

# RIGIDITY OF CRITICAL CIRCLE MAPPINGS I

EDSON DE FARIA AND WELINGTON DE MELO

November, 1997

ABSTRACT. We prove that two  $C^r$  critical circle maps with the same rotation number of *bounded type* are  $C^{1+\alpha}$  conjugate for some  $\alpha > 0$  provided their successive renormalizations converge together at an exponential rate in the  $C^0$  sense. The number  $\alpha$  depends only on the rate of convergence. We also give examples of  $C^\infty$  critical circle maps with the same rotation number that are not  $C^{1+\beta}$  conjugate for any  $\beta > 0$ .

## 1. INTRODUCTION

The purpose of this paper is to study certain rigidity questions concerning critical circle mappings. This study is continued in [4].

In the qualitative theory of smooth dynamical systems, the notions of *rigidity* and *flexibility* play an important role. The smooth systems are usually classified according to the equivalence relation given by topological conjugacies: two smooth maps  $f$  and  $g$  are topologically equivalent if there exists a homeomorphism  $h$  of the ambient space such that  $h \circ f = g \circ h$ . Such a homeomorphism maps orbits of  $f$  onto orbits of  $g$ . One can also consider a stronger equivalence relation given by smooth conjugacies. This leads to a quantitative or geometric classification of smooth dynamical systems, since a smooth conjugacy, being essentially affine at small scales, preserves the small-scale geometric properties of the dynamics. Hence each topological equivalence class is “foliated” by the smooth conjugacy classes and the quotient space is the *moduli* or *deformation space* of the dynamics. The moduli space describes the *flexibility* of the dynamics. When this space reduces to a single point, we are in the presence of *rigidity*.

In general, since eigenvalues at the periodic points are smooth conjugacy invariants, we can hope to find rigidity only in the absence of periodic points. From this viewpoint, the simplest case to consider is that of circle diffeomorphisms. If  $f$  is a circle diffeomorphism without periodic points then  $f$  is combinatorially equivalent to a rigid rotation  $R_\rho : x \mapsto x + \rho \pmod{1}$ , in the sense that for each  $N$ , the first

---

1991 *Mathematics Subject Classification*. Primary 58F03; Secondary 58F23.

*Key words and phrases*. Commuting pairs, real a-priori bounds, rigidity, renormalization. This work has been partially supported by the Pronex Project on Dynamical Systems, by FAPESP Grant 95/3187-4 and by CNPq Grant 30.1244/86-3.

$N$  elements of an orbit of  $f$  are ordered in the circle in the same way as the first  $N$  elements of an orbit of  $R_\rho$ . From Denjoy's theorem it follows that if  $f$  is at least  $C^2$  (or  $C^1$  and its derivative has bounded variation) then  $f$  is topologically conjugate to  $R_\rho$ . By a fundamental result of Herman [10], improved by Yoccoz [19], we have that if the rotation number  $\rho$  satisfies a Diophantine condition such as

$$\left| \rho - \frac{p}{q} \right| \geq \frac{C}{q^{2+\beta}} ,$$

for all rationals  $p/q$ , with  $C > 0$  and  $0 < \beta < 1$ , and if  $f$  is  $C^r$ ,  $r \geq 3$ , then the conjugacy is  $C^1$  (in fact it is  $C^{r-1-\beta-\varepsilon}$  for every  $\varepsilon > 0$ ). On the other hand, Arnold proved that some such condition on the rotation number is essential: there exist real analytic circle diffeomorphisms with irrational rotation number such that the conjugacy with a rigid rotation is not even absolutely continuous with respect to Lebesgue measure.

Maps with periodic points cannot be rigid, but we can analyze the rigidity of some relevant invariant set, such as an attractor of the map. This is the situation studied by Sullivan and McMullen in the context of unimodal maps of the interval. They considered the so-called infinitely renormalizable maps of bounded combinatorial type. For such maps, almost all orbits are asymptotic to a Cantor set which is the closure of the critical orbit. They proved that if two such maps are smooth enough and have the same combinatorics then there exists a  $C^{1+\alpha}$  diffeomorphism of the real line that conjugates the restriction of the maps to the corresponding Cantor attractors. The tools they developed have been of fundamental importance for the proof of our results.

Perhaps the most famous rigidity result in Geometry is the celebrated *Mostow rigidity theorem*. A special case of this theorem states that two compact hyperbolic manifolds of dimension at least 3 which have the same homotopy type are in fact isometric. Here a hyperbolic manifold is the quotient space  $\mathbb{H}^n/\Gamma$  of the hyperbolic space  $\mathbb{H}^n$  by a discrete group  $\Gamma$  of isometries. The hypothesis of the theorem implies the existence of a quasiconformal homeomorphism of the sphere at infinity that conjugates the actions of the two groups there. Such *a-priori* step may be regarded as a *pre-rigidity* result. The rigidity is then obtained by proving that this qc-homeomorphism is in fact conformal, *i.e.* a Moebius transformation.

The situation for critical circle mappings fits perfectly into this framework. A critical circle mapping is a homeomorphism  $f : S^1 \rightarrow S^1$  that is of class  $C^r$ ,  $r \geq 3$ , and has a unique critical point  $c$  around which, in some  $C^r$  coordinate system,  $f$  has the form  $x \mapsto x^p$ , where  $p \geq 3$  is an odd integer called the *power law* of  $f$ . Yoccoz proved in [20] that a critical circle mapping without periodic points is topologically conjugate to an irrational rotation. Later, in an unpublished work, he proved that the conjugacy between two critical circle mappings with the same rotation number is in fact quasisymmetric, *i.e.* there exists a constant  $K \geq 1$  such

that, for all pairs of adjacent intervals  $I_1, I_2$  of equal length  $|I_1| = |I_2|$ , we have

$$\frac{1}{K} \leq \frac{|h(I_1)|}{|h(I_2)|} \leq K .$$

This is in contrast with the diffeomorphism case where, without restriction on the rotation number, the conjugacy may fail to be quasisymmetric (see [15], p. 75). Yoccoz's result, whose proof we present in §4 and Appendix B, is the exact analogue of the pre-rigidity step in the proof of Mostow's theorem.

**Rigidity Conjecture** *If  $f, g$  are  $C^3$  critical circle mappings with the same irrational rotation number of bounded type and the same power-law at the critical point, then there exists a  $C^{1+\alpha}$  conjugacy  $h$  between  $f$  and  $g$  for some universal  $\alpha > 0$ .*

So far we have succeeded in proving this conjecture only when the maps are real-analytic. Our proof involves real techniques developed in this paper, and deformation of complex structures, developed in the next paper.

**1.1 Summary of results.** We now present a quick summary of our results. As already mentioned, we prove two main new theorems concerning critical circle homeomorphisms.

The first theorem brings forth the connection between renormalization and rigidity in the context of circle maps. The proof is given in §4.4.

**First Main Theorem** *Let  $f$  and  $g$  be topologically conjugate  $C^3$  critical circle maps, and let  $h$  be the conjugacy between  $f$  and  $g$  which maps the critical point of  $f$  to the critical point of  $g$ . If the partial quotients of their common rotation number are bounded, and if their renormalizations converge together exponentially fast in the  $C^0$ -topology, then  $h$  is  $C^{1+\alpha}$  for some  $\alpha > 0$ .*

The second theorem shows that the bounded type hypothesis in the Rigidity Conjecture stated above cannot be removed. The proof occupies §5 in its entirety.

**Second Main Theorem** *There exists an uncountable set  $S$  of rotation numbers such that for any  $\rho \in S$  there exist  $C^\infty$  critical circle maps  $f$  and  $g$  with rotation number  $\rho$  with the property that the conjugacy between  $f$  and  $g$  sending the critical point of  $f$  to the critical point of  $g$  is not  $C^{1+\beta}$  for any  $\beta > 0$ .*

The set  $S$  is very small: its Hausdorff dimension is not greater than  $1/2$ . But it does contain Diophantine numbers, in somewhat remarkable contrast with the case of circle diffeomorphisms. The saddle-node surgery procedure we develop here is quite general, and can be used to produce similar counterexamples to the rigidity of infinitely renormalizable unimodal maps with special unbounded combinatorics.

All estimates performed in this paper rely heavily on the *real a-priori bounds* of M. Herman [11] and G. Świątek [17]. These bounds are revisited in §3. Several

technical consequences of the real bounds needed in this paper, such as the  $C^{r-1}$  boundedness of the renormalizations of a  $C^r$  critical circle map, are gathered in Appendix A.

**Acknowledgments.** We are grateful to Dennis Sullivan for suggesting many interesting ideas, and to Curt McMullen for explaining to us various aspects of his beautiful work on renormalization and rigidity. We also want to thank Jean-Christophe Yoccoz for providing us with a sketch of the proof of his unpublished theorem on quasisymmetric conjugacies between critical circle maps. We are indebted also to Marco Martens, Jacek Graczyk, Oscar Lanford, Rafael de la Llave, Michael Lyubich and Henri Epstein, for several useful discussions on these and related matters. This work has been financially sponsored by PRONEX, CNPq and FAPESP. We also would like to thank IHES, MSRI, ETH-Zürich, IMPA and SUNY at Stony Brook for their warm hospitality and generous support.

## 2. PRELIMINARIES

We have three goals in this section. First, we present some of the basic notations commonly used when studying circle maps. Second, we present the notions of commuting pair and renormalization in the context of circle maps, and discuss their relationship. Third, we state the distortion tools that are necessary for proving the real bounds in §3.

**2.1 Critical circle mappings.** Following the tradition in this subject, we identify the unit circle  $S^1$  with the one-dimensional torus  $\mathbb{R}/\mathbb{Z}$ . The obvious advantage of such identification is that it allows us to use additive notation when dealing with circle mappings.

We briefly recall some standard facts concerning circle mappings. Given a homeomorphism  $f : S^1 \rightarrow S^1$ , we denote its rotation number by  $\rho(f)$ . It can be expressed as a continued fraction

$$\rho(f) = [a_0, a_1, \dots, a_n, \dots] = \frac{1}{a_0 + \frac{1}{a_1 + \frac{1}{\dots + \frac{1}{a_n + \frac{1}{\dots}}}}},$$

which can be finite or infinite, depending on whether  $\rho(f)$  is rational or irrational, respectively. The positive integers  $a_n$  are the *partial quotients* of  $\rho(f)$ . They give rise to a sequence of *return times* for  $f$ , recursively defined by  $q_0 = 1$ ,  $q_1 = a_0$  and  $q_{n+1} = a_n q_n + q_{n-1}$  for all  $n \geq 1$  (for which  $a_n$  exists – an assumption that will be implicit henceforth). Given  $x \in S^1$  and  $n \geq 1$ , we denote by  $J_n(x)$  the

closed interval containing  $x$  whose endpoints are  $f^{q_n}(x)$  and  $f^{q_{n-1}}(x)$ . We also let  $I_{n-1}(x) \subseteq J_n(x)$  be the closed interval whose endpoints are  $x$  and  $f^{q_{n-1}}(x)$ . Observe that  $J_n(x) = I_n(x) \cup I_{n-1}(x)$  for all  $n \geq 1$ .

From the dynamics standpoint, we are not interested in *all* circle homeomorphisms, but only in those that possess a unique critical point in  $S^1$ , being local diffeomorphisms everywhere else. These are the so-called *critical circle maps*. More precisely, let  $f : S^1 \rightarrow S^1$  be a  $C^r$  homeomorphism, for some  $r \geq 1$ . We say that  $f$  is a critical circle map if there exists  $c \in S^1$  (the critical point) such that  $f'(c) = 0$  and  $f'(x) \neq 0$  for all  $x \neq c$ . Moreover, we require  $f$  to have a *power-law* at  $c$ . This means that in a suitable  $C^r$  coordinate system around the critical point, our  $f$  is represented by a map of the form  $x \mapsto x|x|^{p-1} + a$ , for some real number  $p > 1$  called the *power-law exponent* of  $f$ . The class of all  $C^r$  critical circle maps will be denoted by  $\text{Crit}^r(S^1)$ .

Since the critical point  $c$  of a critical circle map is a distinguished point on the circle, we will write  $I_n$  and  $J_n$  throughout, instead of  $I_n(c)$  and  $J_n(c)$ , respectively.

**2.2 Commuting pairs.** We will study the successive renormalizations of a critical circle map  $f$ . Here, as in many other settings in dynamics, the word renormalization is taken to mean a (suitably normalized) Poincaré first return map of  $f$  to some neighborhood of its critical point. Abstracting the essential features of such first return maps yields the notion of *commuting pair*, due to O. Lanford [12] and D. Rand [16]. We formulate this notion as follows.

*Definition.* A  $C^r$  *commuting pair* consists of two mappings  $\mathbf{f}_- : [\lambda, 0] \rightarrow \mathbb{R}$ , where  $\lambda < 0$ , and  $\mathbf{f}_+ : [0, 1] \rightarrow \mathbb{R}$ , satisfying the following conditions.

- [P<sub>1</sub>] Both  $\mathbf{f}_-$  and  $\mathbf{f}_+$  are  $C^r$  orientation-preserving homeomorphisms onto their images.
- [P<sub>2</sub>] We have  $\mathbf{f}_-(0) = 1$ ,  $\mathbf{f}_+(0) = \lambda$  and  $0 < \mathbf{f}_-(\lambda) = \mathbf{f}_+(1) < 1$ .
- [P<sub>3</sub>] We have  $D\mathbf{f}_-(x) > 0$  for all  $\lambda \leq x < 0$ , and  $D\mathbf{f}_+(x) > 0$  for all  $0 < x \leq 1$ .
- [P<sub>4</sub>] For each  $1 \leq k \leq r$ , we have  $D^k(\mathbf{f}_+ \circ \mathbf{f}_-)(0) = D^k(\mathbf{f}_- \circ \mathbf{f}_+)(0)$ .

A *critical commuting pair* is a commuting pair such that  $D\mathbf{f}_-(0) = 0 = D\mathbf{f}_+(0)$ .

Although it is more customary to use the symbols  $\xi$  and  $\eta$  instead of  $\mathbf{f}_-$  and  $\mathbf{f}_+$ , respectively, the present notation will be more convenient for our purposes in this paper. It can be proved that, in the presence of the other conditions, P<sub>4</sub> is equivalent to the following.

- [P'<sub>4</sub>] There exist open intervals  $\Delta_- \supseteq [\lambda, 0]$  and  $\Delta_+ \supseteq [0, 1]$ , and  $C^r$  homeomorphic extensions  $\mathbf{F}_- : \Delta_- \rightarrow \mathbb{R}$  and  $\mathbf{F}_+ : \Delta_+ \rightarrow \mathbb{R}$  of  $\mathbf{f}_-$  and  $\mathbf{f}_+$  respectively, satisfying  $\mathbf{F}_+ \circ \mathbf{F}_-(x) = \mathbf{F}_- \circ \mathbf{F}_+(x)$  for all  $x \in \Delta_- \cap \Delta_+$  such that  $\mathbf{F}_\pm(x) \in \Delta_\mp$  (the set of such  $x$  is an open interval around 0).

This justifies the name *commuting pair*. The class of all  $C^r$  critical commuting pairs will be denoted by  $\mathbf{P}^r$ . We shall henceforth identify a commuting pair  $(\mathbf{f}_-, \mathbf{f}_+)$  with a single map  $\mathbf{f} : [\lambda, 1] \rightarrow [\lambda, 1]$ , called the *shadow* of the commuting

pair, defined as follows,

$$\mathbf{f}(x) = \begin{cases} \mathbf{f}_-(x), & \text{when } \lambda \leq x \leq 0 \\ \mathbf{f}_+(x), & \text{when } 0 < x \leq 1. \end{cases}$$

To each commuting pair  $\mathbf{f}$  we associate an element  $a \in \mathbb{N} \cup \{\infty\}$  called the *height* of  $\mathbf{f}$ , in the following way. If there exists  $n \geq 1$  such that  $\mathbf{f}^{n+1}(1) < 0 \leq \mathbf{f}^n(1)$ , then we set  $a = n$ ; otherwise we set  $a = \infty$ . It is clear that  $\mathbf{f}$  has infinite height if and only if there exists  $0 < \bar{x} < 1$  such that  $\mathbf{f}(\bar{x}) = \bar{x}$ .

**2.3 Renormalizing a commuting pair.** Every commuting pair  $\mathbf{f}$  with finite height  $a$  such that  $\mathbf{f}^a(1) > 0$  can be *renormalized*, in the following sense. Let  $\Lambda : \mathbb{R} \rightarrow \mathbb{R}$  be the linear map  $x \mapsto \lambda x$ , let  $\lambda' = \mathbf{f}^a(1)/\lambda < 0$ , and let  $\mathcal{R}\mathbf{f} : [\lambda', 1] \rightarrow [\lambda', 1]$  be the map defined by

$$\mathcal{R}\mathbf{f}(x) = \begin{cases} \Lambda^{-1} \circ \mathbf{f} \circ \Lambda(x), & \text{when } \lambda' \leq x \leq 0 \\ \Lambda^{-1} \circ \mathbf{f}^{a+1} \circ \Lambda(x), & \text{when } 0 < x \leq 1. \end{cases}$$

This map is (the shadow of) a commuting pair  $(\mathcal{R}\mathbf{f}_-, \mathcal{R}\mathbf{f}_+)$ , called the *first renormalization* of  $\mathbf{f}$ . Equivalently,

$$\begin{cases} \mathcal{R}\mathbf{f}_-(x) = \Lambda^{-1} \circ \mathbf{f}_+ \circ \Lambda(x), & \text{for all } \lambda' \leq x \leq 0 \\ \mathcal{R}\mathbf{f}_+(x) = \Lambda^{-1} \circ \mathbf{f}_+^a \circ \mathbf{f}_- \circ \Lambda(x), & \text{for all } 0 \leq x \leq 1. \end{cases}$$

The class of all  $C^r$  critical commuting pairs which are renormalizable in this sense will be denoted  $\mathbf{P}_1^r$ . In this way, we have a well-defined map  $\mathcal{R} : \mathbf{P}_1^r \rightarrow \mathbf{P}^r$ , the so-called *renormalization operator*. More generally, for all  $n \geq 1$  we write  $\mathbf{P}_n^r = \mathcal{R}^{-n}(\mathbf{P}^r)$  for the set of all  $C^r$  critical commuting pairs which can be renormalized  $n$  times. We have  $\mathbf{P}_{n+1}^r \subseteq \mathbf{P}_n^r$  for all  $n$ . We are especially interested in the set of all *infinitely renormalizable* critical commuting pairs, namely

$$\mathbf{P}_\infty^r = \bigcap_{n \geq 1} \mathbf{P}_n^r.$$

Given  $\mathbf{f} \in \mathbf{P}^r$ , let  $a_0 = a$  be its height, and for each  $n \geq 1$  such that  $\mathbf{f} \in \mathbf{P}_n^r$ , let  $a_n$  be the height of  $\mathcal{R}^n(\mathbf{f})$ . This can be a finite or infinite sequence; in any case, using the convention  $1/\infty = 0$ , we define the *rotation number* of  $\mathbf{f}$  to be

$$\rho(\mathbf{f}) = [a_0, a_1, \dots, a_n, \dots] = \frac{1}{a_0 + \frac{1}{a_1 + \frac{1}{\dots + \frac{1}{a_n + \frac{1}{\dots}}}}}$$

In particular,  $\rho(\mathcal{R}f) = [a_1, a_2, \dots]$ , that is, the renormalization operator acts as the Gaussian shift on continued fractions.

