OLD AND NEW ON QUADRATIC SIEGEL DISKS†

SAEED ZAKERI

To Professor S. Shahshahani, with admiration

1. Introduction

A Siegel disk for a rational map of degree two or more of the Riemann sphere is an invariant topological disk on which the map acts as an irrational rotation. Such rotation domains were predicted by Fatou in the 1920's, but nobody knew whether they actually existed until 1942 when Siegel constructed many of them by a highly innovative argument.

Understanding the dynamics of a rational map in the presence of a Siegel disk is often difficult. The orbit structure of such a map is extremely fragile, that is you can easily destroy it by perturbing the map within the family of all rational maps of the same degree. These maps lack the hyperbolicity property that has proved so fruitful in dynamics. Life becomes somewhat easier if you focus only on polynomial maps with Siegel disks, as you can eliminate some irrelevant features. But the characteristic difficulties of the problem remain there, even if you narrow down your study to maps as simple as quadratic polynomials.

That gives me enough of an excuse to consider only quadratic polynomials in the present paper. Some of the results extend with no significant effort to higher degree polynomials or general rational maps, but I will not discuss them. Also missing will be the study of Cremer points, which has many parallels to the Siegel case and is equally interesting and challenging. The paper is divided into two parts of roughly equal length. The first half is a survey of known results, along with many questions which deserve to be addressed in any attempt to understand Siegel disks. The second half is an expository account of a recent joint work with Carsten L. Petersen [28] which gives a complete understanding of quadratic Siegel polynomials for almost every rotation number. Though this part is more technical in nature, I will try to keep the overall exposition as elementary as possible. However, some familiarity with holomorphic dynamics is inevitably required.

† Based on the author's talk at John Milnor's 70th birthday conference "Around Dynamics" at Stony Brook in March 2001.

2. Preliminaries

Consider the quadratic polynomial $P_\theta : z \mapsto e^{2\pi i\theta} z + z^2$ which has a fixed point at the origin with the *multiplier* $P_\theta'(0) = e^{2\pi i\theta}$. Throughout this paper we assume that the *rotation number* $\theta \in \mathbb{T} = \mathbb{R}/\mathbb{Z}$ is irrational. In this case, the fixed point 0 is said to be *irrationally indifferent*. Note that you do not lose generality if you work with quadratics of this special form, since any quadratic polynomial with a fixed point of multiplier $e^{2\pi i\theta}$ is affinely conjugate to P_θ.

The polynomial P_θ is said to be *linearizable* near the fixed point 0 if there exists a holomorphic change of coordinates φ in a neighborhood of 0, called a *linearizing map*, which conjugates P_θ to the rigid rotation $R_\theta : z \mapsto e^{2\pi i\theta} z$. In this case, the maximal linearization domain Δ around 0 is an open simply-connected set called the *Siegel disk* of P_θ. Thus, $P_\theta : \Delta \to \Delta$ acts as an irrational rotation by the angle θ and one has the following commutative diagram:

$$\begin{array}{ccc} (\Delta, 0) & \xrightarrow{P_\theta} & (\Delta, 0) \\ \downarrow \varphi & & \downarrow \varphi \\ (\mathbb{D}, 0) & \xrightarrow{R_\theta} & (\mathbb{D}, 0) \end{array}$$

Such P_θ is called a *Siegel quadratic*. In the complementary case, when P_θ is not linearizable in any neighborhood of 0, we say that P_θ is a *Cremer quadratic* and the fixed point 0 is a *Cremer point*.

Perhaps I should emphasize that the linearizability of P_θ is a topological not an analytic property. By this I mean that the existence of a homeomorphic linearizing map implies the existence of a holomorphic one. In fact, if h is a local homeomorphism satisfying $h \circ P_\theta = R_\theta \circ h$ near 0, then for small $\varepsilon > 0$, $U = h^{-1}(\mathbb{D}(0, \varepsilon))$ is a topological disk containing the fixed point 0 and is invariant under the action of P_θ. Let $\varphi : U \to \mathbb{D}$ denote a conformal isomorphism given by the Riemann mapping theorem which satisfies $\varphi(0) = 0$. Then Schwarz lemma tells us that $\varphi \circ P_\theta \circ \varphi^{-1} : \mathbb{D} \to \mathbb{D}$ is the rigid rotation R_θ, meaning that φ is a holomorphic linearizing map for P_θ. It follows in particular that a Cremer quadratic is not linearizable even in the weaker topological sense.

