infty}{1 - |\sigma_t|_\infty} < \frac{|\mu|_\infty - \delta_0 t + 12t^2}{1 - 12t^2}$$

for all $t \leq t_0 |\mu|_\infty$. Therefore, by the choice of δ,

$$|\tau_{\bar{t} \cdot |\mu|_\infty}| \leq (1 - \delta)|\mu|_\infty.$$

This is impossible because μ is δ-extremal and since τ_t is equivalent to μ as a Beltrami coefficient. □

Let us now sketch the proof of the statement of Lemma 8.1 in the setting of the lamination $\mathcal{L}$ of a quadratic-like mapping.

The lamination $\mathcal{L}$ has a Z cover $\hat{\mathcal{L}}$ whose natural boundary is the solenoid and whose deck transformation becomes the dynamics on the solenoid, which is the mapping $\varprojlim$ (degree two mapping of the circle). We think of $\hat{\mathcal{L}}$ as obtained by naturally attaching an upper half plane to each affine line of the solenoid. We can further unwrap $\hat{\mathcal{L}}$ in the solenoidal direction to obtain a further covering space $\hat{U}$ whose natural boundary is the (real line $\times$ Cantor set) with a 2 generator covering group, which we denote Γ. One generator comes from the deck transformation of $\hat{\mathcal{L}}$ and the other a translation in the $\mathbb{R}$-direction $\times$ (adding one on the 2-adic Cantor set). The details of this action are not important, we only need to work equivariantly with respect to this Γ whatever it is.

Now we consider the Steps 1, 2, 3, 4 of the above proof of Lemma 8.1. Instead of one hyperbolic plane invariant under a Fuchsian group, we now have a Cantor set of hyperbolic planes permuted isometrically by our group Γ. We construct β using the solution of the Beltrami equation. Continuous dependence on the Cantor set direction λ follows if we assume ν is lifted from a Beltrami vector on the lamination with its complex structure which is smooth in the leaf direction and continuous in the transverse direction λ. For the ν lifted is λ-continuous for convergence on compact sets of the open disk D containing support ν, and such continuity gives required λ continuity of the map β. This amounts to the fact that a large Poincaré disk about z near the boundary of D contains most of the mass of the integrals used to construct β, see Theorem 6.2.

The equivariance with respect to Γ is automatic and the other considerations of Steps 1 and 2 go through automatically. Steps 3 and 4 are exactly the same.

Proof of Theorem 8.2: The above discussion proves the analogue of Lemma 8.1 and Theorem 8.2 for Beltrami vectors ν on the lamination which are admissible, i.e., leafwise smooth and transversally continuous, with respect to the sup dilation metric between admissible isotopy classes of admissible conformal structures on the laminations. Now the transversally locally constant part of these objects (denoted by TLC) is what concerns us for Theorem 8.2 because these are precisely the ones coming from the dynamical plane.

Easy vector field approximation arguments using a partition of unity shows leafwise smooth TLC conformal structures and TLC Beltrami vectors are dense and the TLC Teichmüller metric agrees with the lamination Teichmüller metric restricted to the dense TLC subset, see the proof of Theorem 6.3. □

Step 2: The contraction of the renormalization operators

Now we complete the proof of Theorem 1.1. Take two germs of infinitely renormalizable quadratic-like maps $[F]$ and $[\tilde{F}]$ which are combinatorially equivalent and of combinatorial type $\leq N$.

Theorem 8.3. *If $[F]$ and $[\tilde{F}]$ are germs of symmetric quadratic-like maps which are infinitely renormalizable and of the same bounded combinatorial type then $d_{JT}(\mathcal{R}^n([F]), \mathcal{R}^n([\tilde{F}]))$ goes to zero as $n \to \infty$.*

Proof. It is enough to prove that there exists $0 < \lambda < 1$ such that for each pair of germs $[F], [\tilde{F}]$, as in the statement of the theorem, there exists an integer n, that may depend on F and $\tilde{F}$, such that $d_{JT}(\mathcal{R}^n([F]), \mathcal{R}^n([\tilde{F}]) \leq \lambda d_{JT}([F], [\tilde{F}])$. The proof will combine Theorem 8.1, Theorem 8.2 and the complex bounds. To get a definite contraction we start by taking the Beltrami coefficient of a quasiconformal conjugacy which almost realizes the Julia-Teichmüller distance. Using this Beltrami coefficient as a Beltrami vector we construct, by multiplying the Beltrami vector by a positive real parameter, a very long arc of Beltrami coefficient so that the Julia-Teichmüller distance of the end-points is 10 times the constant given by the complex bounds. Next we iterate the arc by the renormalization operator long enough to see, via the complex bounds, a big contraction of the Julia-Teichmüller distance of the end-points. By Theorem 8.1, this implies that the tangent Beltrami vector at the origin of the image path cannot be too efficient. Combining this with Theorem 8.2 we get a definite contraction of the Julia-Teichmüller distance between the original germs under this iteration.

Let us be more precise. Let $\alpha\colon [0,1) \to [0,\infty)$ be defined by $\alpha(x) = \log \frac{1+x}{1-x}$. Let l_0 be the constant given by the complex bounds: for each pair of germs $[F_0], [F_1]$ of the same bounded combinatorial type, there exists n such that $d_{JT}(\mathcal{R}^n([F_0]), \mathcal{R}^n([\tilde{F}_1])) \leq l_0$. Let $0 < d_1 < d_5 < d_{10} < 1$ be such that $\alpha(d_i) = i \times l_0$. In the hypothesis of Almost Geodesic Principle, we take $d = d_{10}$, $1 - \epsilon = \frac{d_1}{d_5}$ and the corresponding δ we call a. Now in the hypothesis of Theorem 8.2 we take $\epsilon = a$, $d = d_{10}$ and the corresponding δ we denote by b. Let $c > 0$ be such that $(1+c)(1-b) < 1$ and $\frac{(1+c)d_5}{d_{10}} < 1$. Let $\lambda = \max\{\frac{\alpha((1-b)(1+c)y)}{\alpha(y)}, \frac{\alpha((1+c)d_5 d_{10}/z)}{\alpha(z)}\ ;\ 0 < y, z \leq d_{10}\}$. Clearly, $0 < \lambda <$

1. Choose a quasiconformal conjugacy between representatives $F\colon U \to V$, $\tilde{F}\colon \tilde{U} \to \tilde{V}$, whose Beltrami coefficient μ satisfy

$$|\mu|_\infty \le (1+c)\cdot|\mu|_{JT}$$

i.e., the logarithm of the conformal distortion of this conjugacy is very near the Julia-Teichmüller distance between $[F]$ and $[\tilde{F}]$. Let us consider the Beltrami coefficient of F, $\mu_1 = d_{10}\frac{\mu}{|\mu|_\infty}$. Let $F_1 = h^{\mu_1}\circ F\circ(h^{\mu_1})^{-1}$. By the complex bounds, there exists an integer n such that $d_{JT}(\mathcal{R}^n([F_1]),\mathcal{R}^n([F])) \le l_0$. This is the same as $|\mathcal{R}^n(\mu_1)|_{JT} \le d_1$, where $\mathcal{R}^n(\mu_1)$ is the restriction of μ_1 to a neighbourhood of the Julia set of $\mathcal{R}^n(F)$ which is a restriction of F^n to a neighbourhood of the critical point of F. If $|\mathcal{R}^n(\mu)|_\infty \le \frac{d_5}{d_{10}}|\mu|_\infty$ then

$$|\mathcal{R}^n(\mu)|_{JT} \le |\mathcal{R}^n(\mu)|_\infty \le \frac{d_5}{d_{10}}|\mu|_\infty \le \frac{(1+c)d_5}{d_{10}}|\mu|_{JT}.$$

Therefore,

$$d_{JT}(\mathcal{R}^n([\tilde{F}]),\mathcal{R}^n([F])) = \alpha(|\mathcal{R}^n(\mu)|_{JT}) \le \lambda\alpha(|\mu|_{JT}) = \lambda(d_{JT}([\tilde{F}],[F]))$$

by the choice of λ. So we can assume that

$$|\mu|_\infty \ge |\mathcal{R}^n(\mu)|_\infty \ge \frac{d_5}{d_{10}}|\mu|_\infty$$

Since $\mathcal{R}^n(\mu_1) = \frac{d_{10}}{|\mu|_\infty}\mathcal{R}^n(\mu)$ we get that

$$d_{10} = |\mu_1|_\infty \ge |\mathcal{R}^n(\mu_1)|_\infty \ge \frac{d_{10}}{|\mu|_\infty}\frac{d_5}{d_{10}}|\mu|_\infty \ge d_5.$$

Since $|\mathcal{R}^n(\mu_1)|_{JT} \le d_1$ (the complex bounds), we have that

$$|\mathcal{R}^n(\mu_1)|_\infty \ge \frac{d_5}{d_1}|\mathcal{R}^n(\mu_1)|_{JT}.$$

Because of the Almost Geodesic Principle, this implies that

$$|\mathcal{R}^n(\mu)|_T < (1-a)|\mathcal{R}^n(\mu)|_\infty.$$

From this last inequality we get

$$|\mathcal{R}^n(\mu)|_{JT} \le (1-b)|\mathcal{R}^n(\mu)|_\infty,$$

see Theorem 8.2. Therefore,

$$|\mathcal{R}^n(\mu)|_{JT} \le (1-b)|\mathcal{R}^n(\mu)|_\infty \le (1-b)|\mu|_\infty \le (1-b)(1+c)|\mu|_{JT}.$$

Thus,

$$\begin{aligned} d_{JT}(\mathcal{R}^n([\tilde{F}]),\mathcal{R}^n([F])) &= \alpha(|\mathcal{R}^n(\mu)|_{JT}) \\ &\le \alpha((1-b)(1+c)|\mu|_{JT}) \\ &\le \lambda\alpha(|\mu|_{JT}) = \lambda d_{JT}([\tilde{F}],[F]). \end{aligned}$$

This completes the proof. □

Next we will use the previous results to prove the contraction of the renormalization operator in the C^r topology for every r. Let $\mathcal{D}_N$ be the set of real analytic maps of the interval $[0,1]$ that are infinitely renormalizable with combinatorial type bounded by N. We will consider several topologies on this space. As usual, we say that f_n converges to f in the C^r topology if $f_n - f$, together with all its derivatives up to order r converges uniformly to zero. We say that f_n converges strongly to f if there exists an open neighbourhood W of the interval $[0,1]$ in the complex plane, and holomorphic extensions F_n of f_n and F of f to W such that F_n converges uniformly to F on W. Clearly, if f_n converges strongly to f then f_n converges C^r to f for all r.

Lemma 8.2. *Suppose that $f_n \in \mathcal{D}_N$ converges uniformly to f and that each f_n extends to a quadratic-like map F_n whose conformal type is uniformly bounded. Then f_n converges strongly to f.*

Proof. Under the hypothesis of the boundedness of the conformal type we have, by trimming the domains of the quadratic-like extension as in the proof of Theorem 4.2d, that there is a neighbourhood of the interval $[0,1]$ contained in the domain of all F_n and that every subsequence of F_n has a subsequence which converges uniformly on this neighbourhood. Since the sequence f_n converges, it follows that the limit of all convergent subsequences of F_n must be equal to f on the interval $[0,1]$. Hence the sequence F_n converges uniformly. □

Lemma 8.3. *Let $F_n\colon U_n \to V_n$ be a sequence of quadratic-like maps normalized so that the dynamical interval is always $[0,1]$. If the conformal type of F_n is bounded by B for all n then there exist a sequence $n_i \to \infty$, a quadratic-like map $F\colon U \to V$ with the same normalization and a neighbourhood W of $[0,1]$ in the complex plane, with the following properties:*

1. *$F_{n_i}|W$ converges to $F|W$ uniformly;*
2. *W is an open neighbourhood of $J(F)$ and of $J(F_{n_i})$ for all i;*
3. *$F|W$ is quadratic-like.*

Proof. Follows from Lemma 4.5. □

Lemma 8.4. *Given $\tilde{\epsilon} > 0$, there exists $\epsilon > 0$ and $L >> 10$ with the following property. If H is a $(1+\epsilon)$-quasiconformal homeomorphism that maps the unit interval in itself fixing the endpoints, whose domain and range contains the disc of radius L then*

$$(*) \qquad |H(z) - z| < \tilde{\epsilon}\ ,\ |H^{-1}(z) - z| < \tilde{\epsilon}$$

for all $z \in \mathbb{D}_{10} = \{z \in \mathbb{C}\,;\ |z| \le 10\}$.

Proof. Follows from Koebe's Distortion Lemma and from compactness property of the set of quasiconformal homeomorphisms of the unit disc. □

Lemma 8.5. *Let $f, g \in \mathcal{D}_N$ be maps of the same combinatorial type. Then $\mathcal{R}^n(f) - \mathcal{R}^n(g)$ converges strongly to zero.*

Proof. If $f, g \in \mathcal{D}_N$ have the same combinatorial type, we define $d_{JT}(f, g)$ to be the Julia-Teichmüller distance between quadratic-like extensions F and G of f and g respectively. By Theorem 8.3 we have that $d_{JT}(\mathcal{R}^n(f), \mathcal{R}^n(g))$ converges to zero. Let W be a neighbourhood of the dynamical interval that is contained in the domains of F_n and G_n. It is enough to prove that for n big enough, $||F_n|W - G_n|W||_0 \leq \frac{1}{2}||F_0|W - G_0|W||_0$. Let $\tilde{\epsilon} > 0$ be such that if $\tilde{F}, \tilde{G} \colon W \to \mathbb{C}$ are arbitrary maps such that $\tilde{G} = H^{-1} \circ \tilde{F} \circ H$ with H a quasiconformal homeomorphism satisfying $(*)$, then $|\tilde{F}(z) - \tilde{G}(z)| \leq \frac{1}{2}||F_0 - G_0||_0$ for all $z \in W$. Let $\epsilon > 0$ be as in Lemma 8.4. Since $d_{JT}(\mathcal{R}^n(f), \mathcal{R}^n(g)) \to 0$ there exists a $(1+\epsilon)$-quasiconformal conjugacy ϕ between F_{n_1} and G_{n_1} defined on a neighbourhood W_1 of the dynamical interval. By rescaling, ϕ defines a $(1+\epsilon)$-quasiconformal homeomorphism H_n that conjugates F_n and G_n in W for $n \geq n_1$ and such that the domain and range of H_n contains a disc whose radius increases exponentially with n. Hence, for $n \geq \hat{n}$, for some integer $\hat{n}$ big enough, the domain and range of H_n contains the disc of radius L. Hence the lemma follows from Lemma 8.4. □

Lemma 8.6. *Given $\epsilon > 0$, there exists $\delta > 0$ such that if $f, g \in \mathcal{D}_N$ have quadratic-like extensions of conformal type bounded by B and $d_{JT}(f, g) \geq \epsilon$, then the C^0 distance between f and g is at least δ.*

Proof. Suppose, by contradiction, that there exist maps f_n, g_n satisfying the conditions of the lemma and such that $f_n - g_n$ converges uniformly to zero. By taking a subsequence we may assume that f_n converges uniformly to a map $f \in \mathcal{D}_N$. By Lemma 8.3, there exist a neighbourhood W of the dynamical interval such that for n big enough, both f_n and g_n have quadratic-like extensions F_n, G_n to W and they converge uniformly to a quadratic-like extension F of f. Hence, for n big enough, we can construct fundamental domains of F_n and G_n and a quasiconformal map between these fundamental domains that conjugates F_n with G_n in the boundary and have very small conformal distortion. By the pullback argument we can extend these maps to quasiconformal conjugacies between F_n and G_n with the same conformal distortion. This is a contradiction that proves the lemma □

Lemma 8.7. *Let f_n and g_n, $n \in \mathbb{Z}$ be infinitely renormalizable maps of the same bounded combinatorial type that have quadratic-like extensions F_n and G_n of conformal type bounded by B. Suppose the dynamical interval is always $[0,1]$ and that $\mathcal{R}(f_n) = f_{n+1}$ and $\mathcal{R}(g_n) = g_{n+1}$ for all $n \in \mathbb{Z}$. Then $g_0 = f_0$.*

Proof. Let us first prove that the Julia-Teichmüller distance between f_n and g_n is equal to zero. Suppose, by contradiction, that $d_{JT}(f_{-n_0}, g_{-n_0}) \geq \epsilon > 0$. Since the Julia-Teichmüller distance cannot increase under renormalization, we have that $d_{JT}(f_{-n}, g_{-n}) \geq \epsilon$ for all $n \geq n_0$. Hence, there exists $\delta > 0$ such that the C_0 distance between f_{-n} and g_{-n} is bigger or equal δ for all $n \geq n_0$.

Let us take a subsequence $n_i \to -\infty$ such that f_{n_i} converges strongly to f and g_{n_i} converges strongly to g. Since f_{n_i} and g_{n_i} have the same combinatorial type then f and g also have the same combinatorial type. Hence, by Lemma 8.5, there exists $k \in \mathbb{N}$, such that the C^0 distance between $\mathcal{R}^k(f)$ and $\mathcal{R}^k(g)$ is smaller than $\frac{\delta}{2}$. On the other hand, the C^0 distance between $\mathcal{R}^k(f)$ and $\mathcal{R}^k(f_{n_i})$ converges to zero as $i \to \infty$ as well as the C^0 distance between $\mathcal{R}^k(g)$ and $\mathcal{R}^k(g_{n_i})$. This is a contradiction that proves that the Julia-Teichmüller distance between f_{-n} and g_{-n} must be zero. Therefore, F_{-n} and G_{-n} are holomorphically conjugate in a (probably small) neighbourhood of the Julia set. By Lemma 8.3, there exist quadratic-like maps $F\colon W_1 \to F(W_1)$ and $G\colon W_2 \to G(W_2)$, normalized so that their dynamical intervals are equal to $[0,1]$, and a sequence $n_i \to \infty$ such that W_1 contains the Julia set of F_{n_i}, W_2 contains the Julia set of G_{n_i}, $F_{n_i}|W_1$ converges uniformly to F and $G_{n_i}|W_2$ converges uniformly to G. Clearly F and G are infinitely renormalizable of the same bounded combinatorial type. Let us prove that the Julia-Teichmüller distance between F and G is zero. Indeed, since F_{n_i} converges uniformly to F, there exists a quasiconformal homeomorphism with small conformal distortion of a neighbourhood of the fundamental domain $F_{n_i}(W_1) \setminus W_1$ on a neighbourhood of the fundamental domain $F(W_1) \setminus W_1$ that conjugates F and F_{n_i} in the boundary of W_1. Using the pullback argument, see Theorem 4.2c, and the Douady-Hubbard construction we get that the Julia-Teichmüller distance between the external class of F_{n_i} and the external class of F goes to zero as $i \to \infty$. Similarly, the Julia-Teichmüller distance between the external class of G_{n_i} and the external class of G goes to zero. Again, from the pullback argument, the external class of F_{-n} is equal to the external class of G_{-n} since F_{-n} and G_{-n} are holomorphically conjugated in a neighbourhood of the Julia set. Hence, the Julia-Teichmüller distance between the external class of F and that of G is zero. This proves (again by the pullback argument) that the Julia-Teichmüller distance between F and G is zero. Therefore, there exist neighbourhoods $U_1 \subset W_1$ of $J(F)$, $U_2 \subset W_2$ of $J(G)$ such that $F(U_1)$ contains the closure of U_1, and a there exists a holomorphic conjugacy $H\colon F(U_1) \to G(U_2)$ between F and G. Using again the pullback argument, we get from this that there exist quasiconformal conjugacies $H_{n_i}\colon U_1 \to U_2$ between F_{n_i} and G_{n_i} and that the conformal distortion of these conjugacies goes to zero as i goes to infinity. Now F_0 is a renormalized map of F_{n_i} with a very long renormalization period. Hence, the dynamical interval of F_0 inside the dynamical interval of F_{n_i} is a very small interval and the restriction of H_{n_i} to this interval is again a conjugacy between F_0 and G_0. If we rescale, we get a sequence of quasiconformal homeomorphisms $\tilde{H}_{n_i}$ which are defined in neighbourhoods of the dynamical interval of f_0 that grows to infinity with i and such homeomorphisms have conformal distortion that goes to zero and they conjugate f_0 and g_0. By taking a converging subsequence we get a conjugacy between f_0 and g_0 which is holomorphic in the whole plane. Since the dynamical interval is $[0,1]$ for both maps, the above conjugacy must be the identity. $\square$

Theorem 8.4. *Let $\mathcal{A} \subset \mathcal{S}_N$ be the set of limit points of the renormalization operator (in the C^0 topology). Given any bi-infinite sequence $\underline{\sigma} = (\dots, \sigma_{-n}, \dots, \sigma_0, \dots, \sigma_n, \dots)$ of unimodal permutations with $|\sigma_i| \le N$, there exists a unique sequence $f_{-n} \in \mathcal{A}$, $n \ge 0$, with the following properties:*

1. $f_{-n} \in \mathcal{D}_{\sigma_{-n},\dots,\sigma_0,\dots}$ *for* $n \ge 0$.
2. $\mathcal{R}(f_{-n}) = f_{-n+1}$.

Proof. The uniqueness of the above sequence of maps was already proved in Lemma 8.7. Let us prove the existence. First, we prove the existence in the case where the sequence $\underline{\sigma}$ is periodic, under the shift map, of period m. In this case, $\mathcal{D}_{\underline{\sigma}}$ is invariant under $\mathcal{R}^m$. Let $g \in \mathcal{D}_{\underline{\sigma}}$ be a map in some Epstein class (say g is a quadratic map). By the complex bounds, there exists a sequence $n_i \to \infty$ such that $\mathcal{R}^{n_i m}(g)$ converges to a map $f_0 \in \mathcal{A}$. By taking the limit of a convergent subsequence of $\mathcal{R}^{(n_i-1)m}$ we get a map $f_{-m} \in \mathcal{D}_{\underline{\sigma}}$ such that $\mathcal{R}^m(f_{-m}) = f_0$. Similarly, from this subsequence, we take a subsequence so that $\mathcal{R}^{(n_i-2)k}$ converges to f_{-2m}. Repeating this argument, we get a sequence $f_{-km} \in \mathcal{D}_{\underline{\sigma}}$ such that $\mathcal{R}^m(f_{-km}) = f_{-(k-1)m}$. From this we get the desired sequence by taking $f_{-n} = \mathcal{R}^{km-n}(f_{-km})$ where k is such that $km - n \ge 0$. This proves the existence in the case of periodic data. If $\underline{\sigma}$ is not periodic, we take a sequence $\underline{\sigma}_i$ of periodic data converging to $\underline{\sigma}$ (in the product topology), and, since all the maps belong to the compact set $\mathcal{A}$, we can take convergent subsequences and get the existence in the non-periodic case.

Proof of Theorem 1.1: Let Σ_N be the space of all bi-infinite sequences $\underline{\sigma} = (\dots, \sigma_{-n}, \dots, \sigma_0, \sigma_1, \dots)$ of unimodal permutations with $|\sigma_i| \le N$. Endowed with the product topology, Σ_N is a compact space. From Theorem 8.4, for each $\underline{\sigma} \in \Sigma_N$ there exists a unique bi-infinite sequence of maps $f_n \in \mathcal{A}$ such that $f_n \in \mathcal{D}_{\sigma_n,\sigma_{n+1},\dots}$ and $\mathcal{R}(f_n) = f_{n+1}$. Let $H\colon \Sigma_N \to \mathcal{A}$ be defined as $H(\underline{\sigma}) = f_0$. H is clearly surjective and, by the uniqueness of Lemma 8.7 and the compactness of $\mathcal{A}$, it follows that H is continuous. Furthermore, from Proposition 1.1, H is injective. Hence H is a homeomorphism conjugating $\mathcal{R}|\mathcal{A}$ to the shift of Σ_N.
Let $f, g \in \mathcal{S}_N$ be of class C^{1+z} and have the same combinatorial type. We claim that $\mathcal{R}^n(f) - \mathcal{R}^n(g)$ converges to zero in the $C^{1+\alpha}$ topology for every $0 < \alpha < 1$. Indeed, if this were not the case, by the compactness properties of Section 2, there exists a subsequence $n_i \to \infty$ such that $\mathcal{R}^{n_i}(f) \to f_0$, $\mathcal{R}^{n_i}(g) \to g_0$ in the $C^{1+\alpha}$ metric and $f_0 \ne g_0$. By taking convergent subsequences, we may assume that for each $m \in \mathbb{N}$, $\mathcal{R}^{n_i-m}(f) \to f_{-m}$ and $\mathcal{R}^{n_i-m}(f) \to g_{-m}$. From Lemma 8.4 we get that $f_0 = g_0$ which is a contradiction and the claim is proved.