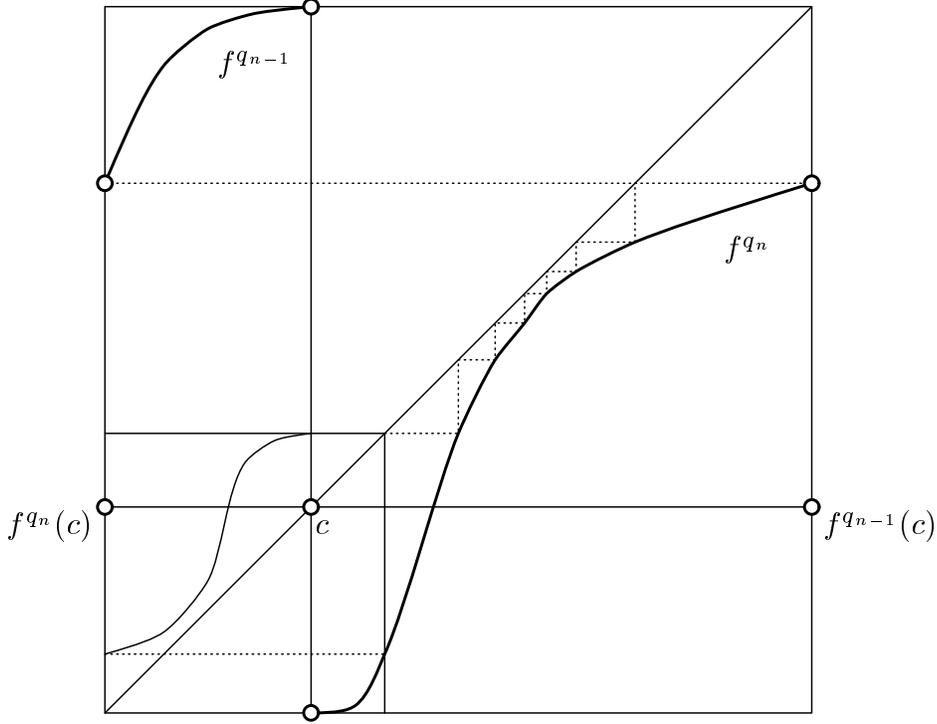


Figure 1. Two consecutive renormalizations of  $f$ .

**2.4 Renormalizing a critical circle map.** Let  $f$  be a critical circle map with critical point  $c$ , and for each  $k \geq 0$  such that  $f^{q_k}(c) \neq c$ , let  $A_k : \mathbb{R} \rightarrow \mathbb{R}/\mathbb{Z}$  be the affine covering map such that  $A_k([0, 1]) = I_k$ , with  $A_k(0) = c$  and  $A_k(1) = f^{q_k}(c)$ . For each  $n \geq 1$  such that  $f^{q_k}(c) \neq c$  for all  $0 \leq k \leq n$ , consider the Poincaré first return map  $f_n : I_n \cup I_{n-1} \rightarrow I_n \cup I_{n-1}$ , namely

$$f_n(x) = \begin{cases} f^q(x) & \text{when } x \in I_n \\ f^Q(x) & \text{when } x \in I_{n-1}, \end{cases}$$

where  $q = q_{n-1}$  and  $Q = q_n$ . Define  $\lambda_n$  to be the largest negative number such that  $A_{n-1}(\lambda_n) = f^Q(c)$  (one sees in fact that  $\lambda_n = -|I_n|/|I_{n-1}|$ ). Then  $A_{n-1}([\lambda_n, 0]) = I_n$  and  $A_{n-1}([0, 1]) = I_{n-1}$ , and we can consider the map  $\mathbf{f}_n : [\lambda_n, 1] \rightarrow [\lambda_n, 1]$  given by  $\mathbf{f}_n = A_{n-1}^{-1} \circ f_n \circ A_{n-1}$ . Here, it is implicit that  $A_{n-1}^{-1}$  is the inverse branch that maps  $I_n \cup I_{n-1}$  onto  $[\lambda_n, 1]$ . This defines (the shadow of) a  $C^r$  critical

commuting pair called *n-th renormalization of f*. It is well-defined provided the rotation number of  $f$  has a continued-fraction development of length at least  $n + 1$  (in particular it is well-defined for all  $n$  when the rotation number of  $f$  is irrational). It is easy to see in this case that  $\mathbf{f}_{k+1} = \mathcal{R}\mathbf{f}_k$  for all  $1 \leq k \leq n - 1$ . Moreover, if  $a_0, a_1, \dots$  are the partial quotients of the rotation number of  $f$  then, from the recurrence relations satisfied by the sequence of return times  $q_k$ , we see at once that the height of  $\mathbf{f}_k$  is equal to  $a_k$ , and that  $\rho(\mathbf{f}_k) = [a_k, a_{k+1}, \dots]$ .

*Remark.* Note that the largest interval containing  $I_n$  on which  $f^{q_{n-1}}$  is a diffeomorphism is the  $[\alpha_n, c]$  where  $f^{a_n - 1 q_{n-1}}(\alpha_n) = f^{q_n}(c)$ . Similarly, the largest interval containing  $I_{n-1}$  on which  $f^{q_n}$  is a diffeomorphism is  $[c, \beta_n]$  where  $f^{q_n}(\beta_n) = f_{n-1}^q(c)$ .

**2.5 The  $C^k$  metrics.** The following is only one of several equivalent ways of defining a  $C^k$  distance between commuting pairs. We normalize our commuting pairs to be defined on  $[0, 1]$ , using for each  $\mathbf{f}$  a fractional linear transformation that maps  $\lambda, 0, 1$  respectively to  $0, \frac{1}{2}, 1$ , and then use the  $C^k$  norm of the difference of the normalized pairs. The  $C^k$  distance between  $\mathbf{f}$  and  $\mathbf{g}$  defined in this fashion is denoted by  $d_k(\mathbf{f}, \mathbf{g})$ .

Let us be a bit more precise. If a function  $\varphi : [0, 1] \rightarrow \mathbb{R}$  has a jump discontinuity at  $x = 1/2$  but is elsewhere  $k$  times continuously differentiable, let  $\|\varphi\|_k = \max\{\|\varphi^-\|_k, \|\varphi^+\|_k\}$  where  $\varphi^-$  is the restriction of  $\varphi$  to  $[0, \frac{1}{2}]$  and  $\varphi^+$  is the restriction of  $\varphi$  to  $[\frac{1}{2}, 1]$ . Given two elements  $\mathbf{f} : [\lambda, 1] \rightarrow \mathbb{R}$  and  $\mathbf{g} : [\mu, 1] \rightarrow \mathbb{R}$  of  $\mathbf{P}^r$ , and given  $0 \leq k \leq r$ , let  $A_\lambda$  be the fractional linear transformation which maps  $0, \lambda, 1$  to  $0, \frac{1}{2}, 1$ , respectively, and let  $A_\mu$  be similarly defined. Then write

$$d_k(\mathbf{f}, \mathbf{g}) = \max\{|\lambda - \mu|, \|A_\lambda \mathbf{f} A_\lambda^{-1} - A_\mu \mathbf{g} A_\mu^{-1}\|_k\}.$$

**Lemma 2.1** *For each  $0 \leq k \leq r$ ,  $d_k$  is a metric.*

*Proof.* The only thing not entirely obvious is that  $d_k(\mathbf{f}, \mathbf{g}) = 0$  implies  $\mathbf{f} = \mathbf{g}$ . But if  $d_k(\mathbf{f}, \mathbf{g}) = 0$  then on one hand  $\lambda = \mu$ , so that  $A_\lambda = A_\mu$ , and on the other hand  $A_\lambda \mathbf{f} A_\lambda^{-1} - A_\mu \mathbf{g} A_\mu^{-1} = 0$ , so that  $\mathbf{f} = A_\lambda^{-1} A_\mu \mathbf{g} A_\mu^{-1} A_\lambda = \mathbf{g}$ .  $\square$

**Proposition 2.2** *Let  $\mathbf{f} : [\lambda, 1] \rightarrow [\lambda, 1]$  and  $\mathbf{g} : [\mu, 1] \rightarrow [\mu, 1]$  be elements of  $\mathbf{P}_\infty^r$ , and suppose there exists a  $C^r$  diffeomorphism  $h : [\lambda, 1] \rightarrow [\mu, 1]$  such that  $h \circ \mathbf{f} = \mathbf{g} \circ h$ . Then for all  $k \leq r - 1$  the distances  $d_k(\mathcal{R}^n(\mathbf{f}), \mathcal{R}^n(\mathbf{g}))$  converge to 0 at an exponential rate.*

*Proof.* Let  $\mathbf{f}_n = \mathcal{R}^n \mathbf{f}$  and  $\mathbf{g}_n = \mathcal{R}^n \mathbf{g}$ . Then  $\mathbf{f}_n = h_n^{-1} \circ \mathbf{g}_n \circ h_n$ , where  $h_n$  is obtained from  $h$  by restriction and affine rescaling. We will see below (after we prove the real bounds for critical circle maps, cf. Theorem 3.1) that  $\{h_n\}$  converges exponentially in the  $C^r$  sense to the space of *affine* maps. Therefore, we have that  $d_{r-1}(\mathbf{f}_n, \mathbf{g}_n) \rightarrow 0$  exponentially fast.  $\square$

**2.6 Distortion tools.** In §3 we will need some distortion tools to get real bounds for critical circle maps. The most basic is the notion of *cross-ratio distortion*. Given intervals  $M \subseteq T$  on the line or circle, their cross-ratio is defined as

$$D(M, T) = \frac{|M||T|}{|L||R|} ,$$

where  $L$  and  $R$  are the left and right components of  $T \setminus M$ . The cross-ratio distortion of a map  $f$  (whose domain contains  $T$ ) on the pair of intervals  $(M, T)$  is

$$B(f; M, T) = \frac{D(f(M), f(T))}{D(M, T)} .$$

Cross-ratios are always increased by a map with negative Schwarzian derivative. More precisely, if  $f$  is  $C^3$  and  $Sf < 0$  then  $B(f; M, T) > 1$ .

**Lemma 2.3** (*Cross-ratio distortion principle*)

Given a map  $f$  as above,  $m \geq 1$  and intervals  $M \subseteq T$  such that  $f^m|_T$  is a diffeomorphism onto its image, we have

$$B(f^m; M, T) \geq \exp\left\{-\sigma \sum_{j=0}^{m-1} |f^j(T)|\right\} ,$$

where  $\sigma > 0$  depends on  $f$  and  $\max_{0 \leq j \leq m-1} |f^j(T)|$ .

For a proof of (a much more general version of) this principle, see [15], p. 287. This fact will be used in combination with the following classical distortion principle. For intervals  $M \subseteq T$  as above we define the *space* of  $M$  inside  $T$  to be the smallest of the ratios  $|L|/|M|$  and  $|R|/|M|$ .

**Lemma 2.4** (*Koebe distortion principle*)

Given  $\ell, \tau > 0$  and a map  $f$  as above, there exists  $K = K(\ell, \tau, f) > 1$  of the form

$$K = \left(1 + \frac{1}{\tau}\right)^2 \exp C\ell ,$$

where  $C$  is a constant depending only on  $f$ , with the following property. If  $T$  is an interval such that  $f^m|_T$  is a diffeomorphism and if  $\sum_{j=0}^{m-1} |f^j(T)| \leq \ell$ , then for each interval  $M \subseteq T$  for which the space of  $f^m(M)$  inside  $f^m(T)$  is at least  $\tau$  and for all  $x, y \in M$  we have

$$\frac{1}{K} \leq \left| \frac{Df^m(x)}{Df^m(y)} \right| \leq K .$$

Once again, see [15], p. 295, for a proof. Used in combination with Lemma 2.3, the Koebe distortion principle allows one to propagate space around under fairly general circumstances.

### 3. THE REAL A-PRIORI BOUNDS

In this section we establish real *a-priori* bounds for critical circle maps, obtaining as a corollary the fact that their renormalizations are pre-compact in the  $C^1$  topology. The results are well-known, and the reader will not fail to notice the overlap with some of the material in [18] and [9].

Let  $f: S^1 \rightarrow S^1$  be a critical circle homeomorphism with critical point  $c$ . The iterates of  $c$  are denoted by  $c_i = f^i(c)$ . Let  $I_n$  be the interval with endpoints  $c$  and  $c_{q_n}$  that contains  $c_{q_{n+2}}$ , as defined in section 2. For simplicity, we write  $I_n^j = f^j(I_n)$  for all  $j$  and  $n$ . The most basic combinatorial fact to be remembered here is that the collection of intervals

$$\mathcal{P}_n = \left\{ I_{n-1}, I_{n-1}^1, \dots, I_{n-1}^{q_n-1} \right\} \cup \left\{ I_n, I_n^1, \dots, I_n^{q_n-1} \right\}$$

constitutes a partition of  $S^1$  modulo endpoints, called the *dynamical partition of level  $n$*  associated to  $f$ . In order to get an actual partition we exclude from each interval in  $\mathcal{P}_n$  its right endpoint, say, according to the standard choice of orientation of  $S^1$ . Let  $\mathcal{P}_n(x)$  denote the atom of the partition  $\mathcal{P}_n$  that contains  $x$  (in particular,  $\mathcal{P}_n(c)$  is either  $I_n$  or  $I_{n-1}$  according to the parity of  $n$ ).

**Theorem 3.1** (*Real Bounds*) *Let  $f \in \text{Crit}^r(S^1)$  be a map with irrational rotation number. There exist constants  $C_0 > 1$  and  $0 < \mu_0 < \mu_1 < 1$  depending only on  $f$  such that*

- (a) *If  $I$  and  $J$  are any two adjacent atoms of  $\mathcal{P}_n$ , then  $C_0^{-1}|J| < |I| < C_0|J|$ ;*
- (b) *For every  $x \in S^1$ , we have  $|\mathcal{P}_n(x)| < \mu_1|\mathcal{P}_{n-1}(x)|$ ;*
- (c) *If the rotation number of  $f$  is of bounded type then  $|\mathcal{P}_n(x)| > \mu_0^n/C_0$ ;*
- (d) *If the rotation number of  $f$  is of bounded type then  $|\mathcal{P}_n(x)| > |\mathcal{P}_{n-1}(x)|/C_0$ ;*
- (e) *If  $0 < i \leq j \leq q_n$  then the distortion of the restriction of  $f^{j-i}$  to  $I_{n-1}^i = f^i(I_{n-1})$  is bounded by  $C_0$ .*

*In particular, the critical commuting pairs  $\mathcal{R}^n f$  form a bounded sequence in the  $C^1$  topology.*

Later in this section we will see that the bounds in this theorem are eventually universal.

**3.1 Bounding space.** In what follows, two positive numbers  $a$  and  $b$  are said to be *comparable modulo  $f$* , or simply *comparable*, if there exists a constant  $C > 1$ , depending only on our map  $f$ , such that  $C^{-1}b \leq a \leq Cb$ . This relation is denoted by  $a \asymp b$ . It is also convenient to write  $a \preccurlyeq b$  to indicate that  $a \leq Cb$ . Comparability modulo  $f$  is reflexive and symmetric, but not transitive since the constants multiply. Hence, if  $b_1 \asymp b_2 \asymp \dots \asymp b_N$ , we can only say that  $b_1 \preccurlyeq b_N$  if  $N$  is bounded (by a constant depending only on  $f$ ).

**Lemma 3.2** For each  $n \geq 0$  there exist  $z_1, z_2, z_3, z_4, z_5 \in S^1$  with  $z_{j+1} = f^{q_n}(z_j)$  such that  $|z_1 - z_2| \asymp |z_2 - z_3| \asymp |z_3 - z_4| \asymp |z_4 - z_5|$ .

*Proof.* Let  $z \in S^1$  be a point such that  $|f^{q_n}(z) - z| \leq |f^{q_n}(x) - x|$  for all  $x \in S^1$ . From Koebe's principle applied successively to  $f^{-q_n}, f^{-2q_n}$  and  $f^{-3q_n}$ , we have

$$|z - f^{q_n}(z)| \gtrsim |f^{-q_n}(z) - z| \gtrsim |f^{-2q_n}(z) - f^{-q_n}(z)| \gtrsim |f^{-3q_n}(z) - f^{-2q_n}(z)|.$$

Moreover, by our choice of  $z$  we have  $|z - f^{q_n}(z)| \lesssim |f^{-3q_n}(z) - f^{-2q_n}(z)|$ . Therefore we can take  $z_5 = f^{q_n}(z), z_4 = z, \dots, z_1 = f^{-3q_n}(z)$  as the desired five points.  $\square$

**Lemma 3.3** Let  $z_1, z_2, \dots, z_5$  and  $w_0, w_1, \dots, w_5$  be points on the circle such that  $z_{j+1} = f^{q_n}(z_j)$  and  $w_{j+1} = f^{q_n}(w_j)$ , and such that  $w_1$  lies on the interval of endpoints  $z_1$  and  $z_2$  in the partition of  $S^1$  determined by the  $z_i$ 's. If  $|z_1 - z_2| \asymp |z_2 - z_3| \asymp |z_3 - z_4| \asymp |z_4 - z_5|$ , then

$$|w_0 - w_1| \gtrsim |w_1 - w_2| \lesssim |w_2 - w_3|. \quad (1)$$

*Proof.* Let  $\ell = \min |z_j - z_{j+1}|$ . If there is a  $j$  with  $1 \leq j \leq 3$  such that  $|w_j - w_{j+1}| \leq \ell/2$ , then we must have  $|w_{j-1} - w_j| \geq \ell/2$  and  $|w_{j+1} - w_{j+2}| \geq \ell/2$  also. But then  $[w_j, w_{j+1}]$  has space on both sides inside  $[w_{j-1}, w_{j+2}]$ . Applying  $f^{-(j-1)q_n}$  to these points and using the Koebe principle, we get (1). If on the other hand there is no  $j$  with that property, then  $|w_1 - w_2| \asymp |w_2 - w_3| \asymp |w_3 - w_4|$ . Again, applying  $f^{-q_n}$  and using Koebe we get (1).  $\square$

**Lemma 3.4** For all  $n \geq 0$  and all  $x \in S^1$ , we have  $|f^{q_n}(x) - x| \asymp |x - f^{-q_n}(x)|$ .

*Proof.* To show that  $|x - f^{-q_n}(x)| \geq C^{-1}|f^{q_n}(x) - x|$ , let  $i \leq q_n$  be such that  $f^i(x) \in [z_1, z_2]$ , where  $z_1, z_2, \dots$  are the points given by Lemma 3.2. Then let  $w_0 = f^{i-q_n}(x), w_1 = f^i(x)$ , etc. We know from Lemma 3.3 that  $|w_0 - w_1| \gtrsim |w_1 - w_2| \lesssim |w_2 - w_3|$ . Applying  $f^{-i}$  to these points and using the Koebe distortion principle, we find a definite *space* around  $[x, f^{q_n}(x)]$  inside  $[f^{-q_n}(x), f^{2q_n}(x)]$ . Therefore  $|x - f^{-q_n}(x)| \gtrsim |f^{q_n}(x) - x|$ . The proof of the opposite inequality is similar.  $\square$

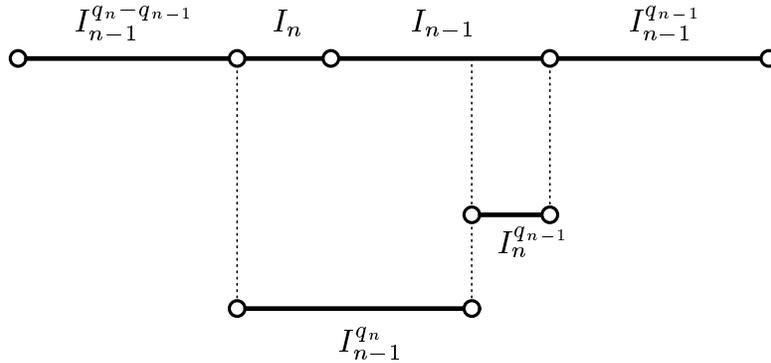


Figure 2. These six intervals are pairwise comparable.