It is well known that to study the behavior of orbits near an irrationally indifferent fixed point, one has to take into account the arithmetical properties of the rotation number θ. Various arithmetical classes of irrational numbers arise this way and play a major role in the theory. Here, I will quickly recall some of them, as they come up in the rest of the paper.

Consider the continued fraction expansion

$$\theta = \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cfrac{1}{a_3 + \cdots}}} = [a_1, a_2, a_3, \ldots],$$

where the a_n are positive integers uniquely determined by θ. The *rational convergents* of θ are defined by

$$\frac{p_n}{q_n} = \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cfrac{1}{\ddots + \cfrac{1}{a_n}}}} .$$

If we set $p_0 = q_{-1} = 0$ and $p_{-1} = q_0 = 1$, the rational convergents are given by the recursions

$$p_n = a_n \, p_{n-1} + p_{n-2}$$
$$q_n = a_n \, q_{n-1} + q_{n-2},$$

for $n \geq 1$. It follows in particular that $q_n > q_{n-1}$ and $q_n > 2q_{n-2}$, and so $q_n \to +\infty$ at least exponentially fast. Since by classical number theory

$$\left| \theta - \frac{p_n}{q_n} \right| \leq \frac{1}{q_n q_{n+1}},$$

we conclude that $p_n/q_n \to \theta$ at least exponentially fast.

For our purposes, the following classes of irrational numbers in $\mathbb{T}$ will be important:

- The class $\mathcal{D}_\nu$ of *Diophantine numbers of order* $\nu \geq 2$. By definition, $\theta \in \mathcal{D}_\nu$ if there is a constant $C > 0$ such that $|\theta - p/q| > C \, q^{-\nu}$ for all rational numbers p/q. One has
$$\theta \in \mathcal{D}_\nu \Leftrightarrow \sup_n \frac{q_{n+1}}{{q_n}^{\nu-1}} < +\infty.$$
The class $\mathcal{D}$ of all Diophantine numbers is the union $\bigcup_{\nu \geq 2} \mathcal{D}_\nu$. Thus
$$\theta \in \mathcal{D} \Leftrightarrow \sup_n \frac{\log q_{n+1}}{\log q_n} < +\infty.$$
- The class of irrationals of *bounded type*. By definition, θ is bounded type if $\sup_n a_n < +\infty$. Since
$$a_{n+1} < \frac{q_{n+1}}{q_n} = a_{n+1} + \frac{q_{n-1}}{q_n} < a_{n+1} + 1,$$
the sequence $\{a_n\}$ is bounded if and only if $\{q_{n+1}/q_n\}$ is. Therefore,
$$\theta \text{ is bounded type } \Leftrightarrow \theta \in \mathcal{D}_2.$$

- The class $\mathcal{Q}$ of irrationals of *quadratic type*. By definition, $\theta \in \mathcal{Q}$ if θ is a root of a quadratic polynomial with integer coefficients. One has

$$\theta \in \mathcal{Q} \Leftrightarrow \{a_n\} \text{ is eventually periodic.}$$

  Eventually periodic means that there is a $p \geq 1$ such that $a_{n+p} = a_n$ holds for all large n.

- The class $\mathcal{H}$ of irrationals of *Herman-Yoccoz type*. By definition, $\theta \in \mathcal{H}$ if every analytic circle diffeomorphism with rotation number θ is analytically linearizable. An explicit arithmetical description for $\mathcal{H}$ is given by Yoccoz [35]. A closely related class $\mathcal{H}' \supset \mathcal{H}$ is defined as follows: $\theta \in \mathcal{H}'$ if every analytic circle diffeomorphism with rotation number θ, with no periodic orbit in a neighborhood of the circle, is analytically linearizable [23].
- The class $\mathcal{B}$ of irrationals of *Brjuno type*. By definition, $\theta \in \mathcal{B}$ if

$$\sum_{n=1}^{\infty} \frac{\log q_{n+1}}{q_n} < +\infty.$$

For $\nu > 2$, we have the proper inclusions

$$\mathcal{Q} \subsetneq \mathcal{D}_2 \subsetneq \mathcal{D}_\nu \subsetneq \mathcal{D} \subsetneq \mathcal{H} \subsetneq \mathcal{B}.$$

It is not hard to check that for $\nu > 2$ the Diophantine class $\mathcal{D}_\nu$ (hence $\mathcal{D}$, $\mathcal{H}$ and $\mathcal{B}$) has full measure, while $\mathcal{D}_2$ (hence $\mathcal{Q}$) has zero measure in the circle $\mathbb{T}$.