If f, g are of the same bounded combinatorial type and have quadratic-like extensions of bounded conformal type then, by Lemma 8.5, $\mathcal{R}^n(f) - \mathcal{R}^n(g)$

converges to zero in the strong topology and hence in the C^r topology for any $r < \infty$. From the complex bounds in Section 5, the same is true if f, g are of the same bounded combinatorial type either belong to some Epstein class or have quadratic-like extensions. Conversely, if $\mathcal{R}^n(f)$ and $\mathcal{R}^n(g)$ converge to zero in the C^0 topology then $\mathcal{R}^n(f)$ and $\mathcal{R}^n(g)$ have the same combinatorial type for some n. This follows from the fact that the C^0 distance between two (once) renormalizable maps of bounded period and different types is bounded from below. □

9 Universality of the Attracting Cantor Set

In Sections 2 and 3 we have seen that the ratio geometry of the attracting Cantor set of an infinitely renormalizable $\mathcal{U}^{1+z}$ map of bounded combinatorics is bounded.

Using the contraction of the renormalization operator we will prove here that under the above hypothesis the asymptotic ratio geometry is in fact constant: it depends only on the combinatorics and not on the map. We will show that the asymptotic ratio geometry is given by so called scaling functions. These are continuous functions defined, not on the attracting Cantor set but on a 'dual' Cantor set, and these maps are independent of the map. This is the rigidity result we had mentioned at the introduction of this chapter: two topologically conjugate infinitely renormalizable maps of bounded combinatorial type have the same asymptotic geometrical structure. Assuming the iterates of two C^{1+z} infinite renormalizable maps of the same bounded combinatorial type contract exponentially fast, we will prove the stronger rigidity result of Theorem 9.4.

Let us start by recalling the notation used in previous sections. Let f be an infinitely renormalizable map of class $\mathcal{U}^{1+z}$ of combinatorial type bounded by N. We denote by $\triangle_n$ the n-th renormalizing interval $[f^{2q(n)}(c), f^{q(n)}(c)]$ and by Ξ_n the collection

$$\{\triangle_n, \triangle_n^1 = f(\triangle_n), \dots, \triangle_n^{q(n)-1} = f^{q(n)-1}(\triangle_n)\}$$

of disjoint intervals, where $q(1), q(2), \dots$ is the sequence of renormalizing return times. Hence, by assumption, $a(n) = \frac{q(n+1)}{q(n)} \le N$ for all n. The attracting Cantor set is

$$\Lambda_f = \bigcap_{n=1}^{\infty} \bigcup_{j=0}^{q(n)-1} \triangle_n^j.$$

Let us first consider the Feigenbaum case. In this case $q(n) = 2^n$ for every n. Let us consider the Cantor set $\Lambda = \{0,1\}^{\mathbb{N}}$. Since for each $0 \le i < 2^n$ the interval $\triangle_n^i$ contains both $\triangle_{n+1}^i$ and $\triangle_{n+1}^{i+2^n}$ and no other interval of Ξ_{n+1} we see that the map $\phi \colon \Lambda \to \Lambda_f$ which associates to each sequence $\omega \colon \mathbb{N} \to \{0,1\}$ in Λ the point

$$\phi(\omega) = \bigcap_{n=1}^{\infty} \triangle_n^{k(n,\omega)} \text{ where } k(n,\omega) = \sum_{j=0}^{n-1} \omega(j) 2^j$$

is a well defined homeomorphism. Under ϕ the interval $\triangle_n^k$ corresponds to the cylinder set of all infinite words ω in Λ with the same initial block of length n; more precisely, $\omega^{(n)} = (\omega(0), \ldots, \omega(n-1))$, where $k = \sum_{j=0}^{n-1} \omega(j) 2^j$, see Figure 9.1. We denote this interval $\triangle_n^k$ by $[\omega(0), \ldots, \omega(n-1)]$. Note that

$$[\omega(0), \ldots, \omega(n-1)] \subset [\omega(0), \ldots, \omega(n-2)].$$

In other words $\omega(0) \in \{0,1\}$ determines whether we are in Δ_0 or in $f(\Delta_0)$ and so on. ϕ is a conjugacy between the restriction of f to the attracting Cantor set and the homeomorphism of Λ which is the translation by one: the adding machine. This homeomorphism is defined as follows. Let $\omega \in \Lambda$ and let j_ω be so that $\omega(j_\omega) = 0$ and $\omega(k) = 1$ for all $0 \le k < j_\omega$. Then

$$(\omega+1)(k) = \begin{cases} 1 & \text{if } k = j_\omega, \\ 0 & \text{if } k < j_\omega, \\ \omega(k) & \text{if } k > j_\omega. \end{cases}$$

If $\omega(j) = 1$ for all j then define $(\omega+1)(j) = 0$ for all j. Clearly $\phi(\omega+1) = f \circ \phi(\omega)$.

Definition. The *dual Cantor set* is the set Λ^* of all left infinite words

$$\{\omega = (\ldots, \omega(n), \ldots, \omega(1), \omega(0))\,;\ \omega(i) \in \{0,1\}\,\}$$

endowed with the product topology. The *ratio functions* are the functions $\Sigma_s \colon \Lambda^* \times \mathbb{N} \to \mathbb{R}$ defined by

$$\Sigma_s(\omega, n) = \frac{\text{length of the interval } [\omega(n-1), \ldots, \omega(0)]}{\text{length of the interval } [\omega(n-1), \ldots, \omega(0), s]}, \qquad s = 0, 1.$$

Here, as before, $[\omega(n-1), \ldots, \omega(0)]$ stands for the interval Δ_n^i which corresponds with this block. Note that $[\omega(n-1), \ldots, \omega(0), s] \subset [\omega(n-1), \ldots, \omega(0)]$.

Remark. The reason that Λ^* is called the dual Cantor set is that if we consider the interval $I_n = [\omega(n-1), \ldots, \omega(0)]$ then $\omega(n-1)$ determines whether I_n is contained in $\triangle_0$ or in $f(\triangle_0)$, i.e., the interval on the largest level. Similarly $\omega(n-2)$ determines the interval on level 1 and so on. So in order to determine where the interval $[\omega(n-1), \ldots, \omega(0)]$ is positioned in the real line, the coefficients $\omega(j)$ with the largest j are most important. So if $\omega \in \{0,1\}^{\mathbb{N}}$ then the sequence of intervals $I_n = [\omega(n-1), \ldots, \omega(0)]$ does in general *not* converge as $n \to \infty$.

Theorem 9.1. *Let f be a $\mathcal{U}^{1+z}$ map of Feigenbaum's combinatorial type. Then*

1. *For every $\omega \in \Lambda^*$ and $s \in \{0,1\}$, $\Sigma_s(\omega, n)$ converges to some real number $\sigma_s(\omega)$.*

2. *The functions* $\sigma_0, \sigma_1 \colon \Lambda^* \to \mathbb{R}$ *are continuous, strictly positive and do not depend on* f*.*

Proof. We start by proving the convergence of the sequences $\Sigma_s(\omega, n)$ for ω in a dense subset of Λ^*, namely, we consider left infinite words ω such that $\omega(j) = 0$ for all j big enough. So let p be such that $\omega(j) = 0$ for all $j > p$. Notice that the intervals $[\omega(m+p), \dots, \omega(0)]$ and $[\omega(m+p), \dots, \omega(0), s]$, $s \in \{0, 1\}$ are contained in $[\omega(m+p), \dots, \omega(p+1)]$ and since $\omega(j) = 0$ for all $j > p$, $\triangle_m = [\omega(m+p), \dots, \omega(p+1)]$. So the end points of the intervals $[\omega(m+p), \dots, \omega(0)]$ and $[\omega(m+p), \dots, \omega(0), s]$ are determined by the first $2p$ iterates of the map $f^{2^m}|\triangle_m$. On the other hand, by the contraction of the renormalization operator, this map, after renormalizing the domain, converges in the C^0 topology to Feigenbaum's fixed point g as $m \to \infty$. Therefore for each p, $\lim_{m\to\infty} \Sigma_s(\omega, m+p)$ exists and is equal to the ratio of the corresponding intervals whose endpoints are contained in the first $2p$ iterates of the critical point by the fixed point of the Renormalization operator in the Feigenbaum case. This proves already that not only the limit exists but also that it is independent of the map. So the functions σ_s are defined in a dense subset of Λ^*.

Let us prove that σ_s extends continuously to all of Λ^* and that it coincides with the limits from Statement 1. So let ω be an infinite word in Λ^*. We claim that for each $\epsilon > 0$ there exists $p > 0$, not depending on ω, with the following property: if $\alpha \in \Lambda^*$ is defined by $\alpha(j) = \omega(j)$ for $j \le p$ and $\alpha(j) = 0$ for $j > p$ then $|\Sigma_s(\omega, p+m) - \Sigma_s(\alpha, p+m)|$ is smaller than ϵ for all $m > 0$.

Let us finish the proof using the claim. From the claim we get that for m big enough, the distance between $\Sigma_s(\omega, m+p)$ and $\sigma_s(\alpha)$ is smaller than 2ϵ. Hence the sequence $\Sigma_s(\omega, n)$ converges since it is a Cauchy sequence. Furthermore, since p does not depend on ω we get that σ_s is uniformly continuous.

It remains to prove the claim. From the real bounds we get that there exists a universal constant $\lambda < 1$ such that the length of the interval $[\omega(m+p), \dots, \omega(0)]$ is smaller than λ^p times the length of the interval $[\omega(m+p), \dots, \omega(p+1)] \supset [\omega(m+p), \dots, \omega(0)]$. In Section 2 we also showed that $f^{2^m - j}$ maps an neighbourhood of $\triangle_m^j = [\omega(m+p), \dots, \omega(p+1)]$ diffeomorphically onto a universally scaled neighbourhood of $\triangle_m$. Hence, by the Koebe Principle, its restriction to the much smaller interval $[\omega(m+p), \dots, \omega(0)]$ has distortion less than $1 + \epsilon$ if p is big enough. This proves the claim and the theorem. (By the real bounds $\Sigma_s(\omega, n)$ is uniformly bounded from above and from zero.) □

Corollary 9.1. *If* f, g *are* $\mathcal{U}^{1+z}$ *maps with Feigenbaum's combinatorics then the attracting Cantor sets of* f *and* g *have the same Hausdorff dimension.*

Now our aim is to show that two Cantor sets with the same scaling functions are 'geometrically the same'. In order to prove this we need the following concept.

Definition. The functions σ_0 and σ_1 are called *scaling functions*. Let $\phi\colon \Lambda \to \tilde{\Lambda} \subset \mathbb{R}$ be a homeomorphism between the two Cantor sets. Let $\Sigma_s\colon \Lambda^* \times \mathbb{N} \to \mathbb{R}$ and $\sigma_s\colon \Lambda^* \to \mathbb{R}$ be the ratio and scaling functions of $\tilde{\Lambda}$. We say that the scaling functions σ_s have the *geometric approximation property* if there exist $C > 0$ and $0 < \lambda < 1$ such that

$$|\sigma_s(\omega) - \Sigma_s(\omega, n)| \leq C\lambda^n, \quad \forall\, n \in \mathbb{N} \text{ and } \forall\, \omega \in \Lambda^*.$$

Theorem 9.2. *Let $\phi_i\colon \Lambda \to \Lambda_i \subset \mathbb{R}$, $i = 1, 2$ be homeomorphisms between Cantor sets as above such that $\phi = \phi_2 \circ \phi_1^{-1}\colon \Lambda_1 \to \Lambda_2$ is monotone. If Λ_1 and Λ_2 have the same scaling functions which are bounded away from zero and they have the geometric approximation property then there exists $\alpha > 0$ and a $C^{1+\alpha}$ diffeomorphism $h\colon \mathbb{R} \to \mathbb{R}$ which maps Λ_1 onto Λ_2.*

Proof. We will show that the homeomorphism $\phi_2 \circ \phi_1^{-1}$ extends to a $C^{1+\alpha}$ diffeomorphism of the real line. Let $0 \leq j < q(n)$ and choose $\epsilon_0, \ldots, \epsilon_{n-1} \in \{0, 1\}$ so that $j = \sum_{k=0}^{n-1} \epsilon_k \cdot q(k)$. Let $\triangle_n^j(i)$, $i = 1, 2$ be the interval of the real line corresponding under ϕ_i to the cylinder set

$$\{\omega \in \Lambda\,;\ \omega(k) = \epsilon_k \text{ for } k = 0, \ldots, n-1\}.$$

Let $\Xi_n(i)$ the collection of intervals $\{\triangle_n^j(i)\}$ and let $F_n(i)$ be the union of these intervals. Then $F_n(i) \setminus F_{n+1}(i)$ are called the gaps of order $n + 1$. From the approximation property,

$$\exp(-C_1\lambda^n) \leq \frac{\Sigma_2(\omega, n)}{\Sigma_1(\omega, n)} \leq \exp(C_1\lambda^n)$$

for some constant C_1. From this and the real bounds the following two statements follow easily.

a. The numbers $\frac{|\triangle_n^j(2)|}{|\triangle_n^j(1)|}$ are bounded and bounded away from zero.

b. There exist constants $C > 0$ and $0 < \lambda < 1$ such that if $J(i) \subset \triangle_n^j(i)$ is either an element of Ξ_{n+p} or a gap of order $n + p$ then

$$\frac{|\triangle_n^j(2)|}{|\triangle_n^j(1)|} \exp(-C\lambda^n) \leq \frac{|J(2)|}{|J(1)|} \leq \frac{|\triangle_n^j(2)|}{|\triangle_n^j(1)|} \exp(C\lambda^n).$$

So let us show that the theorem follows from these two statements. Let $x \in \Lambda_1$ and let $\triangle_n^{j(n)}(1)$ be the sequence of intervals which contain x. Statement b) gives that $\frac{|\triangle_n^{j(n)}(2)|}{|\triangle_n^{j(n)}(1)|}$ is a Cauchy sequence and therefore converges to a number l_x. By Statement a) l_x is bounded and bounded away from zero. This shows that if the map $\phi_2 \circ \phi_1^{-1}$ is differentiable at x then its derivative must be l_x. Let us now investigate how l_x varies with x. So let $\alpha > 0$ be such that $|\triangle_n^{j(n)}(i)|^\alpha \geq \text{constant} \cdot \lambda^n$. Then, for $x, y \in \Lambda_1$, we have that

$$|l_x - l_y| \leq \text{constant} \cdot |x - y|^\alpha. \tag{9.1}$$

Indeed, let n be the biggest integer such that the interval $\triangle_n^i$ contains both x and y. From Statement b) we get that

$$\exp(-C\lambda^n)\frac{|\triangle_n^i(2)|}{|\triangle_n^i(1)|} < l_z < \exp(C\lambda^n)\frac{|\triangle_n^i(2)|}{|\triangle_n^i(1)|}$$

for $z \in \{x, y\}$. Therefore,

$$l_y - l_x \leq (\exp(C\lambda^n) - \exp(-C\lambda^n))\frac{|\triangle_n^i(2)|}{|\triangle_n^i(1)|}.$$

Using Statement a) we get from the above inequality that

$$\begin{aligned}|l_y - l_x| &\leq \text{constant} \cdot (\exp(C\lambda^n) - \exp(-C\lambda^n)) \leq \text{constant} \cdot \lambda^n \\ &\leq \text{constant} \cdot |\triangle_n^i|^\alpha.\end{aligned}$$

Since $|x - y| \leq \text{constant} \cdot |\triangle_n^i|$ this gives (9.1).

Let us extend ϕ as follows to $\mathbb{R}$. Take two gaps $G(1) = [a, b]$ and $G(2) = [\phi(a), \phi(b)]$ of Λ_1 respectively Λ_2. Next take a C^∞ diffeomorphism $\phi\colon G(1) \to G(2)$ whose left derivative at a is l_a, whose right derivative at b is l_b and for which

$$|\phi'(x) - \phi'(y)| < K.|x - y| \tag{9.2}$$

where $K = 2 \cdot \max\{l_a, l_b, \frac{|G(2)|}{|G(1)|}\}$. It is not hard to see that this is possible.

So let us show that ϕ is differentiable. So take $x, y \in \Lambda_1$. If $G_i(1)$ (respectively $G_i(2) = \phi(G_i(1))$) are the gaps of Λ_1 (Λ_2) between y and x ($\phi(x)$ and $\phi(y)$) then because Λ_i has Lebesgue measure zero, $|y - x| = \sum_{i=0}^\infty |G_i(1)|$ (respectively $|\phi(y) - \phi(x)| = \sum |G_i(2)|$). From Statement b) we get that

$$\exp(-2C\lambda^n) \cdot l_x < \frac{|G_i(2)|}{|G_i(1)|} < \exp(2C\lambda^n) \cdot l_x$$

This clearly implies that

$$\Big|\frac{|\phi(y) - \phi(x)|}{|y - x|} - l_x\Big| \leq \text{constant} \cdot \lambda^n \tag{9.3}$$

where n is an integer such that $x, y \in \triangle_n^j(i)$. Using (9.2) and (9.3) we get easily that ϕ is differentiable at $x \in \Lambda_1$ and its derivative is equal to l_x. By (9.1) and (9.2) we get that ϕ is $C^{1+\alpha}$. □

Theorem 9.3. *Let f be a $\mathcal{U}^{1+z}$ map with the same combinatorics as the Feigenbaum's fixed point. Suppose that the iterates of f under the renormalization operator $\mathcal{R}$ converge C^0 to the fixed point g with rate $\rho < 1$, i.e., $|R^n(f) - g|_0 <$ constant $\cdot\, \rho^n$. Then the scaling function of the attracting Cantor set Λ_f has the geometric approximation property.*

Proof. It is enough to follow the same steps as in the proof of Theorem 9.1. □

Let us now consider the general situation. Suppose that f and g are infinitely renormalizable maps with the same bounded combinatorial type. Let $q(1), q(2), \dots$ be the sequence of return times and let Λ be the set of all sequences $\omega\colon \mathbb{N} \to \{0, 1, \dots, T-1\}$ such that $\omega(n) \in \{0, 1, \dots, \frac{q(n)}{q(n-1)}\}$. In Λ we consider the product topology and, as before, we consider the homeomorphism between Cantor sets: $\phi_f\colon \Lambda \to \Lambda_f$ defined by

$$\phi_f(\omega) = \bigcap_{n=0}^{\infty} \triangle_n^{k(n)}(f), \text{ where } k(n) = \sum_{j=0}^{n-1} \omega(j) q(j).$$

As before we define ϕ_g using the renormalizing intervals for g. Since f and g have the same combinatorics we have that $\phi_g \circ \phi_f^{-1}\colon \Lambda_f \to \Lambda_g$ is a monotone homeomorphism. Again, we can define for each $s < T$ the ratio function $\Sigma_{s,f}\colon \Lambda \times \mathbb{N} \to \mathbb{R}$ by the formula

$$\Sigma_{s,f}(\omega, n) = \frac{\text{length of the interval } [\omega(n-1), \dots, \omega(0)]}{\text{length of the interval } [\omega(n-1), \dots, \omega(0), t]}$$

where $t \in \{0, \dots, \frac{q(n)}{q(n-1)} - 1\}$ is equal to s mod $\frac{q(n)}{q(n-1)} - 1$.

Theorem 9.4. *Let f and g be $\mathcal{U}^{1+z}$ infinite renormalizable maps with the same bounded combinatorial type. If there exist constants $C > 0$ and $0 < \lambda < 1$ such that $|R^n(f) - R^n(g)| \leq C \cdot \lambda^n$ then there exists a $C^{1+\alpha}$ diffeomorphism $h\colon \mathbb{R} \to \mathbb{R}$ that maps the attracting Cantor set of f onto the attracting Cantor set of g.*

Proof. The same as the proof of Theorem 9.2 and of Theorem 9.3. □Now,

Theorem 2.1 follows from Theorem 1.1 and Theorem 9.4. Finally we want to

show that the add-one map on the Cantor set extends to an expanding $C^{1+\alpha}$ diffeomorphism. Let us do this for the Feigenbaum's combinatorics. Let f be a $\mathcal{U}^{1+z}$ unimodal map of Feigenbaum's combinatorial type. Let $\phi\colon \Lambda \to \Lambda_f$ be the homeomorphism described in the beginning of this section. Here $\Lambda = \{0, 1\}^{\mathbb{N}}$. The adding map add: $\Lambda \to \Lambda$ is defined by $\text{add}(\omega)(n) = \omega(n+1)$. Hence the add-one map is a two to one continuous map of Λ onto Λ. We denote by the same name the two to one continuous map $\phi \circ \text{add} \circ \phi^{-1}$.

Theorem 9.5. *The add-one map $\phi \circ \text{add} \circ \phi^{-1}\colon \Lambda_f \to \Lambda_f$ extends to a $C^{1+\alpha}$ expanding map S of a neighbourhood of Λ_f in $\mathbb{R}$.*

Proof. We construct a $C^{1+\alpha}$ extension of the add-one map to a neighbourhood of Λ_f in the same way as we proceeded in the proof of Theorem 9.2. Let S be the $C^{1+\alpha}$ extension of the add-one map. Since S^n is $C^{1+\alpha}$ and maps the interval corresponding to the cylinder $\{\omega \in \Lambda;\ \omega(n-1) = \epsilon_{n-1}, \dots, \omega(0) = \epsilon_0\}$ diffeomorphically onto $\triangle_0$, we get, from the bounded geometry of the attracting Cantor set that the distortion of this map is bounded independently of n.

Since the lengths of the above intervals go to zero with n we conclude that the derivative of S^n is bigger than one in all intervals of the above type. This proves that S is an expanding map. □

Exercise 9.1. Let $J \subset \mathbb{R}$ be an interval and J_0, J_1 be two disjoint subintervals of J. Let $f\colon J_0 \cup J_1 \to J$ be a $C^{1+\alpha}$ expanding map that maps J_i diffeomorphically onto J for $i = 0, 1$. Show that the set $\Lambda_f = \{x \in J_0 \cup J_1\, ;\ f^n(x) \in J_0 \cup J_1 \text{ for all } n\}$ is an invariant Cantor set. Show that the scaling function of Λ_f is a well defined Hölder continuous function which has the geometric approximation property.

Exercise 9.2. Let $\sigma\colon \Lambda^* \to \mathbb{R}$ be a Hölder continuous function. Show that there exists an embedding $\phi\colon \Lambda \to \tilde{\Lambda} \subset \mathbb{R}$ such that: i) the convex hull of the image of the n-cylinders are disjoint intervals; ii) the scaling function of $\tilde{\Lambda}$ is equal to σ. Show that the add-one map extends to a $C^{1+\alpha}$ expanding map on a neighbourhood of $\tilde{\Lambda}$. So, in contrast to the situation for attracting Cantor sets of infinitely renormalizable maps, in the case of expanding Cantor sets, there is no rigidity: each scaling function can appear!

For some ideas in a similar situation for critical circle maps see Feigenbaum (1988).

10 Some Further Remarks and Open Questions

One of the first questions that springs to mind is why the proof of the theorems in this chapter require so many deep results from complex analysis and the proof of the rigidity result for the circle case can be done in a real setting. A possible answer is that in the case of diffeomorphisms the canonical maps are well known, the rigid rotations, whereas in the unimodal case the canonical maps are non-trivial analytic maps whose existence must be proved. It would be interesting to find a different proof of this theorem that would not rely so much on the theory of quadratic-like maps. Such a proof could give more general results that might include maps of the type $f(x) = \phi(|x|^r)$ where ϕ is a smooth diffeomorphism and $r > 1$ is a real number. Numerical experiments indicate that the rigidity results hold for these maps.