We arrive at the following fundamental fact first proved by Świątek [17] and Herman [11].

**Lemma 3.5** *Any two adjacent intervals in the dynamical partition of level  $n$  of  $f$  are comparable.*

*Proof.* First we prove that all intervals in Figure 2 are pairwise comparable, through the following steps.

- (a) From Lemma 3.4, we know that  $|I_{n-1}| \asymp |I_{n-1}^{q_{n-1}}|$  and  $|I_{n-1}^{q_n - q_{n-1}}| \asymp |I_{n-1}^{q_n}|$ .
- (b) Since the dynamical symmetric of  $I_n$ , namely the interval  $I_n^{-q_n}$ , is contained in  $I_{n-1}$ , we also have  $|I_n| \asymp |I_{n-1}|$ .
- (c) Since the dynamical symmetric of  $I_{n-1}$ , namely  $I_{n-1}^{-q_{n-1}}$  is contained in  $I_n \cup I_{n-1}^{q_n - q_{n-1}}$ , we have  $|I_{n-1}| \asymp |I_{n-1}^{q_n - q_{n-1}}|$ . Moreover, since  $I_{n-1}^{q_n} \subseteq I_n \cup I_{n-1}$ , items (a) and (b) yield  $|I_{n-1}^{q_n - q_{n-1}}| \asymp |I_{n-1}|$ . Therefore  $|I_{n-1}| \asymp |I_{n-1}^{q_n - q_{n-1}}|$ .
- (d) Next, we claim that  $|I_n| \asymp |I_n^{q_n}|$ . To see why, consider the diffeomorphism

$$f^{q_n - q_{n-1}} : I_{n-1} \cup I_{n-1}^{q_n} \rightarrow I_{n-1}^{q_n - q_{n-1}} \cup I_{n-1}^{q_n}.$$

By the cross-ratio inequality (Lemma 2.3) applied to  $M = I_n^{q_n}$  and  $T = I_{n-1} \cup I_{n-1}^{q_n}$ , we have  $|I_n^{q_n}| \asymp |I_n| \asymp |I_n|$ . Conversely, considering the diffeomorphism

$$f^{q_n} : I_{n-1}^{q_n - q_{n-1}} \cup I_n \rightarrow I_{n-1}^{q_n} \cup I_n^{q_n}$$

and applying the cross-ratio inequality to  $M = I_n^{q_n}$  and  $T = I_{n-1}^{q_n - q_{n-1}} \cup I_n$ , we get

$$|I_n| \asymp |I_n^{q_n}| \asymp |I_{n-1}^{q_n + q_{n-1}}| \asymp |I_n^{q_n}|.$$

This proves our claim.

- (e) Finally, we claim that  $|I_{n-1}| \asymp |I_n|$ , thereby reversing the inequality in (b). It is here that we use the critical point in a crucial way. Let  $\theta_n = |I_n|/|I_{n-1}|$ ; we already know that  $\theta_n \asymp 1$ . Look at the intervals  $I_{n-1}^1$ ,  $I_n^1$  and  $I_{n-1}^{q_n - q_{n-1} + 1}$ , all near the critical value of  $f$ . By an argument similar to the one in (c), we have  $|I_{n-1}^1| \asymp |I_{n-1}^{q_n - q_{n-1} + 1}|$ . Moreover,  $|I_n^1| \asymp \theta_n^p |I_{n-1}^1|$ , where  $p > 1$  is the power-law of  $f$  at the critical point. Hence these three intervals have a cross-ratio comparable to  $\theta_n^p$ . On the other hand the map  $f^{q_n - 1}$  carries them diffeomorphically onto  $I_{n-1}^{q_n - 1}$ ,  $I_n^{q_n - 1}$  and  $I_{n-1}^{q_n}$ , respectively, whose cross-ratio is comparable to  $|I_n^{q_n - 1}|/|I_{n-1}^{q_n - 1}|$ , which in turn is comparable to  $\theta_n$ . Applying the Koebe distortion principle, we see that  $\theta_n^p \asymp \theta_n$ , and so  $\theta_n \asymp 1$  as claimed.

This proves that all six intervals in Figure 2 are comparable. To derive the remaining comparability relations, propagate this information using Koebe's distortion principle.  $\square$

*Proof of Theorem 3.1.* Part (a) is Lemma 3.5 above. The remaining statements are straightforward consequences of (a).  $\square$

**3.2 Beau property of renormalization.** The bounds obtained in the proof of Theorem 3.1 depended on  $f$ , more precisely on the space that each atom of  $\mathcal{P}_n$  enjoys relative to its two neighbors in  $\mathcal{P}_n$ . We now concentrate in proving that such bounds eventually become universal. It suffices to prove that the space in question is eventually universal. Bounds of this type are called *beau* by Sullivan.

**Lemma 3.6** *There exists  $n_0 = n_0(f)$  such that for all  $n \geq n_0$  the first return map  $f_n : J_n \rightarrow J_n$  satisfies  $Sf_n(x) < 0$  for all  $x \in J_n$ .*

*Proof.* This is proved in Theorem A.4 of Appendix A.  $\square$

**Lemma 3.7** *Given  $\varepsilon > 0$ , there exists  $n_1 = n_1(f, \varepsilon) > n_0(f)$  such that the following holds for all  $n \geq n_1$ . Let  $\Delta \in \mathcal{P}_n$ , let  $k \geq 1$  be an integer such that  $f^j(\Delta)$  is contained in an element of  $\mathcal{P}_n$  for all  $1 \leq j \leq k$ , and let  $\Delta^*$  be the union of  $\Delta$  with its left and right neighbors in  $\mathcal{P}_n$ . Then we have  $f^k|_{\Delta^*} = \phi_1 \circ \phi_2 \circ \phi_3$  where  $\phi_1$  and  $\phi_3$  are diffeomorphisms with distortion bounded by  $1 + \varepsilon$  and  $\phi_2$  is either the identity or a map with negative Schwarzian derivative. In particular, if  $\varepsilon$  is small enough and if  $I_{n-1} \neq \Delta \neq I_n$ , then the distortion of  $f^k|_{\Delta}$  is bounded from below by one-half.*

*Proof.* Let  $n_1 > n_0$  be such that  $\mu_0^{n_1 - n_0} \ll \varepsilon$ , where  $\mu_0$  is the constant of Theorem 3.1. For  $n \geq n_1$ ,  $\Delta$  and  $k$  as in the statement, let  $J \in \mathcal{P}_{n_0}$  be such that  $\Delta \subseteq J$ , let  $J^*$  be the union of  $J$  with its two neighbors in  $\mathcal{P}_{n_0}$ , and note that the space of  $\Delta^*$  inside  $J^*$  is bounded from below by  $C|J^*|/|\Delta^*|$ , for some constant  $C > 0$ . Let  $m \geq 0$  be the smallest integer such that  $f^m(J) \subseteq J_{n_0}$ . Then for all  $j \leq m$  the map  $f^j|_{J^*}$  is a diffeomorphism onto its image and, by Theorem 3.1 (b) and the Koebe distortion principle, its distortion on  $\Delta^*$  is bounded by

$$\left(1 + C \frac{|\Delta^*|}{|J^*|}\right)^2 \exp\left\{C \frac{|\Delta^*|}{|J^*|}\right\} \leq \exp\{C\mu_0^{n_1 - n_0}\} \leq 1 + \varepsilon .$$

Now, there are two possibilities. The first is that  $m \geq k$ ; in this case we can take  $\phi_1 = f^k|_{\Delta^*}$  and  $\phi_2 = \phi_3 = \text{identity map}$ . The second is that  $m < k$ . In this case we consider the first return map  $f_{n_0} : J_{n_0} \rightarrow J_{n_0}$  and let  $\ell \geq 0$  be the *largest* such that

$$f^k = f^{k_1} \circ f_{n_0}^\ell \circ f^{k_3} ,$$

where  $k_1 \geq 0$  and  $k_3 \geq 0$ . We then take  $\phi_1 = f^{k_1 - 1}$ ,  $\phi_2 = f \circ f_{n_0}^\ell$  and  $\phi_3 = f^{k_3}|_{\Delta^*}$  (if  $k_1 = 0$  we take instead  $\phi_2 = f_{n_0}^\ell$  and  $\phi_1 = \text{identity}$ ). By Lemma 3.6,  $S\phi_2 < 0$ , and by the above remarks the distortions of both  $\phi_1$  and  $\phi_3$  are bounded by  $1 + \varepsilon$  in the appropriate domains.  $\square$

**Proposition 3.8** *All bounds in Theorem 3.1 are beau. In other words, there exist universal constants  $K_0 > 0$  and  $0 < \lambda_0 < \lambda_1 < 1$  and some  $\bar{n} = \bar{n}(f) > 0$  such*

that for all  $n \geq \bar{n}$  the constants  $C_0$ ,  $\mu_0$  and  $\mu_1$  in Theorem 3.1 can be replaced by  $K_0$ ,  $\lambda_0$  and  $\lambda_1$ , respectively.

*Proof.* This is straightforward from Lemma 3.7.  $\square$

*Remark.* From now on, whenever we say that a constant “depends only on the real bounds”, we mean that the said constant is a universal function of constants  $K_0$ ,  $\lambda_0$  and  $\lambda_1$  given by this proposition.

#### 4. HOW SMOOTH IS THE CONJUGACY?

Now we turn to the first main result in this paper. The theorem states that if the successive renormalizations of two critical circle maps with the same rotation number of bounded type converge together at an exponential rate, then such maps are  $C^{1+\alpha}$  conjugate, for some universal  $\alpha > 0$ . First, in order to get bounds that do not depend on the maximum of the partial quotients of the rotation number, we need to perform some “saddle-node” estimates and constructions.

**4.1 Saddle-node geometry.** Let  $a$  be a positive integer and let  $\Delta_1, \Delta_2, \dots, \Delta_{a+1}$  be consecutive intervals on the line or circle. By an *almost parabolic map* of length  $a$  and fundamental domains  $\Delta_j$ ,  $1 \leq j \leq a$ , we mean a negative-Schwarzian diffeomorphism

$$f : \Delta_1 \cup \Delta_2 \cup \dots \cup \Delta_a \rightarrow \Delta_2 \cup \Delta_3 \cup \dots \cup \Delta_{a+1}$$

such that  $f(\Delta_j) = \Delta_{j+1}$ .

The basic geometric estimate on almost parabolic maps is due to J.-C. Yoccoz.

**Yoccoz’s Lemma** *Let  $f : I \rightarrow J$  be an almost parabolic map of length  $a$  and fundamental domains  $\Delta_j$ ,  $1 \leq j \leq a$ . If  $|\Delta_1| \geq \sigma|I|$  and  $|\Delta_a| \geq \sigma|I|$ , then*

$$\frac{1}{C_\sigma} \frac{|I|}{\min\{j, a-j\}^2} \leq |\Delta_j| \leq C_\sigma \frac{|I|}{\min\{j, a-j\}^2},$$

where  $C_\sigma > 1$  depends only on  $\sigma$ .

For a proof, see Appendix B. We will use Yoccoz’s estimates to compare two almost parabolic maps.

**Proposition 4.1** *Let  $f$  and  $g$  be two almost parabolic maps with the same length  $a$  defined on the same interval. Then for all  $x \in \Delta_1(f) \cap \Delta_1(g)$  and all  $0 \leq k \leq a/2$  we have*

$$|f^k(x) - g^k(x)| \leq C \|f - g\|_{C^0} k^3. \quad (2)$$

*Proof.* First note, using the mean-value theorem, that

$$\begin{aligned} |f^k(x) - g^k(x)| &= \left| \sum_{j=0}^{k-1} (f^{k-j-1}(f(g^j(x))) - f^{k-j-1}(g^{j+1}(x))) \right| \\ &\leq \sum_{j=0}^{k-1} |Df^{k-j-1}(\xi_j)| |f(g^j(x)) - g(g^j(x))| , \end{aligned}$$

where  $\xi_j$  lies between  $f(g^j(x))$  and  $g^{j+1}(x)$ . Hence we have

$$|f^k(x) - g^k(x)| \leq \|f - g\|_{C^0} \sum_{j=0}^{k-1} |Df^{k-j-1}(\xi_j)| . \quad (3)$$

Let us estimate each summand in the right-hand side of (3). Let  $m = m(j)$  be such that  $\xi_j \in \Delta_{j+m}$ , and assume also that  $j + m \leq a/2$ . This last condition is always satisfied if the central fundamental domain of  $g$  lies to the left of the central fundamental domain of  $f$  (if this is not the case, then reverse the roles of  $f$  and  $g$  in (3) and throughout). Using Yoccoz's Lemma, we see that

$$|Df^{k-j-1}(\xi_j)| \asymp \frac{(j+m)^2}{(a-k-m+1)^2} \leq \left( \frac{j+m}{j+1} \right)^2 . \quad (4)$$

Hence, it suffices to estimate  $m$  as a function of  $j$ . For this purpose, let  $n = n(j)$  be such that  $g^{j+1}(x) \in [f^{j+n-1}(x), f^{j+n}(x)]$ . We claim that  $m \leq n+1$ . There are two possibilities. The first is that  $f(g^j(x)) \geq g^{j+1}(x)$ : in this case we see easily that

$$\xi_j \in [g^{j+1}(x), f(g^j(x))] \subseteq [f^{j+n-1}(x), f^{j+n+1}(x)]$$

and so  $m \leq n+1$ . The second is that  $f(g^j(x)) < g^{j+1}(x)$ . In this case we have  $\xi_j < g^{j+1}(x) < f^{j+n}(x) \in \Delta_{j+n+1}$ , so once again  $m \leq n+1$ . This proves our claim.

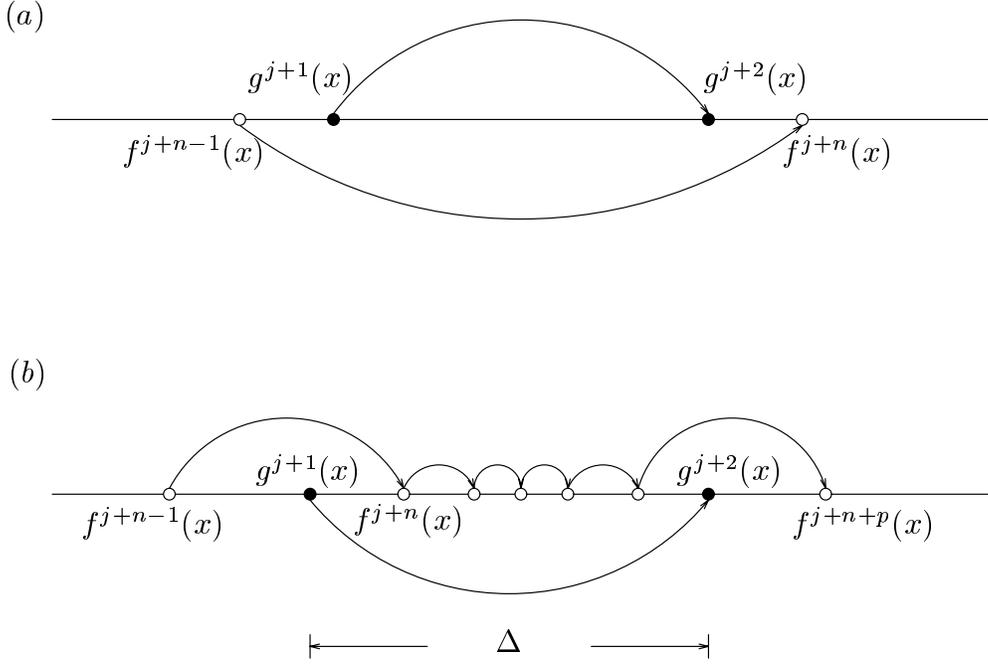


Figure 3. Bounding  $n$  as a function of  $j$ .

So now we must bound  $n$  as a function of  $j$ . Again, there are two cases to consider.

- (a) We have  $[g^{j+1}(x), g^{j+2}(x)] \subseteq [f^{j+n-1}(x), f^{j+n}(x)]$ . In this case (Figure 3a) Yoccoz's Lemma gives us

$$\frac{1}{j^2} \leq \frac{C}{(j+n)^2},$$

which implies  $n \leq Cj$ .

- (b) We have  $g^{j+2}(x) > f^{j+n}(x)$ . In this case (Figure 3b),  $f^{j+n}(x)$  is the first point in the  $f$ -orbit of  $x$  that lands inside the interval  $\Delta = [g^{j+1}(x), g^{j+2}(x)]$ . Let  $p$  be such that  $f^{j+n+i}(x) \in \Delta$  for  $i = 0, 1, \dots, p-1$  but  $f^{j+n+p}(x) \notin \Delta$ . Then we have  $\Delta \subseteq [f^{j+n-1}(x), f^{j+n+p}(x)]$ , and this time Yoccoz's Lemma gives us

$$\frac{1}{j^2} \leq C \left( \frac{1}{(j+n)^2} + \frac{1}{(j+n+1)^2} + \dots + \frac{1}{(j+n+p)^2} \right) \leq \frac{C}{j+n}.$$

Therefore  $n \leq Cj^2$  in this case.

In either case we see that  $m \leq Cj^2$ . Carrying this information back to (4), we deduce that

$$|Df^{k-j-1}(\xi_j)| \leq Cj^2. \quad (5)$$

Substituting (5) into (3), we arrive at (2), and the proof is complete.  $\square$

**4.2 A criterion for smoothness.** One key ingredient in the proof of our First Main Theorem is a slight extension of a result originally due to Carleson [2], namely Proposition 4.3 below. To formulate it, we need an auxiliary definition.

*Definition.* A *fine grid* is a sequence  $\{\mathcal{Q}_n\}_{n \geq 0}$  of finite partitions of  $S^1$  which satisfies

- (a) Each  $\mathcal{Q}_{n+1}$  is a strict refinement of  $\mathcal{Q}_n$ ;
- (b) There exists  $a > 0$  such that each  $I \in \mathcal{Q}_n$  is the disjoint union of at most  $a$  atoms of  $\mathcal{Q}_{n+1}$ ;
- (c) There exists  $c > 0$  such that  $c^{-1}|I| \leq |J| \leq c|I|$  for each pair of adjacent atoms  $I, J \in \mathcal{Q}_n$ .

For example, the dynamical partitions  $\{\mathcal{P}_n\}$  of a critical circle map with rotation number of bounded type always form a fine grid, by Theorem 3.1. We note the following easy lemma concerning a fine grid  $\{\mathcal{Q}_n\}$ .

**Lemma 4.2** *If  $I \in \mathcal{Q}_n$ ,  $J \in \mathcal{Q}_{n+1}$  and  $J \subseteq I$ , then  $(1 + c^{-1})|J| \leq |I| \leq ac^a|J|$ . In particular, there exist  $C_0 > 1$  and  $0 < \lambda_0 < \lambda_1 < 1$  such that  $C_0^{-1}\lambda_0^n \leq |I| \leq C_0\lambda_1^n$ , for all  $I \in \mathcal{Q}_n$ .  $\square$*

The constants  $a, c, C_0, \lambda_0, \lambda_1$ , are the *fine constants* of  $\{\mathcal{Q}_n\}$ .