3. Siegel vs. Cremer

The so-called "center problem" proposed by Poincaré asks when a holomorphic germ $z \mapsto e^{2\pi i\theta}z + O(z^2)$ is linearizable near its indifferent fixed point 0. Here is a brief chronology of various attempts to solve this problem which I have taken from Milnor's book [22]:

- (1912) Kasner conjectured that every holomorphic germ with an irrationally indifferent fixed point must be linearizable.
- (1917) Pfeiffer constructed the first example of a non-linearizable holomorphic germ with an irrationally indifferent fixed point [29].
- (1919) Julia "proved" that indifferent fixed points of rational maps are *never* linearizable [18]. However, his proof was incorrect.
- (1927) Cremer proved that for a generic θ, every indifferent fixed point of a rational map with rotation number θ is non-linearizable [6].
- (1942) Siegel proved the first positive result by showing that if $\theta \in \mathcal{D}$, then every germ $z \mapsto e^{2\pi i\theta}z + O(z^2)$ is linearizable near the origin [31].

Siegel's result was generalized in 1969 by Brjuno who showed the linearization is possible when θ belongs to the larger class $\mathcal{B}$ [3]. The question of optimality of the Brjuno's condition remained open until 1988 when Yoccoz proved that whenever $\theta \notin \mathcal{B}$, there exists a non-linearizable germ $z \mapsto e^{2\pi i\theta} z + O(z^2)$ [34]. Shortly after, he strengthened this by proving that when $\theta \notin \mathcal{B}$, the quadratic polynomial P_θ has periodic orbits in every neighborhood of the fixed point 0 [36]. Evidently, the existence of such *small cycles* is an obstruction to linearizability.

Theorem 3.1 (Siegel, Brjuno, Yoccoz)**.** *The quadratic polynomial P_θ is linearizable near* 0 *if and only if $\theta \in \mathcal{B}$.*

This theorem is a definitive answer to the local linearization problem for quadratic polynomials. However, as is often the case, the global dynamical picture is what you really want to understand. From this perspective, the theorem gives no clue as to what is going on in a large scale, especially in the presence of a critical point.

To get started, consider the quadratic P_θ for a fixed rotation number $\theta \in \mathcal{B}$. Let $c = -e^{2\pi i\theta}/2$ be the unique critical point of P_θ and $\beta = 1 - e^{2\pi i\theta}$ be the fixed point other than 0. This fixed point is always repelling (it pushes away nearby points) since $|P'_\theta(\beta)| > 1$. Define the *filled Julia set* of P_θ as

$$K = K(P_\theta) = \{z \in \mathbb{C} : \text{the forward orbit } \{P_\theta^{\circ n}(z)\}_{n \geq 0} \text{ is bounded}\},$$

and the *Julia set* $J = J(P_\theta)$ as the topological boundary of K. From classical Fatou-Julia theory [22] we know that both K and J are non-empty (for example they contain c and β), totally invariant, compact and connected subsets of the plane. The interior of K decomposes into infinitely many connected components consisting of the Siegel disk Δ and its iterated preimages. Hence, for a $z \in \mathbb{C}$ there are three possibilities: (i) $z \notin K$ in which case $P_\theta^{\circ n}(z) \to \infty$ as $n \to \infty$; (ii) $z \in \operatorname{int}(K)$ in which case there exists a smallest integer $k \geq 0$ such that $P_\theta^{\circ k}(z) \in \Delta$; (iii) $z \in \partial K = J$ in which case the behavior of the orbit of z is highly non-trivial and often difficult to understand. Fig. 1 shows the filled Julia set K for the golden mean rotation number $(\sqrt{5}-1)/2 = [1, 1, 1, \ldots]$.