In Section 8 we have proved the contraction of the renormalization operator but without a rate. We do not know yet how to prove the exponential contraction of the renormalization operator. This result would imply a stronger rigidity statement as we pointed out in Section 9. The following question seems to be an important step towards proving the existence of the above rate of contraction. Is the stable set, with respect to the renormalization operator, of an infinite renormalizable map of bounded combinatorial type a smooth submanifold of the Banach space of mappings that have a holomorphic extension to a neighbourhood of the dynamical interval in $\mathbb{C}$?

From the results of Section 4, it follows that for the quadratic family $f_\mu(x) = \mu x(1-x)$, $0 < \mu \leq 4$, the set C_N of parameter values for which the corresponding map is infinitely renormalizable of combinatorial type bounded by N is a Cantor set. We do not know however if this Cantor set has zero Lebesgue measure. This and many other rigidity statements on the parameter space would follow from the hyperbolicity of the renormalization operator at the attracting set. Of course this hyperbolicity of the renormalization operator would also extend the validity of the rigidity Theorem 9.2 to C^{1+z} maps.

In Section 5 bounds are given for maps of bounded combinatorial type which are either in some Epstein class or symmetric and quadratic-like. It is not known whether these bounds can be obtained in any of the following three situations:

1. the initial map is just real analytic and of bounded combinatorial type;
2. unbounded combinatorial type (even in some Epstein class);
3. a quadratic-like map without the symmetry assumption.

Edson de Faria (1992), in his PhD thesis at CUNY, has used many of the ideas of this chapter to study renormalization of critical circle maps. He proved the following rigidity theorem. He considers real analytic circle homeomorphisms belonging to some class analogous to the Epstein class from Section 1 and which have a unique critical point which is cubic. If two such maps have the same rotation number of bounded type, then they are $C^{1+\alpha}$ conjugate for each $\alpha \in (0, 1)$.

The same kind of rigidity question can also be formulated for non-renormalizable maps. In particular, if the ω-limit of the critical point of a unimodal map is a minimal Cantor set, one can investigate the smoothness properties of the conjugacies between two such maps. For example, Lyubich and Milnor (1991) have shown metric universalities of these Cantor sets for unimodal maps of Fibonacci type which are C^2 and have a quadratic critical point. (Fibonacci maps were described in Section II.3.b.) To describe these results, let $S(n)$ be the n-th Fibonacci number. For maps f within this class $|f^{S(n)}(c) - c|$ tends to zero in a universal way:

$$\frac{|f^{S(n)}(c) - c|}{|f^{S(n-1)}(c) - c|} \cdot 2^{n/3}$$

converges exponentially fast to some positive and finite constant a. So this universality is slightly different from the one discussed in this chapter: here the parameter a still depends on the map f. This implies that the conjugacy between the Cantor sets of two such maps is in general not differentiable. So the asymptotic geometry is not rigid: its moduli space has real dimension equal to one. Lyubich and Milnor's proof uses many of the ideas developed in this chapter. In particular, they develop for these maps a renormalization theory. The renormalized maps are no longer unimodal maps but are maps defined in a finite number of disjoint intervals and having a unique critical point. The renormalization consists in taking the first return map to the interval that contains the critical point and restricting it to the components of the domain that intersect the critical orbit.

Nowicki and Van Strien (1991b) have shown that if f satisfies $Sf < 0$ and the critical point is of order $l > 2$ then $\frac{|f^{S(n)}(c) - c|}{|f^{S(n-1)}(c) - c|}$ is bounded from below and above. In particular, a Fibonacci map with a quadratic critical point and a Fibonacci map with a fourth order critical point are not Hölder and therefore not quasisymmetrically conjugate!

It is likely that a more general theory will emerge for maps for which the closure of the critical orbit is a minimal Cantor set. Ultimately, this theory may become equally well developed as the one for circle diffeomorphisms. Lyubich's results about the non-existence of absorbing Cantor attractors are steps in this direction, see the discussion in Section V.7 and Lyubich (1992b).

Świątek (1992b) has proved the following deep rigidity result for the quadratic family $f_\mu(x) = \mu x(1-x)$, $\mu \in [0,4]$. He proves that if two such maps are topologically conjugate then they are quasi-symmetrically conjugate. As a consequence, we get from the arguments in Section 4 that the set of parameter values for which the map has the same non-periodic kneading sequence has only one element. In particular, the set of parameter values for which the corresponding map has an attracting periodic point is open and dense.

McMullen has announced an alternative proof of the contraction of the renormalization operator using the complex bounds of Section 5 and some rigidity ideas from the theory of Kleinian groups.

Chapter VII.

Appendix

The purpose of this appendix is to present some basic definitions, theorems and background material for this book. We assume the reader to be familiar with manifolds.

1 Some Terminology in Dynamical Systems

Let us first give a short introduction to some general terminology used in dynamical systems, see also Smale (1967), Devaney (1986) and Palis and de Melo (1982). If $f\colon X \to X$ is a continuous map of a metric space X, the *iterates* of f are the maps f^n defined inductively by $f^0 = \mathrm{id}_X$, $f^1 = f$, $f^{n+1} = f^n \circ f$. If f is a homeomorphism then we can also define the map $f^{-n} = (f^{-1})^n$ for $n \in \mathbb{N}$. If f is not invertible then we define $f^{-n}(y) = \{x\,;\ f^n(x) = y\}$ for $n \in \mathbb{N}$. The *full orbit* of a point $x \in X$ is the set $O_f(x) = \{f^n(x); n \in \mathbb{Z}\}$ and the forward orbit is the set $\{f^n(x); n \in \mathbb{N}\}$. Here, $\mathbb{Z}$ denotes the set of integers and $\mathbb{N}$ the set of natural numbers (including 0). Mostly one is interested in recurrent behaviour, so in periodic, recurrent or non-wandering points. Let us define these notions. We say that x is a *periodic point* of period n if $f^n(x) = x$ and $f^i(x) \neq x$ for $1 \leq i < n$, i.e., the forward orbit of x has exactly n points. The ω-*limit set* of the orbit of $x \in X$ is the set

$$\omega(x) = \{y \in X\,;\ \exists \text{ a sequence } n_i \to \infty \text{ with } f^{n_i}(x) \to y\},$$

i.e., it is the set of accumulation points of the sequence of forward iterates of a point in this orbit. Similarly, the α-*limit set* of x is

$$\alpha(x) = \{y \in X\,;\ \exists\ y_i \to y \text{ and } n_i \to \infty \text{ with } f^{n_i}(y_i) \to x\}.$$

If X is a compact metric space then the ω-limit set of any orbit is a non-empty compact set. If the orbit is periodic then it coincides with its ω-limit set and a non-periodic point for which $x \in \omega(x)$ is called *recurrent*. The Ω-*set*, or the *set of non-wandering points* of $f\colon X \to X$ is the set of points x for which there exists a sequence $x_n \to x$ and $k(n) \to \infty$ with $f^{k(n)}(x_n) \to x$ (equivalently, for each neighbourhood U of x some forward iterate $f^n(U)$ has a non-empty intersection with U, i.e., $U \cap f^n(U) \neq \emptyset$. A set $A \subset X$ is *transitive* if $f(A) \subset A$ and if there exists a point $x \in A$ whose forward orbit is dense in A. Quite often such a set is a Cantor set, i.e., *perfect* (compact and without isolated points) and *totally disconnected* (each connected subset consists of at most one point).

Furthermore, we say that two maps $f, g\colon X \to X$ are *topologically conjugate* if there exists a homeomorphism $h\colon X \to X$ such that $h \circ f = g \circ h$. This implies that $h \circ f^n = g^n \circ h$ for every integer n. The map h, which is called the *conjugacy* between f and g, maps orbits of f onto orbits of g.

2 Some Background in Topology

2.1. General definitions

The topology of a space X can be generated by taking a collection of subsets S of X, called a *subbasis*: simply take as open sets in X all sets which can be formed by taking unions and finite intersections of sets in S together with $\emptyset$ and X. Given topological spaces X_i where i belongs to some index set, let $X = \prod_i X_i$ and $\pi_i\colon X \to X_i$ be the natural projections. X has a natural topology with a subbasis consisting of 'cylinder' sets of the form $\pi_i^{-1}(U_i)$ where U_i is open in X_i.

Theorem 2.1. **(Tychonoff)** *Let X_i be a family of compact sets then $X = \prod_i X_i$ is compact.*

If $\sim$ is an equivalence relation on X then let $\pi\colon X \to X/\sim$ be the natural projection. $X/\sim$ has a topology: U is open in $X/\sim$ if and only if $\pi^{-1}(U)$ is an open subset of X. In general $X/\sim$ is not Hausdorff, where we say that a topological space Y is called *Hausdorff* if each two distinct points x, y have disjoint neighbourhoods. Y is called *connected* if there exists no disjoint open sets U, V with $U \cup V = Y$ and with U and V both non-empty. It is called *locally compact* if it is Hausdorff and each point has a compact neighbourhood.

Let $C(X, Y)$ be the space of continuous maps from X to Y where X, Y are topological spaces. If X, Y are metric spaces then a collection of $\mathcal{F} \subset C(X, Y)$ is called *equicontinuous* if for each $\epsilon > 0$ there exists $\delta > 0$ such that for each $x, y \in X$ with $d(x, y) < \epsilon$ one has $d(f(x), f(y)) < \delta$ for all $f \in \mathcal{F}$. If Y is a metric space, the space $C(X, Y)$ is endowed with the so-called *supremum metric* defined by $d(f, g) = \sup_{x \in X} d(f(x), g(x))$. If $d(f_n, g) \to 0$ then we say that f_n tends *uniformly* to g. More generally, if X, Y are topological spaces then there is the *compact-open* topology on $C(X, Y)$ defined by the subbasis with sets of the form

$$(K, U) = \{f \in C(X, Y)\,;\; f(K) \subset U\}$$

with K compact and U open.

Theorem 2.2. **(Ascoli)** *If X is locally compact, X, Y are metric spaces and $\mathcal{F}$ is an equicontinuous family in $C(X,Y)$ such that for each x the closure of the set $\{f(x);\ f \in \mathcal{F}\}$ is compact in Y, then the closure of $\mathcal{F}$ is compact in $C(X,Y)$.*

2.2. Covering spaces

We say that curves $\alpha, \alpha' \colon [0,1] \to X$ are *homotopic* (relative to endpoints) if there exists a continuous map $\gamma\colon [0,1] \times [0,1] \to X$ such that for $\gamma_s(t) = \gamma(t,s)$ one has $\gamma_0 = \alpha$ and $\gamma_1 = \alpha'$ and such that the curves γ_s all have the same endpoints for each s. In particular, if α is a closed curve with initial and endpoint equal to x_0 then γ_s are also closed curves with endpoints x_0. The map γ is a called a *homotopy* between α and α'. Given x_0 let $\pi_1(X, x_0)$ be the space of homotopy classes of closed curves with $\gamma(0) = \gamma(1) = x_0$. This space can be made into a group by defining for each $[\gamma_1], [\gamma_2] \in \pi_1(X, x_0)$ their sum $[\gamma_1] + [\gamma_2]$ to be the homotopy class of the curve:

$$\gamma(t) = \begin{cases} \gamma_1(2t) & \text{for } t \in [0,1/2] \\ \gamma_2(2t-1) & \text{for } t \in [1/2,1]. \end{cases}$$

A topological space X is *simply connected* if $\pi_1(X, x_0) = 0$. We say that a continuous map $\pi\colon \tilde{X} \to X$ is a *covering map* if each $y \in X$ has a neighbourhood V such that π maps each component of $\pi^{-1}(V)$ homeomorphically onto V. Often we simply say that $\tilde{X}$ covers X. The main property of a covering map is that curves and homotopies can be lifted: for each continuous $\alpha\colon [0,1] \to X$ there exists a continuous $\tilde{\alpha}\colon [0,1] \to \tilde{X}$ with $\pi\tilde{\alpha} = \alpha$. Moreover, if γ is a homotopy between α and α' then there exists a homotopy $\tilde{\gamma}$ between $\tilde{\alpha}$ and some lift $\tilde{\alpha}'$ of α'. A cover $\pi\colon \tilde{X} \to X$ is *universal* if $\tilde{X}$ is simply connected.

Theorem 2.3. *Let X be a topological space which is path connected. Then one has the following.*

1. *There is a universal cover $\pi\colon \tilde{X} \to X$.*

2. *Each continuous map $f\colon Y \to X$ has a lift. More precisely, for each $y \in Y$ and each $\tilde{x} \in \pi^{-1}(f(y))$ there exists a unique continuous map $\tilde{f}\colon Y \to \tilde{X}$ with $\pi \circ \tilde{f} = f$ and $\tilde{f}(y) = \tilde{x}$.*

3. *If $f\colon \tilde{Y} \to Y$ is a covering and Y is simply connected then f is a homeomorphism. In particular, the universal covering is unique up to a homeomorphism.*

Proof. Fix a point x_0 and let X' be the space of curves $\gamma\colon [0,1] \to X$ with $\gamma(0) = x_0$ with the compact-open topology on $C([0,1], X)$. We say that such curves γ, γ' are equivalent, $\gamma \sim \gamma'$, if $\gamma(1) = \gamma'(1)$. Then $\tilde{X} = X'/\sim$ with the projection map $\pi([\gamma]) = \gamma(1)$ is a universal covering. The last two statements follow easily. □

Because of the previous theorem one can lift $\pi\colon \tilde{X} \to X$ to a map $\tilde{\pi}\colon \tilde{X} \to \tilde{X}$. Since π is a covering map, $\tilde{\pi}$ is also a covering map and since $\tilde{X}$ is simply connected a homeomorphism. Let Γ be the space of such lifts,

$$\Gamma = \{\tilde{\pi}\,;\ \tilde{\pi} \text{ is a lift of } \pi\colon \tilde{X} \to X\}.$$

This space Γ is a group under composition and called the *group of deck transformations.*

Remark. 1. If $\tilde{f}_1, \tilde{f}_2$ are lifts of $f\colon Y \to X$ then $\tilde{f}_1 = \tilde{\pi} \circ \tilde{f}_2$ for some $\tilde{\pi} \in \Gamma$.

2. Provided X is locally simply connected, the group Γ acts discontinuously on $\tilde{X}$ (each point has a neigbourhood U such that $\gamma(U) \cap U \neq \emptyset$ for only a finite number elements γ of Γ). Moreover, the space X is homeomorphic to $\tilde{X}/\Gamma$.

3. The group Γ is isomorphic to $\pi_1(X)$.

4. Usually a structure on X can be lifted to a structure on $\tilde{X}$ which is invariant under the group of deck transformations. For example, a manifold structure on X defines a unique manifold structure on $\tilde{X}$ such that π is local diffeomorphism and Γ is a group of diffeomorphisms; conversely a manifold structure on $\tilde{X}$ which is invariant by Γ defines a manifold structure on X. Similarly, given a Riemannian metric on X there exists a unique Riemannian metric on $\tilde{X}$ so that π is a local isometry and elements of Γ are isometries; conversely, any Riemannian metric on $\tilde{X}$ such that the elements of Γ are isometries define a Riemannian metric on X. In Section 4 of this appendix we shall use similar observations in the case of a complex structure.

3 Some Results from Analysis and Measure Theory

Often we will want to measure the size of subsets of a topological space. In order to do this in a reasonable way we only consider *Borel sets* of a topological space. This is the smallest collection of subsets of X which contains the open subsets and which is closed under taking complements and unions of countably many members of this collection. (Such a set is called a σ-algebra: countable unions, intersections and differences of elements of this collection are also contained in this collection.) These Borel sets of X will be called *measurable.* We say that f is *measurable* if $f^{-1}(A)$ is measurable whenever A is measurable. (So if f is continuous this is certainly the case.) A function μ which associates to each Borel set a positive number in such a way that for each countable collection of disjoint sets A_i one has

$$\mu(\cup A_i) = \sum_i \mu(A_i)$$

is called a *measure.* It is a *probability measure* if $\mu(X) = 1$ and a *σ-finite measure* if X can be written as the countable union of sets with finite measure. A measurable function $f\colon X \to \mathbb{R}$ is *integrable* with respect to μ (in the sense of

Lebesgue) if there exists a sequence of functions f_n of the form $f_n = \sum_{i=0}^{n} \lambda_i 1_{A_i}$ where 1_A is the indicator function of the set A and $\lambda_i \in \mathbb{R}$ and A_i are measurable sets such that

$$\lim_{n\to\infty} f_n(x) = f(x) \text{ for } \mu\text{-almost every } x \in X$$

and

$$\lim_{n,m\to\infty} \int_X |f_n - f_m| d\mu = 0.$$

Here the last integral is defined by writing $|f_n - f_m|$ in the form $\sum_{i=0}^{N} \lambda'_i 1_{A'_i}$ and letting the integral of this last function be $\sum_{i=0}^{N} \lambda'_i \mu(A'_i)$. Let L^1 be space of functions which are integrable and for which $\int_X |f|\, d\mu$ is finite.

Moreover, $f\colon [a,b] \to \mathbb{R}$ is called *absolutely continuous* if for every $\epsilon > 0$ there exists $\delta > 0$ such that, if $I_i \subset [a,b]$ are disjoint intervals with total length at most δ then $\sum_i |f(I_i)| \le \epsilon$. It has *bounded variation* if there exists $K < \infty$ such that for any disjoint covering I_i of $[a,b]$ the total length of $f(I_i)$ is always at most K. The smallest upper bound is denoted by $\mathrm{Var}(f)$. It is not hard to show that every absolutely continuous function is of bounded variation. Furthermore we say that f has the *null property* if $f(A)$ has Lebesgue measure zero whenever A has Lebesgue measure zero.

Theorem 3.1. *Let $f\colon [a,b] \to \mathbb{R}$. Then the following are equivalent.*

a) f is absolutely continuous;

b) f is continuous, of bounded variation and f has the null property;

c) f is the indefinite integral of a Lebesgue integrable function.

d) f is almost everywhere differentiable, f' is Lebesgue integrable and f is the indefinite integral of f'.

Proof. See for example Natanson (1955) or Rooij and Schikhof (1982). □

A measure ν is *absolutely continuous* with respect to a measure μ if $\nu(A) = 0$ for each measurable set A for which $\mu(A) = 0$.

Theorem 3.2. **(Radon-Nikodym)** *If μ is a σ-finite measure then ν is absolutely continuous with respect to μ if and only if there exists a function $f\colon X \to \mathbb{R}$ which is μ-integrable on all Borel sets A with μ-finite measure and*

$$\nu(A) = \int_A f d\mu.$$

Proof. See Rudin (1966).

We shall be mainly working with $X = [0,1]$ or $X = S^1$ and then there is a unique measure, the *Lebesgue measure* λ, which associates a real number to each measurable subset of X such that $\lambda(I)$ coincides with the length of I when I is an interval. This measure has the following remarkable property: if A has positive Lebesgue measure then Lebesgue almost every point of A is a *density point*. Here we say that x is a density point of A if

$$\frac{\lambda(A \cap (x-\epsilon, x+\epsilon))}{2\epsilon} \to 1$$

as $\epsilon \to 0$.

Theorem 3.3. **(Lebesgue Density Theorem)** *If $A \subset \mathbb{R}$ is a measurable set then almost every point of A is a density point. If $A \subset X \subset \mathbb{R}$ are measurable sets and for almost all $x \in X$,*

$$\limsup_{\epsilon \to 0} \frac{\lambda(A \cap (x-\epsilon, x+\epsilon))}{2\epsilon} > 0$$

then A has full Lebesgue measure in X.

Proof. See for example Natanson (1955) or Rooij and Schikhof (1982). □

Finally we define the *essential supremum* of a real valued measurable function f to be the infimum of all k such that the set $\{z\,;\ |f(z)| > k\}$ has measure zero.

4 Some Results from Ergodic Theory

As before let X be some topological space and $f\colon X \to X$ some continuous map. Often one is only interested in the typical behaviour of points and in order to make this precise one uses a measure on the space. This type of question is studied in ergodic theory, see for example the monographs by Petersen (1983), Walters (1982) and Mãné (1987).

Often we are also interested in dynamical measures in particular measures which are *invariant* under f. These are measures μ such that $f_*\mu = \mu$. This means that

$$\mu(f^{-1}(A)) = \mu(A)$$

for each measurable set A. Let us make an intuitive remark on this definition. If $X = [0,1]$ and $\mu(X) = 1$ one can think of X as a metal bar with mass 1 and non-homogeneous mass density determined by μ. Then $\mu(f^{-1}(A)) = \mu(A)$ means that if one folds X via the f-map one obtains again the original mass distribution on X! Often one would like to have that X cannot be split into smaller dynamically invariant sets. One way in which this can be formalized is by requiring that μ is an *ergodic measure*. This means that for each measurable set A for which $f^{-1}(A) = A$ one has $\mu(A) = 0$ or $\mu(X \setminus A) = 0$. This definition can also be used if μ is not an invariant measure for f. If the Lebesgue measure is ergodic for a certain map f, then we simply say that f is *ergodic*.

Furthermore we say that μ is a *probability measure* if $\mu(X) = 1$ and a *finite measure* if $\mu(X) < \infty$. If some property holds for a set $A \subset X$ for which $\mu(X \setminus A) = 0$ then we say that it holds for μ-almost all x.

By the Ergodic Theorem one can use ergodic invariant measures to take time averages of certain functions.

Theorem 4.1. **(Birkhoff Ergodic Theorem)** *If μ is an invariant probability measure of $f\colon X \to X$ and if $g\colon X \to \mathbb{R}$ is continuous then for μ-almost all x the limit*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} g(f^i(x))$$

exists. If μ is also ergodic then

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} g(f^i(x)) = \int_X g\, d\mu$$

for μ-almost all x.

Proof. See, for example, Mañé (1987) and Petersen (1983). □

It follows that, if μ is an ergodic invariant probability measure, for each open set A,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} 1_A(f^i(x)) = \mu(A)$$

for μ-almost all x, where 1_A is the indicator function on A. There is just one snag: μ could be just associate mass $1/n$ to each point on a periodic orbit of period n and then the statement μ-almost all x is not very interesting. Therefore we will be mainly interested in *absolutely continuous* measures. These are measures μ for which $\mu(A) = 0$ whenever the Lebesgue measure of A is zero.

Another way of confining ourselves to physically relevant measures is to require that they are the limits of experimentally found measures. More precisely, we say that a sequence of measures μ_n converges in the weak sense to the measure μ if $\int g\mu_n \to \int g\mu$ for each continuous function g. If μ satisfies the conclusion of the Birkhoff Ergodic Theorem for almost all $x \in X$, or equivalently, if

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} f_*^i \delta_x \to \mu$$

for almost every x then it is called a *Bowen-Ruelle-Sinai* or *physical measure.* Here δ_x is the Dirac measure.

We should note that the Birkhoff theorem does not hold if μ is an infinite invariant measure. However, some analogue statements still hold in that case. A map is called *conservative* with respect to some measure if there exists no set A of positive measure such that $f^n(A) \cap A = \emptyset$ for all n. If it is not conservative then it is called *dissipative.*

5 Some Background in Complex Analysis

In this section we shall give a short introduction into conformal and quasiconformal maps. For further background the reader should consult the books by Ahlfors, Lehto and Gardiner mentioned in the list of references.

5.1. The Cauchy Integral Formula

Let $\mathbb{C}$ denote the set of complex numbers. As usual, $a \in \mathbb{C}$ can be written as $a = Re(a) + i \cdot Im(a)$ where $Re(a)$ and $Im(a)$ are real numbers called the real and imaginary parts of a. The complex number $\overline{a} = Re(a) - i \cdot Im(a)$ is the complex conjugate of a and $|a| = \sqrt{a \cdot \overline{a}}$ is called the absolute value or the norm of a. Here $\cdot$ denotes complex multiplication.