**Proposition 4.3** *Let  $h : S^1 \rightarrow S^1$  be a homeomorphism and let  $\{\mathcal{Q}_n\}_{n \geq 0}$  be a fine grid.*

- (a) *If there exists  $C > 0$  such that*

$$\left| \frac{|I|}{|J|} - \frac{|h(I)|}{|h(J)|} \right| \leq C ,$$

*for each pair of adjacent atoms  $I, J \in \mathcal{Q}_n$ , for all  $n \geq 0$ , then  $h$  is quasimetric.*

- (b) *If there exist constants  $C > 0$  and  $0 < \lambda < 1$  such that*

$$\left| \frac{|I|}{|J|} - \frac{|h(I)|}{|h(J)|} \right| \leq C\lambda^n , \tag{6}$$

*for each pair of adjacent atoms  $I, J \in \mathcal{Q}_n$ , for all  $n \geq 0$ , then  $h$  is a  $C^{1+\alpha}$ -diffeomorphism for some  $\alpha > 0$ .*

The proof of Proposition 4.3 will depend on the following fact from elementary real analysis. If  $\phi$  is a real-valued function in an interval or oriented arc on the circle, let  $D^+\phi(x) = \lim_{t \downarrow 0} (\phi(x+t) - \phi(t))/t$  be the *right derivative* of  $\phi$  at  $x$ , if the limit exists.

**Lemma 4.4** *Let  $\phi_n : [0, 1] \rightarrow \mathbb{R}$  be a sequence of continuous, right differentiable mappings such that  $D^+\phi_n$  converges uniformly to an  $\alpha$ -Hölder continuous function  $\varphi : [0, 1] \rightarrow \mathbb{R}$ , and such that each  $D^+\phi_n$  is Riemann-integrable. If  $\phi_n$  converges uniformly to  $\phi$ , then  $\phi$  is  $C^{1+\alpha}$  and  $D\phi = \varphi$ .  $\square$*

*Proof of Proposition 4.3.* We will prove (b) only, the proof of (a) being somewhat easier. Let  $\phi_n$  be the piecewise affine  $C^0$ -approximations to  $h$  determined by the vertices of  $\mathcal{Q}_n$ . Then  $\phi_n$  is differentiable on the right, and  $D^+\phi_n$  is a step function. First we show that  $\{D^+\phi_n\}_{n \geq 0}$  is a uniform Cauchy sequence, and then that the limit is Hölder continuous. Take an atom  $I$  of  $\mathcal{Q}_n$ , and consider the decomposition

$$I = J_1 \cup J_2 \cup \dots \cup J_p ,$$

with  $J_k \in \mathcal{Q}_{n+1}$  consecutive and pairwise disjoint and  $p \leq a$ . Then  $D^+\phi_n$  is constant on  $I$  and  $D^+\phi_{n+1}$  is constant on each  $J_k$ , say

$$\begin{cases} D^+\phi_n(t) = \sigma = \frac{|\phi_n(I)|}{|I|} & (t \in I) \\ D^+\phi_{n+1}(t) = \sigma_k = \frac{|\phi_{n+1}(J_k)|}{|J_k|} & (t \in J_k) \end{cases} .$$

Thus, we have

$$\sigma|I| = \sum_{k=1}^p \sigma_k |J_k| ,$$

and in particular  $\sigma' = \min \sigma_k \leq \sigma \leq \max \sigma_k = \sigma''$ . Also,  $\sigma'/\sigma'' \leq \sigma/\sigma_k \leq \sigma''/\sigma'$  for all  $k$ . Since by assumption  $|1 - (\sigma_{k+1}/\sigma_k)| \leq C\lambda^{n+1}$ , an easy telescoping trick gives us

$$\frac{\sigma''}{\sigma'} \leq (1 + C\lambda^{n+1})^a \leq 1 + C\lambda^{n+1} .$$

A similar lower bound holds true for  $\sigma'/\sigma''$ . Therefore we have

$$1 - C\lambda^n \leq \frac{\sigma}{\sigma_k} \leq 1 + C\lambda^n , \quad (7)$$

for all  $k = 1, 2, \dots, p$ . This shows that the sequence  $\{D^+\phi_n\}_{n \geq 0}$  is uniformly bounded, and moreover that for all  $m \geq n \geq 0$  and all  $t \in S^1$ , we have

$$|D^+\phi_m(t) - D^+\phi_n(t)| \leq C \sum_{j=n}^{m-1} \lambda^j < \frac{C}{1-\lambda} \lambda^n . \quad (8)$$

Hence  $\{D^+\phi_n\}_{n \geq 0}$  is a uniform Cauchy sequence as claimed. Let  $\varphi = \lim D^+\phi_n$ , and let  $\alpha > 0$  be such that  $\lambda_0^\alpha = \lambda$ . We prove  $\varphi$  is  $\alpha$ -Hölder as follows. It suffices

to consider points  $x, y \in S^1$  whose distance is smaller than  $\inf_{I \in \mathcal{Q}_0} |I|$ . Take the smallest  $n$  such that  $x$  and  $y$  belong to distinct elements of  $\mathcal{Q}_n$ . Then either  $n = 0$  or  $x$  and  $y$  lie in a common element of  $\mathcal{Q}_{n-1}$ . Either way we have by (7)

$$|D^+ \phi_n(x) - D^+ \phi_n(y)| \leq C \lambda^n . \quad (9)$$

Combining (8) and (9), we deduce that

$$\begin{aligned} |\varphi(x) - \varphi(y)| &\leq |\varphi(x) - D^+ \phi_n(x)| + |D^+ \phi_n(x) - D^+ \phi_n(y)| + |D^+ \phi_n(y) - \varphi(y)| \\ &\leq \frac{C}{1-\lambda} \lambda^n + C \lambda^n + \frac{C}{1-\lambda} \lambda^n \leq C \lambda_0^{n\alpha} \\ &\leq C |x - y|^\alpha , \end{aligned}$$

and so  $\varphi$  is  $\alpha$ -Hölder as claimed.  $\square$

*Remark.* In the language of conditional expectations, the sequence  $\{D^+ \phi_n\}_{n \geq 0}$  satisfies  $\mathbb{E}(D^+ \phi_n | \mathcal{B}_n) = D^+ \phi_{n+1}$ , where  $\mathcal{B}_n$  is the  $\sigma$ -algebra generated by  $\mathcal{Q}_n$ , and therefore constitutes a *martingale*. Thus, the existence of a pointwise *a.e.* limit  $\varphi$ , merely as an integrable function, is a special case of J. Doob's *martingale convergence theorem*, see [1], p. 490.

**4.3 A suitable fine grid.** The dynamical partitions  $\mathcal{P}_n$  of a critical circle map  $f$  do *not* determine a fine grid, unless the rotation number of  $f$  is of bounded type. We will however use these dynamical partitions to build a fine grid  $\{\mathcal{Q}_n\}$  for our map  $f$ . The construction requires some preliminary definitions.

An element  $I \in \mathcal{P}_n$  is a *saddle-node atom* if it is the disjoint union of some number  $a \geq 1000$  of atoms of  $\mathcal{P}_{n+1}$ .

Given two atoms  $\mathcal{P}_{n+1} \ni J \subseteq I \in \mathcal{P}_n$ , the *order* of  $J$  inside  $I$  is one plus the smallest number of atoms of  $\mathcal{P}_{n+1}$  on the right and left components of  $I \setminus J$ .

Note that inside a saddle-node atom  $I \in \mathcal{P}_n$  there are exactly two atoms of  $\mathcal{P}_{n+1}$  of order  $k$  for each  $k \leq a/2$ . Let  $N \geq 0$  be such that  $2^{N+1} < a/2$ . For each  $0 \leq i \leq N$ , we define  $M_i$ , the  *$i$ -th central interval* of  $I$ , to be the convex-hull  $[J, J^*] \subseteq I$  of the union of both atoms  $J, J^*$  of order  $2^i$ . Note that these central intervals are nested (see Figure 4). The left and right components of  $M_i \setminus M_{i+1}$ , respectively  $L_i$  and  $R_i$ , are the *lateral intervals* of  $I$ . The central interval  $M_N$  is also called the *final interval* of  $I$ . The lateral intervals together with the final interval form a special partition of  $I$ , the *balanced partition* of  $I$ .

*Remark.* It follows from Yoccoz's lemma that  $|L_i| \asymp |M_{i+1}| \asymp |R_i|$  for all  $i$ .

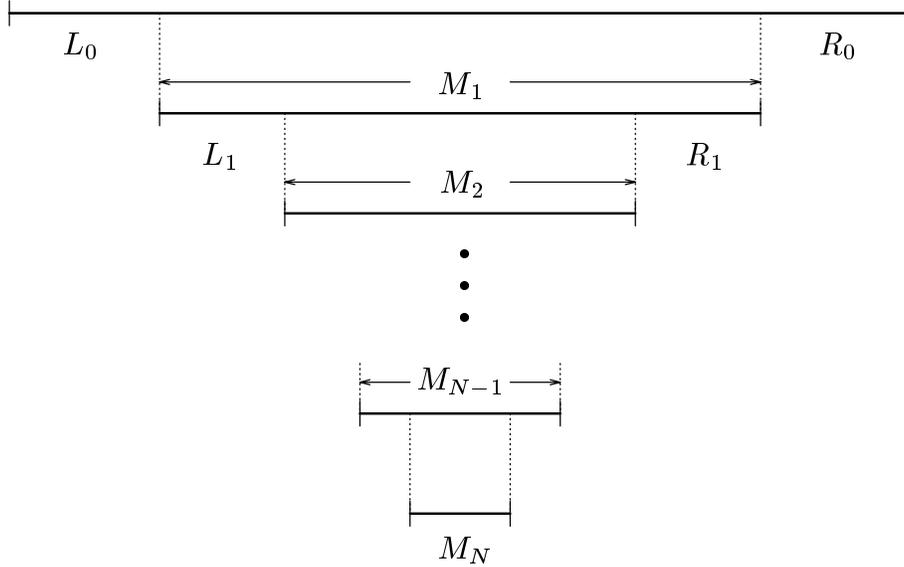


Figure 4. Central and lateral intervals of a saddle-node atom.

Now we define an auxiliary partition  $\tilde{\mathcal{P}}_n$ , for each  $n \geq 1$ . The atoms of  $\tilde{\mathcal{P}}_n$  are all atoms of  $\mathcal{P}_n$  which are not saddle-node, together with the atoms of the balanced partitions of all saddle-node atoms of  $\mathcal{P}_n$ . The partition  $\mathcal{Q}_n$  that we want is constructed from  $\tilde{\mathcal{P}}_n$  and  $\mathcal{P}_n$  as follows.

**Proposition 4.5** *There exists a fine grid  $\{\mathcal{Q}_n\}$  in  $S^1$  with the following properties.*

- (a) *Every atom of  $\mathcal{Q}_n$  is the union of at most 3 atoms of  $\mathcal{Q}_{n+1}$ .*
- (b) *Every atom  $\Delta$  of  $\mathcal{Q}_n$  is a union of atoms of  $\mathcal{P}_m$  for some  $m \leq n$ , and there are four possibilities:*
  - (b<sub>1</sub>)  $\Delta$  *is a single atom of  $\mathcal{P}_m$ ;*
  - (b<sub>2</sub>)  $\Delta$  *is a central interval of  $\tilde{\mathcal{P}}_m$ ;*
  - (b<sub>3</sub>)  $\Delta$  *is the union of at least two atoms of  $\mathcal{P}_{m+1}$  contained in a single atom of  $\tilde{\mathcal{P}}_m$ .*
  - (b<sub>4</sub>)  $\Delta$  *is a union of intervals which are simultaneously atoms of  $\mathcal{P}_m$  and  $\tilde{\mathcal{P}}_m$ .*

*Proof.* The proof is by induction on  $n$ . The first partition  $\mathcal{Q}_1$  consists of all atoms of  $\mathcal{P}_1$  which are not saddle-node atoms together with the intervals  $L_0$ ,  $M_1$  and  $R_0$  of each saddle-node interval  $I \in \mathcal{P}_1$  ( $I = L_0 \cup M_1 \cup R_0$ ). It is clear that each atom of  $\mathcal{Q}_1$  falls within one of the categories (b<sub>1</sub>)-(b<sub>4</sub>) above.

Assuming  $\mathcal{Q}_n$  defined, define  $\mathcal{Q}_{n+1}$  as follows. Take an atom  $I \in \mathcal{Q}_n$  and consider the four cases below.

- (1) If  $I$  is a single atom of  $\mathcal{P}_m$  then one of two things can happen:
  - (i)  $I$  is a saddle-node atom: In this case write  $I = L_0 \cup M_1 \cup R_0$  as above and take  $L_0$ ,  $R_0$  and  $M_1$  as atoms of  $\mathcal{Q}_{n+1}$ . Note that the lateral intervals  $L_0$  and  $R_0$  are atoms of type (b<sub>1</sub>), while the central interval  $M_1$  is of type (b<sub>2</sub>).

- (ii)  $I$  is not a saddle-node atom: In this case write  $I = L \cup M \cup R$  where  $L$  and  $R$  are the atoms of  $\mathcal{P}_{m+1}$  adjacent to the endpoints of  $I$  and  $M$  is the union of the other atoms of  $\mathcal{P}_{m+1}$  inside  $I$ . Add these three intervals to  $\mathcal{Q}_{n+1}$ , noting that  $L$  and  $R$  are of type  $(b_1)$ , while  $M$  is of type  $(b_4)$ .
- (2) If  $I$  is a central interval of  $\tilde{\mathcal{P}}_m$  which is not the final interval, consider the next central interval of  $\tilde{\mathcal{P}}_m$  inside  $I$ , say  $M$ , and the two corresponding lateral intervals  $L$  and  $R$  such that  $I = L \cup M \cup R$ , and declare  $L$ ,  $R$  and  $M$  members of  $\mathcal{Q}_{n+1}$ . Note that  $L$  and  $R$  are of type  $(b_3)$ , while  $M$  is of type  $(b_2)$ .
- (3) If  $I$  is a union of  $p \geq 2$  consecutive atoms  $\Delta_1, \dots, \Delta_p$  of  $\tilde{\mathcal{P}}_{m+1}$  inside a single atom of  $\mathcal{P}_m$ , divide it up into three approximately equal parts. More precisely, write  $p = 3q + r$  and, when  $r = 0$  or  $1$ , consider  $I = L \cup M \cup R$  where

$$L = \bigcup_{j=1}^q \Delta_j, \quad M = \bigcup_{j=q+1}^{p-q} \Delta_j, \quad R = \bigcup_{j=p-q+1}^p \Delta_j.$$

When  $r = 2$ , consider  $I = L \cup M \cup R$  where

$$L = \bigcup_{j=1}^{q+1} \Delta_j, \quad M = \bigcup_{j=q+2}^{p-q-1} \Delta_j, \quad R = \bigcup_{j=p-q}^p \Delta_j.$$

Note that  $M$  is empty when  $p = 2$ . In any case, we obtain two or three new atoms of  $\mathcal{Q}_{n+1}$  which are either single atoms of  $\mathcal{P}_{m+1}$ , and therefore of type  $(b_1)$ , or once again intervals of type  $(b_3)$ .

- (4) If  $I$  is a union of intervals which are simultaneously atoms of  $\mathcal{P}_m$  and  $\tilde{\mathcal{P}}_m$ , divide it up exactly as in (3), obtaining either two or three new atoms of  $\mathcal{Q}_{n+1}$  which are either single atoms of  $\mathcal{P}_m$ , and therefore of type  $(b_1)$ , or once again intervals of type  $(b_4)$ .

This completes the induction. That  $\{\mathcal{Q}_n\}_{n \geq 0}$  constitutes a fine grid follows easily from the real bounds and the remark preceding this proposition.  $\square$

An immediate consequence of the mere existence of such a fine grid is the fact that any two critical circle maps with the same rotation number are quasisymmetrically conjugate.

**Corollary 4.6** *Let  $f$  and  $g$  be critical circle maps with the same irrational rotation number, and let  $h$  be the conjugacy between  $f$  and  $g$  that maps the critical point of  $f$  to the critical point of  $g$ . Then  $h$  is quasisymmetric.*

*Proof.* Apply Proposition 4.3 (a) to the fine grid constructed above.  $\square$

**4.4 Proof of the First Main Theorem.** The strategy for the proof of our First Main Theorem is as follows. Given two critical circle maps  $f$  and  $g$  with the same rotation number, consider the special partitions  $\mathcal{Q}_n = \mathcal{Q}_n(f)$  and  $\tilde{\mathcal{Q}}_n = \mathcal{Q}_n(g)$  given by Proposition 4.5. The conjugacy  $h$  is an isomorphism between the corresponding fine grids, and we want to show that the coherence property (6) holds for  $h$  and  $\{\mathcal{Q}_n\}$ . To do this, we use the fact that the renormalizations  $f_n$  and  $g_n$  are exponentially close to prove that the vertices of  $\mathcal{Q}_{n+p}$  are exponentially close to the corresponding vertices of  $\tilde{\mathcal{Q}}_{n+p}$ , provided  $p$  is a small fraction of  $n$ . This property is proved in two steps: first for those vertices of  $\mathcal{Q}_n$  that lie in the domain of  $f_n$ , a step that works for arbitrary rotation numbers, and then propagated via Koebe's distortion principle to the other vertices. The first step holds without restriction on the rotation number, but the second does not. And it could not, indeed, as the counterexamples of section 5 will show.

In what follows we use the notation  $x_n = x_n(f) = f^{q_n}(x)$ . If the renormalizations  $f_n$  and  $g_n$  converge together exponentially fast then  $x_n(f)/x_n(g) \rightarrow 1$  and  $|x_n(f) - x_n(g)|/|x_n(f)| \rightarrow 0$  exponentially fast as well.

*Definition.* Let  $f_m : J_m \rightarrow J_m$  be the  $m$ -th renormalization of  $f$  and let  $k$  be an integer such that  $|k| \leq a_m/2$ . The *restricted domain* of  $f_m^k$ , denoted  $D_{m,k}$ , is defined as follows.

$$D_{m,k} = \begin{cases} I_m \cup [f^{\frac{a_m}{2}-k}(x_{m-1}), x_{m-1}], & \text{when } k > 0 \\ (I_m \setminus [x_m, f_m(x_{m+1})]) \cup [c, f_m^{\frac{a_m}{2}}(x_{m-1})], & \text{when } k = -1 \\ (I_m \setminus [x_m, f_m(x_{m+1})]) \cup [c, f_m^{-\frac{a_m}{2}-k}(x_{m+1})], & \text{when } k < -1. \end{cases}$$

**Lemma 4.7** For all  $x \in D_{m,k}$  we have  $|Df_m^k(x)| \leq C$ , where  $C > 0$  depends only on the real bounds.

*Proof.* Use Theorem 3.1 and Yoccoz's Lemma.  $\square$

**Lemma 4.8** Let  $v_0$  be a vertex of  $\mathcal{P}_{n+p}$  such that  $v_0 \in J_n$ . Then there exist  $n \leq m \leq n+p$  and  $1 \leq N \leq p$  such that  $v_0$  can be represented in the form

$$v_0 = \phi_1 \circ \phi_2 \circ \cdots \circ \phi_N(x_m),$$

where  $\phi_j = f_{m_j}^{k_j}$  for some  $n \leq m_j \leq n+p$  and  $|k_j| \leq a_{m_j}/2$ , and where the point  $\phi_{j+1} \circ \cdots \circ \phi_N(x_m)$  belongs to the restricted domain of  $\phi_j$  for each  $j$ . Moreover, if  $v_0$  is also a vertex of  $\mathcal{Q}_{n+p}$ , then  $|k_j| \leq 2^p$  for all  $j$ .