The filled Julia set K is full in the sense that $\widehat{\mathbb{C}} \smallsetminus K$ has a single connected component Ω containing ∞. This *basin of infinity* consists of all points which tend to ∞ under the iterations of P_θ; it is simply-connected and totally invariant. There exists a unique conformal isomorphism $\psi : \widehat{\mathbb{C}} \smallsetminus \overline{\mathbb{D}} \xrightarrow{\cong} \Omega$ satisfying $\psi(\infty) = \infty$ and $\psi'(\infty) = 1$ which conjugates the squaring map to the action of P_θ on Ω:

$$\psi(z^2) = P_\theta(\psi(z)).$$

The image $\mathcal{R}_t = \psi\{re^{2\pi i t} : r > 1\}$ of the radial line at angle $t \in \mathbb{T}$ is called the *external ray* at angle t. We say that $\mathcal{R}_t$ *lands* at $p \in J$ if $p = \lim_{r \to 1} \psi(re^{2\pi i t})$. The external rays define an analytic foliation of the basin of infinity that is invariant under the action of P_θ since $P_\theta(\mathcal{R}_t) = \mathcal{R}_{2t}$ (see Fig. 1).

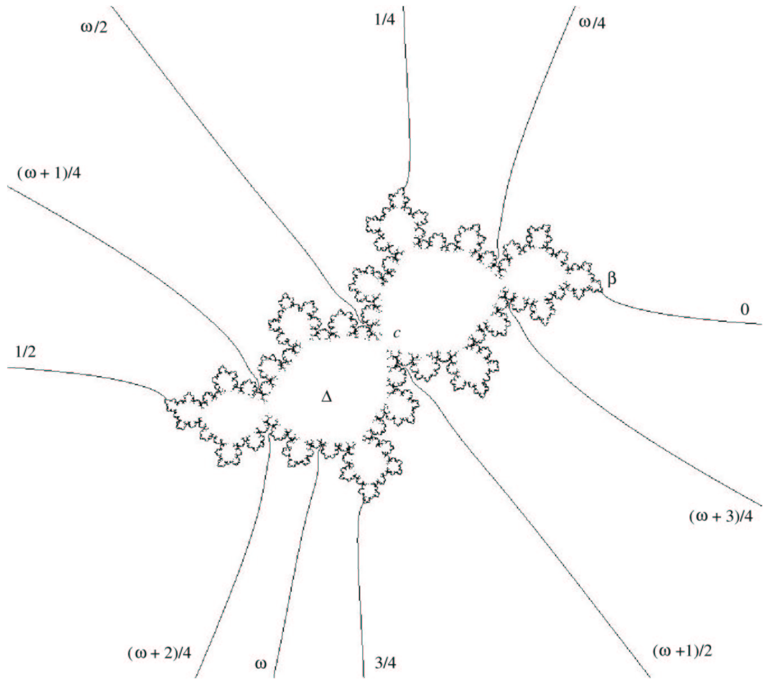

Figure 1. *The filled Julia set of the golden mean quadratic* P_θ*, with* $\theta = (\sqrt{5}-1)/2 = [1, 1, 1, \ldots]$*. The Siegel disk* Δ *is the large domain to the left of the critical point* c *at the center of the picture. The external rays at angles* $t = 0, 1/2, 1/4$ *and* $3/4$ *landing at the fixed point* β *and its preimages are shown. Also shown is the ray* $\mathcal{R}_\omega$ *landing at the critical value* $P_\theta(c)$*, its two preimages* $\mathcal{R}_{\omega/2}, \mathcal{R}_{(\omega+1)/2}$ *landing at the critical point* c*, and four further preimages. Here* $\omega = [2^0, 2^1, 2^1, 2^2, 2^3, 2^5, \ldots] \approx 0.709803$*, where the powers of* 2 *form the Fibonacci sequence.*

Understanding the landing pattern of these external rays is important for the following reason. Call two angles $s, t \in \mathbb{T}$ equivalent if $\mathcal{R}_s$ and $\mathcal{R}_t$ land at the same point of the Julia set J. We call this relation *ray equivalence*. If J is locally-connected, the theorem of Carathéodory in conformal mapping theory guarantees that ψ extends continuously to the boundary circle; in particular, all external rays land. It easily follows that J is homeomorphic to the quotient of $\mathbb{T}$ by this ray equivalence relation. Thus, knowing how the external rays land allows you to build the Julia set up to homeomorphism at least when J is locally-connected.