For each $\mathbb{R}$-linear map $L\colon \mathbb{C} \to \mathbb{C}$ there exists complex numbers $a, b \in \mathbb{C}$ such that

$$L(v) = a \cdot v + b \cdot \overline{v} \text{ for all } v \in \mathbb{C}.$$

In other words, the real vector space of $\mathbb{R}$-linear maps of $\mathbb{C}$ into $\mathbb{C}$ is isomorphic to $\mathbb{C} \times \mathbb{C}$ as a real vector space. Furthermore, the linear map L preserves orientation if and only if $|a| > |b|$ and it is an isomorphism if and only if $|a| \neq |b|$. If L is orientation preserving then L maps the unit circle in an ellipse whose major axis is in the direction of $\sqrt{\frac{b}{a}}$ and whose eccentricity (ratio of the length of the major axis by that of the minor axis) is equal to

$$\frac{1 + |\frac{b}{a}|}{1 - |\frac{b}{a}|}.$$

Similar statements hold for orientation reversing maps except that the direction of the axis are rotated. In particular, we say that the linear map L is *conformal* (i.e., preserves angles) if and only if either $a \neq 0$ and $b = 0$ or $a = 0$ and $b \neq 0$.

If $f\colon U \subset \mathbb{C} \to \mathbb{C}$ is a differentiable at z then we write

$$Df(z)(v) = \partial f(z) \cdot v + \bar{\partial} f(z) \cdot \bar{v}$$

and call the complex numbers $\partial f(z)$ and $\bar{\partial} f(z)$ the ∂-derivative respectively the $\bar{\partial}$-derivative of f at z.

Definition. A map f as above is said to be *holomorphic* if $\bar{\partial} f \equiv 0$ (these are the Cauchy-Riemann equations). Similarly, f is called *anti-holomorphic* if $\partial f \equiv 0$. We say that f is *conformal* if and only if it is either holomorphic or anti-holomorphic and the derivative is non-zero at every point.
Finally we say that $f\colon U \to f(U)$ is *quasiconformal* if

a) f is an orientation preserving homeomorphism between the open sets U and $f(U)$;

b) the real part $Re(f)$ and imaginary part $Im(f)$ of f are absolutely continuous on almost all verticals and on almost all horizontals in the sense of Lebesgue;

c) there exists $k < 1$ such that for

$$\mu_f(z) = \frac{\bar{\partial} f(z)}{\partial f(z)}$$

one has

$$|\mu_f(z)| \le k \text{ for almost all } z \in U.$$

The *conformal distortion* of the quasiconformal homeomorphism f is the essential supremum $K(f)$ of

$$\frac{1+|\mu_f(z)|}{1-|\mu_f(z)|}.$$

Notice that any C^1 diffeomorphism of a compact domain that contains U restricts to a quasiconformal homeomorphism of U.

The most basic result in complex analysis is the Cauchy Theorem:

Theorem 5.1. **(Cauchy Integral Formula)** *If f is holomorphic on a domain which contains a closed r-disc around z_0 then*

$$|w-z_0|<r \quad \text{implies} \quad f(w)=\frac{1}{2\pi i}\int_{|z-z_0|=r}\frac{f(z)}{z-w}dz.$$

From this formula we immediately get

Corollary 5.1. *f is an analytic map, i.e., it is the sum of a convergent power series $f(w)=\sum_{n=0}^{\infty}a_n(w-z_0)^n$ for w near z_0.*

Proof. It is enough to expand $\dfrac{1}{z-w}$ in a power series and integrate term by term. □

Corollary 5.2. (Liouville's Theorem) *If $f\colon \mathbb{C}\to\mathbb{C}$ is holomorphic and bounded then f is constant.*

Proof. $f'(w)=\partial f(w)=\dfrac{1}{2\pi i}\int_{|z-w|=r}\dfrac{f(z)}{(z-w)^2}\,dz$. This integral tends to zero as $r\to\infty$. Hence $f'\equiv 0$. □

Corollary 5.3. *If $f_n\colon U\to\mathbb{C}$ are holomorphic maps and $f_n\to f$ uniformly on compact subsets of U then f is holomorphic and each derivative of f_n converges uniformly on compact subsets of U to the corresponding derivative of f.*

Proof. $f_n'(w)=\dfrac{1}{2\pi i}\int_{|z-w|=r}\dfrac{f_n(z)}{(z-w)^2}\,dz$. Hence f_n' converges uniformly to $\phi(w)=\dfrac{1}{2\pi i}\int_{|z-w|=r}\dfrac{f(z)}{(z-w)^2}\,dz$. So f is differentiable and $Df(z)(v)=\partial\phi(z)\cdot v$. This implies that f is holomorphic and $\partial f=\phi$. Similarly we get the convergence of the higher order derivatives. □

Corollary 5.4. *Let U be an open domain in $\mathbb{C}$, $H(U,\mathbb{C})$ the vector space of holomorphic mappings on U, $K\subset U$ a compact set and $C^0(K,\mathbb{C})$ the Banach space of continuous maps on K with the supremum norm. Then the restriction map $r\colon H(U,\mathbb{C})\to C^0(K,\mathbb{C})$ is compact. This means that if $B\subset H(U,\mathbb{C})$ is a subset of uniformly bounded functions then the closure of $r(B)$ is a compact subset of $C^0(K,\mathbb{C})$.*

Proof. From the Cauchy Integral Formula it follows that the derivative of maps in B are uniformly bounded in a neighbourhood of K. Hence the Corollary follows from Ascoli's Theorem. □

Another simple consequence of the Cauchy Integral Formula, via Corollary 1, is a local normal form of holomorphic functions. If $f\colon U \to \mathbb{C}$ is a non-constant holomorphic map and $z_0 \in U$ then there exists $k \in \mathbb{N}$, a neighbourhood W of z_0 and holomorphic diffeomorphism $\phi\colon (W; z_0) \to (\phi(W), 0)$ and $\psi\colon (f(W); f(z_0)) \to (\psi(f(W)), 0)$ such that

$$\psi \circ f \circ \phi^{-1}(z) = z^k \text{ for all } z \in \phi(W).$$

In particular, f has only isolated zeros.

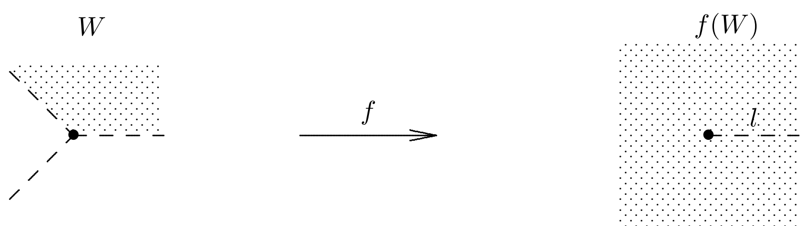

Fig. 5.1: f has a critical point of order k. Then let l be an arc as on the right. Then f^{-l} consists of k arcs and f restricted to each component of $U \setminus f^{-1}(l)$ is a holomorphic diffeomorphism onto $f(W) \setminus l$.

5.2 Hyperbolic Geometry

Definition. Let S be a topological space which is connected, Hausdorff and which has a countable neighbourhood basis. A *holomorphic atlas* on S is a collection of homeomorphisms $\phi_i\colon U_i \to \phi_i(U_i) \subset \mathbb{C}$, called *charts*, such that U_i are open subsets of S which together cover S, $\phi_i(U_i)$ are open sets in $\mathbb{C}$ and the overlapping maps

$$\phi_j \circ \phi_i^{-1}\colon \phi_i(U_i \cap U_j) \to \phi_j(U_i \cap U_j)$$

are holomorphic. A *Riemann surface structure* on S is a maximal atlas. (Here maximality is meant with respect to inclusion: one atlas contains the other if each chart of the second atlas is a chart of the first one. A map $f\colon S_1 \to S_2$ between two Riemann surfaces is *holomorphic* if for each $z_0 \in S_1$, $\psi \circ f \circ \phi^{-1}$ is holomorphic near $\phi(z_0)$, where ϕ is a chart of S_1 near z_0 and ψ is a chart of S_2 near $f(z_0)$.) The complex plane and all connected open subsets of the plane are

examples of Riemann surfaces. The simplest example of a compact Riemann surface is the Riemann sphere $\overline{\mathbb{C}} = \mathbb{C} \cup \{\infty\}$ which has a holomorphic atlas with two charts:

$$\overline{\mathbb{C}} \setminus \{\infty\} \to \mathbb{C} \text{ defined by } z \mapsto z$$

and

$$\overline{\mathbb{C}} \setminus \{0\} \to \mathbb{C} \text{ defined by } z \mapsto 1/z \text{ and } \infty \mapsto 0.$$

From Liouville's Theorem (Corollary 2 above) and since holomorphic maps have only isolated zeros, it is easy to conclude that $f\colon \bar{\mathbb{C}} \to \bar{\mathbb{C}}$ is holomorphic if and only if $f(z) = \dfrac{P(z)}{Q(z)}$ where P and Q are polynomial maps. In particular, f is a holomorphic diffeomorphism if and only if $f(z) = \dfrac{az+b}{cz+d}$ with $a, b, c, d \in \mathbb{C}$ and $ad - bc \neq 0$. These are called the *Möbius transformation* and the group (under composition) of the Möbius transformations if denoted by $M(\bar{\mathbb{C}})$. One can see very easily that this group is generated by the following family of maps

$$z \mapsto \frac{1}{z},\, z \mapsto az,\, z \mapsto z + b$$

with $b \in \mathbb{C}$ and $a \in \mathbb{C} \setminus \{0\}$. Notice that a Möbius transformation on $\mathbb{C}$ which is not the identity has at most two fixed points. It is *hyperbolic* if it has one attracting and one repelling fixed point. It is *parabolic* if it has a unique fixed point. Finally, it is *elliptic* if it is conjugate to a rotation (as a map on $\bar{\mathbb{C}}$).

Proposition 5.1.

1. *A Möbius transformation maps a line either onto a line or onto a circle and the image of a circle is either a circle or a line.*

2. *A Möbius transformation preserves the cross-ratio of four points: if*

$$C(z_1, z_2, z_3, z_4) = \frac{(z_3 - z_1)(z_4 - z_2)}{(z_2 - z_1)(z_4 - z_3)}$$

and ϕ is a Möbius transformation then

$$C(\phi(z_1), \phi(z_2), \phi(z_3), \phi(z_4)) = C(z_1, z_2, z_3, z_4).$$

3. *Given any three distinct points z_1, z_2, z_3 in the Riemann sphere, there exists a unique Möbius transformation ϕ such that $\phi(z_1) = 0$, $\phi(z_2) = 1$, $\phi(z_3) = \infty$. Hence any three points can be mapped to any other three points by a Möbius transformation.*

4. *The Möbius transformation Φ which maps the points $1, i, -1$ to the points $0, 1, \infty$ is a holomorphic diffeomorphism between the unit disc $\mathbb{D}$ and the upper half-plane $\mathbb{H} = \{z = x + iy \in \mathbb{C};\ y > 0\}$.*

5. *Each Möbius transformation ψ which map $\mathbb{D}$ onto $\mathbb{D}$ is of the form $\psi(z) = (az + \overline{c})/(cz + \overline{a})$ where $|a|^2 - |c|^2 = 1$.*

6. *The subgroup Γ of all Möbius transformations which map $\mathbb{D}$ onto $\mathbb{D}$ acts transitively in $\mathbb{D}$. More precisely, let C be a circle or a straight line that contains $z, w \in \mathbb{D}$ that is orthogonal to the unit circle $\mathbb{D}$. Let $\{z_\infty, w_\infty\} = C \cap \partial\,\mathbb{D}$ be such that z lies between z_∞ and w in $C \cap \overline{\mathbb{D}}$ as in Figure 5.2. Then the Möbius transformation ψ that maps $z_\infty \mapsto z_\infty$, $z \mapsto w$ and $w_\infty \mapsto w_\infty$ maps $\mathbb{D}$ diffeomorphically onto $\mathbb{D}$.*

Proof. Elementary. Let us prove for example the last statement. From 1) it follows C is mapped onto C. Since the unit circle $\mathbb{D}$ is orthogonal to C and $\phi(z_\infty) = z_\infty$ and $\phi(w_\infty) = w_\infty$ the conformality of ϕ, Statement 1) also implies that $\phi(\partial\,\mathbb{D}) = \partial\,\mathbb{D}$. □

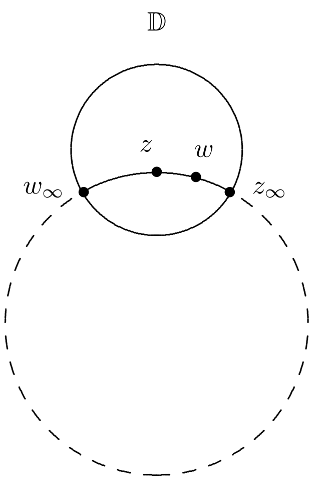

Fig. 5.2: The points z_∞, z, w and w_∞.

Since the subgroup of maps in $M(\mathbb{D})$ which fix the origin is the group of rotations, we have

Corollary 5.5. *Up to a multiplicative factor there is a unique Riemannian metric on $\mathbb{D}$ such that the elements of $M(\mathbb{D})$ are isometries of this metric.*

Proof. Since the required Riemannian metric has to be invariant under rotations, it is a positive multiple of the Euclidean inner product at the origin. This defines the norm $|\cdot|_0$ at $0 \in \mathbb{D}$. The norm $|v|_z$ is by definition equal to $|\partial\psi(z) \cdot v|_0$ where $\psi \in M(\mathbb{D})$ is such that $\psi(z) = 0$. This is well defined because if $\psi_1 \in M(\mathbb{D})$ also has the property that $\psi_1(z) = 0$ then $\psi_1 \circ \psi^{-1}$ is a rotation and hence $\partial\psi(z) = e^{i\theta} \cdot \partial\psi_1(z)$. □

Definition. The *Poincaré* or *hyperbolic* metric on $\mathbb{D}$ is the Riemannian metric on $\mathbb{D}$ having the Möbius transformations in $M(\mathbb{D})$ as isometries and such that $|v|_0 = 2 \cdot |v| = 2 \cdot \sqrt{v \cdot \bar{v}}$. (The choice of the factor two is to make the curvature of $\mathbb{D}$ equal to -1.)

Remark. 1. Notice that if
$$\phi(z) = \frac{z+x}{1+\bar{x}z},$$
where $x \in \mathbb{D}$ and where $\bar{x} \in \mathbb{D}$ is the complex conjugate of $x \in \mathbb{D}$, then $\phi \in M(\mathbb{D})$ and $\phi(0) = x$. Since $D\phi(0) \cdot v = (1 - |x|^2) \cdot v$ and $||v||_0 = ||D\phi(0) \cdot v||_x$, we have that
$$||v||_x = \frac{1}{1-|x|^2}||v||_0 = \frac{2}{1-|x|^2}|v|.$$
This gives an expression for the Poincaré metric of $\mathbb{D}$. In particular, the conformal map $\mathbb{D} \to \mathbb{D}$ defined by $z \mapsto \bar{z}$ is also an isometry of this metric.

2. Similarly, there is a unique Riemannian metric $|| \cdot ||$ on the upper-half plane $\mathbb{H}$ which is invariant under the group of Möbius transformations which map $\mathbb{H}$ onto $\mathbb{H}$ and such that $||v||_i = |v|$. As before it follows that $||v||_z$ is equal to $\frac{1}{y}|v|$ if $z = x + iy \in \mathbb{H}$ where $z, y \in \mathbb{R}$. This is called the Poincaré metric on $\mathbb{H}$ and by Statement 4 of Proposition 5.1 it is isometric to the hyperbolic metric on $\mathbb{D}$ defined in the previous remark. (In fact, the derivative of the Möbius transformation from $\mathbb{D}$ to $\mathbb{H}$ from this statement is equal to $2i$ at the point 0.)

Proposition 5.2. 1. *The geodesics of the hyperbolic metric of $\mathbb{D}$ are the straight lines and circles perpendicular to $\partial\mathbb{D}$.*

2. *If $d_P(z, w)$ is the infimum of the hyperbolic length of all C^1 curves from z to w then*
$$d_P(z, w) = \log \frac{|w - z_\infty|}{|z - z_\infty|} \frac{|w_\infty - z|}{|w_\infty - w|} \qquad (*)$$
where z_∞ and w_∞ are the intersection of the unit circle with the circle (or line) that contains z, w and is orthogonal to the unit circle. Here z_∞ is the point of intersection such that the arc bounded by z and z_∞ does not contain w.

Proof. The map $z \mapsto \bar{z}$ is an isomorphism. Because there is precisely one geodesic through a point in a given direction, the geodesic connecting 0 to $w \in (0, 1)$ is the curve $(0, w)$. Since distinct points $x, y \in \mathbb{D}$ can be mapped by a Möbius transformations in $M(\mathbb{D})$ to 0 respectively $w \in (0, 1)$ the first statement follows. Since the cross-ratio is invariant under Möbius transformations, the right hand side of $(*)$ is invariant. As the Möbius transformations that preserve $\mathbb{D}$ are isometries under the Poincaré metric, both sides of $(*)$ are invariant under Möbius transformations. It is therefore enough to verify the equation for $z = 0$ and $w \in (0, 1)$ (and therefore $z_\infty = -1$, $w_\infty = 1$). In this case the right hand side of the above equation is equal to $\log \frac{1+|w|}{1-|w|}$. Let us compute the left hand side. Since we already know that the geodesic connecting 0 to w is $(0, w)$,
$$d_P(0, w) = 2\int_0^w \frac{ds}{1-s^2} = \log \frac{1+|w|}{1-|w|}.$$
□

Theorem 5.2. **(Schwarz Lemma)** *If $f: \mathbb{D} \to \mathbb{D}$ is a holomorphic map then either f strictly contracts the hyperbolic metric or f is a Möbius transformation.*

Proof. Suppose f is not a Möbius transformation. Let $\psi_i \in M(\mathbb{D})$, $i = 1, 2$ be such that $\psi_1(z) = 0$ and $\psi_2(f(z)) = 0$. Then $g = \psi_2 \circ f \circ \psi_1^{-1}: \mathbb{D} \to \mathbb{D}$ is holomorphic, maps 0 into 0 and is not a rotation. By the classical Schwarz Lemma, $|\partial g(0)| < 1$. Since ψ_i are isometries of the hyperbolic metric,

$$|\partial f(z)v|_{f(z)} < |v|_z. \quad \square$$

Corollary 5.6. 1. *If $\phi: \mathbb{D} \to \mathbb{D}$ is a holomorphic diffeomorphism then ϕ is a Möbius transformation.*

2. *If $\phi: \mathbb{D} \to \mathbb{D}$ is a hyperbolic isometry then either $z \mapsto \phi(z)$ or $z \mapsto \overline{\phi(z)}$ is a Möbius transformation.*

Remark. Let $\phi: \mathbb{D} \to \mathbb{D}$ be a Möbius transformation. If ϕ has a fixed point $x \in \mathbb{D}$ then, if $\psi \in M(\mathbb{D})$ is the Möbius transformation that maps x to 0, $\psi \circ \phi \circ \psi^{-1}$ is a rotation. In particular, ϕ has at most one fixed point in $\mathbb{D}$ (unless it is the identity). If ϕ has a fixed point then we say that ϕ is an *elliptic* isometry of $\mathbb{D}$.

If $\phi \in M(\mathbb{D})$ has no fixed point (in $\mathbb{D}$) then, since ϕ has at least one fixed point and at most two fixed points in $\bar{\mathbb{C}}$, it follows that ϕ has at least one and at most two fixed points in $\partial\mathbb{D}$.

If ϕ has two fixed points in the boundary then the geodesic connecting these points is invariant under ϕ. So if we take ψ to be the Möbius transformation that maps these fixed points to $0, \infty$ then $\psi \circ \phi \circ \psi^{-1}$ is the map $z \mapsto \lambda \cdot z$ for some $\lambda \in \mathbb{R}$. In this case we say that ϕ is a *hyperbolic* isometry of $\mathbb{D}$.

The last possibility is that ϕ has a unique fixed point in $\partial\mathbb{D}$. Then there exists a Möbius transformation $\psi: \mathbb{D} \to \mathbb{H}$ such that $\psi \circ \phi \circ \psi^{-1}$ is the map $z \mapsto z + 1$ and ϕ is called a *parabolic* isometry of $\mathbb{D}$.

5.3. The Uniformization Theorem

Theorem 5.3. **(Uniformization Theorem)** *If S is a simply connected Riemann surface then S is holomorphically diffeomorphic to either $\mathbb{D}$, $\mathbb{C}$ or to $\bar{\mathbb{C}}$.*

Remark. A special case of Theorem 3 is: if U is a simply connected subset of $\mathbb{C}$ (which is not equal to $\mathbb{C}$) then there exists a holomorphic diffeomorphism $\phi: U \to \mathbb{D}$. Any one of such maps is called a *Riemann mapping* of U and the Riemann mapping is completely determined by the image of a point and the image of a direction by the derivative at that point. Furthermore, if the boundary of U is a Jordan curve then ϕ extends continuously to a homeomorphism of $\bar{U}$ onto $\bar{\mathbb{D}}$. In this case ϕ is completely determined by the image of three distinct points in ∂U. In this case the proof is much easier. The proof of Theorem 5.3 can be found in Ahlfors (1973).

Let S be a Riemann surface and $\pi\colon \tilde{S} \to S$ be its universal covering map. It is clear that there exists a unique Riemann surface structure on $\tilde{S}$ that makes π holomorphic; hence $\tilde{S}$ is holomorphically diffeomorphic to either $\mathbb{C}$, $\bar{\mathbb{C}}$ or $\mathbb{D}$. In the first case we say that S is *parabolic*, in the second case it is *elliptic* and in the last case it is a *hyperbolic Riemann surface*.

Before we give examples of hyperbolic Riemann surfaces we need the following important result.

Theorem 5.4. **(Schwarz Reflection Principle)** *Let $U \subset \mathbb{C}$ be a domain which is symmetric with respect to complex conjugation. Let $U_+ = U \cap \mathbb{H}$ and $f\colon U_+ \to \mathbb{H}$ be a holomorphic map such that the imaginary part of f extends continuously to the zero function on $U \cap \mathbb{R}$. Then f extends to a holomorphic function on U which satisfies the symmetry relation $f(\bar{z}) = \overline{f(z)}$.*

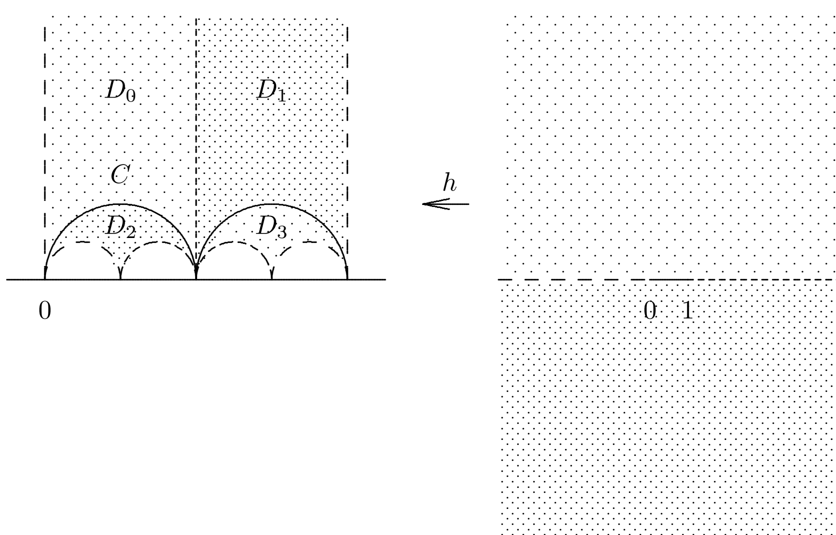

Fig. 5.3: The fundamental domain of the covering of the three punctured sphere.