*Proof.* Let  $n \leq m_1 \leq n+p$  be largest with the property that  $v_0 \in J_{m_1} \setminus J_{m_1+1}$ , and let  $0 < i \leq a_{m_1}$  be such that  $f_{m_1}^i(v_0) \in J_{m_1+1}$ . If  $i \leq a_{m_1}/2$  then let

$k_1 = -i$ ; otherwise let  $k_1 = i - a_{m_1}$ . We get  $\phi_1 = f_{m_1}^{k_1}$  and a new vertex  $v_1 = f_{m_1}^{-k_1}(v_0) \in J_{m_1+1}$ . If  $v_1 \in J_{n+p}$  then  $v_1 = f_{n+p}(x_{n+p-1})$  necessarily, and we can stop. On the other hand, if  $v_1 \notin J_{n+p}$ , then once again there exists  $m_2$  in the range  $m_1 < m_2 < n + p$  such that  $v_1 \in J_{m_2} \setminus J_{m_2+1}$ , and we can proceed inductively. At the end of this process we get sequences  $m_1 < m_2 < \dots < m_N \leq n + p$  (so  $N \leq p$ ) and  $v_1, v_2, \dots, v_N$  with  $v_j \in J_{m_j} \setminus J_{m_j+1}$ , and for each  $j$  an integer  $k_j$  with  $|k_j| \leq a_{m_j}/2$  such that  $v_{j+1} = f_{m_j}^{-k_j}(v_j)$ . The last vertex  $v_N$  is necessarily  $x_m$  for some  $m \leq n + p$ . Hence it suffices to take  $\phi_j = f_{m_j}^{k_j}$  to get the desired representation. The last statement of the Lemma is immediate from the way the partitions  $\mathcal{Q}_n$  were constructed.  $\square$

Now we want to estimate  $|v_0 - w_0|$ , where  $v_0 \in J_n(f)$  is a vertex of  $\mathcal{Q}_{n+p}(f)$  and  $w_0$  is the corresponding vertex of  $\mathcal{Q}_{n+p}(g)$ . Here we assume  $p \leq \sigma n$  for some small  $\sigma > 0$  to be chosen later. By Lemma 4.8 above, there exist points  $x_m = x_m(f)$ ,  $y_m = x_m(g)$  and a number  $N \leq p$  such that

$$|v_0 - w_0| = |\phi_1 \circ \phi_2 \circ \dots \circ \phi_N(x_m) - \psi_1 \circ \psi_2 \circ \dots \circ \psi_N(y_m)| ,$$

where  $\phi_j = f_{m_j}^{k_j}$  and  $\psi_j = g_{m_j}^{k_j}$ , with  $n \leq m_j \leq n + p$  and  $|k_j| \leq 2^p$ .

For all  $n$ , the return map  $f_n$  is an almost parabolic map on  $J_{n-1} \setminus J_n$ , and similarly for  $g_n$ . Our hypothesis is that  $\|f_n - g_n\|_0 \leq C\lambda^n$  for all  $n$ , for some  $0 < \lambda < 1$ . If  $x$  is a point in the first fundamental domain of both  $f_{m_j}$  and  $g_{m_j}$ , then by Proposition 4.1 we have

$$\begin{aligned} |\phi_j(x) - \psi_j(x)| &\leq C\lambda^{m_j} k_j^3 \leq C\lambda^n 2^{3p} \\ &\leq C(\lambda 2^{3\sigma})^n = C\theta^n , \end{aligned}$$

where  $\theta = \lambda 2^{3\sigma} < 1$  if  $\sigma$  is small enough.

Using this, and since  $|x_m - y_m| \leq C\lambda^n$ , we see that

$$\begin{aligned} |\phi_N(x_m) - \psi_N(y_m)| &\leq |\phi_N(x_m) - \psi_N(x_m)| + |\psi_N(x_m) - \psi_N(y_m)| \\ &\leq C\theta^n + \|D\psi_N\|_0 |x_m - y_m| \leq C(\theta^n + \lambda^n) \leq C\theta^n , \end{aligned}$$

because  $\|D\psi_N\|_0$  is bounded, by Lemma 4.7. Proceeding inductively, we get

$$\begin{aligned} |\phi_1 \circ \dots \circ \phi_N(x_m) - \psi_1 \circ \dots \circ \psi_N(y_m)| &\leq (C + C^2 + \dots + C^N) \theta^n \\ &\leq C^p \theta^n \leq (C^\sigma \theta)^n , \end{aligned}$$

so making  $\sigma$  still smaller,  $|v_0 - w_0|$  is exponentially small as desired.

## 5. COUNTEREXAMPLES TO $C^{1+\alpha}$ RIGIDITY

Our purpose now is to construct  $C^\infty$  counterexamples to the conjectured  $C^{1+\alpha}$  rigidity of critical circle maps. We will consider critical circle maps whose rotation number  $\rho(f) = [a_0, a_1, \dots, a_n, \dots]$  satisfies

$$\begin{cases} \limsup \frac{1}{n} \log a_n = \infty \\ a_n \geq 2 \end{cases} \quad \text{for all } n, \quad (10)$$

The class of all rotation numbers satisfying (10) will be denoted by  $S$ . It can be shown that the Hausdorff dimension of  $S$  is less than or equal to  $1/2$ , see [6]. On the other hand,  $S$  contains Diophantine numbers: for example, the number  $\rho$  whose partial quotients are  $a_n = 2^{2^n}$  is Diophantine and satisfies (10).

**Theorem 5.1** *For every  $\rho \in S$  there exist  $C^\infty$  critical circle maps  $f, g$  with  $\rho(f) = \rho(g) = \rho$  such that  $f$  and  $g$  are not  $C^{1+\beta}$  conjugate for any  $\beta > 0$ .*

The proof will make use of a  $C^\infty$  surgery procedure that we explain next. These counterexamples have one additional feature: their successive renormalizations do converge together at an exponential rate. This will be clear from the construction.

**5.1 Saddle-node surgery.** Given  $f$  as above and a fixed  $n \geq 1$ , let  $J_n = J_n(f) = [f^{q_n}(c), f^{q_{n-1}}(c)] \subseteq S^1$  be the  $n$ -th renormalization interval of  $f$ . When  $a_n$  is very large, the first return map  $f_n : J_n \rightarrow J_n$  is an almost parabolic map of length  $a_n$ .

Let  $\Delta_1$  be the fundamental domain of this almost parabolic map which is adjacent to  $x_{n-1} = f^{q_{n-1}}(c)$ , and let  $\Delta_j = f_n^{j-1}(\Delta_1)$ ,  $j \leq a_n$ . Let  $z \in \Delta_1$  be the point such that  $f_n^{a_n}(z) = x_{n+2} = f^{q_{n+2}}(c)$ , that is,  $z = f^{q_{n+2}-a_n q_n}(c)$ . Note that since  $a_n \geq 2$ ,  $x_{n+2}$  is not an endpoint of  $f_n^{a_n}(\Delta_1)$ , and so by the real bounds it splits  $f_n^{a_n}(\Delta_1)$  into two intervals of comparable lengths. Hence the same holds for  $z$ . Namely,  $z$  splits  $\Delta_1$  into two intervals  $L, R$  with  $|L| \asymp |R|$ . In particular we have  $\tau|\Delta_1| \leq |L| \leq (1-\tau)|\Delta_1|$  (and similarly for  $R$ ) for some constant  $\tau$  depending on the real bounds; we use this fact in the proof of Proposition 5.2 below.

Consider now another critical circle map  $\tilde{f}$  with the same rotation number as  $f$ , the interval  $\tilde{J}_n = J_n(\tilde{f})$ , the first return map  $\tilde{f}_n : \tilde{J}_n \rightarrow \tilde{J}_n$ , the point  $\tilde{z} = \tilde{f}^{q_{n+2}-a_n q_n}(\tilde{c})$  and the corresponding intervals  $\tilde{L}, \tilde{R}$ . Also, let  $N = \lceil a_n/2 \rceil$ .

*Definition.* The number

$$\left| \frac{|f_n^{N-1}(L)|}{|f_n^{N-1}(R)|} - \frac{|\tilde{f}_n^{N-1}(\tilde{L})|}{|\tilde{f}_n^{N-1}(\tilde{R})|} \right|$$

is called the  $n$ -th order discrepancy between  $f$  and  $\tilde{f}$ .

**Proposition 5.2** *Given a  $C^\infty$  critical circle map  $f$  with  $\rho(f) \in S$ , consider a function  $\sigma(n) \rightarrow \infty$  such that*

$$\limsup \frac{1}{n\sigma(n)} \log a_n = \infty .$$

*Then for all  $n \geq 1$ , there exists a critical circle map  $\tilde{f} = F(n; f)$  with the same rotation number and critical point as  $f$  and having the following properties.*

- (a) *We have  $\tilde{f}^j(c) = f^j(c)$  for  $0 \leq j \leq q_{n+1}$ ; in particular,  $J_n(\tilde{f}) = J_n = J_n(f)$ .*
- (b) *We have  $\tilde{f} = \Phi \circ f$ , where  $\Phi$  is a  $C^\infty$  diffeo such that*

$$\|\Phi^{\pm 1} - \text{id}_{S^1}\|_{C^k} \leq B_k |J_n|^{\sigma(n)-k+1}$$

*for all  $k$ , where  $B_k > 0$  is constant depending only on  $k$ .*

- (c) *The  $n$ -th order discrepancy between  $f$  and  $\tilde{f}$  is  $\geq C|J_n|^{2\sigma(n)}$ .*
- (d) *We have  $J_{n+1}(\tilde{f}) = J_{n+1}(f)$  and  $\tilde{f}_{n+1} = f_{n+1}$ ; in particular,  $m$ -th order discrepancy between  $f$  and  $\tilde{f}$  is equal to zero for all  $m > n$ .*

*Proof.* We modify  $f$  inside  $f^{-1}(\Delta_1)$  using a  $C^\infty$  bump function so as to move  $z$  by a distance  $\geq C|\Delta_1|^{1+\sigma(n)}$  inside  $\Delta_1$ . This we do as follows.

Let  $\varphi : [0, 1] \rightarrow [0, 1]$  be a  $C^\infty$  perturbation of the identity such that  $|\varphi(x) - x| \geq |\Delta_1|^{\sigma(n)}$  for all  $\tau \leq x \leq 1 - \tau$  (and  $\tau$  as above), and such that  $|D^k \varphi(x)| \leq B_k |\Delta_1|^{\sigma(n)}$  for all  $0 \leq x \leq 1$  and all  $k \geq 2$ . Define  $\phi : \Delta_1 \rightarrow \Delta_1$  by  $\phi = A \circ \varphi \circ A^{-1}$  where  $A$  is the affine orientation-preserving map that carries  $[0, 1]$  onto  $\Delta_1$ . Note that  $|\phi(z) - z| \geq |\Delta_1|^{1+\sigma(n)}$ . Moreover, since  $D^k \phi = |\Delta_1|^{1-k} D^k \varphi$ , we have

$$\|\phi^{\pm 1} - \text{id}_{\Delta_1}\|_{C^k} \leq B_k |\Delta_1|^{\sigma(n)-k+1}$$

for all  $k$ . Define  $\psi : \Delta_{a_n} \rightarrow \Delta_{a_n}$  as the conjugate of  $\phi^{-1}$  by the diffeo  $f_n^{a_n-1} : \Delta_1 \rightarrow \Delta_{a_n}$ , namely

$$\psi = f_n^{a_n-1} \circ \phi^{-1} \circ (f_n^{a_n-1})^{-1} . \quad (11)$$

Using the  $C^m$  Approximation Lemma (see Appendix A), we see from (11) that

$$\|\psi^{\pm 1} - \text{id}_{\Delta_{a_n}}\|_{C^{k-1}} \leq C \|\phi^{\pm 1} - \text{id}_{\Delta_1}\|_{C^k} \leq B_k |\Delta_1|^{\sigma(n)-k+1} .$$

Define  $\Phi : S^1 \rightarrow S^1$  to be equal to  $\phi$  on  $\Delta_1$ , to  $\psi$  on  $\Delta_{a_n}$  and to the identity everywhere else. The critical circle map we look for is  $\tilde{f} = \Phi \circ f$ . Note that  $\|\Phi^{\pm 1} - \text{id}_{S^1}\|_{C^k} \leq B_k |\Delta_1|^{\sigma(n)-k+1}$  for all  $k$ ; since  $|\Delta_1| \asymp |J_n|$  by the real bounds, this proves (b). It is also clear from the construction that property (a) holds too. It follows in particular that the first  $n+1$  partial quotients of the rotation number

of  $\tilde{f}$  agree with those of  $f$ . More remarkable is that, because what  $\phi$  does is undone by  $\psi$ , we have

$$\begin{cases} \tilde{f}^{q_n} |I_{n+1} = f^{q_n} |I_{n+1} \\ \tilde{f}^{q_{n+1}} |I_n = f^{q_{n+1}} |I_n \end{cases} .$$

In other terms,  $\tilde{f}_n = f_n$ , the  $n$ -th renormalizations agree. Therefore all subsequent renormalizations agree as well. This shows that  $\rho(\tilde{f}) = \rho(f)$  and also proves (d).

It remains to prove (c), so we estimate the  $n$ -th order discrepancy between  $f$  and  $\tilde{f}$  from below. Since  $|z - \tilde{z}| \geq |\Delta_1|^{1+\sigma(n)}$ , a simple calculation yields

$$\left| \frac{|L|}{|R|} - \frac{|\tilde{L}|}{|\tilde{R}|} \right| \geq C |\Delta_1|^{\sigma(n)} \geq C |J_n|^{2\sigma(n)} , \quad (12)$$

provided  $n$  is sufficiently large. Since, by the real bounds, the map  $f_n^{N-1} : \Delta_1 \rightarrow \Delta_N$  has bounded distortion, and since  $\tilde{f}_n = f_n$ , (12) gives us

$$\left| \frac{|f_n^{N-1}(L)|}{|f_n^{N-1}(R)|} - \frac{|\tilde{f}_n^{N-1}(\tilde{L})|}{|\tilde{f}_n^{N-1}(\tilde{R})|} \right| \geq C |J_n|^{2\sigma(n)} ,$$

and this proves (c).  $\square$

**5.2 The counterexamples.** We now iterate the procedure given by Proposition 5.2 to prove our Second Main Theorem (that is, Theorem 5.1). We start with a  $C^\infty$  map  $f$  with  $\rho(f) \in S$  as before and select  $n_1 < n_2 < \dots$  such that

$$\lim_{i \rightarrow \infty} \frac{1}{n_i \sigma(n_i)} \log a_{n_i} = \infty , \quad (13)$$

where  $\sigma(n)$  is as in Proposition 5.2. Now we generate a sequence  $g_0, g_1, \dots, g_i, \dots$  recursively, starting with  $g_0 = f$ , and taking, for all  $i \geq 0$ ,  $g_{i+1} = F(n_{i+1}, g_i)$ , where  $F(\cdot, \cdot)$  is as given in Proposition 5.2. Each  $g_i$  is a  $C^\infty$  critical circle map with  $\rho(g_i) = \rho(f)$ , and  $g_{i+1} = \Phi_{i+1} \circ g_i$ , where  $\Phi_{i+1}$  is a  $C^\infty$  diffeo with

$$\|\Phi_{k+1}^{\pm 1} - \text{id}_{S^1}\|_{C^k} \leq B_k \theta^{n_i(\sigma(n_i) - k + 1)} , \quad (14)$$

for all  $k$ , where  $0 < \theta < 1$  is a constant depending only on the real bounds. From (14) it follows that  $\Phi = \lim \Phi_i \circ \dots \circ \Phi_1$  exists as a  $C^\infty$  diffeo, and therefore so does  $g = \lim g_i = \Phi \circ f$  as a critical circle map.

Using properties (c) and (d) of Proposition 5.2 for each  $g_i$ , we deduce that the  $n_i$ -th order discrepancy between  $f$  and  $g$  satisfies

$$\left| \frac{|f_{n_i}^{N_i-1}(L_{n_i})|}{|f_{n_i}^{N_i-1}(R_{n_i})|} - \frac{|g_{n_i}^{N_i-1}(\tilde{L}_{n_i})|}{|g_{n_i}^{N_i-1}(\tilde{R}_{n_i})|} \right| \geq C |J_{n_i}|^{2\sigma(n_i)} , \quad (15)$$

where  $N_i = \lceil a_{n_i}/2 \rceil$ , etc.

Now, let  $h : S^1 \rightarrow S^1$  be the conjugacy between  $f$  and  $g$  mapping the critical point  $c$  to itself. Suppose  $h$  were  $C^{1+\beta}$  for some  $\beta > 0$ . Then the left-hand side of (15) would be  $\leq C |f_{n_i}^{N_i-1}(\Delta_1^{(n_i)})|^\beta$ , where  $\Delta_1^{(n_i)} = L_{n_i} \cup R_{n_i}$ . But by Yoccoz's Lemma, we have

$$|f_{n_i}^{N_i-1}(\Delta_1^{(n_i)})| \asymp \frac{1}{N_i^2} |J_{n_i}| \asymp \frac{1}{a_{n_i}^2} |J_{n_i}|. \quad (16)$$

Combining the above with (15) and (16), we would get the inequality

$$a_{n_i}^{2\beta} |J_{n_i}|^{2\sigma(n_i)-\beta} \leq C.$$

But by the real bounds  $|J_n| \geq C\mu^n$  for all  $n$ , where  $0 < \mu < 1$ . Therefore, taking logarithms, we would have

$$\frac{\beta \log a_{n_i}}{n_i \sigma(n_i)} \leq \log \frac{1}{\mu},$$

but this clearly contradicts (13).  $\square$

## APPENDIX A. COMPACTNESS OF RENORMALIZATIONS

The real a-priori bounds proved in the section 3 have produced a very important corollary, namely, that the renormalizations of an arbitrary  $C^3$  critical circle map are uniformly bounded in the  $C^1$  topology. In this appendix we will use further a-priori estimates, this time involving the Schwarzian derivative, to prove that such renormalizations are uniformly bounded in the  $C^{r-1}$  topology when the critical circle map is  $C^r$ . Some technical tools are necessary.

**A.1 The  $C^m$ -Approximation Lemma.** In what follows,  $m \geq 1$  will be a fixed integer and  $I, J \subseteq \mathbb{R}$  fixed closed intervals. We denote by  $C^m(I)$  the Banach space of  $C^m$ -mappings  $f : I \rightarrow \mathbb{R}$  with the norm  $\|f\|_m = \max\{\|D^i f\|_0 : 0 \leq i \leq m\}$ , where  $\|\phi\|_0 = \sup_{x \in I} |\phi(x)|$ . Sometimes, when we need to emphasize the domain of  $f$ , we write  $\|f\|_{I,m}$  instead of  $\|f\|_m$ . We consider also the closed, convex subset  $C^m(I, J) \subseteq C^m(I)$  consisting of those  $f$ 's such that  $f(I) \subseteq J$ .