4. A survey of what is known

What follows is a survey of some of the results on the *global* dynamics of Siegel quadratics. They are presented as attempts to answer four questions that one may consider "fundamental" in this subject. Let me say right at the outset that despite important progress in recent years, none of these fundamental questions is yet fully addressed in a satisfactory way.

• "What can be said about the topology of the boundary $\partial\Delta$? How tame or wild a continuum can it be?"

In the 1980's, Douady and Sullivan asked if the boundary of every Siegel disk of a rational map must be a Jordan curve [8]. This problem is still open, even for quadratic polynomials:

Question 1. Is the boundary $\partial\Delta$ always a Jordan curve?

I should mention here Herman's construction of a smooth diffeomorphism of the 2-sphere which is analytically conjugate to an irrational rotation of a topological disk whose boundary is a pseudo-circle [11]. Apparently he believed that Douady-Sullivan's question has a negative answer for polynomials of high degrees [14].

In [30] Rogers studied the topological structure of the boundaries of Siegel disks. He showed that if $\partial\Delta$ fails to be a Jordan curve, it must be either essentially tame (something like a Jordan curve with a sequence of topologist's sine curves implanted on it) or a horribly complicated continuum. This is the content of part (ii) of the following result.

Theorem 4.1 (Rogers)**.**

(i) *If $\partial\Delta$ is arcwise-connected (in particular, if it is locally-connected), then it is a Jordan curve.*
(ii) *Let $\psi : \Delta \xrightarrow{\cong} \mathbb{D}$ be any conformal isomorphism. Then either ψ extends continuously to $\partial\Delta$ or else $\partial\Delta$ is an indecomposable continuum.*

Recall that a continuum is *indecomposable* if it is not the union of two proper subcontinua. If $\partial\Delta$ fails to be a Jordan curve for some θ, it would be natural to ask which of the alternatives in (ii) can actually take place:

Question 2. Can $\partial\Delta$ ever be an indecomposable continuum?

Another terrifying (or beautiful, depending of what kind of mathematician you are) scenario is that $\partial\Delta$ may separate the plane into more than two connected components.

Question 3. Can $\mathbb{C} \smallsetminus \partial\Delta$ have more than two connected components?

A positive answer would imply $\partial\Delta$ is indecomposable and equal to the entire Julia set J! In this case, one would obtain an example of "Lakes of Wada" [15], with the Siegel disk Δ and its iterated preimages as the lakes, the basin of infinity Ω as the ocean, and the Julia set J as the dry land. I must admit that such a scenario sounds very unlikely, but keep in mind that Lakes of Wada do occur in the simplest nonlinear dynamical systems [16]. In any case, if you were to rule out the existence of Lakes of Wada, the most natural strategy would be to show that the repelling fixed point β is off $\partial\Delta$, for this would immediately imply $\partial\Delta \neq J$.

Question 4. Is it true that the repelling fixed point β of P_θ is always off $\partial\Delta$?

A natural approach to this question would be to consider the perturbation $P_{\theta,r}(z) = re^{2\pi i\theta}z + z^2$ for $0 < r < 1$, which has an attracting fixed point at the origin and a repelling fixed point at $\beta_r = 1 - re^{2\pi i\theta}$. The question would then be whether the distance of β_r to the linearization domain of 0 for $P_{\theta,r}$ is greater than some $\varepsilon > 0$ independent of r.

• "What can be said about the metric structure of $\partial\Delta$? When is it a quasicircle? What is its Hausdorff dimension?"

Recall that a *quasicircle* is the image of the round circle under a quasiconformal map of the plane. Alternatively, one can characterize quasicircles among Jordan curves by Ahlfors's *bounded turning* condition [1]: Each pair of points z, w on a Jordan curve γ divide it into two closed arcs. Let $[z, w]$ denote the arc with smaller Euclidean diameter. Then γ is a quasicircle if there exists a constant $C > 0$ such that $\text{diam}[z, w] < C|z - w|$ for every pair of points $z, w \in \gamma$. Quasicircles first appeared in conformal dynamics as limit sets of quasi-Fuchsian groups. The introduction of quasiconformal maps to iteration theory revived the interest in them. They occur, for example, as the Julia set of $z \mapsto z^2 + c$ for small $|c|$.