Corollary 5.7. $S = \overline{\mathbb{C}} \setminus \{z_1, z_2, z_3\}$ *is a hyperbolic Riemann surface.*

Proof. We may assume that $z_1 = 0,$, $z_2 = 1$ and $z_3 = \infty$. Let $D_0 \subset \mathbb{H}$ be the simply connected domain bounded by the vertical lines through 0 and 1 and by the circle Δ with radius 1/2 and centred at 1/2, as in Figure 5.3.

Let $\phi_0\colon D_0 \to \mathbb{H}$ be the Riemann mapping whose extension to the boundary maps $0 \mapsto 0$, $1 \mapsto 1$ and $\infty \to \infty$. Let D_1 be the reflection of D_0 with respect to the vertical line through 1 and D_2 be the reflection of D_0 with respect to the geodesic C with endpoints 0 and 1. Using the Schwarz Reflection Principle. we can extend ϕ_0 holomorphically to D_1 and D_2 mapping these domains holomorphically onto the lower half-plane. Now the boundary of D_2 consists of Δ and two circles perpendicular to the real axis, one going through 0 and 1/2 and one going through 1/2 and 1. Reflecting D_2 in the first circle, we get a new domain D_3. The holomorphic extension of ϕ_0 from the Schwarz Reflection Principle, maps D_3 again to the upper half-plane. Continuing this construction we get a holomorphic covering

$$h\colon \mathbb{H} \to \bar{\mathbb{C}} \setminus \{0, 1, \infty\} = \bar{\mathbb{C}} \setminus \{z_1, z_2, z_3\}$$

that extends ϕ_0. □

Corollary 5.8. *Every open set $U \subset \bar{\mathbb{C}}$ such that the cardinality of $\mathbb{C} \setminus U$ is ≥ 3 is a hyperbolic Riemann surface.*

Proof. If this were not the case then there would exist a holomorphic covering map $\pi\colon \mathbb{C} \to U$. On the other hand, $U \subset \bar{\mathbb{C}} \setminus \{z_1, z_2, z_3\}$ and there exists a holomorphic covering map $\tilde{\pi}\colon \mathbb{D} \to \bar{\mathbb{C}} \setminus \{z_1, z_2, z_3\}$. The inclusion $i\colon U \to \bar{\mathbb{C}} \setminus \{z_1, z_2, z_3\}$ lifts to a holomorphic map $\tilde{i}\colon \mathbb{C} \to \mathbb{D}$. By Liouville's Theorem this map is constant, a contradiction. □

Definition. Let $\mathcal{F}$ be a family of holomorphic maps on some Riemann surface S. We say that this family is *normal* if any sequence of maps in $\mathcal{F}$ has a subsequence which converges uniformly on compact subsets of S.

Corollary 5.9. *Let S be a Riemann surface and $\mathcal{F}$ a family of holomorphic maps from S to the Riemann sphere $\bar{\mathbb{C}}$. If there are three distinct points $z_1, z_2, z_3 \in \bar{\mathbb{C}}$ that are omitted by the image of each map in $\mathcal{F}$ then $\mathcal{F}$ is a normal family.*

Let S be a hyperbolic Riemann surface and $\pi\colon \mathbb{D} \to S$ a holomorphic covering map. The set $\Gamma = \{\phi\colon \mathbb{D} \to \mathbb{D};\ \phi \text{ is a homeomorphism with } \pi\phi = \pi\}$ is called the group of deck transformations. Since π is holomorphic, the maps in Γ are also holomorphic. Therefore Γ consists of isometries of the hyperbolic metric, there exists a unique Riemannian metric on S, called the Poincaré or the hyperbolic metric of S, such that π is a local isometry. As a consequence of Theorem 2 we immediately get the following statement.

Schwarz Lemma. *If $f\colon S_1 \to S_2$ is a holomorphic map between hyperbolic Riemann surfaces then either f strictly contracts the Poincaré metric or it is a covering map.*

Example. Suppose S is homeomorphic to an annulus. Then the fundamental group $\pi_1(S)$ is isomorphic to $\mathbb{Z}$. Therefore the group Γ of deck transformations, which is isomorphic to $\pi_1(S)$, is generated by a unique Möbius transformation ϕ. There are three cases:

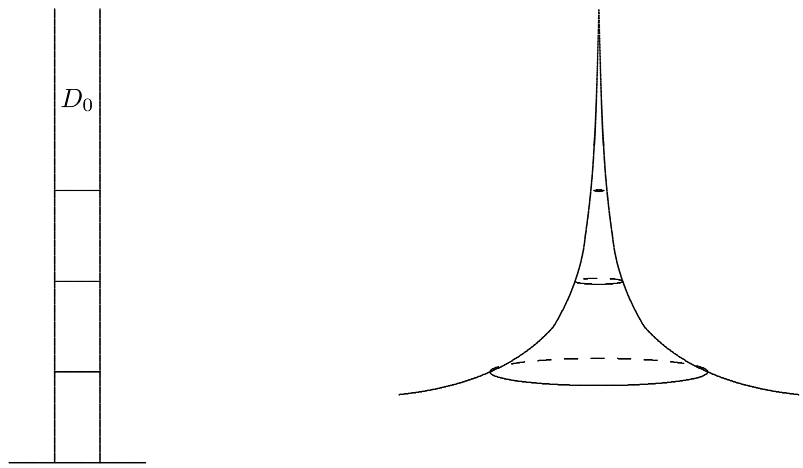

Fig. 5.4: On the right of the figure we see an isometric embedding of the surface S in $\mathbb{R}^3$ from Example b). The length of the closed curves indicated, which are the images of the higher and higher horizontal segments in the fundamental domain D_0, go to zero as they approach the cusp.

a) If S is not hyperbolic (so S is covered by $\mathbb{C}$), then we may assume that $\phi(z) = z + 1$ up to conjugation with a Möbius transformation in $\mathbb{C}$. The image by the projection map $\pi\colon \mathbb{C} \to S = \mathbb{C}/\Gamma$ of the horizontal lines gives a foliation on S by closed curves having all the same length. So $S = \mathbb{C}/\Gamma$ is isometric to an infinite cylinder.

b) If S is hyperbolic (so S is covered by $\mathbb{H}$), but the group of deck transformation is generated by a parabolic transformation ϕ, we may assume by conjugating the group with a Möbius transformation that $\phi(z) = z + 1$. The image by the projection map $\pi\colon \mathbb{H} \to S = \mathbb{H}/\Gamma$ of the horizontal lines gives a foliation on S by closed curves such that the length of the curves goes to zero in one direction

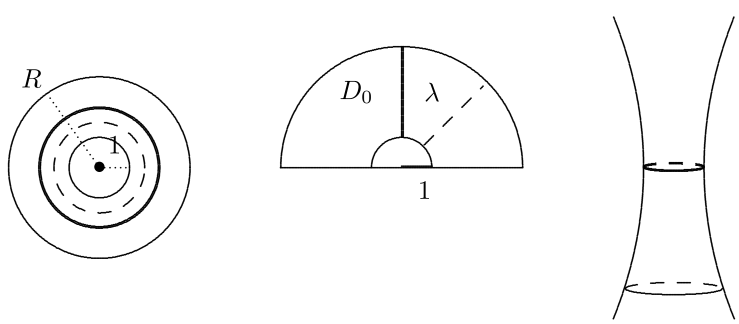

Fig. 5.5: On the right of the figure we see an isometric embedding of the surface S in $\mathbb{R}^3$ from Example c). The closed curves in S are also drawn in the fundamental domain in the figure in the middle (where they are straight lines and where $\log R = \lambda$) and on the left (where they are circles).

and to infinity in the other. So $S = \mathbb{C}/\Gamma$ is isometric to the surface on the right in Figure 5.4.

c) If $\phi\colon \mathbb{H} \to \mathbb{H}$ is a hyperbolic transformation we may assume $\phi(z) = \lambda z$ where $\lambda > 1$. The image by the projection map $\pi\colon \mathbb{H} \to S = \mathbb{H}/\Gamma$ of the horizontal lines gives a foliation on S by closed curves such that the length of the curves goes to infinity in both directions. So $S = \mathbb{C}/\Gamma$ is isometric to the surface drawn in Figure 5.5. Moreover, S is holomorphically equivalent to

$$A_R = \{z\,;\ 1 < |z| < R\},$$

where

$$(*) \qquad \log R = \frac{1}{\log(\text{hyperbolic length of } \gamma)}$$

and γ is the unique simple closed hyperbolic geodesic of S (it is the image of the imaginary axis). The number from $(*)$ is called the *modulus* of the annulus S. From this definition it follows that this modulus is a conformal invariant: A_{R_1} and A_{R_2} are conformally diffeomorphic iff $R_1 = R_2$ (any conformal diffeomorphism between A_{R_1} and A_{R_2} is an isometry for the Poincaré metric and therefore preserves the length of the simple closed geodesics).

Proposition 5.3. *Let A, A_1, A_2 be hyperbolic Riemann surfaces homeomorphic to the annulus. Then*

1) *if $f\colon A_1 \to A_2$ is a holomorphic covering map of degree n then*

$$\text{modulus of } A_2 = \frac{1}{n}(\text{ modulus of } A_1);$$

2) *if A_1 is a proper subset of A_2 and the generators of $\pi_1(A_1)$ also generate A_2 then*

$$\text{modulus } A_1 < \text{ modulus } A_2;$$

3) *if $A \supset A_1 \cup A_2$, $A_1 \cap A_2 = \emptyset$ and the generators of the fundamental groups $\pi_1(A_1)$ and $\pi_1(A_2)$ both generate $\pi_1(A)$ then*

$$\text{modulus of } A \geq (\text{ modulus of } A_1) + (\text{ modulus of } A_2).$$

Proof. The first two statements follow easily from the Schwarz Lemma. The last statement is more difficult to prove, see Theorem 4 in Chapter I of Ahlfors (1966). □

From the Riemann Mapping Theorem and the above discussion, it follows that any simply connected domain in the plane and any doubly connected domain may be mapped conformally onto the disc $\mathbb{D}$ respectively on some annulus A_R. Therefore a univalent map on the disc $\mathbb{D}$ (here we say that a map is called *univalent* if it is holomorphic and injective) may have very large nonlinearity. However, the results below say that this nonlinearity is concentrated near the boundary of the domain.

Theorem 5.5. **(Koebe's Distortion Theorem)** *For each* $\lambda \in (0,1)$ *there exists* $B(\lambda) \geq 1$ *such that if* $F\colon \mathbb{D} \to \mathbb{C}$ *is univalent then*

$$\left| \frac{\partial F(z)}{\partial F(w)} \right| \leq B(\lambda) \text{ for all } z, w \in \mathbb{D}_\lambda = \{x \in \mathbb{C}\,;\ |x| < \lambda\}.$$

Furthermore, $B(\lambda) \to 1$ *if* $\lambda \to 0$.

Proof. See Ahlfors (1973, pp. 84). □

Corollary 5.10. *For each* $\lambda \in (0, 1/2)$ *and each* $a > 1$ *there exists* $D(\lambda, a) \geq 1$ *such that if* $F\colon A_a \to \mathbb{C}$ *is a univalent function then*

$$\left| \frac{\partial F(z)}{\partial F(w)} \right| \leq D(\lambda, a) \text{ for all } z, w \in \tilde{A}_a$$

where $\tilde{A}_a = \{z\,;\ 1 + \lambda a < |z| < 1 + (1 - \lambda)a\}$.

Theorem 5.6. **(Koebe's** $1/4$**-Theorem)** *If* $F\colon \mathbb{D} \to \mathbb{C}$ *is univalent and* $|\partial F(0)| = 1$ *then* $F(\mathbb{D})$ *contains a disc of radius* $1/4$ *centred at* $F(0)$.

Proof. Ahlfors (1973, pp. 84). □

5.4 Deformations of complex structures; Teichmüller spaces

The basic tool for understanding the space of all conformal structures on a given surface is the Measurable Riemann Mapping Theorem below.

Definition. A *Beltrami coefficient* on an open set $U \subset \bar{\mathbb{C}}$ is a measurable function $\mu\colon U \to \mathbb{C}$ such that $|\mu|$ has essential supremum $k < 1$. (This means that the set $\{z \in U\,;\ |\mu(z)| > k\}$ has zero Lebesgue measure and that k is the smallest number with this property.) If $f\colon U \to f(U)$ is a quasiconformal homeomorphism then $\mu_f(z) = \dfrac{\bar{\partial} f(z)}{\partial f(z)}$ is a Beltrami coefficient. It is called the Beltrami coefficient of f. A Beltrami coefficient can be seen as a field of ellipses: given a Beltrami coefficient μ, we can associate a field of ellipses (up to homothety). Indeed, if $\mu(z) = 0$, the ellipse $E(z)$ is a circle. If $\mu(z) \neq 0$ then the direction of the major axis of $E(z)$ is given by $\sqrt{\nu(z)}$ and the eccentricity (i.e., the ratio of major axis by the minor axis) is $\frac{1+|\mu(z)|}{1-|\mu(z)|}$. Conversely, any measurable field of ellipses, with bounded eccentricity, defines a Beltrami coefficient. The next result tells us that we can always integrate such a Beltrami coefficient:

Theorem 5.7. **(Measurable Riemann Mapping Theorem)** *Let* $\mu\colon \bar{\mathbb{C}} \to \mathbb{C}$ *be a Beltrami coefficient. Then there exists a unique quasiconformal homeomorphism* $f\colon \bar{\mathbb{C}} \to \bar{\mathbb{C}}$ *which satisfies the Beltrami differential equation*

$$\frac{\bar{\partial} f(z)}{\partial f(z)} = \mu(z) \text{ for almost all } z \in \bar{\mathbb{C}}$$

and is normalized such that $f(0) = 0$, $f(1) = 1$ *and* $f(\infty) = \infty$.

Remark. It follows that if $\mu\colon U \to \mathbb{C}$ is a Beltrami coefficient, there exists a quasiconformal homeomorphism $f\colon U \to f(U)$ with Beltrami coefficient μ. Indeed, we just extend μ to $\bar{\mathbb{C}}$ (for example by taking $\mu(z) = 0$ for $z \notin U$) and use the theorem. The homeomorphism will depend on the extension of μ and any two such homeomorphisms f_1 and f_2 differ by a holomorphic homeomorphism $\phi\colon f_1(U) \to f_2(U)$ (that is, $\phi \circ f_1 = f_2$).

The result below describes the dependence of the solutions of the Beltrami equations on the parameters.

Theorem 5.8. **(Ahlfors-Bers Theorem)** *Let $\Lambda \subset \mathbb{C}^n$ be an open set and $\mu\colon \bar{\mathbb{C}} \times \Lambda \to \mathbb{D}$ be a measurable function satisfying:*

a) $|\mu(z,\lambda)| \leq k < 1$ *for all $\lambda \in \Lambda$ and for almost all $z \in \bar{\mathbb{C}}$;*

b) $\lambda \mapsto \mu(z,\lambda)$ *is holomorphic in λ for almost all $z \in \bar{\mathbb{C}}$.*

Then there exists a unique function $F\colon \bar{\mathbb{C}} \times \Lambda \to \bar{\mathbb{C}}$ such that

1. $F(0,\lambda) = 0$, $F(1,\lambda) = 1$, $F(\infty,\lambda) = \infty$;
2. *For every $\lambda \in \Lambda$ the map $z \mapsto F(z,\lambda)$ is a quasiconformal homeomorphism whose Beltrami coefficient is $\mu(\cdot,\lambda)$;*
3. $\lambda \mapsto F(z,\lambda)$ *is holomorphic for almost every z.*

If instead of b) one has that $\lambda_n \to \lambda$ implies $\mu(\cdot,\lambda_n) \to \mu(\cdot,\lambda)$ almost everywhere, then F satisfies Statements 1, 2 and is continuous in both variables.

Let us give some useful properties of quasiconformal homeomorphisms.

Proposition 5.4. 1. *If $\mathcal{A}$ is a family of quasiconformal homeomorphisms of $\bar{\mathbb{C}}$ whose conformal distortions are uniformly bounded by K then*

a) *every uniform limit of a sequence in $\mathcal{A}$ is either constant or it is a K-quasiconformal homeomorphism;*

b) *every sequence in $\mathcal{A}$ has a subsequence that converges uniformly.*

2. $f\colon U \subset \bar{\mathbb{C}} \to V$ *is a 1-quasiconformal homeomorphism if and only if f is holomorphic.*

3. *Every quasiconformal homeomorphism $\phi\colon \mathbb{D} \to \mathbb{D}$ extends to a quasiconformal homeomorphism of $\mathbb{C}$.*

4. $f\colon \bar{\mathbb{C}} \to \bar{\mathbb{C}}$ *is a K-quasiconformal homeomorphism if and only if for every annulus $A \subset \bar{\mathbb{C}}$*

$$\frac{1}{K}\text{ modulus of } f(A) \leq \text{ modulus of } A \ \leq K \text{ modulus of } f(A).$$

5. *If* $f\colon U \to V$ *and* $g\colon V \to U$ *are quasiconformal homeomorphism then* $g \circ f$ *is quasiconformal and*

$$\mu_{g\circ f}(z) = \frac{\mu_f(z) + \mu_g(f(z)) \cdot \dfrac{\overline{\partial f(z)}}{\partial f(z)}}{1 + \mu_g(f(z)) \cdot \overline{\mu_f(z)} \cdot \dfrac{\overline{\partial f(z)}}{\partial f(z)}}.$$

In particular, if g *is holomorphic then* $\mu_{g\circ f} = \mu_f$ *and if* f *is holomorphic then* $\mu_{g\circ f} = \mu_g(f(z)) \cdot \dfrac{\overline{\partial f(z)}}{\partial f(z)}$.

Let us introduce some notation. Given an open subset U of $\bar{\mathbb{C}}$, we denote by $\mathcal{B}(U)$ the set of Beltrami coefficient on U. So $\mathcal{B}(U)$ is the unit ball in $L^\infty(U)$. We define a new metric on $\mathcal{B}(U)$ called the Poincaré metric of $\mathcal{B}(U)$, as follows:

$$d_B(\mu_1, \mu_2) = \text{ess sup}_{z\in U} d_P(\mu_1(z), \mu_2(z))$$

where $d_P(\mu_1(z), \mu_2(z))$ is the Poincaré distance between two points in $\mathbb{D}$.

Definition. A *Beltrami path* on U is a path $t \mapsto \mu_t \in \mathcal{B}(U)$ such that for almost all $z \in U$, $t \mapsto \mu_t(z)$ is a hyperbolic geodesic on $\mathbb{D}$.

Notice that a Beltrami path μ_t is a straight line in the metric space $\mathcal{B}(U)$ with the metric d_B. This means that for all $t_0 < t_1 < \cdots < t_n$

$$d_B(\mu_{t_1}, \mu_{t_n}) = \sum_{i=0}^{n-1} d_B(\mu_{t_i}, \mu_{t_{i+1}}).$$

In particular, μ_t is a minimal geodesic: it is the curve of smallest length joining its endpoints. Here we are using the following concept. In a metric space M with distance function d, the length of a curve $\alpha\colon [0,1] \to M$ is defined as the supremum of $\sum_{i=0}^{n-1} d(\alpha(t_{i+1}), \alpha(t_i))$ over all partitions $0 = t_0 < t_1 < \cdots < t_n = 1$ of the interval $[0,1]$. From the triangle inequality it follows that the length of a curve is at least as big as the distance between the endpoints. If it is equal then the curve is called a *geodesic*.

If μ_t is a Beltrami path then its tangent vector at $t = t_0$,

$$\nu_{t_0}(z) = \frac{d}{dt}\mu_t(z)\Big|_{t=t_0},$$

is an essentially bounded measurable function called a *Beltrami vector*. Conversely, any function $\nu \in L^\infty(U)$ is tangent to a unique Beltrami path. The following lemma will be used in extending these notions to Riemann surfaces.

Lemma 5.1. 1. *Let* $f\colon U \to V$ *be a quasiconformal map. If* $\mu \in \mathcal{B}(V)$ *then its pullback* $f^*\mu$ *defined by*

$$(*) \qquad (f^*\mu)(z) = \frac{\mu_f(z) + \mu(f(z)) \cdot \dfrac{\overline{\partial f(z)}}{\partial f(z)}}{1 + \mu(f(z)) \cdot \overline{\mu_f(z)} \cdot \dfrac{\overline{\partial f(z)}}{\partial f(z)}}$$

is a Beltrami coefficient on U.

2. *The map $f^*\colon \mathcal{B}(V) \to \mathcal{B}(U)$ defined by $(*)$ is an isometry of the Poincaré metric and maps Beltrami paths into Beltrami paths.*

Proof. The result follows easily from the fact that for each $z \in U$ the map

$$\mathbb{D} \ni u \mapsto \frac{\mu_f(z) + u \cdot \dfrac{\overline{\partial f(z)}}{\partial f(z)}}{1 + u \cdot \overline{\mu_f(z)} \cdot \dfrac{\overline{\partial f(z)}}{\partial f(z)}}$$

is in $M(\mathbb{D})$. □

Remark. 1. If $f\colon U \to V$ is a quasiconformal homeomorphism, then the field of ellipses associated to $f^*(\mu)$ is mapped, by the derivative of f into the field of ellipses associated to μ. Therefore, the invariance of a Beltrami coefficient under a dynamical system or under a Fuchsian group (i.e., $f^*\mu = \mu$ for the dynamical system f or each map from the group) is the same as the invariance of the corresponding field of ellipses under the derivative of the corresponding maps. In particular, a Beltrami coefficient in a Riemann surface S is the same as a pair $(|\mu|, l)$, where $|\mu|\colon S \to [0,1)$ is a measurable function whose essential supremum is smaller than one and l is a measurable line field in the set of points of S where the previous function is non-zero.

2. If f is holomorphic then

$$(f^*\mu) = \mu(f(z)) \cdot \frac{\overline{\partial f(z)}}{\partial f(z)} \text{ a.e.}$$

and this pullback map is well defined even if f is not injective.

3. $\mu_f = f^*(0)$. In other words, the Beltrami coefficient of a quasiconformal homeomorphism is the pullback of the Beltrami coefficient identically zero.

4. If $f\colon U \to V$, $g\colon V \to W$ are quasiconformal homeomorphisms then

$$(g \circ f)^* = f^* \circ g^*.$$

5. The tangent action of f on the Beltrami vector is given by

$$(Tf^*(\mu))(\nu)(z) = \frac{\left(\dfrac{\overline{\partial f(z)}}{\partial f(z)}\right)(1 - |\mu_f|^2)}{\left(1 + \mu(f(z)) \cdot \overline{\mu_f(z)} \cdot \dfrac{\overline{\partial f(z)}}{\partial f(z)}\right)^2}\nu(f(z)).$$

So if ν is a Beltrami vector tangent to the Beltrami path μ_t at $t = t_0$ and $\mu_{t_0} = \mu$ then $(Tf^*(\mu))(\nu)$ is the Beltrami vector tangent to the Beltrami path $f^*(\mu_t)$ at $t = t_0$. Notice that if f is holomorphic then

$$(T^*f(\mu))(\nu)(z) = \nu(f(z)) \cdot \frac{\overline{\partial f(z)}}{\partial f(z)} \text{ a.e.}$$

Let us show how to use the Beltrami paths to construct deformations of holomorphic dynamical systems.

Theorem 5.9. *Let $F\colon U \to V$ be a holomorphic map between open sets such that $V \supset U$. Let $\mu_t\colon V \to \mathbb{C}$ be a Beltrami path in V with $\mu_0 \equiv 0$, satisfying the condition*

$$F^*(\mu_t) = \mu_t \text{ for all } t.$$

Let $\phi_t\colon V \to V_t \subset \mathbb{C}$ be a continuous family of quasiconformal homeomorphisms with $\phi_0 = id$ such that the Beltrami coefficient of ϕ_t is μ_t. Then

$$F_t = \phi_t \circ F \circ \phi_t^{-1}$$

is a continuous family of holomorphic maps.