Recall Leibnitz's formula for the  $k$ -th derivative of a product of two functions,

$$D^k(uv) = \sum_{j=0}^k \binom{k}{j} D^j u D^{k-j} v,$$

from which it is clear that

$$\|uv\|_m \leq 2^m \|u\|_m \|v\|_m \quad (17)$$

whenever  $u, v \in C^m(I)$ . Something similar holds for the composition of two  $C^m$  mappings. Namely, we have Faa-di-Bruno's formula (cf. [10], p. 42)

$$D^k(f \circ g) = \sum_{j=1}^k B_{j,k}(D^1g, D^2g, \dots, D^jg) D^{k-j+1}f \circ g ,$$

where each  $B_{j,k}$  is a homogeneous polynomial of degree  $j$  on  $j$  variables whose coefficients are non-negative numbers depending only on  $k$  and  $j$ . It readily follows from this formula that if  $\psi \in C^m(I, J)$  and  $\phi \in C^m(J)$  then

$$\|\phi \circ \psi\|_m \leq A(m) \|\phi\|_m \sum_{k=1}^m \|\psi\|_m^k , \quad (18)$$

where  $A(m) = \max_{1 \leq k \leq m} \max_{1 \leq j \leq k} B_{j,k}(1, 1, \dots, 1)$ .

Another well-known fact we will need below is the following (cf. [5], Th. 3.1). Suppose  $m > 1$  and consider the composition operator  $(f, g) \mapsto f \circ g$  as a map  $\Theta : C^m(J) \times C^{m-1}(I, J) \rightarrow C^{m-1}(I)$ . Then  $\Theta$  is  $C^1$  and its Fréchet derivative is given by

$$D\Theta(f, g)(u, v) = u \circ g + v Df \circ g . \quad (19)$$

Note that  $C^m(J) \times C^{m-1}(I, J) \subseteq C^m(J) \times C^{m-1}(I)$ ; we consider this last product endowed with the norm

$$|(f, g)|_{I, J, m} = \max\{\|f\|_{J, m}, \|g\|_{I, m-1}\} .$$

**Lemma A.1** *For each  $M > 0$ , there exists  $c(M) > 0$  such that, if  $f_1, g_1 \in C^m(J)$  and  $f_2, g_2 \in C^{m-1}(I, J)$  and if  $|(f_1, f_2)|_{I, J, m} < M$  and  $|(g_1, g_2)|_{I, J, m} < M$ , then*

$$\|f_1 \circ f_2 - g_1 \circ g_2\|_{m-1} \leq c(M) |(f_1 - g_1, f_2 - g_2)|_{I, J, m} .$$

*Proof.* By the mean value theorem,

$$\|f_1 \circ f_2 - g_1 \circ g_2\|_{m-1} \leq \sup_{(\phi, \psi)} \|D\Theta(\phi, \psi)\| |(f_1 - g_1, f_2 - g_2)|_{I, J, m} ,$$

where the supremum is taken over all  $(\phi, \psi)$  in the line segment joining  $(f_1, f_2)$  to  $(g_1, g_2)$  inside  $C^m(J) \times C^{m-1}(I, J)$ , and where

$$\|D\Theta(\phi, \psi)\| = \sup \{ \|D\Theta(\phi, \psi)(u, v)\|_{m-1} : |(u, v)|_{I, J, m} \leq 1 \}$$

is the operator-norm of  $D\Theta(\phi, \psi)$ . Using (19), and then (17) and (18), we have

$$\begin{aligned} \|D\Theta(\phi, \psi)(u, v)\|_{m-1} &\leq \|u \circ \psi\|_{m-1} + \|v D\phi \circ \psi\|_{m-1} \\ &\leq (\|u\|_{m-1} + 2^{m-1} \|v\|_{m-1} \|D\phi\|_{m-1}) A(m-1) \sum_{k=1}^{m-1} \|\psi\|_{m-1}^k . \end{aligned}$$

From this, and taking into account that  $\|u\|_{m-1} \leq \|u\|_m \leq |(u, v)|_{I, J, m}$  as well as  $\|v\|_{m-1} \leq |(u, v)|_{I, J, m}$ , we deduce that

$$\|D\Theta(\phi, \psi)\| \leq A(m-1) (1 + 2^{m-1}\|D\phi\|_{m-1}) \sum_{k=1}^{m-1} \|\psi\|_{m-1}^k.$$

Finally, since  $\|D\phi\|_{m-1} \leq \|\phi\|_m$  and  $|(\phi, \psi)|_{I, J, m} < M$ , we get

$$\sup_{(\phi, \psi)} \|D\Theta(\phi, \psi)\| \leq A(m-1) (1 + 2^{m-1}M) \sum_{k=1}^{m-1} M^k = c(M). \quad \square$$

Let us denote by  $\mathbb{B}^m(I; M)$  the ball of radius  $M$  centered at the origin in  $C^m(I)$ .

**Lemma A.2** (*The  $C^m$ -Approximation Lemma*)

For each  $M > 0$ , there exist constants  $\varepsilon_M > 0$  and  $C_M > 0$  such that the following holds for all  $\varepsilon \leq \varepsilon_M$ . Let  $\Delta_1, \Delta_2, \dots, \Delta_{n+1}$  be closed intervals on the line or on the circle, and for each  $1 \leq i \leq n$  let  $f_i, g_i \in C^m(\Delta_i, \Delta_{i+1})$  be such that

(a) For all  $1 \leq j \leq k \leq n$ , we have  $f_k \circ f_{k-1} \circ \dots \circ f_j \in \mathbb{B}^m(\Delta_j; M)$ ;

(b) We have  $\sum_{i=1}^n \|f_i - g_i\|_m < \varepsilon$ .

Then for all  $k \leq n$  we have  $g_k \circ g_{k-1} \circ \dots \circ g_1 \in \mathbb{B}^{m-1}(\Delta_1; 2M)$ , and moreover

$$\|f_k \circ f_{k-1} \circ \dots \circ f_1 - g_k \circ g_{k-1} \circ \dots \circ g_1\|_{m-1} \leq C_M \sum_{j=1}^k \|f_j - g_j\|_m.$$

*Proof.* In the notation of Lemma A.1, let us write

$$C_M = \max\{1, c(2M), c(2M)c(3M)\}$$

and  $\varepsilon_M = M/C_M$ . We proceed by induction on  $k$ . When  $k = 1$ , we have  $\|f_1 - g_1\|_m \leq \varepsilon$  and there is nothing to prove. Suppose the assertion is valid for all  $j < k$ , and write

$$\begin{aligned} \|f_k f_{k-1} \dots f_1 - g_k g_{k-1} \dots g_1\|_{m-1} &\leq \\ &\sum_{j=1}^k \|f_k \dots f_{j+1} g_j g_{j-1} \dots g_1 - f_k \dots f_{j+1} f_j g_{j-1} \dots g_1\|_{m-1}. \end{aligned} \quad (20)$$

Since  $|f_j, g_{j-1} \circ \dots \circ g_1|_{\Delta_1, \Delta_j, m} < 2M$  and also  $|g_j, g_{j-1} \circ \dots \circ g_1|_{\Delta_1, \Delta_j, m} < 2M$ , it follows from Lemma A.1 that

$$\|f_j g_{j-1} \dots g_1 - g_j g_{j-1} \dots g_1\|_{m-1} \leq c(2M) \|f_j - g_j\|_m,$$

for  $j = 1, \dots, k$ . In particular, by the induction hypothesis, we have for all  $1 \leq j \leq k-1$

$$\|f_j g_{j-1} \cdots g_1\|_{m-1} \leq \|g_j g_{j-1} \cdots g_1\|_{m-1} + \varepsilon_M c(2M) < 3M .$$

Taking this back to (20) and applying Lemma A.1 again, we get

$$\begin{aligned} \|f_k f_{k-1} \cdots f_1 - g_k g_{k-1} \cdots g_1\|_{m-1} \\ \leq c(2M) \|f_k - g_k\|_m + c(2M)c(3M) \sum_{j=1}^{k-1} \|f_j - g_j\|_m \\ \leq C_M \sum_{j=1}^k \|f_j - g_j\|_m , \end{aligned}$$

and this shows also that  $\|g_k g_{k-1} \cdots g_1\|_{m-1} \leq M + \varepsilon_M C_M < 2M$ , thereby completing the induction.  $\square$

**A.2 Koebe principle revisited.** We present a generalization of the classical Koebe non-linearity principle. This principle states that if a  $C^3$  diffeomorphism has non-negative Schwarzian derivative on an open interval, then its non-linearity on any smaller closed subinterval with *space* on both sides is bounded. The generalized version below seems to be new. We denote by  $S\phi$  the Schwarzian derivative of  $\phi$ .

**Lemma A.3** *Given positive constants  $B$  and  $\tau$ , there exists  $K_{\tau, B} > 0$  such that the following holds. If  $\phi$  is a  $C^3$ -diffeomorphism of an interval  $I \supseteq [-\tau, 1 + \tau]$  into the reals and if  $S\phi(t) \geq -B$  for all  $t \in I$ , then for all  $t \in [0, 1]$  we have*

$$\left| \frac{\phi''(t)}{\phi'(t)} \right| \leq K_{\tau, B} .$$

*Proof.* Writing  $y = \phi''/\phi'$ , so that  $S\phi = y' - \frac{1}{2}y^2$ , we have the differential inequality

$$y' \geq \frac{1}{2}y^2 - B . \tag{21}$$

Let  $0 \leq t_0 \leq 1$  be a point where  $|y(t)|$  attains its maximum in  $[0, 1]$  and suppose  $y_0 = y(t_0)$  is such that  $|y_0| > \sqrt{2B} = \beta$ . If  $z(t)$  is the solution of the differential equation corresponding to (21) with initial condition  $z(t_0) = y_0$ , then by a well-known comparison theorem we must have  $y(t) \geq z(t)$  for all  $t \geq t_0$  and  $y(t) \leq z(t)$  for all  $t \leq t_0$ . Now, if  $y_0 > \beta$  then integration of the ODE leads to

$$z(t) = \beta \frac{(y_0 + \beta) + (y_0 - \beta)e^{\beta(t-t_0)}}{(y_0 + \beta) - (y_0 - \beta)e^{\beta(t-t_0)}} .$$

Since this solution explodes at time

$$t_1 = t_0 + \frac{1}{\beta} \log \left( \frac{y_0 + \beta}{y_0 - \beta} \right),$$

so does  $y(t)$ . Hence  $t_1 \notin I$ , i.e.  $t_1 - t_0 > \tau$ , which gives us

$$\frac{\phi''(t_0)}{\phi'(t_0)} = y_0 < \beta \frac{e^{\beta\tau} + 1}{e^{\beta\tau} - 1}.$$

If instead  $y_0 < -\beta$ , then we get

$$z(t) = \beta \frac{(\beta + y_0) - (\beta - y_0)e^{\beta(t-t_0)}}{(\beta + y_0) + (\beta - y_0)e^{\beta(t-t_0)}},$$

and arguing as before for  $t \leq t_0$  gives us

$$\frac{\phi''(t_0)}{\phi'(t_0)} = y_0 > -\beta \frac{e^{\beta\tau} + 1}{e^{\beta\tau} - 1}.$$

Therefore the lemma is proved if we take

$$K_{\tau,B} = \beta \frac{e^{\beta\tau} + 1}{e^{\beta\tau} - 1}. \quad \square$$

*Remark.* As  $B \rightarrow 0$ ,  $K_{\tau,B} \rightarrow 2/\tau$  and we recover the classical Koebe principle.

**A.3 Bounding the  $C^2$  norms.** As before, let  $f \in \text{Crit}^r(S^1)$ ,  $r \geq 3$ , be a critical circle map with critical point  $c$  of power-law  $p > 1$ . Conjugating  $f$  by a suitable  $C^r$ -diffeomorphism, we may assume that there exists a neighborhood  $\mathcal{U} \subseteq \mathbb{R}/\mathbb{Z}$  of  $c$  such that

$$f(x) = (x - c)|x - c|^{p-1} + a$$

for all  $x \in \mathcal{U}$ , where  $a$  is a constant. This will be our standing hypothesis on  $f$ , and we will sometimes say that  $f$  is a *canonical* circle map. Note in this case that for all  $x \in \mathcal{U} \setminus \{c\}$ , the Schwarzian derivative of  $f$  equals

$$Sf(x) = -\frac{p^2 - 1}{2(x - c)^2}. \quad (22)$$

We are interested in the maps  $f^{q_{n-1}-1}: I_n^1 \rightarrow I_n^{q_{n-1}}$  and  $f^{q_n-1}: I_{n-1}^1 \rightarrow I_{n-1}^{q_n}$ , for a fixed  $n \geq 1$ . They extend as diffeomorphisms to maximal open intervals  $J_{n,1}^- \supseteq I_n^1$  and  $J_{n,1}^+ \supseteq I_{n-1}^1$  respectively. When linearly rescaled to unit size, these diffeomorphisms are called the *coefficients* of the  $n$ -th renormalization of  $f$ .

Let us be more precise. Consider the  $n$ -th renormalization of  $f$ , namely the commuting pair  $\mathbf{f}_n : [\lambda_n, 1] \rightarrow \mathbb{R}$  defined in section 3. We write  $J_{n,i}^- = f^{i-1}(J_{n,1}^-)$  for each  $1 \leq i \leq q = q_{n-1}$  and  $J_{n,j}^+ = f^{j-1}(J_{n,1}^+)$  for each  $1 \leq j \leq Q = q_n$ . We also write  $J_{n,0}^- = f^{-1}(J_{n,1}^-)$  and  $J_{n,0}^+ = f^{-1}(J_{n,1}^+)$ . For each  $0 \leq j \leq Q$ , let  $\Lambda_j : \mathbb{R} \rightarrow \mathbb{R}/\mathbb{Z}$  be the affine (orientation-preserving) covering map such that  $\Lambda_j([0, 1]) = I_{n-1}^j$ . Let  $\Delta_n^-$  be the component of  $\Lambda_1^{-1}(J_{n,1}^-)$  that contains the interval  $[\lambda_n, 0]$ , and let  $\Delta_n^+$  be the component of  $\Lambda_1^{-1}(J_{n,1}^+)$  that contains the interval  $[0, 1]$ . Then define

$$\begin{cases} \mathcal{F}_n^- = \Lambda_0^{-1} f^{q-1} \Lambda_1 : \Delta_n^- \rightarrow \mathbb{R} \\ \mathcal{F}_n^+ = \Lambda_0^{-1} f^{Q-1} \Lambda_1 : \Delta_n^+ \rightarrow \mathbb{R} \end{cases} .$$

These are the  $n$ -th renormalization coefficients of  $f$ . Consider also the so-called *folding factors* of  $\mathbf{f}_n$ , namely the maps

$$\begin{cases} \varphi_n^- = \Lambda_1^{-1} f \Lambda_0 : \Lambda_0^{-1}(J_{n,0}^-) \rightarrow \mathbb{R} \\ \varphi_n^+ = \Lambda_1^{-1} f \Lambda_0 : \Lambda_0^{-1}(J_{n,0}^+) \rightarrow \mathbb{R} \end{cases} .$$

Each of these maps is a homeomorphism with a unique critical point at zero. One verifies at once that the maps  $\mathbf{F}_n^- = \mathcal{F}_n^- \circ \varphi_n^-$  and  $\mathbf{F}_n^+ = \mathcal{F}_n^+ \circ \varphi_n^+$  are  $C^r$  extensions of  $\mathbf{f}_n^-$  and  $\mathbf{f}_n^+$ , respectively.

It will be useful to express the coefficients  $\mathcal{F}_n^\pm$  as long compositions of rescaled diffeomorphisms in the following way. We will give the explicit decomposition for  $\mathcal{F}_n^+$ . A similar decomposition can be worked out for  $\mathcal{F}_n^-$ . Let us denote by  $\Delta_{n,j}^+$  the component of  $\Lambda_j^{-1}(J_{n,j}^+)$  containing the unit interval. Note in particular that  $\Delta_n^+ = \Delta_{n,1}^+$ . For each  $j$  in the range  $0 \leq j \leq Q - 1$ , let

$$f_j = \Lambda_{j+1}^{-1} f \Lambda_j : \Delta_{n,j}^+ \rightarrow \Delta_{n,j+1}^+ .$$

We call such maps the *elementary factors* of  $\mathcal{F}_n^+$ . Each  $f_j$  is a  $C^r$  diffeomorphism such that  $f_j([0, 1]) = [0, 1]$  (see Figure 5). We have of course  $\varphi_n^+ = f_0$ , but more importantly

$$\mathcal{F}_n^+ = (\Lambda_0^{-1} \circ \Lambda_Q) \circ (f_{Q-1} \circ \cdots \circ f_j \circ \cdots \circ f_1) . \quad (23)$$

We note also that for all  $t \in \Delta_{n,j}^+$

$$Sf_j(t) = Sf(\Lambda_j(t)) [D\Lambda_j(t)]^2 = Sf(\Lambda_j(t)) |I_{n-1}^j|^2 , \quad (24)$$

by the chain rule for the Schwarzian derivative.

*Notation.* Given  $J = [a, b] \subseteq \mathbb{R}$  and  $\tau > 0$ , we denote by  $J^\tau$  the interval  $[c, d] \supseteq J$  such that  $(a - c)/(b - a) = (d - b)/(b - a) = \tau$ . Note that  $J$  has *space* equal to  $\tau$  inside  $J^\tau$ .

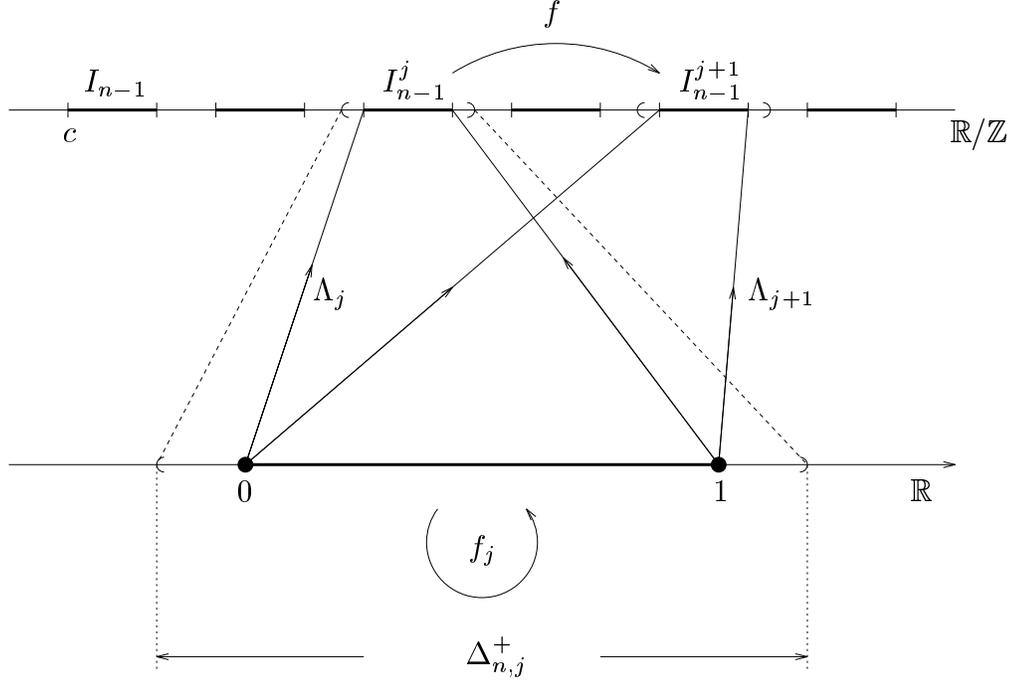


Figure 5. The elementary factors of  $\mathcal{F}_n^+$ .