The most studied and best understood Siegel quadratics are those with bounded type rotation numbers. Their Siegel disks turn out to be well-behaved from both the topological and metric point of view.

Theorem 4.2. *Suppose* $\theta \in \mathcal{D}_2$.

(i) (Ghys-Douady-Herman-Shishikura-Światek) *The boundary* $\partial\Delta$ *is a quasicircle containing the critical point* c [8].
(ii) (Petersen) *The Julia set* J *is locally-connected and has measure zero* [26].
(iii) (McMullen) *The Hausdorff dimension of* J *is less than* 2 [21].
(iv) (Graczyk-Jones) *The Hausdorff dimension of* $\partial\Delta$ *is greater than* 1 [9].

Fig. 2 shows the details of the golden mean Siegel disk. I will not say anything about the proof of (iii) and (iv) here. However, later in §5, I will describe in detail the proof of a much more general result which includes (i) and (ii) as a special case. Part (i) has been recently generalized to cubic Siegel polynomials [37].

The converse of (i) above is also true. It has been attributed to Herman, but the only written proof that I am aware of is due to Petersen [27].

Theorem 4.3 (Petersen). *If* $\partial\Delta$ *is a quasicircle containing the critical point* c*, then* $\theta \in \mathcal{D}_2$.

The assumption $c \in \partial\Delta$ is quite essential, as $\partial\Delta$ can be a quasicircle for some unbounded type rotation numbers and in these cases $c \notin \partial\Delta$; compare Theorem 4.7 below.

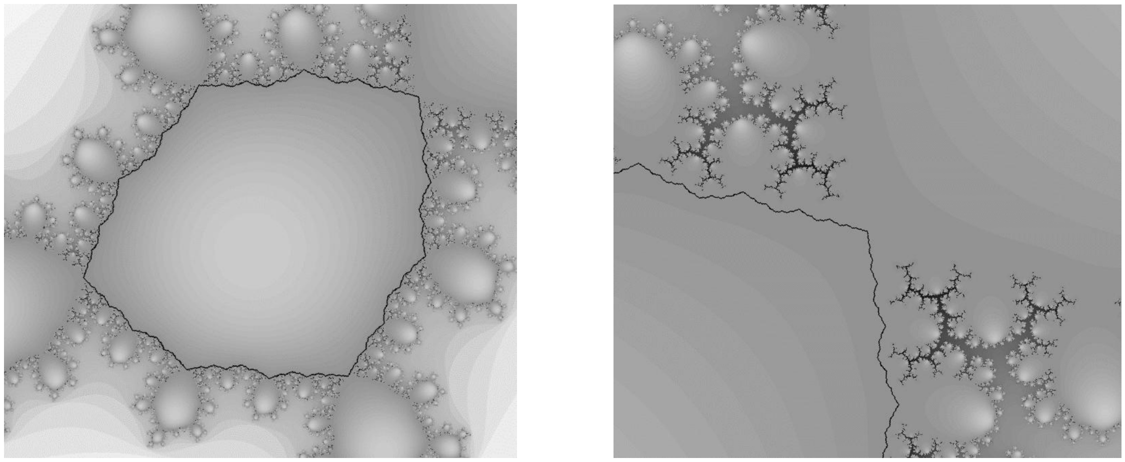

FIGURE 2. *Siegel disk of the golden mean quadratic and its magnification near the critical point.*

When the rotation number θ is of quadratic type, more can be said about $\partial\Delta$. This case has been the subject of extensive numerical studies by physicists since the early 1980's, largely due to the fact that the corresponding quadratics exhibit interesting self-similarity and universality phenomena.

Theorem 4.4 (McMullen). *If $\theta \in \mathcal{Q}$, then the boundary $\partial\Delta$ is self-similar about the critical point c.*

Roughly speaking, the theorem says that for some suitable $\lambda > 1$, the successive blow-ups of $\partial\Delta$ near c under $L : z \mapsto \lambda(z - c) + c$ converge to a well-defined curve which is invariant under L (compare Fig. 2).