Proof.

$$\begin{aligned} F_t^*(0) &= (\phi_t^{-1})^* F^* \phi_t^*(0) \\ &= (\phi_t^{-1})^* F^* \mu_t \\ &= (\phi_t^{-1})^*(\mu_t) = 0. \end{aligned}$$

Hence F_t is locally 1-quasiconformal and hence conformal. □

Let us give another important application of the Measurable Riemann Mapping Theorem to holomorphic dynamical systems. Let $F\colon U \to V$ be a holomorphic proper map where U and V are simply connected domains and V contains the closure of U. Such a map is called *polynomial-like*. The *degree* of such a map is the cardinality of $F^{-1}(y)$ for every y which is not a critical value. Let $J(F) = \{z\,;\ f^n(z) \in U \text{ for all } n \geq 0\}$. This set is called the *filled Julia set* of the polynomial-like map F. Douady and Hubbard (1985a) proved

Theorem 5.10. **(Straightening Theorem)** *Let $F\colon U \to V$ be a polynomial-like map and d be the degree of F. Then there exists a polynomial map P of degree d, a neighbourhood W of $J(F)$ such that $F\colon W \to F(W)$ is a polynomial-like map and there exists a quasiconformal homeomorphism $\phi\colon F(W) \to P(\phi(W))$ that conjugates $F \mid W$ to $P \mid \phi(W)$.*

Proof. Let $\tilde{V} \subset V$ be a simply connected domain whose boundary is a smooth curve in $V \setminus \mathrm{cl}\,(U)$. Take $W = F^{-1}(\tilde{V})$. Then $F\colon W \to F(W)$ is also a polynomial map of degree d. Let ϕ be a smooth diffeomorphism of a small neighbourhood of the closure of $F(W) \setminus W$ onto a neighbourhood of the closure of $\mathbb{D}_{2^d} \setminus \mathbb{D}_2$ that maps the boundary of W to the boundary of $\mathbb{D}_2$ and conjugates $F|\partial W$ to the restriction of $z^d|\partial\,\mathbb{D}_2$. Let S be the space obtained by taking the disjoint union of $F(W)$ and $\mathbb{D}_2$ modulo the equivalence relation that identifies $z \in F(W)$ to $\phi(z) \in \mathbb{D}_2$. Clearly S is homeomorphic to the sphere. Let π be the projection map and let $U_1 = \pi(F(W))$ and $U_2 = \pi(\bar{\mathbb{C}} \setminus \mathbb{D}_2)$. Let $\tilde{\phi}_1 = \pi^{-1}|U_1$ and $\tilde{\phi}_2 = \pi^{-1}|U_2$ be the charts. The change of coordinates is ϕ and so we have a smooth structure on S. Take in $\bar{\mathbb{C}} \setminus \mathbb{D}_2$ the Beltrami coefficient $\mu_2 \equiv 0$ and in $F(W) \setminus W$ be $\mu_1 = \phi^*(0)$. Next extend it dynamically:

$F^*(\mu_1) = \mu_1$ on $F(W) \setminus J(F)$. Finally define $\mu_1 \equiv 0$ on $J(F)$. Let ψ be the quasiconformal homeomorphism from $F(W)$ to its image whose Beltrami coefficient is μ_1. Let $\phi_1 = \psi \circ \tilde{\phi}_1$ and $\phi_2 = \tilde{\phi}_2$. These maps define a holomorphic atlas on S. Take $z \in S$. If z is in the domain of ϕ_1 and $\phi_1(z) \in W$ then take $P(z) = \tilde{\phi}_1^{-1}(F(\tilde{\phi}_1(z))$. Otherwise $P(z) = \tilde{\phi}_2^{-1}([\tilde{\phi}_2(z)]^d)$. Now P is a holomorphic map of degree d (here we use that ϕ can also be defined as a conjugacy on a small neighbourhood of the closure of $F(W) \setminus W$ to show that P is well-defined in the intersections of these domains). Since S is homeomorphic to the sphere, it is conformally the Riemann sphere. Therefore P is rational map of degree d. Moreover, $\phi_2^{-1}(\infty)$ is both a fixed point and a critical point of degree d for P. Hence P is a polynomial. □

The deformation theory for Riemann surfaces is similar to the above deformation of holomorphic dynamical systems. Let S be a Riemann surface. A *Beltrami coefficient* μ on the Riemann surface S is a collection of Beltrami coefficients $\mu_i \colon \phi_i(U_i) \to \mathbb{C}$ one for each chart $\phi_i \colon U_i \to \phi_i(U_i) \subset \mathbb{C}$ of S, called the expression of μ in ϕ_i, satisfying the compatibility condition

$$(\phi_j \circ \phi_i^{-1})^* \mu_j = \mu_i \text{ on } \phi_i(U_i \cap U_j)$$

and the boundedness condition

$$|\mu_i(z)| < k \text{ a.e. } z \in \phi_i(U_i)$$

where $k < 1$ is independent of i.

A homeomorphism f between two Riemann surfaces S_1 and S_2 is K*-quasiconformal* if all local expressions of f are K-quasiconformal. The pullback map $f^* \colon \mathcal{B}(S_2) \to \mathcal{B}(S_1)$ is defined by taking the pullback of the local expressions. The Poincaré metric on $\mathcal{B}(S)$ is defined as

$$d_B(\mu_1, \mu_2) = \sup_i \text{ess sup}_{z \in \phi_i(U_i)} d_P(\mu_{1,i}(z), \mu_{2,i}(z)).$$

(Note that $d_P(\mu_{1,i}(z), \mu_{2,i}(z)) = d_P(\mu_{1,j}(z'), \mu_{2,j}(z'))$ if $\phi_i^{-1}(z) = \phi_j^{-1}(z')$. So the supremum over i above may be taken over any atlas.) Note that f^* is an isometry, see Statement 2 of Lemma 5.1.

We say that a one parameter family μ_t of Beltrami coefficients is a *Beltrami path* on S if for each chart $\phi_i \colon U_i \to \phi_i(U_i)$, $t \mapsto \mu_{i,t}$ is a Beltrami path in $\phi_i(U_i)$ where $\mu_{i,t}$ is the expression of μ_t in this chart. Similarly, we define a Beltrami vector ν at the Beltrami coefficient μ as a collection $\nu_i \colon \phi_i(U_i) \to \mathbb{C}$ of Beltrami vectors such that

$$T(\phi_j \circ \phi_i^{-1})^*(\mu_j)(\nu_j) = \nu_i$$

and such that the essential supremum $|\nu_i|$ is uniformly bounded. This invariance condition can be written in local coordinates as follows

$$\frac{\bar{\partial}(\phi_j \circ \phi_i^{-1})}{\partial(\phi_j \circ \phi_i^{-1})} \cdot \nu_j(\phi_j \circ \phi_i^{-1}(z)) = \nu_i(z).$$

As before, each Beltrami vector ν at μ determines a unique Beltrami path μ_t with $\mu_0 = \mu$ and conversely.

By the Measurable Riemann Mapping Theorem, each Beltrami coefficient μ on S determines a new Riemann surface S_μ and a quasiconformal map $f\colon S \to S_\mu$ such that $f^*(0) = \mu$. Indeed, let $\mathcal{A}$ be an atlas on S and for each $\phi_i \in \mathcal{A}$ let $\mu_i\colon \phi_i(U_i) \to \mathbb{C}$ be the expression of μ. Let $\psi_i\colon \phi_i(U_i) \to \psi(\phi_i(U_i))$ be a quasiconformal map whose Beltrami coefficient is μ_i, i.e., $\psi_i^*(0) = \mu_i$. Then $\tilde{\phi}_i = \psi_i \circ \phi_i$ is an atlas $\tilde{\mathcal{A}}$ of S whose overlapping maps are 1-quasiconformal and therefore holomorphic maps. Then let S_μ be the topological space S endowed with the complex structure defined by $\tilde{\mathcal{A}}$. By taking f to be the identity map we get $f^*(0) = \mu$. Hence a Beltrami path μ_t on a Riemann surface S defines a one parameter family of Riemann surfaces S_{μ_t}. If $\mu_0 = 0$ then this is a deformation of the original complex structure on S. Conversely, given any Riemann surface S_1 and a quasiconformal homeomorphism f from S to S_1 one gets a Beltrami coefficient $\mu = f^*(0)$ on S. Hence, we can identify the space of $\mathcal{B}(S)$ with the space $C(S)$ of all complex structures on the topological space which are quasiconformally homeomorphic to the original structure.

This discussion is very similar to the corresponding deformation of dynamical systems. Indeed consider the universal holomorphic covering map $\pi\colon \mathbb{D} \to S$. Then the group of deck transformations is a group of conformal homeomorphisms and the Beltrami coefficients μ on S correspond precisely to the Beltrami coefficients $\tilde{\mu}$ on $\mathbb{D}$ such that $A^*\tilde{\mu} = \tilde{\mu}$ for all $A \in \Gamma$. Each such Beltrami coefficient determines a quasiconformal homeomorphism $\psi\colon \mathbb{D} \to \mathbb{D}$ and the above compatibility condition implies that $\Gamma_\mu = \psi \circ \Gamma \circ \psi^{-1} = \{\psi \circ \gamma \circ \psi^{-1}\,;\ \gamma \in \Gamma\}$ is also a group of Möbius transformations.

In general, given a Beltrami coefficient μ, a complex structure S_μ can be holomorphically diffeomorphic to the original structure S. This is the case if S is simply connected, see Theorem 5.3. In fact we are going to distinguish the Beltrami coefficients using a stronger equivalence relation. Two Beltrami coefficients μ_1 and μ_2 are equivalent in the sense of Teichmüller, $\mu_1 \sim_T \mu_2$ if $\mu_1 = \phi^*\mu_2$ where $\phi\colon S \to S$ is a quasiconformal homeomorphism which is *isotopic* to the identity. The space of equivalence classes of Beltrami coefficients on S is called the Teichmüller space of S and denoted by $T(S)$. Let $\pi\colon \mathcal{B}(S) \to T(S)$ be the corresponding projection. The space $T(S)$ has a natural metric:

$$d_T([\mu_1],[\mu_2]) = d_B(\pi^{-1}([\mu_1]), \pi^{-1}([\mu_2])).$$

Note that this is precisely the minimum of the Poincaré distance $d_P(\mu_1, \phi^*(\mu_2))$ over all quasiconformal homeomorphisms $\phi\colon S \to S$ which are isotopic to the identity.

Example. Let $S = A_R$ with $R > 1$. Then the Teichmüller space of S is homeomorphic to the real line. Indeed, let $m\colon \mathcal{B}(S) \to (0,\infty)$ be the modulus of S_μ. Then $m(\mu_1) = m(\mu_2)$ if and only if there exists a quasiconformal homeomorphism $\phi\colon S \to S$ with $\phi^*(\mu_1) = \mu_2$. This last statement is equivalent to $\mu_1 \sim_T \mu_2$ because all homeomorphisms of the annulus are isotopic to the identity. So m induces an bijection between $T(S)$ and $(0,\infty)$. It is easy to see that m is a homeomorphism.

It is easy to show that a Beltrami path μ_t is a minimal geodesic in $\mathcal{B}(S)$ with the metric d_B in the sense discussed before. However, the projection in the Teichmüller space $T(S)$ is general not a geodesic of the Teichmüller metric (for example, a Beltrami path could be tangent to a fibre of π). One of the main goals of the Teichmüller theory is to characterize the Beltrami vectors that generate Beltrami paths which give geodesics in the Teichmüller space. In Chapter VI we will give this characterization in the case of compact Riemann surfaces.

From this characterization one can conclude that the Teichmüller space of a compact surface of genus g is homeomorphic to the open unit ball in $\mathbb{R}^{6g-6}$.

6 Some Results from Functional Analysis

In the last chapter we will also need the following two theorems.

Theorem 6.1. **(Riesz Representation Theorem)** *Let X be a locally compact metric space and let Λ be a positive linear functional on the space of all continuous functions with compact support. Then there exists a unique positive Borel measure on X such that*

$$\Lambda f = \int_X f \, d\mu$$

for all continuous functions with compact support.

Theorem 6.2. **(Hahn-Banach Extension Theorem)** *If M is a subspace of a normed linear space X and if f is a bounded linear functional on M, then f can be extended to a bounded linear functional F on X such that $||F|| = ||f||$.*

VIII.

Bibliography

This is not intended to be a complete bibliography on the subject of one-dimensional dynamical systems. We have only included some references to the combinatorial theory of interval maps and to physics papers on this subject.

This is not intended to be a complete bibliography on the subject of one-dimensional dynamical systems. We have only included some references to the combinatorial theory of interval maps and to physics papers on this subject.

J. Aaronson, M. Denker and M. Urbański (1992): Ergodic theory for Markov fibred systems and parabolic rational maps. Preprint

R.L. Adler (1973): F-expansions revisited. Lect. Notes in Math., vol. 318. Springer, Berlin New York, pp. 1–5

R.L. Adler, A.G. Konheim and M.H. MacAndrew (1965): Topological entropy. Trans. A.M.S. **114***, 309–319*

L.V. Ahlfors (1966): Lectures on Quasiconformal Mappings. *Van Nostrand, Princeton, N.J. Reprinted in 1987 by Wadsworth & Brooks/Cole*

L.V. Ahlfors (1973): Conformal Invariants, Topics in Geometric Function Theory, *McGraw-Hill, New York*

L.V. Ahlfors (1979): Complex Analysis. *McGraw-Hill, New York*

L.V. Ahlfors and L. Bers (1960): Riemann's mapping theorem for variable metrics. Ann. Math. **72***, 385–404*

D.J. Allwright (1978): Hypergraphic functions and bifurcations in recurrence relations. SIAM J. App. Math. **34***, 687–691*

L. Alsedà and J. Llibre (1985): A note on the set of periods for continuous maps of the circle which have degree one. Proc. A.M.S. **93***, 133–138*

L. Alsedà, J. Llibre and M. Misiurewicz (1987): Periodic orbits of Y. Preprint Barcelona

L. Alsedà, J. Llibre and M. Misiurewicz (1990): Combinatorial Dynamics and Entropy in Dimension One. *To appear*

L. Alsedà, J. Llibre, M. Misiurewicz and C. Tresser (1989): Periods and entropy for Lorenz-like maps. Ann. Inst. Fourier, Grenoble **39***, 929–952*

L. Alsedà, J. Llibre and R. Serra (1984): Minimal periodic orbits for continuous maps of the interval. Trans. A.M.S. **286***, 595–627*

L. Alsedà and F. Mañosas (1990): Kneading theory and rotation intervals for a class of circle maps of degree one. Nonlinearity **3***, 413–452*

V.I. Arnol'd (1961): Small denominators I. On the mapping of a circle into itself. Izv. Akad. Nauk. Math. Serie **25**, *21–86 [Transl. A.M.S. Serie 2,* **46**, *(1965)]*

V.I. Arnol'd (1983): Geometric Methods in the Theory of Ordinary Differential Equations. *Springer, Berlin New York*

S.Kh. Aranson and V.Z. Grines (1986): Topological classification of flows on closed two-dimensional manifolds. Russian Math. Surv. **41**, *183–208*

V. Baladi and G. Keller (1990): Zeta functions and transfer operators for piecewise monotonic transformations. Commun. Math. Phys. **127**, *459–477*

S. Baldwin (1987): Generalizations of a theorem of Sarkovskii on orbits of continuous real-valued functions. Discrete Math. **67**, *111–127*

S. Baldwin (1988): An extension of Sarkovskii's theorem to the n-od. Preprint Auburn Univ.

S. Baldwin (1990): A complete classification of the piecewise monotone functions of the interval. Trans. Amer. Math. Soc. **319**, *155–178*

J. Barkmeijer (1988): Dynamics and topological invariants of circle endomorphisms. Thesis, Groningen

T. Bedford, M. Keane and C. Series (1991): Ergodic Theory, Symbolic Dynamics and Hyperbolic Spaces. *Oxford University Press.*

M. Benedicks and L. Carleson (1985): On iterations of $1 - ax^2$ *on* $(-1, 1)$. *Ann. Math.* **122**, *1–25*

M. Benedicks and L. Carleson (1991): The dynamics of the Hénon map. Ann. Math. **133**, *73–169*

M. Benedicks and M. Misiurewicz (1989): Absolutely continuous invariant measures for maps with flat tops. Publ. I.H.E.S. **69**, *203–213*

M. Benedicks and L.-S. Young (1992): Absolutely continuous invariant measures and random perturbations for certain one-dimensional maps. Ergod. Th. & Dynam. Sys. **12**, *13–37*

C. Bernhardt (1982): Rotation intervals of a class of endomorphisms of the circle. Proc. London Math. Soc. **XLV**, *258–280*

C. Bernhardt (1987): The ordering on permutations induced by continuous maps of the real line. Ergod. Th. & Dynam. Sys. **7**, *155–160*

D. Berry and B.D. Mestel (1991): Wandering intervals for Lorenz maps with bounded nonlinearity. Bull. London Math. Soc. **23**, *183–189*

P. Blanchard (1984): Complex analytic dynamics on the Riemann sphere. Bull. A.M.S. **11**, *85–141*

M.L. Blank (1991): Some counter-examples to the theory of chaotic mappings. Preprint Moscow Univ.

L.S. Block (1979): Simple periodic orbits of mappings of the interval. Trans. A.M.S. **254**, *391–398*

L.S. Block and W.A. Coppel (1986): Stratifications of continuous maps of an interval. Trans. A.M.S. **297**, *587–604*

L.S. Block and W.A. Coppel (1989): Approximating entropy of maps of the interval. Banach Center Publications (of the Semester of Dynamical Systems in Warsaw, 1986) **23**, *237–241*

L.S. Block and W.A. Coppel (1992): Dynamics in One Dimension. *Lect. Notes in Math., vol. 1513. Springer, Berlin New York*

L.S. Block, J. Guckenheimer, M. Misiurewicz and L.S. Young (1980): Periodic points and topological entropy of one dimensional maps. Lect. Notes in Math., vol. 819. Springer, Berlin New York, pp. 18–34

L.S. Block and J. Keesling (1991): Computing the topological entropy of maps of the interval with three monotone pieces. Preprint

A.M. Blokh (1983): Decomposition of dynamical systems on an interval. Russ. Math. Surv. **38**, *133–134 (translation from Russian)*

A.M. Blokh (1986/7): On dynamical systems on one dimensional branched manifolds I,II, III. Functional Anal. and Appl. (Kharkov) (1986), **46**, *8–18/ (1987),* **47**, *67–77/ (1987),* **48**, *32–46*

A.M. Blokh (1991a): On some properties of graph maps: spectral decomposition, Misiurewicz conjecture and abstract sets of periods. Preprint Bonn

A.M. Blokh (1991b): The 'spectral' decomposition for one-dimensional maps. Preprint Middletown

A.M. Blokh and M.Yu. Lyubich (1986): Attractors of transformations of an interval. Banach Center Publications (of the Semester of Dynamical Systems in Warsaw, 1986) **23**, *427–442*

A.M. Blokh and M.Yu. Lyubich (1987): Attractors of maps of the interval. Funct. Anal. and Appl. **21**, *70–71 (in Russian)*

A.M. Blokh and M.Yu. Lyubich (1989a): Ergodicity of transitive maps of the interval. Ukrainian Math. J. **41**, *985–988*

A.M. Blokh and M.Yu. Lyubich (1989b): Ergodic properties of maps of the interval. Func. Anal and Appl. **23**, *59–60*

A.M. Blokh and M.Yu. Lyubich (1989c): On the decomposition of one-dimensional attractors of unimodal maps of the interval. Algebra and analysis (Leningrad Math. J.), **1**, *128–145*

A.M. Blokh and M.Yu. Lyubich (1989d): Non-existence of wandering intervals and structure of topological attractors of one dimensional dynamical systems II. The smooth case. Ergod. Th. & Dynam. Sys. **9**, *751–758*

A.M. Blokh and M.Yu. Lyubich (1990): Measure and dimension of solenoidal attractors of one-dimensional dynamical systems. Commun. Math. Phys. **127**, *573–583*

A.M. Blokh and M.Yu. Lyubich (1991): Measurable dynamics of S-unimodal maps of the interval. Ann. Sc. E.N.S. 4e *série,* **24**, *545–573*

R. Bowen (1971): Entropy of group endomorphisms and homogeneous spaces. Trans. A.M.S. **153**, *401–414*

R. Bowen (1975): A horseshoe with positive measure. Invent. Math. **29**, *203–204*

R. Bowen (1977a): Entropy for maps of the interval. Topology **16**, *465–467*

R. Bowen (1977b): Bernoulli maps of the interval. Israel J. of Math. **28**, *161–168*

R. Bowen (1979): Invariant measures for Markov maps of the interval. Commun. Math. Phys. **69**, *1–17*

C. Boyd (1985): On the structure of the family of Cherry fields on the torus. Ergod. Th. & Dynam. Sys. **5**, *27–46*

P.L. Boyland (1986): Bifurcations of circle maps: Arnol'd tongues, bistability and rotation intervals. Commun. Math. Phys. **106**, *353–381*

P.L. Boyland (1987): Braid types and a topological method of proving positive entropy. Preprint

P.L. Boyland (1988): An analog of Sarkovskii's theorem for twist maps. In: "Hamiltonian Dynamical Systems", ed. K.R. Meyer and D.G. Saari, Contemp. Math. **81**, *119–133*

B. Branner and J.H. Hubbard (1991): The iteration of cubic polynomials. To appear in Acta Math.

K.M. Brucks, M. Misiurewicz and C. Tresser (1991): Monotonicity properties of the family of trapezoidal maps. Commun. Math. Phys. **137***, 1–12*

K.M. Brucks, M.V. Otera-Espinar and C. Tresser (1991): Dense orbits of critical points for the tent map. Contemporary Math. **117***, 57–61*

K.M. Brucks, M.V. Otera-Espinar and C. Tresser (1992): Homeomorphic restrictions of smooth endomorphisms of an interval. Ergod. Th. & Dynam. Sys. **12***, 429–440*

H. Bruin (1992a), Numerical determination of the continued fraction expansion of the rotation number. Physica D **59***, 158–168*

H. Bruin (1992b): A topological condition for the existence of invariant measures for unimodal maps. Preprint Delft

H. Bruin (1992c): Combinatorics of the kneading map. In preparation

M. Campanino and H. Epstein (1981): On the existence of Feigenbaum's fixed point. Commun. Math. Phys. **79***, 261–302*

M. Campanino, H. Epstein and D. Ruelle (1982): On Feigenbaum's functional equation $g \circ g(\lambda x) + \lambda g(x) = 0$. *Topology* **21***, 125–129*

A. Candel (1991): Uniformization of foliations by surfaces. Thesis, Washington University, St. Louis, U.S.A.

E. Cartan (1937): Leçons sur la Théorie des Espaces a Conexion Projective. *Cahiers scientifique, Fasc XVII, Gautheir-Villars, Paris*

A.J. Casson and S.A. Bleiler (1988): Automorphisms of Surfaces after Nielsen and Thurston. *London Math. Soc. Student Texts* **9**. *Cambridge Univ. Press*

A. Chenciner, J.-M. Gambaudo and C. Tresser (1984): Une remarque sur las structure des endomorphismes de degré 1 du cercle. C.R. Acad. Sci. Paris, **299***, série I, 145–148*

T. Cherry (1938): Analytic quasi-periodic curves of discontinuous type on a torus. Proc. London Math. Soc. **44***, 175–215*

P. Collet (1984): Ergodic properties of some unimodal mappings of the interval. Preprint Institute Mittag-Leffler **11**

P. Collet (1988): Hausdorff dimension of the singularities for invariant measures of expanding dynamical systems. In: "Dynamical Systems Valpariso 1986", Lect. Notes in Math., vol. 1331. Springer, Berlin New York, pp. 47–58

P. Collet, J.P. Crutchfield and J.-P. Eckmann (1983): Computing the topological entropy of maps. Commun. Math. Phys. **88***, 257–262*

P. Collet and J.-P. Eckmann (1980): Iterated Maps of the Interval as Dynamical Systems. *Birkhauser, Boston*

P. Collet and J.-P. Eckmann (1983): Positive Liapunov exponents and absolute continuity for maps of the interval. Ergod. Th. & Dynam. Sys. **3***, 13–46*

P. Collet, J.-P. Eckmann and H. Koch (1980): Period doubling bifurcations for families of maps on $\mathbb{R}^n$. *J. Stat. Phys.* **25***, 1–15*

P. Collet, J.-P. Eckmann and O.E. Lanford (1980): Universal properties of maps on an interval. Commun. Math. Phys. **76***, 211–254*

P. Collet, J.L. Lebowitz and A. Porzio, The dimension spectrum of some dynamical systems. Preprint

G. Contreras (1991): On the C^2*-creation of links of critical points. Preprint PUC, Brazil*

I.P. Cornfield, S.V. Fomin and Ya. G. Sinai (1982): Ergodic Theory. *Grundlehren, vol. 245. Springer, Berlin New York*

P. Coullet and C. Tresser (1978) Itération d'endomorphismes et groupe de renormalisation, J. de Physique **C5***, 25*

P. Cvitanović (1989): Universality in Chaos. *Adam Hilger, Bristol*

P. Cvitanović and B. Söderberg (1988): Scaling laws for mode lockings in circle maps. Physica Scripta **32***, 263*

A.M. Davie (1992): private communication

A. Denjoy (1932): Sur les courbes definies par les equations differentielles a la surface du tore. J. Math. Pure et Appl. **11***, série 9, 333–375*

A. Denjoy (1975): Évocation de l'homme et de l'œuvre. Astérisque **28-29**

M. Denker and G. Keller (1986): Rigorous statistical procedures for data from dynamical systems. J. of Stat. Ph. **44***, 67–93*

M. Denker, G. Keller and M. Urbański, On the uniqueness of equilibrium states for piecewise monotone mappings. To appear in Studia Math.