**Theorem A.4** (The  $C^2$  bounds)

Let  $f \in \text{Crit}^3(S^1)$  be a critical circle map with arbitrary irrational rotation number, let  $\mathbf{f}_n : [\lambda_n, 1] \rightarrow \mathbb{R}$  be the  $n$ -th renormalization of  $f$ , and let  $\mathcal{F}_n^\pm : \Delta_n^\pm \rightarrow \mathbb{R}$  be the coefficients of  $\mathbf{f}_n$ . Also, let  $f_j : \Delta_{n,j}^+ \rightarrow \Delta_{n,j+1}^+$  be the elementary factors of  $\mathcal{F}_n^+$ . There exist positive constants  $B$  and  $\tau$  depending only on the real bounds for  $f$  such that the following statements hold for all  $n \geq 1$ .

- (a) We have  $\Delta_n^- \supseteq [\lambda_n, 0]^\tau$  and  $\Delta_n^+ \supseteq [0, 1]^\tau = [-\tau, 1 + \tau]$ .
- (b) For all  $0 \leq j \leq Q$ , we have  $\Delta_{n,j}^+ \supseteq [0, 1]^\tau$ .
- (c) We have  $|S\mathcal{F}_n^-(t)| \leq B$  for all  $t \in \Delta_n^-$  and  $|S\mathcal{F}_n^+(t)| \leq B$  for all  $t \in \Delta_n^+$ .
- (d) More generally, for all  $1 \leq j < k \leq Q$ , we have  $|S(f_k \circ \dots \circ f_j)(t)| \leq B$  for all  $t \in \Delta_{n,j}^+$ .
- (e) The  $C^2$  norms of the restrictions  $\mathcal{F}_n^-|_{[\lambda_n, 0]^{\tau/2}}$  and  $\mathcal{F}_n^+|_{[0, 1]^{\tau/2}}$  are bounded by  $B$ .
- (f) More generally, for all  $1 \leq j < k \leq Q$ , the  $C^2$  norm of the restriction of  $f_k \circ \dots \circ f_j$  to the interval  $f_{j-1} \circ \dots \circ f_1([0, 1]^{\tau/2})$  is bounded by  $B$ .
- (g) The  $C^2$  norms of  $\mathbf{f}_n^-$  and  $\mathbf{f}_n^+$  are bounded by  $B$ .

Moreover, if  $n$  is sufficiently large then both coefficients have negative Schwarzian derivatives at all points of their respective domains.

The proof will use the following lemma concerning the dynamical partitions  $\mathcal{P}_n$ . Let us denote by  $d(c, I)$  the distance between an interval  $I \subseteq S^1$  and the critical point  $c$ . For each  $n \geq 1$ , let

$$S_n = \sum_{I \in \mathcal{P}_n \setminus \{I_{n-1}, I_n\}} \left( \frac{|I|}{d(c, I)} \right)^2. \quad (25)$$

**Lemma A.5** *The sequence  $S_n$  is bounded (by a constant depending only on  $f$ ).*

*Proof.* Recall that  $\mathcal{P}_{n+2}$  is a strict refinement of  $\mathcal{P}_n$ . From the real bounds, we know that there exists a constant  $0 < \lambda < 1$  depending only on  $f$  such that, if  $I$  is in  $\mathcal{P}_n$  and  $J \subseteq I$  is in  $\mathcal{P}_{n+2}$ , then  $|J| \leq \lambda|I|$ . Hence

$$\sum_{I \supseteq J \in \mathcal{P}_{n+2}} |J|^2 \leq \left( \max_{I \supseteq J \in \mathcal{P}_{n+2}} |J| \right) |I| \leq \lambda |I|^2.$$

Since we also have  $d(c, J) \geq d(c, I)$  whenever  $J \subseteq I$ , it follows that

$$\begin{aligned} S_{n+2} &\leq \lambda S_n + \sum_{\mathcal{P}_{n+2} \ni J \subseteq I_{n-1} \setminus I_{n+1}} \left( \frac{|J|}{d(c, J)} \right)^2 + \sum_{\mathcal{P}_{n+2} \ni J \subseteq I_n \setminus I_{n+2}} \left( \frac{|J|}{d(c, J)} \right)^2 \\ &\leq \lambda S_n + \lambda \left( \frac{|I_{n-1}|}{|I_{n+1}|} \right)^2 + \lambda \left( \frac{|I_n|}{|I_{n+2}|} \right)^2. \end{aligned}$$

From this and the facts that  $|I_{n-1}| \asymp |I_{n+1}|$  and  $|I_n| \asymp |I_{n+2}|$ , we get  $S_{n+2} \leq \lambda S_n + \mu$ , where  $\mu$  is a constant depending only on  $f$ . But then, by induction,

$$S_{2n} \leq \lambda^{n-1} S_2 + \frac{\mu}{1-\lambda}, \quad S_{2n+1} \leq \lambda^n S_1 + \frac{\mu}{1-\lambda},$$

and therefore  $S_n$  is bounded as claimed.  $\square$

*Proof of Theorem A.4.* It is enough to prove this theorem under the assumption that  $f$  is canonical. The existence of  $\tau > 0$  such that (a) and (b) hold is a consequence of the real bounds. Hence we proceed to prove (c) for  $\mathcal{F}_n^+$ , the proof for  $\mathcal{F}_n^-$  being completely similar. Making  $\tau$  smaller if necessary and using the classical Koebe non-linearity principle, we can assume that there exists  $C > 0$  depending only on the real bounds for  $f$  such that

$$|D(f_j \cdots f_1)(t)| \leq C, \quad (26)$$

for all  $t \in [-\tau, 1 + \tau]$  and all  $j = 1, \dots, Q - 1$ .

Let  $\mathcal{V} \subseteq S^1$  be an open set whose closure does not contain  $c$  and such that  $\mathcal{U} \cup \mathcal{V} = S^1$ . Also, let  $M = \sup_{x \in \mathcal{V}} |Sf(x)|$ . We assume that  $n$  is so large that the largest interval in  $\mathcal{P}_n$  has length smaller than the Lebesgue number of the covering  $\{\mathcal{U}, \mathcal{V}\}$ . Together with (23) and (24), iterated use of the chain rule for the Schwarzian yields

$$\begin{aligned} S\mathcal{F}_n^+(t) &= S(f_{Q-1} \cdots f_j \cdots f_1)(t) \\ &= \sum_{j=1}^{Q-1} Sf_j(f_{j-1} \cdots f_1(t)) [D(f_{j-1} \cdots f_1)(t)]^2 \\ &= \sum_{j=1}^{Q-1} Sf(\Lambda_j f_{j-1} \cdots f_1(t)) |I_{n-1}^j|^2 [D(f_{j-1} \cdots f_1)(t)]^2 . \end{aligned}$$

We split this last sum into  $\Sigma_1(t) + \Sigma_2(t)$ , where  $\Sigma_1(t)$  is the sum over all  $j$ 's such that  $I_{n-1}^j \subseteq \mathcal{U}$  and  $\Sigma_2(t)$  is the sum over the remaining terms (i.e. those with  $I_{n-1}^j \subseteq \mathcal{V}$ ). Then we have on one hand

$$|\Sigma_2(t)| \leq C^2 M \sum_{I_{n-1}^j \subseteq \mathcal{V}} |I_{n-1}^j|^2 \leq C^2 M \max_{1 \leq j \leq Q-1} |I_{n-1}^j| . \quad (27)$$

On the other hand, since  $d(c, J_{n,j}^+) \asymp d(c, I_{n-1}^j)$  for all  $j$ , we have by (22)

$$|\Sigma_1(t)| \leq C^2 \sum_{I_{n-1}^j \subseteq \mathcal{U}} \frac{|I_{n-1}^j|^2}{[d(c, J_{n,j}^+)]^2} \leq C' S_n , \quad (28)$$

where  $C'$  is another constant depending only on  $f$  and  $S_n$  is given by (25). From (27) and (28) it follows that  $|S\mathcal{F}_n^+(t)|$  is uniformly bounded, and this proves (c). Moreover, since by (27)  $\Sigma_2(t)$  goes to zero with  $n$  while  $\Sigma_1(t)$  is always negative and bounded away from zero, we deduce that  $S\mathcal{F}_n^+(t) < 0$  for all  $n$  sufficiently large. The proof of (d) is entirely analogous.

To prove (e), let  $B_0$  be the upper-bound that we have just obtained for  $|S\mathcal{F}_n^+|$ . Applying Lemma A.3 to  $\mathcal{F}_n^+$ , we get for all  $t \in [0, 1]^{\tau/2}$

$$\left| \frac{D^2 \mathcal{F}_n^+(t)}{D\mathcal{F}_n^+(t)} \right| \leq K_{\tau_0, B_0} ,$$

where  $\tau_0 = \tau/2(1 + \tau)$  is the space of  $[0, 1]^{\tau/2}$  inside  $[0, 1]^\tau$ . Therefore  $\|D^2 \mathcal{F}_n^+\|_0 \leq K_{\tau_0, B_0} \|D\mathcal{F}_n^+\|_0 \leq CK_{\tau_0, B_0}$ , by (26) above. This shows that the  $C^2$  norm of  $\mathcal{F}_n^+$  is

bounded as claimed. A similar argument proves (f). Finally, (g) follows from (e) and the fact that the folding factors  $\varphi_n^\pm$  are linear blow-ups of a fixed power-law map. The theorem is therefore proved if we take  $B$  to be the largest of all the upper-bounds obtained in the argument.  $\square$

*Remark.* We can go a bit further in (e), (f) and (g) and bound also the  $C^3$  norms. For this purpose, it suffices to note for instance that

$$D^3 \mathcal{F}_n^+(t) = D\mathcal{F}_n^+(t) \left( S\mathcal{F}_n^+(t) + \frac{3}{2} \left[ \frac{D^2 \mathcal{F}_n(t)}{D\mathcal{F}_n(t)} \right]^2 \right),$$

and then use (c) and (e). However, this argument does not generalize to get bounds for higher derivatives. Our bootstrap argument in the next section will follow a different route, based on the  $C^m$  Approximation Lemma.

**A.4 Bounding the  $C^{r-1}$  norms.** We will show that the sequence of renormalizations of a  $C^r$  critical circle map is bounded in the  $C^{r-1}$  sense. The limits fall into (a compact subset of) a special family of analytic critical commuting pairs known as the Epstein class. Moreover, we will prove that such limits are attained at an exponential rate in the  $C^{r-1}$  topology. The rate of convergence turns out to depend only on the rotation number of the given critical circle map.

An *Epstein map* is a homeomorphism  $\varphi : I \rightarrow J$  between closed intervals on the real line such that  $\varphi^{-1}$  is the restriction of an analytic univalent map  $\Phi : \mathbb{C}(J') \rightarrow \mathbb{C}(I')$ , where  $I' \supseteq I$  and  $J' \supseteq J$  are open intervals. Here we use the notation  $\mathbb{C}(\Delta) = (\mathbb{C} \setminus \mathbb{R}) \cup \Delta$ . For example, every fractional linear transformation in  $PSL_2(\mathbb{R})$  is an Epstein map when restricted to an interval on the line which does not contain any of its poles. Further examples include polynomial or rational diffeomorphisms with real coefficients.

*Definition.* A commuting pair  $\mathbf{f}$  is said to be an *Epstein commuting pair* if  $\mathbf{f}^+ = \varphi^+ \circ Q$  and  $\mathbf{f}^- = \varphi^- \circ Q$ , where  $\varphi^+, \varphi^-$  are Epstein maps and  $Q$  is the power-law map  $x \mapsto x^p$  (for some  $p > 1$ ).

**Theorem A.6** *Let  $r \geq 3$  and let  $f$  be a  $C^r$  critical circle map with arbitrary irrational rotation number. Then the sequence of renormalizations  $\{\mathcal{R}^n(f)\}$  is bounded in the  $C^{r-1}$  metric and converges  $C^{r-1}$  exponentially fast to the Epstein class.*

The idea behind the proof of Theorem A.6 is quite simple. In the long composition defining the  $n$ -th renormalization of a critical circle map, we replace the factors away from the critical point by suitable fractional linear approximations, which are all Epstein maps. The factors which are close to the critical point are already Epstein because the map is assumed to be a power-law there. Therefore the entire new composition is an Epstein map. The Moebius approximations have

to be carefully chosen, however, so that the total error involved, estimated with the help of the  $C^m$  Approximation Lemma, be exponentially small in  $n$  (the step of renormalization). We now present the technical result which is needed.

**Lemma A.7** *Given  $r \geq 3$  and an orientation preserving  $C^r$ -diffeomorphism  $\phi : I \rightarrow \mathbb{R}$  of a closed interval  $I$  onto its image, there exist constants  $\ell_\phi > 0$  and  $K_\phi > 0$  with the following property. For each closed interval  $\Delta \subseteq I$  of length  $|\Delta| \leq \ell_\phi$ , there exists a fractional linear transformation  $T_\Delta \in PSL_2(\mathbb{R})$  with  $T_\Delta(\Delta) = \phi(\Delta)$  such that,*

- (a)  $\sup_{x \in \Delta} |D^k \phi(x) - D^k T_\Delta(x)| \leq K_\phi |\Delta|^{3-k}$  for  $k = 0, 1, 2$ .
- (b)  $\sup_{x \in \Delta} |D^k T_\Delta(x)| \leq K_\phi$  for all  $1 \leq k \leq r$ .

*Proof.* Let  $\ell_\phi$  be the constant

$$\ell_\phi = \min \left\{ 1, \inf_{x \in I} \left| \frac{\phi'(x)}{\phi''(x)} \right| \right\} .$$

Take any closed interval  $\Delta \subseteq I$  with  $|\Delta| \leq \ell_\phi$ , and let  $x_0$  be the left endpoint of  $\Delta$ . Let  $T$  be the unique fractional linear transformation with the same 2-jet as  $\phi$  at  $x_0$ . Thus, if  $T(x) = (a(x - x_0) + b)/(c(x - x_0) + d)$  with  $ad - bc = 1$ , then the coefficients are uniquely determined by the conditions

$$\begin{aligned} T(x_0) &= \frac{b}{d} = \phi(x_0) , \\ T'(x_0) &= \frac{1}{d^2} = \phi'(x_0) , \\ T''(x_0) &= \frac{-2c}{d^3} = \phi''(x_0) . \end{aligned} \tag{29}$$

Moreover, for all  $k \geq 1$ ,

$$D^k T(x) = \frac{(-1)^{k+1} k! c^{k-1}}{[c(x - x_0) + d]^{k+1}} . \tag{30}$$

Since  $|x - x_0| \leq |\Delta| \leq \ell_\phi \leq |\phi'(x_0)|/|\phi''(x_0)| = |d|/2|c|$  for all  $x \in \Delta$ , we have

$$\frac{1}{2}|d| \leq |c(x - x_0) + d| \leq \frac{3}{2}|d| \tag{31}$$

for each such  $x$ . Combining (30) with the lower bound in (31), we get

$$|D^k T(x)| \leq \frac{2^{k+1} k! |c|^{k-1}}{|d|^{k+1}} = \frac{4k! |\phi''(x_0)|^{k-1}}{|\phi'(x_0)|^{k-2}} ,$$

for all  $x \in \Delta$  and all  $k \geq 1$ , and consequently

$$\sup_{x \in \Delta} |D^k T(x)| \leq C_0 = \max_{1 \leq k \leq r} \sup_{x \in I} \left\{ \frac{4k! |\phi''(x)|^{k-1}}{|\phi'(x)|^{k-2}} \right\}, \quad (32)$$

when  $1 \leq k \leq r$ . In particular, from

$$D^2 \phi(x) - D^2 T(x) = \int_{x_0}^x D^3 \phi(t) dt - \int_{x_0}^x D^3 T(t) dt,$$

we deduce that

$$\begin{aligned} |D^2 \phi(x) - D^2 T(x)| &\leq \|D^3 \phi\|_0 |x - x_0| + \frac{24 |\phi''(x_0)|^2}{|\phi'(x_0)|} |x - x_0| \\ &\leq (\|D^3 \phi\|_0 + C_0) |\Delta|, \end{aligned}$$

for all  $x \in \Delta$ . Integrating this inequality twice, using (29), we get

$$\sup_{x \in \Delta} |D^k \phi(x) - D^k T(x)| \leq C_1 |\Delta|^{3-k}, \quad (33)$$

for  $k = 0, 1, 2$ , where  $C_1 = C_0 + \|D^3 \phi\|_0$ .

Looking at (32) and (33), we see that  $T$  is almost what we want, but not quite because in general it does not map  $\Delta$  onto  $\phi(\Delta)$ . To correct this flaw, we replace  $T$  by  $T_\Delta = A \circ T$ , where  $A$  is the unique affine, orientation-preserving map that carries  $T(\Delta)$  onto  $\phi(\Delta)$ . We have

$$A(t) - t = \left[ \frac{|\phi(\Delta)|}{|T(\Delta)|} - 1 \right] (t - T(x_0)), \quad (34)$$

for all  $t \in T(\Delta)$ , because  $\phi(x_0) = T(x_0)$ . Let  $\mu = |\phi(\Delta)|/|T(\Delta)|$ . Since by (33) we have  $||\phi(\Delta)| - |T(\Delta)|| \leq 2C_1 |\Delta|^3$ , and since by the upper-bound in (31) we have

$$\frac{|T(\Delta)|}{|\Delta|} \geq \inf_{x \in \Delta} \frac{1}{[c(x - x_0) + d]^2} \geq \frac{4}{9d^2} = \frac{4}{9} |\phi'(x_0)|,$$

it follows that

$$|\mu - 1| \leq \frac{9C_1}{2|\phi'(x_0)|} |\Delta|^2.$$

Thus we see that, for all  $t \in T(I)$ ,

$$|A'(t) - 1| = |\mu - 1| \leq \frac{9C_1}{2 \inf_{x \in I} |\phi'(x)|} |\Delta|^2 = C_2 |\Delta|^2,$$

On the other hand, since  $|T(\Delta)| \leq \|D\phi\|_0|\Delta| + 2C_1|\Delta|^3$ , and since  $|\Delta| \leq 1$ , it follows from (34) that

$$|A(t) - t| \leq C_2 (\|D\phi\|_0 + 2C_1) |\Delta|^3 = C_3 |\Delta|^3 .$$

Therefore

$$|\phi(x) - T_\Delta(x)| \leq |\phi(x) - T(x)| + |T(x) - A(T(x))| \leq (C_1 + C_3) |\Delta|^3 , \quad (35)$$

and moreover, using the fact that  $D^k T_\Delta(x) = \mu D^k T(x)$  for all  $k$ ,

$$\begin{aligned} |D^k \phi(x) - D^k T_\Delta(x)| &\leq |D^k \phi(x) - D^k T(x)| + |\mu - 1| |D^k T(x)| \\ &\leq C_1 |\Delta|^{3-k} + C_0 C_2 |\Delta|^2 \\ &\leq (C_1 + C_0 C_2) |\Delta|^{3-k} , \end{aligned} \quad (36)$$

for all  $x \in \Delta$  and  $k = 1, 2$ . Finally, for all  $k \geq 1$  we have

$$|D^k T_\Delta(x)| \leq (1 + C_2 |\Delta|^2) |D^k T(x)| \leq (1 + C_2) C_0 . \quad (37)$$

Part (a) now follows from (35) and (36), while part (b) follows from (37), provided we take  $K_\phi = \max\{C_1 + C_3, C_1 + C_0 C_2, (1 + C_2) C_0\}$ .  $\square$

*Proof of Theorem A.6.* We now expand the outline given above and present a complete proof of Theorem A.6. In the proof, we will denote by  $C_0, C_1, \dots$  positive constants depending only on the real bounds for  $f$ . As before, we may assume from the start that  $f$  is *canonical*, and accordingly we consider the covering  $\{\mathcal{U}, \mathcal{V}\}$  of  $S^1$  defined in the proof of Theorem A.4. Since the folding factors of  $\mathbf{f}_n$  are power-law maps, and therefore already Epstein, it suffices to prove that the *coefficients* of  $\mathbf{f}_n$  can be approximated by Epstein maps, up to an error exponentially small in  $n$  in the  $C^{r-1}$  topology. We will do this for  $\mathcal{F}_n^+$ , the proof for  $\mathcal{F}_n^-$  being the same.