Before I move on, let me include the following amusing question, apparently asked by Carleson, about the golden mean Siegel quadratic:

Question 5. Let $\theta = (\sqrt{5} - 1)/2 = [1, 1, 1, \ldots]$. Is it true that

$$|c| = \max_{z \in \partial\Delta} |z| \quad \text{and} \quad |P_\theta(c)| = \min_{z \in \partial\Delta} |z|?$$

Computer experiments seem to confirm this and one might suspect that something deep is going on here, but so far there has been no rigorous explanation for it.

You may wonder how regular the boundary of a Siegel disk can get. An easy application of Schwarz reflection principle shows that the boundary of a Siegel disk can never be real analytic. However, Perez-Marco has constructed examples of holomorphic germs with Siegel disks whose boundaries are smooth or even quasi-analytic [24]. It would be interesting to show that some of these germs can be realized as a quadratic polynomial.

Question 6. Does there exist a rotation number $\theta \in \mathcal{B}$ for which $\partial\Delta$ is smooth or even quasi-analytic?

Another metric quantity associated with $\partial\Delta$ is its Hausdorff dimension. When $\theta \in \mathcal{D}_2$, we know that $1 < \dim(\partial\Delta) < 2$ (the lower bound is Theorem 4.2(iv) and the upper bound follows from the fact that $\partial\Delta$ is a quasicircle). It would be interesting to understand the behavior of the dimension function.

Question 7. What can be said about the function $\theta \mapsto \dim(\partial\Delta)$ when $\theta \in \mathcal{B}$? Can it take the values 1 or 2?

Evidently, a positive answer to Question 6 provides an example with dimension 1. On the other hand, there are indications that $\partial\Delta$ might actually have dimension 2 (see the discussion leading to Question 17).

- "What prevents P_θ from being linearizable in a domain larger than Δ?"

It was widely believed in the early 1980's that the boundary of a Siegel disk of a rational map must contain a critical point. This was thought to be *the* obstruction to extending the linearizing map beyond the boundary of such Siegel disks. In [10] Herman provided some support for this common belief. In particular, he showed

Theorem 4.5 (Herman)**.** *If $\theta \in \mathcal{H}$, the critical point c belongs to $\partial\Delta$.*

Herman's original proof deals with the Diophantine class $\mathcal{D}$, but his proof works equally well for the larger class $\mathcal{H}$. I should point out that this theorem can now be proved with minimal effort using Perez-Marco's idea of "hedgehogs" (see [23] and [38]). Carleson and Jones have given a harmonic analysis proof for the weaker fact that $c \in \partial\Delta$ for almost every θ, but their proof does not specify such angles [5].

It had been observed earlier by Douady and Sullivan that a Siegel disk with no critical point on its boundary would cause topological complications by forcing the Julia set to be non locally-connected [8]:

Theorem 4.6 (Douady-Sullivan)**.** *If $c \notin \partial\Delta$, then J cannot be locally-connected.*

In fact, assuming J is locally-connected, consider the set $E \subset \mathbb{T}$ of the angles of all the external rays that land on $\partial\Delta$. This E is a compact non-empty subset of the circle which is mapped homeomorphically onto itself by the doubling map $t \mapsto 2t \pmod{\mathbb{Z}}$. It is then easy to see, using the expanding property of the doubling map, that E has to be a finite set, which is a contradiction since $\partial\Delta$ is infinite.

The following theorem of Herman which appeared later in 1986 put things in much better perspective and also showed the above theorem is non-void [13]:

Theorem 4.7 (Herman)**.** *There exist rotation numbers $\theta \in \mathcal{B} \smallsetminus \mathcal{H}$ for which $\partial\Delta$ is a quasicircle, but the entire orbit $\{P_\theta^{\circ n}(c)\}_{n\geq 0}$ is disjoint from $\partial\Delta$. In particular, the Julia set of P_θ is not locally-connected.*

Inspired by Theorem 4.5 and Theorem 4.7, we ask:

Question 8. Is it true that $c \in \partial\Delta$ if and only if $\theta \in \mathcal{H}$?