B. Derrida, A. Gervois, and Y. Pomeau (1978): Iterations of endomorphisms of the real axis and representations of numbers. Ann. Inst. H. Poincaré, Physique Théorique **29***, 305*

B. Derrida, A. Gervois, and Y. Pomeau (1979): Universal metric properties of bifurcations of endomorphisms. J. Phys. **A 12***, 269–296*

R.L. Devaney (1986): An Introduction to Chaotic Dynamical Systems. *Addison Wesley, Menlo Park*

J. Dieudonné (1974): Éléments d'Analyse. **4**. *Gauthier-Villars, Paris*

A. Dold (1972): Lectures on Algebraic Topology. *Springer, Berlin New York*

A. Douady (1984): Systemes dynamiques holomorphes. Seminaire Bourbaki, **599**

A. Douady and J.H. Hubbard (1984): Étude dynamique des polynômes complexes. Publ. Math. d'Orsay 84-82

A. Douady and J.H. Hubbard (1985a): On the dynamics of polynomial-like mappings. Ann. Sc. E.N.S. 4^e *série,* **18***, 287–343*

A. Douady and J.H. Hubbard (1985b): Proof of Thurston's topological characterization of rational functions. To appear in Acta Math.

D. McDuff (1981): C^1*-minimal subsets of the circle. Ann. Inst. Fourier, Grenoble,* **XXXI***, 151–155*

J.-P. Eckmann (1986): The Mechanism of Feigenbaum universality. In: "Proc. of the International Congress of Math. Berkeley", pp. 1263–1267

J.-P. Eckmann and H. Epstein (1986): On the existence of fixed points of the composition operator for circle maps. Commun. Math. Phys. **107***, 213–231*

J.-P. Eckmann and H. Epstein (1990): Bounds on the unstable eigenvalue for period doubling. Commun. Math. Phys. **128***, 427–435*

J.-P. Eckmann and P. Wittwer (1987): A complete proof of Feigenbaum conjectures. J. of Stat. Phys. **46***, 455–475*

H. Epstein (1986): New proofs of the existence of the Feigenbaum functions. Commun. Math. Phys. **106***, 395–426*

H. Epstein and J. Lascoux (1981): Analyticity properties of the Feigenbaum function. Commun. Math. Phys. **81***, 437–453*

K.J. Falconer (1985): The Geometry of Fractal Sets. *Cambridge Univ. Press*

E. de Faria (1992): Proof of universality for critical circles mappings. Thesis, Cuny, New York

A. Fathi, F. Laudenbach and V. Poenaru (1979): Travaux de Thurston sur les surfaces. Astérisque **66-67**

P. Fatou (1919/20): Sur les equations fonctionnelles. Bull. Soc. Math. France (1919), **47***, 161–271/ (1920),* **48***, 33–94/ (1920),* **48***, 208–314*

M.J. Feigenbaum (1978): Quantitative universality for a class of nonlinear transformations. J. Stat. Phys. **19**, *25–52*

M.J. Feigenbaum (1979): The universal metric properties of non-linear transformations, J. Stat. Phys. **21**, *669–706*

M.J. Feigenbaum (1988): Presentation functions and scaling function theory for circle maps. Nonlinearity **1**, *577–602*

A. Finzi (1950): Problème de la géneration d'une transformation d'une courbe fermée. Ann. Sci. Ec. Norm. Sup. 3[e] *série,* **67**, *243–305*

S.R. Foguel (1969): Ergodic Theory of Markov Processes. *Van Nostrand, Math Studies* **21**. *New York*

H. Furstenberg (1961): Strict ergodicity and transformations of the torus. Amer. J. of Math. **83**, *573–601*

R. Galeeva (1990): Kneading sequences of piecewise linear bimodal maps. Preprint

R. Galeeva and C. Tresser (1991): Piecewise linear discontinuous double coverings of the circle. To appear in Proc. A.M.S.

J.-M. Gambaudo (1987): Ordre, désordre, et frontière des systèmes Morse-Smale. Thèse Univ. de Nice

J.-M. Gambaudo and C. Tresser (1991): A monotonicity property in one dimensional dynamics. Preprint

J.-M. Gambaudo, O. Lanford III and C. Tresser (1984): Dynamique symbolique des rotations. C.R. Acad. Sci. Paris, **299**, *série I, 823–826*

J.-M. Gambaudo, S. van Strien and C. Tresser (1989): Hénon-like maps with strange attractors: there exist C^∞ Kupka-Smale diffeomorphisms on S^2 with neither sinks nor sources. Nonlinearity **2**, *287–304*

J.-M. Gambaudo, S. van Strien and C. Tresser (1990): Vers un ordre de Sarkovskii pour les plongements du disque preservant l'orientation. C.R. Acad. Sci. Paris, **310**, *série I, 291–294*

F.P. Gardiner (1987): Teichmüller Theory and Quadratic Differentials. *John Wiley & Sons, New York*

F.P. Gardiner (1992): Lacunary series as quadratic differentials. to appear in the "Proceedings of the Symposium in honor of Wilhelm Magnus", Polytechnic Institute of Brooklyn, June 1992.

F.P. Gardiner and D.P. Sullivan (1992): Symmetric structures on a closed curve. Am. J. of Math. **114**. *683–736*

C. Godbillon (1983): Dynamical Systems on Surfaces. *Universitext. Springer, Berlin New York*

P. Góra and A. Boyarsky (1991): Computing the topological entropy of general one-dimensional maps. Trans. Amer. Math. Soc., To appear

P. Góra and B. Schmitt (1989): Un example de transformation dilatante et C^1 par marceaux de l'intervalle, sans probabilité absolument continue invariante. Ergod. Th. & Dynam. Sys. **9**, *101–113*

J. Graczyk (1991a): Harmonic scaling for smooth families of diffeomorphisms of the circle. Nonlinearity **4**, *935–959*

J. Graczyk (1991b): Hölder continuous conjugation for circle diffeomorphisms. Preprint Warsaw

J. J. Graczyk (1991c): Dynamics of nondegenerate upper maps. Preprint Queen's University, Ontario

J. Graczyk (1992): Families of diffeomorphisms. Preprint

J. Graczyk and G. Świątek (1991): Critial circle maps near bifurcation. Preprint StonyBrook

J. Graczyk and G. Świątek (1992): Singular measures in circle dynamics. Preprint StonyBrook

J. Graczyk, G. Świątek, F.M. Tangerman and J.J.P. Veerman (1992): Scalings in circle maps III. Preprint StonyBrook

J. Guckenheimer (1976): A strange, strange attractor. In: "The Hopf Bifurcation and its Applications", J.E. Marsden and M. McCracken, Applied Math. Sciences **19**. *Springer, Berlin New York, pp. 368–381*

J. Guckenheimer (1977): On bifurcations of maps of the interval. Invent. Math. **39**, *165–178*

J. Guckenheimer (1979): Sensitive dependence on initial conditions for one dimensional maps. Commun. Math. Phys. **70**, *133–160*

J. Guckenheimer (1980): The growth of topological entropy for one dimensional maps. In: "Global Theory of Dynamical Systems, Northwestern 1979", Lect. Notes in Math., vol. 819. Springer, Berlin New York, pp. 216–223

J. Guckenheimer (1984): Renormalization of one dimensional mappings and strange attractors. Proc. Lefschetz Symp. (Mexico City, 1984), Contemp. Math. **58**. *A.M.S. (1987), Providence 143–160*

J. Guckenheimer (1987): Limits sets of S-unimodal maps with zero entropy. Commun. Math. Phys. **110**, *655–659.*

J. Guckenheimer and S. Johnson (1990): Distortion of S-unimodal maps. Ann. Math. **132**, *71–130*

J. Guckenheimer, J. Moser and S. Newhouse (1980): Bifurcation of Dynamical Systems. *Birkhauser, Boston*

H.W. Guggenheimer (1977): Differential Geometry. *Dover, New York*

C. Gutierrez (1986): Smoothing continuous flows one two-manifolds and recurrences. Ergod. Th. & Dynam. Sys. **6**, *17–44*

C. Gutierrez (1987): A counter-example to a C^2 closing lemma. Ergod. Th. & Dynam. Sys. **7**, *509–530*

C.R. Hall (1981): A C^∞ Denjoy counterexample. Ergod. Th. & Dynam. Sys. **1**, *261–272*

B.R. Henry (1973): Escape from the unit interval under the transformation $x \mapsto \lambda x(1-x)$. Proc. A.M.S. **41**, *146–150*

M. Herman (1977): Mésure de Lebesgue et nombre de rotation. In: "Proc. Symp. Geometry and Topology", Lect. Notes in Math., vol. 597. Springer, Berlin New York, pp. 271–293

M.R. Hermann (1979): Sur la conjugaison differentiable des difféomorphismes du cercle a des rotations. Publ. Math. I.H.E.S. **49**, *5–234*

M.R. Hermann (1976): Sophus Lie's 1884 differential invariant paper. In: "Lie Groups", vol. III, Math. Sci. Press, Brookline, MA

M.W. Hirsch and C.C. Pugh (1970): Stable manifolds and hyperbolic sets. In: "Global Analysis", Proc. Symp. Pure Math. **14**, *133–164*

K. Hockett and P. Holmes (1988): Bifurcations to badly ordered orbits in one parameter families of circle maps. Proc. A.M.S. **102**, *1031–1051*

K. Hockett and P. Holmes (1988): Bifurcations to rotating Cantor sets in maps of the circle. Nonlinearity **1**, *603–616*

F. Hofbauer (1979): On intrinsic ergodicity of piecewise monotonic transformations with positive entropy. Israel J. of Math. **34**, *213–237*

F. Hofbauer (1980): The topological entropy of the transformation $x \mapsto ax(1-x)$. *Monatsh. Math.* **90**, *117–141*

F. Hofbauer (1981): The structure of piecewise monotonic transformations. Ergod. Th. & Dynam. Sys. **1**, *159–178*

F. Hofbauer (1982): Kneading invariants and Markov diagrams. In: "Ergodic Theory and Related Topics, Proceedings", ed. H. Michel, Akademie-Verlag, Berlin, pp. 85–95

F. Hofbauer (1985): Periodic points for piecewise monotonic transformations. Ergod. Th. & Dynam. Sys. **5**, *237–256*

F. Hofbauer (1986): Piecewise invertible dynamical systems. Prob. Th. Rel. Fields **72**, *359–386*

F. Hofbauer (1987): Generic properties of invariant measures for simple piecewise monotonic transformations. Israel J. of Math. **59**, *64–80*

F. Hofbauer (1991a): Hausdorff dimension and pressure for piecewise monotonic maps of the interval. Preprint Wien

F. Hofbauer (1991b): Hausdorff and conformal measures for piecewise monotonic maps of the interval. Preprint Wien

F. Hofbauer and G. Keller (1982a): Equilibrium states for piecewise monotonic transformations. Ergod. Th. & Dynam. Sys. **2**, *23–43*

F. Hofbauer and G. Keller (1982b): Ergodic properties of invariant measures of piecewise monotonic transformations. Math. Z. **180**, *119–140*

F. Hofbauer and G. Keller (1984): Zeta-functions and transfer-operators for piecewise linear transformations. J. Reine und Angew. Math. **352**, *100–113*

F. Hofbauer and G. Keller (1990a): Quadratic maps without asymptotic measure. Commun. Math. Phys. **127**, *319–337*

F. Hofbauer and G. Keller (1990b): Some recent results about S-unimodal maps. Ann. Inst. Henri Poincaré, Physique Théorique **53**, *413–425*

F. Hofbauer and M. Urbański (1991): Fractal properties of invariant subsets for piecewise monotonic maps on the interval. Preprint Wien

F. Hofbauer and P. Raith (1989): Topological transitive subsets of piecewise monotonic maps, which contain no periodic points. Mh. Math. **107**, *217–239*

F. Hofbauer and P. Raith (1991): The Hausdorff dimension of an ergodic invariant measure for piecewise monotonic maps of the interval, Preprint Wien

P. Holmes and P. Whitley (1984): Bifurcations of one- and two-dimensional maps. Phil. Trans. R. Soc. **A 311**, *43–102*

S. Hurder (1991): Exceptional minimal sets of $C^{1+\alpha}$*-group actions on the circle. Ergod. Th. & Dynam. Sys.* **11**, *455–467*

M.C. Irwin (1974): On the smoothness of the composition map. Quart. J. Math. Oxford **23**, *113–133*

A.F. Ivanov (1992): An example of infinitely many sinks for smooth interval maps. Preprint

E.A. Jackson (1989/1990): Perspectives of Nonlinear Dynamics. *(Two volumes). Cambridge Univ. Press*

M.V. Jakobson (1971): On smooth mappings of the circle into itself. Math. USSR. Sb. **14**, *161–185*

M.V. Jakobson (1981): Absolutely continuous invariant measures for one-parameter families of one-dimensional maps. Commun. Math. Phys. **81**, *39–88*

M.V. Jakobson (1987): Families of one-dimensional maps and nearby diffeomorphisms. Proc of the International Congress of Mathematicians 1986, vol. 2, Berkeley, Cal. USA 1150–1160

M.V. Jakobson and G. Świątek (1991a): Metric properties of non-renormalizable S-unimodal maps, I. Induced expansion and invariant measures. Preprint I.H.E.S.

M.V. Jakobson and G. Świątek (1991b): Quasisymmetric conjugacies between unimodal maps. Preprint StonyBrook

Y. Jiang (1991): Dynamics of certain smooth one-dimensional mappings. Preprint StonyBrook

Y. Jiang, T. Morita and D. Sullivan (1992): Expanding direction of the period doubling operator. Commun. Math. Phys. **144**, *509-520.*

S.D. Johnson (1986): Continuous measures in one dimension. Preprint Cornell University

S.D. Johnson (1987): Singular measures without restrictive intervals. Commun. Math. Phys. **110**, *185–190*

S.D. Johnson (1991): Absorbing cantor sets and trapping structures. Ergod. Th. & Dynam. Sys. **11**, *731–736*

L.B. Jonker (1979): Periodic orbits and kneading invariants. Proc. London Math. Soc. **39**, *428–450*

L.B. Jonker (1990): Scaling in Arnol'd tongues for differentiable homeomorphisms of the circle. Commun. Math. Phys. **129**, *1–25*

L.B. Jonker and D.A. Rand (1981): Bifurcation in one-dimension. I. The non-wandering set. Invent. Math. **62**, *347–365; II. A versal model for bifurcations. Invent. Math.* **63**, *1–16*

G. Julia (1918): Memoire sur l'iteration des fonctions rationnelles. J. Math. **8**, *47–245*

M. Kac (1946): On the distribution of values of sums of the type $\sum f(2^k t)$. Ann. Math. **47**, *33–49*

Y. Katznelson (1977): Sigma-finite invariant measures for smooth maps of the circle. J. d'Analyse Math. **31**, *1–18*

Y. Katznelson (1979): The action of diffeomorphisms of the circle on the Lebesgue measure. J. d'Analyse Math. **36**, *156–166*

Y. Katznelson and D. Ornstein (1989a): The differentiability of conjugation of certain diffeomorphisms of the circle. Ergod. Th. & Dynam. Sys. **9**, *643–680*

Y. Katznelson and D. Ornstein (1989b): The absolute continuity of the conjugation of certain diffeomorphisms of the circle. Ergod. Th. & Dynam. Sys. **9**, *681–690*

M. Keane (1977): Non-ergodic interval exchange transformations. Israel J. of Math. **26**, *188–196*

G. Keller (1982): Stochastic stability in some chaotic dynamical systems. Monatsh. Math. **94**, *313–33*

G. Keller (1984): On the rate of convergence to equilibrium in one–dimensional systems. Commun. Math. Phys. **96**, *181-193*

G. Keller (1985): Generalized bounded variation and applications to piecewise monotonic transformations. Zeitschr. Wahrsch. verw. Gebiete **69**, *461–479*

G. Keller (1987): Invariant measures and Lyapounov exponents for S-unimodal maps. Preprint Maryland

G. Keller (1989a): Lifting measures to Markov extensions. Monatsh. Math. **108**, *183–200*

G. Keller (1989b): Markov extensions, zeta functions, and Fredholm theory for piecewise invertible dynamical systems. Trans. A.M.S. **314**, *433–497*

G. Keller (1990a): Exponents, attractors, and Hopf decompositions for interval maps. Ergod. Th. & Dynam. Sys. **10**, *717–744*

G. Keller (1990b): On the distribution of periodic orbits for interval maps. In: "Stochastic Modelling in Biology", ed. P. Tautu, World Scientific, pp. 412–419

G. Keller (1992): Exponential weak Bernoulli mixing for Collet-Eckmann maps. Preprint Erlangen

G. Keller and T. Nowicki (1991): Spectral theory, zeta functions and the distributions of periodic points for Collet-Eckmann maps. To appear in Commun. Math. Phys.

G. Keller and T. Nowicki (1992): Fibonacci maps re(al)visited. Preprint Erlangen

H. Keynes and D. Newton (1976): A minimal non-uniquely ergodic interval exchange transformation. Math. Z. **148**, *101–105*

K.M. Khanin (1990): Thermodynamic formalism for critical circle mappings. In: "Chaos", ed. D.K. Campbell, American Institute of Physics, New York, pp. 71–90

K.M. Khanin (1991): Universal estimates for critical circle mappings. Chaos, Amer. Inst. of Physics., 181–186

K.M. Khanin and Ya.G. Sinai (1987): A new proof of M. Herman's theorem. Commun. Math. Phys. **112**, *89–101*

S.F. Koljada (1989): One-parameter families represented by integrals with negative Schwarzian derivative violating monotone bifurcations, Ukr. Mat. Z. **41**, *258–261 (in Russian)*

A. Kondah and T. Nowicki (1990): Conditions d'existence d'une probabilité invariante absolument continue pour des endomorphismes de l'intervalle. C.R. Acad. Sci. Paris, **311**, *série I, 641–644*

K. Krzyzewski (1978): Some results on expanding maps. In: "Dynamical Systems Warsaw", Astérisque **51**, *205–218*

O.E. Lanford III (1982): A computer-assisted proof of the Feigenbaum conjecture. Bull. A.M.S. **6**, *427–434*

O.E. Lanford III (1984a): A shorter proof of the existence of Feigenbaum fixed point. Commun. Math. Phys. **96**, *521–538*

O.E. Lanford III (1984b): Functional equations for circle homeomorphisms with golden rotation number. J. Statist. Phys. **34**, *57–73*

O.E. Lanford III (1986): Computer-assisted proofs in analysis. Proc. of the international congress of math. Berkeley, Cal. USA 1385–1394

O.E. Lanford III (1987): Renormalization group methods for circle mappings. In: "Proc. of the 1985 Groningen Conf. on Statistical Mechanics and Field Theory". Proc. VIIIth Int. Congr. on Mathematical Physics, World Scientific, Singapore.

O.E. Lanford III (1991): Critical circle mappings. To appear

O.E. Lanford III (1992): private communication

A. Lasota and J.A. Yorke (1973): On the existence of invariant measures for piecewise monotonic transformations. Trans. A.M.S. **186**, *481–488*

P. Lavaurs (1992): Le lieu de connexitè des polynômes du troisième degré n'est pas localement connexe. Preprint

F. Ledrappier (1981): Some properties of absolutely continuous invariant measures of an interval. Ergod. Th. & Dynam. Sys. **1**, *77–93*

F. Ledrappier and M. Misiurewicz (1985): Dimension of invariant measures for maps with exponent zero. Ergod. Th. & Dynam. Sys. **5**, *595–610*

O. Lehto (1987): Univalent Functions and Teichmüller Spaces. *Springer, Berlin New York*

O. Lehto and K.I. Virtanen (1973): Quasiconformal Mappings in the Plane. *Springer, Berlin New York*

M. Levi (1981): Qualitative analysis of the periodically forced relaxation oscillations. Memoirs of the A.M.S. **244**

G. Levitt (1983): Feuilletages des surfaces. Thèse de Doctorat, Université de Paris VII

T. Li and J.A. Yorke (1975): Period three implies chaos. Amer. Math. Monthly, **82**, *985–992*

T. Li and J.A. Yorke (1978): Ergodic transformations from an interval into itself. Trans. A.M.S. **235**, *183–192*

M.Yu. Lyubich (1982): On typical behavior of the trajectory of a rational mapping of the sphere. Dokl. Acad. Nauk SSSR (1982), **268**, *29–32. Transl in English: Soviet Math. Dokl. (1983),* **27**, *22–25*

M.Yu. Lyubich (1986): The dynamics of rational transforms: The topological picture. Russian Math. Surv. **41**, *43–117*

M.Yu. Lyubich (1989): Non-existence of wandering intervals and structure of topological attractors of one dimensional dynamical systems. I. The case of negative Schwarzian derivative. Ergod. Th. & Dynam. Sys. **9**, *737–750*

M.Yu. Lyubich (1990): Ergodic theory for smooth one-dimensional dynamical systems. Preprint StonyBrook

M.Yu. Lyubich (1991): On the Lebesgue measure of the Julia set of a quadratic polynomial. Preprint StonyBrook

M.Yu. Lyubich (1992a): Combinatorics, geometry and attractors of quasi-quadratic maps. Preprint StonyBrook

M.Yu. Lyubich (1992b): Milnor's attractors, persistent recurrence and renormalization. To appear in "Topological Methods in Modern Mathematics, A Symposium in Honor of John Milnor's 60th Birthday".