As in the previous section, let  $f_j : \Delta_{n,j}^+ \rightarrow \Delta_{n,j+1}^+$ ,  $1 \leq j \leq Q - 1$ , be the elementary factors of  $\mathcal{F}_n^+$ . For each  $1 \leq j \leq Q$  we define

$$\Delta_j = f_{j-1} \circ \dots \circ f_2 \circ f_1 \left( [0, 1]^{\tau/4} \right) \subseteq \Delta_{n,j}^+ ,$$

where  $\tau$  is the constant of Theorem A.4. Note that  $f_j(\Delta_j) = \Delta_{j+1}$ . Let  $\Delta'_j = \Lambda_j(\Delta_j)$ , and observe also that  $I_{n-1}^j \subseteq \Delta'_j \subseteq J_{n,j}^+$ .

We introduce individual Epstein approximations  $g_j$  to each  $f_j$ . There are two cases to consider. It may happen that  $\Delta'_j \subseteq \mathcal{U}$ , in which case we simply take  $g_j = f_j$ . Otherwise, we have  $\Delta'_j \subseteq \mathcal{V}$ . In this case, we let  $T_j : \Delta'_j \rightarrow \Delta'_{j+1}$  be the Moebius approximation to  $f|_{\Delta'_j}$  that we get applying Lemma A.7 to the restriction of  $f$  to  $\mathcal{V}$ , and then take  $g_j = \Lambda_{j+1}^{-1} \circ T_j \circ \Lambda_j$ . Note that  $g_j(\Delta_j) = \Delta_{j+1}$ .

CLAIM 1. We have  $\|f_j - g_j\|_r \leq C_0 |I_{n-1}^j|^2$  for all  $j$ .

This is obvious when  $I_{n-1}^j \subseteq \mathcal{U}$ . When  $I_{n-1}^j \subseteq \mathcal{V}$ , we have  $|I_{n-1}^j| \asymp |I_{n-1}^{j+1}|$ , because the derivative of  $f$  on  $\mathcal{V}$  is bounded away from zero, and we also have  $|\Delta'_j| \asymp |I_{n-1}^j|$ . Moreover, for all  $1 \leq s \leq r$  and all  $x \in \Delta_j$ ,

$$D^s f_j(x) - D^s g_j(x) = \frac{|I_{n-1}^j|^s}{|I_{n-1}^{j+1}|} (D^s f(\Lambda_j(x)) - D^s T_j(\Lambda_j(x))) .$$

Therefore the claim follows from Lemma A.7 (treat the cases  $s = 1, 2$  separately).

Now, recall from Theorem A.4 that for all  $1 \leq j < k \leq Q - 1$  we have

$$\|f_k \circ \cdots \circ f_j\|_2 \leq B .$$

CLAIM 2. If  $n$  is sufficiently large then for all  $1 \leq j < k \leq Q - 1$  we have

$$\|f_k \circ \cdots \circ f_j - g_k \circ \cdots \circ g_j\|_1 \leq C_1 \max_{0 \leq i \leq Q} |I_{n-1}^i| . \quad (38)$$

Take  $n_0$  so large that  $C_0 \max |I_{n_0-1}^j| < \varepsilon_B$ , where  $\varepsilon_B$  is the constant given by Lemma A.2 when we take  $M = B$ . Then from Claim 1 and (38), the hypotheses of Lemma A.2 are satisfied, and we get for all  $n \geq n_0$

$$\begin{aligned} \|f_k \circ \cdots \circ f_j - g_k \circ \cdots \circ g_j\|_1 &\leq C_B \sum_{i=j}^k \|f_i - g_i\|_2 \leq C_0 C_B \sum_{i=j}^k |I_{n-1}^i|^2 \\ &\leq C_0 C_B \max_{0 \leq i \leq Q} |I_{n-1}^i| , \end{aligned}$$

where  $C_B$  is the constant of Lemma A.2 for  $M = B$ . This proves the claim.

In order to bootstrap these  $C^1$  estimates up to  $C^{r-1}$  estimates, we apply the  $C^m$  Approximation Lemma once more, this time reversing the roles of  $f_j$  and  $g_j$ , and with  $m = r$ . Thus, we need to verify the hypotheses of that lemma in this new situation.

CLAIM 3. For all  $1 \leq j < k \leq Q - 1$ , we have  $\|g_k \circ \cdots \circ g_j\|_r \leq C_2$ .

For brevity, write  $G_{jk} = g_k \circ \cdots \circ g_j$ . Then  $G_{jk}^{-1}$  is univalent on  $\mathbb{C}(\Delta_{jk})$ , where  $\Delta_{jk}$  is an interval containing  $G_{jk}(\Delta_j)$  with definite space on both sides, by our choice of  $\tau$ . Using Koebe's one-quarter theorem, it is not difficult to see that the domain  $\Omega_{jk} = G_{jk}^{-1}(\mathbb{C}(\Delta_{jk}))$  contains a rectangle  $W_j = \Delta_j^\alpha \times [-\beta, \beta]$ , and that  $d(\partial W_j, \partial \Omega_{jk}) \geq \gamma$  where  $\alpha, \beta$  and  $\gamma$  are positive constants depending only on  $\tau$

and the real bounds for  $f$ . Hence, from the complex Koebe's distortion theorem, we get

$$\left| \frac{G'_{jk}(z)}{G'_{jk}(w)} \right| \leq \exp \left\{ \frac{2}{\gamma} \operatorname{diam}(W_j) \right\} \leq C_3 ,$$

for all  $z, w \in W_j$ . This together with the mean-value theorem gives us  $|G'_{jk}(z)| \leq C_4$ , and therefore also  $|G_{jk}(z)| \leq C_5$ , for all  $x \in W_j$ . Now we use Cauchy's integral formula to bound all higher derivatives of  $G_{jk}$ . We have for all  $x \in \Delta_j$  and all  $s \geq 1$

$$|D^s G_{jk}(x)| = \frac{s!}{2\pi} \left| \int_{\partial W_j} \frac{G_{jk}(z)}{(z-x)^{s+1}} dz \right| \leq \frac{C_5 s!}{\pi} (\beta + (1+2\alpha)|\Delta_j|) \delta_j^{-s-1} ,$$

where  $\delta_j = \inf_{x \in \Delta_j} d(x, \partial W_j) = \min\{\alpha|\Delta_j|, \beta\} \geq \delta = \min\{a, b\}$ . Therefore  $|D^s G_{jk}(x)| \leq C_6 s! \delta^{-s-1}$ . This shows that  $\|G_{jk}\|_r$  is bounded as claimed.

From Claims 1 and 3, the hypotheses of Lemma A.2 are therefore satisfied, and we have

$$\begin{aligned} \|f_k \circ \cdots \circ f_j - g_k \circ \cdots \circ g_j\|_{r-1} &\leq C_{C_2} \sum_{i=j}^k \|f_i - g_i\|_r \leq C_0 C_{C_2} \sum_{i=j}^k |I_{n-1}^i|^2 \\ &\leq C_0 C_{C_2} \max_{0 \leq i \leq Q} |I_{n-1}^i| , \end{aligned}$$

this time for all  $n$  large enough so that  $C_0 \max |I_{n-1}^j| < \varepsilon_{C_2}$ , where  $C_{C_2}$  and  $\varepsilon_{C_2}$  are the constants of Lemma A.2 for  $M = C_2$ . Since  $\max |I_{n-1}^j|$  decreases exponentially with  $n$ , we are done.  $\square$

## APPENDIX B. PROOF OF YOCOZ'S LEMMA

The main geometric idea behind the proof of Yoccoz's Lemma is to use the negative Schwarzian property of  $f$  to *squeeze* the graph of  $f$  between the graphs of two Moebius transformations. The required estimate for  $f$  will then follow from the corresponding estimate for Moebius transformations, which we now state and prove.

Consider the fractional linear transformation  $T(x) = x/(1+x)$ , and given  $\varepsilon > 0$ , let  $T_\varepsilon(x) = T(x) - \varepsilon$ . We are interested in certain quantitative aspects of the orbit  $x_n = T_\varepsilon^n(x_0)$  for  $x_0 = 1$ . Observe that this sequence is strictly decreasing.

**Lemma B.1** *Let  $N > 0$  be such that  $x_{N+1} \leq 0 < x_N$ . Then we have  $N \asymp 1/\sqrt{\varepsilon}$  and moreover  $x_n - x_{n+1} \asymp 1/n^2$  for  $n = 0, 1, \dots, N$ .*

*Proof.* Writing  $\delta_n = T^n(x_0) - T_\varepsilon^n(x_0)$ , we have

$$\delta_n = \varepsilon + \frac{\delta_{n-1}}{(1 + \frac{1}{n})(1 + \frac{1}{n} - \delta_{n-1})} \tag{39}$$

for all  $n = 1, 2, \dots, N + 1$ . We claim that

$$\frac{n\varepsilon}{6} \leq \delta_n \leq n\varepsilon . \quad (40)$$

The last inequality is clear. To prove the first, we note from (39) that

$$\delta_n \geq \varepsilon + \left( \frac{n}{n+1} \right)^2 \delta_{n-1} .$$

By induction, this gives us

$$\delta_n \geq \frac{\varepsilon}{(n+1)^2} (1^2 + 2^2 + \dots + n^2) = \frac{\varepsilon}{(n+1)^2} \frac{n(n+1)(2n+1)}{6} \geq \frac{n\varepsilon}{6} ,$$

which proves the claim. Now, from the fact that  $x_{N+1} \leq 0 < x_N$  we have the inequalities

$$\delta_N < \frac{1}{N+1} , \quad \delta_{N+1} \geq \frac{1}{N+2} .$$

Then, using (40), we get

$$\frac{1}{(N+1)(N+2)} \leq \varepsilon < \frac{6}{N(N+1)} , \quad (41)$$

which proves the first assertion.

Next, note that since  $[x_{N+1}, x_N] \subseteq [T_\varepsilon(0), T_\varepsilon^{-1}(0)] = [-\varepsilon, \varepsilon/(1-\varepsilon)]$ , we have

$$\varepsilon < x_N - x_{N+1} < 3\varepsilon \quad (42)$$

Hence, by (41), we get  $x_N - x_{N+1} \asymp 1/N^2$  and the second assertion is proved when  $n = N$ . To prove it in general using this information, observe that

$$x_n - x_{n+1} = \frac{x_{n-1} - x_n}{(1+x_{n-1})(1+x_n)} = \frac{x_{n-1} - x_n}{(1 + \frac{1}{n} - \delta_{n-1})(1 + \frac{1}{n+1} - \delta_n)}$$

implies

$$x_n - x_{n+1} \geq \frac{n}{n+2} (x_{n-1} - x_n) .$$

By induction, this gives on one hand

$$x_n - x_{n+1} \geq \frac{2}{(n+1)(n+2)} (x_0 - x_1) \geq \frac{1}{(n+1)(n+2)} ,$$

and on the other hand, using (41) and (42),

$$x_n - x_{n+1} \leq (x_N - x_{N+1}) \prod_{j=1}^{N-n} \left( \frac{n+j+2}{n+j} \right) < \frac{54}{(n+1)(n+2)} .$$

This proves the second assertion in all cases.  $\square$

Now recall that  $f : \Delta_1 \cup \Delta_2 \cup \dots \cup \Delta_a \rightarrow \mathbb{R}$  satisfies  $f(\Delta_j) = \Delta_{j+1}$  for all  $j$ . Without loss of generality, we can assume that  $f(x) < x$  for all  $x$ . Thus, if we call  $x_0$  the right endpoint of  $\Delta_1$  and write  $x_j = f^j(x_0)$ , we have  $\Delta_j = [x_j, x_{j-1}]$  for all  $j$ . Since our map  $f$  is a negative-Schwarzian diffeomorphism, there exists a unique  $z$  in the domain of  $f$  such that  $\varepsilon = |f(z) - z| \leq |f(x) - x|$  for all  $x$ . Since the statement we want to prove is invariant under affine changes of coordinates, we may assume also that  $z = 0$  and  $x_0 = 1$ . In this setting, we want to prove that  $|\Delta_j| \asymp 1/j^2$  for all  $j$  such that  $\Delta_j \subseteq [0, 1]$ . Note that  $f'(0) = 1$ .

Next, let  $A$  be the Moebius transformation on the line such that  $A(x_0) = f(x_0)$  and  $A(0) = f(0)$  and  $A'(0) = f'(0) = 1$ . This determines  $A$  uniquely, and in fact

$$A(x) = \frac{x}{1 + \lambda x} - \varepsilon ,$$

for some  $\lambda > 0$ . Since  $Sf < 0$ , we see that  $A(x) \leq f(x)$  for all  $x \in [0, 1]$ .

Likewise, let  $B$  be the Moebius transformation such that  $B(x_a) = f(x_a)$ ,  $B(0) = f(0)$  and  $B'(0) = f'(0) = 1$ . This determines  $B$  uniquely, and in fact

$$B(x) = \frac{x}{1 + \mu x} - \varepsilon ,$$

for some  $\mu > 0$ . This time, since  $x_a < 0$  and  $Sf < 0$ , we have  $f(x) \leq B(x)$  for all  $x \in [0, 1]$ . In particular,  $\lambda > \mu$ . It is easy to see that  $\lambda/\mu \leq c_\sigma$ , where  $c_\sigma$  depends only on the constant  $\sigma$  in the statement.

**Lemma B.2** *Let  $x \in [0, 1]$  and  $k > 0$  be such that  $A(x) < B^k(x)$ . Then  $k \leq 1 + \lambda/\mu$ .*

*Proof.* By induction we have

$$B^k(x) \leq \frac{x}{1 + (k-1)\mu x} - \varepsilon .$$

Therefore  $A(x) < B^k(x)$  implies  $(k-1)\mu x < \lambda x$ .  $\square$

Now, let us write  $\alpha_n = A^n(x_0)$  and  $\beta_n = B(x_0)$ . By Lemma B.2, the number of  $\beta_j$ 's inside each interval of the form  $[\alpha_{n+1}, \alpha_n]$  is bounded independently of  $n$ . Moreover, since  $\alpha_n < x_n < \beta_n$  for all  $n$ , the number of  $x_j$ 's inside each  $[\alpha_{n+1}, \alpha_n]$

is also bounded independently of  $n$ . To prove that  $|\Delta_j| \asymp 1/j^2$ , we proceed as follows. Let  $\ell > 0$  be such that  $\beta_{\ell+1} \leq x_j \leq \beta_\ell \leq x_{j-1}$ . Then Lemma B.2 says that  $\ell \leq Cj$ , and we have also

$$|\beta_{\ell+1} - \beta_\ell| < |B(x_{j-1}) - x_{j-1}| < |x_j - x_{j-1}|.$$

Since by Lemma B.1 we have

$$|\beta_{\ell+1} - \beta_\ell| \asymp \frac{1}{\ell^2} \geq \frac{1}{Cj^2},$$

it follows that  $|\Delta_j| = |x_j - x_{j-1}| \geq 1/Cj^2$ .

To prove an inequality in the opposite direction, let  $m$  be the largest integer such that  $\alpha_m > x_{j-1}$ . Then, again by Lemma B.2, we have  $j \leq Cm$ . Since  $A(x) < f(x) < x$  for all  $x$ , we also have  $\Delta_j \subseteq [\alpha_{m+2}, \alpha_m]$ . Using Lemma B.1 once more, we deduce that

$$|\Delta_j| \leq \frac{C}{m^2} \leq \frac{C}{j^2}.$$

This completes the proof of Yoccoz's Lemma.  $\square$

#### REFERENCES

- [1] P. Billingsley, *Probability and measure*, John Wiley & Sons, New York, 1986.
- [2] L. Carleson, *On mappings conformal at the boundary*, J. Anal. Math. **19** (1967), 1–13.
- [3] E. de Faria, *Asymptotic rigidity of scaling ratios for critical circle mappings*, to appear in Ergod. Th. & Dynam. Sys., IMS Stony Brook preprint 96/13 (1996).
- [4] E. de Faria and W. de Melo, *Rigidity of critical circle mappings II*, IMS Stony Brook preprint 97/17 (1997).
- [5] J. Franks, *Manifolds of  $C^r$  mappings and applications to differentiable dynamics*, Studies in Analysis, Advances in Mathematics Supplementary Studies, vol. 4, Academic Press, 1979, pp. 271–290.
- [6] I. J. Good, *The fractional dimension theory of continued fractions*, Proc. Cambridge Phil. Soc. **37** (1941), 199–228.
- [7] J. Graczyk, *Harmonic scalings for smooth families of diffeomorphisms of the circle*, Nonlinearity **4** (1990), 935–954.
- [8] ———, *Hölder classes for circle maps*, to appear in Proc. Amer. Math. Soc..
- [9] J. Graczyk and G. Świątek, *Critical circle maps near bifurcation*, Commun. Math. Phys. **176** (1996), 227–260.
- [10] M. Herman, *Sur la conjugaison différentiable des difféomorphismes du cercle a des rotations*, Publ. Math. IHES **49** (1979), 5–234.
- [11] ———, *Conjugaison quasi-symétrique des homéomorphismes du cercle a des rotations* (1988), manuscript.
- [12] O. Lanford, *Renormalization group methods for critical circle mappings with general rotation number*, VIIIth International Congress on Mathematical Physics (Marseille, 1986), World Sci. Publishing, Singapore, 1987, pp. 532–536.
- [13] M. Lyubich, *Combinatorics, geometry and attractors of quasi-quadratic mappings*, Ann. of Math. **140** (1994), 347–404.

- [14] C. McMullen, *Renormalization and 3-manifolds which fiber over the circle*, Annals of Math Studies **142** (1996), Princeton University Press, Princeton.
- [15] W. de Melo and S. van Strien, *One dimensional dynamics*, Springer-Verlag, Berlin and New York, 1993.
- [16] D. Rand, *Global phase-space universality, smooth conjugacies and renormalization: the  $C^{1+\alpha}$  case*, Nonlinearity **1** (1988), 181–202.
- [17] G. Świątek, *Rational rotation numbers for maps of the circle*, Commun. Math. Phys. **119** (1988), 109–128.
- [18] ———, *One-dimensional maps and Poincaré metric*, Nonlinearity **4** (1992), 81–108.
- [19] J. C. Yoccoz, *Conjugaison différentiable des difféomorphismes du cercle dont le nombre de rotation vérifie une condition Diophantienne*, Ann. Sci. de l'Ec. Norm. Sup. **17** (1984), 333–361.
- [20] ———, *Il n'y a pas de contre-exemple de Denjoy analytique.*, C. R. Acad. Sci. Paris **298** (1984), 141–144.

INSTITUTO DE MATEMÁTICA E ESTATÍSTICA, UNIVERSIDADE DE SÃO PAULO,  
 RUA DO MATÃO 1010, BUTANTÃ CEP05508-900 SÃO PAULO SP - BRASIL  
*E-mail address:* `edson@ime.usp.br`

INSTITUTO DE MATEMÁTICA PURA E APLICADA, ESTRADA DONA CASTORINA 110,  
 JARDIM BOTÂNICO, CEP22460 RIO DE JANEIRO RJ - BRASIL  
*E-mail address:* `demelo@impa.br`