M.Yu. Lyubich and J. Milnor (1991): The Fibonacci unimodal map. Preprint StonyBrook

R.S. MacKay (1982): Renormalization in area preserving maps. Thesis, Princeton (Univ. Microfilms Int., Ann Arbor, Michigan)

R.S. MacKay (1987): Introduction to the dynamics of area-preserving maps. In: "Proceeding of the U.S. Particle Accelerator School, SLAC, 1985", A.I.P. Conf. Proc. **153**, *534–602*

R.S. MacKay (1988): A simple proof of Denjoy's theorem. Math. Proc. Camb. Philos. Soc. **103**, *299–303*

R.S. MacKay and C. Tresser (1986): Transition to topological chaos for circle maps. Physica D **19**, *206–237. Errata Physica D* **29**, *427*

R.S. MacKay and C. Tresser (1987): Some flesh on the skeleton: the bifurcations of bimodal maps. Physica D **27**, *412–422.*

R.S. MacKay and C. Tresser (1988): Boundary of chaos for bimodal maps of the interval. J. London Math. Soc. **37**, *164–181*

R.S. MacKay and J.B.J. Zeijts (1988): Period doubling for bimodal maps: a horseshoe for a renormalisation operator. Nonlinearity **1**, *253–277*

M.I. Malkin (1989): On continuity of entropy of discontinuous mappings of the interval. Selecta Math. Sovietica **8**, *131–139*

R. Mañé (1985/7): Hyperbolicity, sinks and measure in one dimensional dynamics. Commun. Math. Phys. **100** *(1985), 495–524 and Erratum. Commun. Math. Phys.* **112***, (1987), 721–724*

R. Mañé (1987): Ergodic Theory and Differentiable Dynamics. *Ergebnisse der Mathematik und ihrer Grenzgebiete, vol. 8. Springer, Berlin New York*

R. Mañe, P. Sad and D. Sullivan (1983): On the dynamics of rational maps. Ann. Sci. Ec. Norm. Sup. **16***, 193–217*

M. Martens (1990): Interval dynamics. Thesis, Delft

M. Martens (1992a): The existence of σ-finite invariant measures Applications to interval dynamics. Preprint StonyBrook

M. Martens (1992b): Cantor attractors of unimodal maps. In preparation

M. Martens (1992c): Distortion results and the ergodic classification of unimodal maps. Preprint StonyBrook

M. Martens and W. de Melo (1992): Finiteness of periodic attractors for Lorenz maps. In preparation

M. Martens, W. de Melo, P. Mendes and S. van Strien (1990): On Cherry flows. Ergod. Th. & Dynam. Sys. **10***, 531–554*

M. Martens, W. de Melo and S. van Strien (1992): Julia-Fatou-Sullivan theory for real one-dimensional dynamics. Acta Math. **168***, 273–318*

O. Martio and J. Sarvas (1978/9): Injectivity theorems in plane and space. Ann. Acad. Sci. Fenn. A I Math. **4***, 384–401*

W.S. Massey (1991): A Basic Course in Algebraic Topology*. Grad. Texts in Math.* **127***, Springer, Berlin New York*

R.M. May (1976): Simple mathematical models with very complicated dynamics. Nature **261***, 459–466*

C. McMullen (1991): Rational maps and Kleinian groups. In: "Proceedings of the International Congress of Mathematicians", Springer, Berlin New York, pp. 889–899

P.D. McSwiggen (1992): Diffeomorphisms of the torus with wandering domains. Preprint Cincinatti

W. de Melo (1987): A finiteness problem for one dimensional maps. Proc. A.M.S. **101***, 721–727*

W. de Melo and C. Pugh (1991): On Brunovsky's hypothesis. Preprint Berkeley

W. de Melo and S. van Strien (1988): One-dimensional dynamics: the Schwarzian derivative and beyond. Bull. A.M.S. **18***, 159–162*

W. de Melo and S. van Strien (1989): A structure theorem in one-dimensional dynamics. Ann. Math. **129***, 519–546*

P. Mendes (1991): A metric property of Cherry vector fields on the torus. J. Diff. Eq. **89***, 305–316*

M. Metropolis, M.L. Stein and P.R. Stein (1973): On finite limit sets for transformations of the unit interval. J. Combin. Theory **15***, 25–44*

J. Milnor (1983): The monotonicity theorem for real quadratic maps. Mathematische Arbeitstagung Bonn 1983

J. Milnor (1985): On the concept of attractor. Commun. Math. Phys. **99***, 177–195*

J. Milnor (1990): Remarks on iterates cubic maps. Preprint StonyBrook

J. Milnor (1992): Local connectivity of Julia sets. Preprint StonyBrook

J. Milnor and W. Thurston (1977): On iterated maps of the interval: I,II. Preprint Princeton. Published in: "Dynamical Systems: Proc. Univ, of Maryland 1986-87", (1988), Lect. Notes in Math., vol. 1342. Springer, Berlin New York, pp. 465–563

C. Mira (1986): Chaotic Dynamics. *World Scientific Publishing, Co. Pte. Ltd, Singapore*

M. Misiurewicz (1979): Horseshoes for mappings of the interval. Boll. Acad. Pol. Sci., Sér. Sci. Math. **27**, *167–169*

M. Misiurewicz (1981): Absolutely continuous measures for certain maps of an interval. Publ. Math. I.H.E.S. **53**, *17–51*

M. Misiurewicz (1981): Structure of mappings of an interval with zero entropy. Publ. Math. I.H.E.S. **53**, *5–16*

M. Misiurewicz (1982): Attracting Cantor set of positive measure for a C^∞ *map of an interval. Ergod. Th. & Dynam. Sys.* **2**, *405–415*

M. Misiurewicz (1986): Rotation interval for a class of maps of the real line into itself. Ergod. Th. & Dynam. Sys. **6**, *117–132*

M. Misiurewicz (1989): Jumps of entropy in one dimension. Fund. Math. **132**, *215–226*

M. Misiurewicz and Z. Nitecki (1991): Combinatorial patterns for maps of the interval. *Memoirs of A.M.S. vol. 94.*

M. Misiurewicz and S.V. Shlyachkov (1989): Entropy of piecewise continuous intervals maps. Preprint Warsaw

M. Misiurewicz and W. Szlenk (1980): Entropy of piecewise monotone mappings. Studia Math. **67**, *45–63*

C.C. Moore and C. Schochet (1988): Global analysis on foliated spaces. Mathematical Sciences Research Institute Publications **9**, *Springer, Berlin New York*

L. Mora and M. Viana (1989): Abundance of strange attractors. Preprint IMPA

P.C. Moreira and A.A. Gaspar Ruas (1992): Metric properties of Cherry flows. J. Diff. Eq. **97**, *16–26*

J. Moser (1966): A rapidly convergent iteration method and non-linear differential equations, part II. Ann. Scuola Norm.Sup. di Pisa **20**, *499–535*

J. Moser (1990): On commuting circle mappings and simultaneous Diophantine approximations. Math. Z. **205**, *105–121*

P.J. Myrberg (1959/63): Iteration der reellen Polynome zweiten Grades. Ann. Acad. Sci. Fenn. (1959), **256 A**, *1–10 and (1963),* **336 A**, *1–18*

I.P. Natanson (1955): Theory of Functions of a Real Variable. *Frederick Ungar Publish. Co.*

J. von Neumann and S.M. Ulam (1947): On combination of stochastic and deterministic processes. Bull. A.M.S. **33**, *1120*

S. Newhouse (1991): On some results of Hofbauer on maps of the interval. In: "Dynamical Systems and Related Topics", Advances Series in Dynamical Systems, World Scientific, Singapore, pp. 407–421

S. Newhouse, J. Palis and F. Takens (1983): Bifurcations and stability of families of diffeomorphisms. Publ. I.H.E.S. **57**, *5–71*

Z. Nitecki (1970): Nonsingular endomorphisms of the circle. In: "Global Analysis", Proc. Symp. in Pure Math. **14** *, 203–220*

Z. Nitecki (1982): Topological dynamics on the interval. Ergod. Th. & Dynam. Sys. **II**, *Proc. of the Special Year, Maryland, 1979-1980, Progress in Math. Birkhäuser, Boston 1–73*

A. Norton and J. Velling (1991): Denjoy-like theorems for diffeomorphisms of the 2-torus. Preprint

T. Nowicki (1985a): On some dynamical properties of S-unimodal maps on an interval. Fund. Math. **126**, *27–43*

T. Nowicki (1985b): Symmetric S-unimodal mappings and positive Liapunov exponents. Ergod. Th. & Dynam. Sys. **5**, *611–616*

T. Nowicki (1988): A positive Liapunov exponent for the critical value of an S-unimodal mapping implies uniform hyperbolicity. Ergod. Th. & Dynam. Sys. **8**, *425–435*

T. Nowicki (1991): Some dynamical properties of S-unimodal maps. To appear in Fund. Math.

T. Nowicki and F. Przytycki (1989): The conjugacy of Collet-Eckmann's map of the interval with the tent map is Hölder continuous. Ergod. Th. & Dynam. Sys. **9**, *377–386*

T. Nowicki and S. van Strien (1988): Absolutely continuous invariant measures for C^2-unimodal maps satisfying the Collet-Eckmann condition. Invent. Math. **93**, *619–635*

T. Nowicki and S. van Strien (1990a): Hyperbolicity properties of C^2 multi-modal Collet-Eckmann maps without Schwarzian derivative conditions. Trans. A.M.S. **321**, *793–810*

T. Nowicki and S. van Strien (1990b): Non-hyperbolicity and invariant measures for unimodal maps. Ann. Inst. Henri Poincaré, Physique Théorique, **53**, *427–429*

T. Nowicki and S. van Strien (1991a): Absolutely continuous invariant measures under a summability condition. Invent. Math. **105**, *123-136.*

T. Nowicki and S. van Strien (1991b): Unpublished, see Keller and Nowicki (1992).

H.E. Nusse (1983): Chaos, yet no chance to get lost. Thesis, Utrecht

H.E. Nusse (1987): Chaotic maps with rational zeta functions. Trans. A.M.S. **304**, *705–719*

H.E. Nusse (1988): Qualitative analysis of the dynamics and stability properties for Axiom A maps. J. of Math. Analysis and Appl. **136**, *74–106*

H.E. Nusse and J.A. Yorke (1988a): Period halving for $x_{n+1} = M \cdot F(x_n)$ where F has negative Schwarzian derivative. Physics Letters A **127**, *328–334*

H.E. Nusse and J.A. Yorke (1988b): Is every approximate trajectory of some process near an exact trajectory of a nearby process. Commun. Math. Phys. **114**, *363–379*

S. Ostlund, D. Rand, J. Sethna and E. Siggia (1983): Universal properties of the transition from quasi-periodicity to chaos in dissipative systems, Physica D **8**, *303–342*

J. Palis, W. de Melo (1982): Geometric Theory of Dynamical Systems, an introduction. *Springer, Berlin New York*

W. Pałuba (1989): The Lipschitz condition for the conjugacies of Feigenbaum-like mappings. Fund. Math. **132**, *227–258*

W. Parry (1966): Symbolic dynamics and transformations of the unit interval. Trans. A.M.S. **122**, *368–378*

K. Petersen (1983): Ergodic Theory. *Cambridge Univ. Press*

G. Pianigiani (1980): First return map and invariant measures. Israel J. of Math. **35**, *32–48*

G. Pianigiani (1979): Absolutely continuous invariant measures on the interval for the process $x_{n+1} = Ax_n(1 - x_n)$. Boll. Un. Mat. Ital. **16**, *364–378*

A.A. Pinto and D.A. Rand (1992): Global phase universality, smooth conjugacies and renormalization: 2. The $C^{k+\alpha}$ case using rapid convergence of Markov families. Nonlinearity **4**, *49–79*

H. Poincaré (1881/6): Mémoire sur les courbes définies par une équation différrentielle I, II, III, IV. J. Math. Pures et Appl. (1881), **7***, 375–422/ (1882),* **8***, 251–286/ (1885),* **1***, 167–244/ (1886),* **2***, 151–217*

C. Preston (1983): Iterates of Maps on an Interval*. Lect. Notes in Math., vol. 999. Springer, Berlin New York*

C. Preston (1988): Iterates of Piecewise Monotone Mappings of an Interval*. Lect. Notes in Math., vol. 1347. Springer, Berlin New York*

F. Przytycki (1991): Lyapounov characteristic exponents are non-negative. Preprint

P. Raith (1989): Hausdorff dimension for piecewise monotonic maps. Studia Math. **64***, 17–33*

P. Raith (1991): Continuity of the Hausdorff dimension for piecewise monotonic maps. Preprint Wien

D.A. Rand (1987): Universality and renormalisation in dynamical systems. In: "New Directions in Dynamical Systems", Cambridge Univ. Press, pp. 1–56

D.A. Rand (1988): Global phase space universality, smooth conjugacies and renormalisation I. The $C^{1+\alpha}$ case. Nonlinearity **1***, 181–202*

D.A. Rand (1989): The singularity spectrum $f(\alpha)$ for cookie-cutters. Ergod. Th. & Dynam. Sys. **9***, 527–541*

D.A. Rand (1992): Existence, non-existence and universal breakdown of dissipative golden invariant tori, Nonlinearity **5***, 639–706*

M. Rees (1984): Ergodic rational maps with dense critical point forward orbit. Ergod. Th. & Dynam. Sys. **4***, 311–322*

M. Rees (1986): Positive measure sets of ergodic rational maps. Ann. Sci. Ec. Norm. Sup. 4^e *série* **19***, 383–407*

A. Renyi (1957): Representation of real numbers and their ergodic properties. Acta. Math. Akad. Sc. Hungar. **8***, 477–493*

V.A. Rohlin (1962): On the fundamental ideas of measure theory. A.M.S. Transl. Series 1, **10***, 1–54.*

V.A. Rohlin (1966): Selected topics from the metric theory of dynamical systems. A.M.S. Transl. Series 2, **49***, 171–240*

A.C.M. van Rooij and W.H. Schikhof (1982): A Second Course of Real Functions*. Cambridge Univ. Press*

W. Rudin (1966): Real and Complex Analysis*. McGraw-Hill, New York*

D. Ruelle (1977): Applications conservant une mesure absolument continue par rapport à dx sur $[0, 1]$*. Commun. Math. Phys.* **55***, 47–52*

D. Ruelle (1979): Ergodic theory of differentiable dynamical systems. Publ. Math. I.H.E.S. **50***, 27–58*

D. Ruelle (1978): Dynamical systems with turbulent behavior. In: "Mathematical Problems in Theoretical Physics", Lect. Notes in Physics **80***. Springer, Berlin New York, pp. 341–360*

D. Ruelle (1979): Sensitive dependence on initial conditions and turbulent behavior of dynamical systems. In: "Bifurcation Theory and its Applications in Scientific Disciplines". New York Acad. of Sci. **317**

D. Ruelle (1991): Spectral properties of a class of operators associated with maps in one dimension. Ergod. Th. & Dynam. Sys. **11***, 757–768*

M.R. Rychlik (1983): Bounded variation and invariant measures. Studia Math. **76***, 69–80*

M.R. Rychlik (1988): Another proof of Jakobson's theorem and related results. Ergod. Th. & Dynam. Sys. **8***, 93–109*

R. Sacksteder (1965): Foliations and pseudogroups. Amer. J. Math. **87**, *79-102*

A. Sataev (1975): On the number of invariant measures for flows on orientable surfaces. Math. USSR Izv. **9**, *813–830*

A. Schwartz (1963): A generalization of a Poincaré-Bendixson theorem to closed two-dimensional manifolds. Amer. J. Math. **85**, *453–458*

A. Sarkovskii (1960): Necessary and sufficient conditions for convergence of one-dimensional iteration processes. Ukrain. Math. Zh. **12**, *484–489*

A. Sarkovskii (1961): On the reducibility of a continuous function of a real variable and the structure of the stationary points of the corresponding iteration process. Soviet Math. Dokl. **2**, *1062–1064*

A. Sarkovskii (1964): Coexistence of cycles of a continuous map of the line into itself. Ukrain. Math. Zh. **16**, *61–71*

A.N. Sarkovskii and A.F. Ivanov (1983): C^∞*-mappings of an interval with attracting cycles of arbitrary large periods. Ukr. Math Z.* **4**, *537–539*

A.N. Sarkovskii, Kolyada, Siwak, Federenko (1989): Dynamics of One-Dimensional Maps. *Kiev Naukova Dumka (in Russian)*

A.N. Sarkovskii, I.L. Maistrenko and E.I. Romanenko (1986): Difference Equations and Their Applications. *Kiev Naukova Dumka (in Russian)*

H.G. Schuster (1989): Deterministic Chaos, an Introduction. *VCH Publishers, Weinheim*

C. Series (1983): The geometry of Markoff numbers. The Mathematical Intelligencer **7**, *20–29*

M. Shishikura (1987): On the quasiconformal surgery of rational functions. Ann. Sci. Ec. Norm. Sup. 4^e *série,* **20**, *1–29*

M. Shub (1969): Endomorphisms of compact differentiable manifolds. Amer. J. Math. **91**, *175–199*

M. Shub and D. Sullivan (1985): Expanding endomorphisms of the circle revisited. Ergod. Th. & Dynam. Sys. **5**, *285–289*

Ya.G. Sinai and K.M. Khanin (1989): Smoothness of conjugacies of diffeomorphisms of the circle with rotations. Russian Math. Surv. **44**, *69–99*

D. Singer (1978): Stable orbits and bifurcations of maps of the interval. SIAM J.-Appl.Math. **35**, *260–267*

S. Smale (1967): Differentiable dynamical systems. Bull. A.M.S. **73**, *747–817*

J. Stark (1988): Smooth conjugacy and renormalisation for diffeomorphisms of the circle. Nonlinearity **1**, *541–575*

P. Stefan (1977): A theorem of Sarkovskii on the existence of periodic orbits of continuous endomorphisms on the real line. Commun. Math. Phys. **54**, *237–248*

A. Stirnemann (1992a): Renormalization for golden circles. Preprint ETH, Zürich

A. Stirnemann (1992b): A renormalization proof of Siegel's theorem. Preprint Exeter

E. Straube (1981): On the existence of invariant absolutely continuous measures. Commun. Math. Phys. **81**, *27–30*

S. van Strien (1981): On the bifurcations creating horseshoes. Lect. Notes in Math., vol. 898. Springer, Berlin New York, pp. 316–351

S. van Strien (1987): Smooth dynamics on the interval. In: "New Directions in Dynamical Systems", Cambridge Univ. Press, pp. 57–119

S. van Strien (1990): Hyperbolicity and invariant measures for general C^2 *interval maps satisfying the Misiurewicz condition. Commun. Math. Phys.* **128**, *437–496*

S. van Strien (1991): Interval maps. In: "Structures in Dynamics", North-Holland, Amsterdam, pp. 111–160

D. Sullivan (1980): Discrete conformal groups and measurable dynamics. In: "The Mathematical Heritage of Henri Poincaré". Proc. of Symp. in Pure Math. **39**, *Part 1, A.M.S. Providence, pp. 169–185*

D. Sullivan (1983): Conformal dynamical systems. In: "Proc. Int. Conf. Dyn. Sys., Rio, IMPA, 1981", Lect. Notes in Math., vol. 1007. Springer, Berlin New York, pp. 725–752

D. Sullivan (1985): Quasiconformal homeomorphisms and dynamics I: a solution of Fatou-Julia problem on wandering domains. Ann. Math. (1985), **122**, *401–418, and II: Structural stability implies hyperbolicity for Kleinian groups. Acta Math. (1985),* **155**, *243–260*

D. Sullivan (1986): Quasiconformal homeomorphisms in dynamics, topology, and geometry. Proc. International Congress of Mathematicians. Amer. Math. Soc., pp. 1216–1228

D. Sullivan (1988): Differentiable structures on fractal like sets, determined by intrinsic scaling functions on dual Cantor sets. In: "Nonlinear Evolution and Chaotic Phenomena" Plenum, New York

D. Sullivan (1989): Bounded structure of infinitely renormalizable mappings. In: "Universality and Chaos", ed. P. Cvitanović, 2nd edition, Adam Hilger, Bristol

D. D. Sullivan (1991): The universalities of Milnor, Feigenbaum and Bers. To appear in "Topological Methods in Modern Mathematics", Proceedings of the Symposium held in honor of John Milnor's 60th birthday, SUNY at Stony Brook, June 14-21, 1991

D. Sullivan (1992): Bounds, quadratic differentials, and renormalization conjectures. In: "A.M.S. Centennial Publications, vol. 2, Mathematics into the Twenty-first Century (1988)".

D. Sullivan and W.P. Thurston (1986): Extending holomorphic motions. Acta Math. **157**, *243–257*

G. Świątek (1988): Rational rotation numbers for maps of the circle. Commun. Math. Phys. **119**, *109–128*

G. Świątek (1989): Endpoints of rotation intervals for maps of the circle. Ergod. Th. & Dynam. Sys. **9**, *173–190*

G. Świątek (1992a): One-dimensional maps and Poincaré metric. Nonlinearity. **5**, *81-108*

G. Świątek (1992b): Hyperbolicity is dense in the real quadratic family. Preprint StonyBrook

B. Szewc (1984): The Perron-Frobenius operator in spaces of smooth functions on an interval. Ergod. Th. & Dynam. Sys. **4**, *613–643*

W. Szlenk (1979): Some dynamical properties of certain differentiable mappings. Bol. de la Soc. Mat. Mex. **24**, *57–87*

F. Tangerman and F. Przytycki (1992): Cantor sets in the line: scaling functions and the smoothness of the shift map. Preprint StonyBrook

F.M. Tangerman and J.J.P. Veerman (1991): Scalings in circle maps II. Commun. Math. Phys. **141**, *279-291*

M. Thaler (1983): Transformations on $[0,1]$ *with infinite invariant measures. Israel J. Math.* **46**, *67–96*

W. Thurston (1984): The combinatorics of iterated rational maps. Preprint Princeton

M. Tsujii (1992a): On positive Lyapounov exponent and weak regularity of critical values of one-dimensional mappings I. Preprint Kyoto

M. Tsujii (1992b): A proof of Benedicks-Carleson-Jacobson theorem for the quadratic family. Preprint Kyoto

M. Tsujii (1992c): Positive Lyapounov exponents in families of one dimensional dynamical systems. Preprint Kyoto

M. Tsujii (1992d): Weak regularity of Lyapounov exponents in one-dimensional dynamics. Preprint Kyoto

M. Tsujii (1992e): Rotation number and one-parameter families of circle diffeomorphisms. Ergod. Th. & Dynam. Sys. **12**, *359–363*

M. Urbański (1987): Invariant subsets of expanding mappings of the circle. Ergod. Th. & Dynam. Sys. **7**, *627–645*

E. Vargas (1991): Bifurcation frequency of unimodal maps. Commun. Math. Phys. **141**, *633–650*

E. Vargas (1991): Markov partition in non-hyperbolic interval dynamics. Commun. Math. Phys. **138**, *521–535*

J.J.P. Veerman (1989): Irrational rotation numbers. Nonlinearity **2**, *419–428*

J.J.P. Veerman and F.M. Tangerman (1990a): Scalings in circle maps I. Commun. Math. Phys. **134**, *89-107*

J.J.P. Veerman and F.M. Tangerman (1990b): A remark on Herman's theorem for circle diffeomorphisms. Preprint StonyBrook

E.B. Vul, Ya.G. Sinai, K.M. Khanin (1984): Feigenbaum universality and the thermodynamic formalism. Russ. Math. Surv. **39**, *1–40*

P. Walters (1982): An Introduction to Ergodic Theory. *Graduate Texts in Math., vol. 79. Springer, Berlin New York*

J. Willms (1988): Asymptotic behaviour of iterated piecewise monotonic maps. Ergod. Th. & Dynam. Sys. **8**, *111-131*

J.-C. Yoccoz (1984a): Conjugaison differentiable des difféomorphismes du cercle dont le nombre de rotation vérifie une condition Diophantienne. Ann. Sci. Ec. Norm. Sup. **17**, *333–361*

J.-C. Yoccoz (1984b): Il n'y a pas de contre-example de Denjoy analytique. C.R. Acad. Sci. Paris, **298**, *série I, 141–144*

J.-C. Yoccoz (1991): On the local connectivity of the Mandelbrot set. To appear

L.S. Young (1979): A closing lemma on the interval. Invent. Math. **54**, *179–187*

L.S. Young (1981): On the prevalence of horseshoes. Trans. A.M.S. **263**, *75–88*

L.S. Young (1991): Decay on correlations for certain quadratic maps. Preprint

A. Zdunik (1984): Entropy of transformations of the unit interval. Fund. Math. **124**, *235–241*

K. Ziemian (1985): Almost sure invariance principle for some maps of an interval. Ergod. Th. & Dynam. Sys. **5**, *625–640*

K. Ziemian (1989): Refinement of Shannon-McMillan-Breimann theorem for some maps of an interval. Studia Math. **83**, *271–285*

Index