Recall that the *omega-limit set* $\omega(c)$ is the set of accumulation points of the forward orbit $\{P_\theta^{\circ n}(c)\}_{n\geq 0}$ of the critical point. By classical Fatou-Julia theory, $\partial\Delta \subset \omega(c) \subset J$. Moreover,

Theorem 4.8 (Mañé)**.** *For every $\theta \in \mathcal{B}$, the critical point c is recurrent: $c \in \omega(c)$.*

In fact, there is nothing to prove if $c \in \partial\Delta$. Assuming $c \notin \partial\Delta$ and c is not recurrent, it follows from the work of Mañé in [20] that the invariant set $\partial\Delta$ is expanding. But this easily leads to a contradiction; see [38] for details.

Question 9. How large $\omega(c)$ can get? Can it be equal to J? Can it contain any periodic point?

Herman's Theorem 4.7 shows that the critical point may not be the obstruction to extending the linearization to a domain larger than Δ. In such cases, the obstruction may be the existence of periodic orbits in every neighborhood of $\partial\Delta$. This is certainly the case whenever $c \notin \partial\Delta$ and θ belongs to the arithmetical class $\mathcal{H}'$ [38]. It is interesting to pinpoint the precise condition on θ for which this takes place:

Question 10. For what rotation numbers $\theta \in \mathcal{B}$ is $\partial\Delta$ accumulated by periodic orbits? Does this set contain $\mathcal{B} \smallsetminus \mathcal{H}$?

Somewhat curiously, this question has to do with accessibility of the critical point. The subject is a neat example of how combinatorics, plane set topology and complex analysis interact, so let me say a few words on it.

Fix a $\theta \in \mathcal{B}$ and consider the Julia set J of the quadratic P_θ. A point in J is called *accessible* if it is the landing point of some external ray (see the end of §3 for definitions). A point in J is *biaccessible* if it is the landing point of two or more external rays. Alternatively, $z \in J$ is biaccessible if $J \smallsetminus \{z\}$ is disconnected. As an example, the fixed point β is always accessible but never biaccessible, since it is the landing point of the unique ray $\mathcal{R}_0$. On the other hand, if a ray $\mathcal{R}_t$ lands at the critical point c, then the symmetric ray $\mathcal{R}_{t+1/2}$ must also land at c. It follows that c is biaccessible if it is accessible at all.

The following theorem determines all possible biaccessible points in the Julia set of a Siegel quadratic [38]. The proof uses the theory of "hedgehogs" in an essential way.

Theorem 4.9 (Zakeri)**.** *Fix $\theta \in \mathcal{B}$ and suppose $z \in J$ is biaccessible. Then z is pre-critical, that is there exists an integer $n \geq 0$ such that $P_\theta^{\circ n}(z) = c$.*

In particular, the critical point c is either inaccessible or the landing point of *exactly two* external rays. To see this, note that if $\mathcal{R}_t$ lands at c, so does $\mathcal{R}_{t+1/2}$. If a third ray $\mathcal{R}_s$ landed at c, the critical value $P_\theta(c)$ would be the landing point of the distinct rays $\mathcal{R}_{2t}$ and $\mathcal{R}_{2s}$. By the above theorem, the critical value would then have to be pre-critical, forcing c to be periodic. This would be impossible since $\partial\Delta \subset \omega(c)$.

Naturally, one wants to know what two rays, if any, would land at c. Here a purely combinatorial scheme comes into play. Let $m : \mathbb{T} \to \mathbb{T}$ denote the doubling map $m(t) = 2t \pmod{\mathbb{Z}}$. Given any closed semicircle $A_t = [t/2, (t+1)/2]$, the restriction $m|_{A_t}$ can be extended in an obvious way to a degree 1 monotone map m_t of the circle by sending the entire interval $\mathbb{T} \smallsetminus A_t$ to the single point t. Let $\rho(t)$ be the rotation number of m_t, that is the limit

$$\lim_{n\to\infty} \frac{\widetilde{m}_t^{\circ n}(x_0)}{n} \pmod{\mathbb{Z}} \tag{4.1}$$

where $\widetilde{m}_t : \mathbb{R} \to \mathbb{R}$ is any lift of m_t and x_0 is any real number. The graph of the function $\rho : \mathbb{T} \to \mathbb{T}$ is a "devil's staircase." In fact, given a rational number p/q in the reduced form, we have

$$\rho^{-1}(p/q) = [t^-, t^+], \quad \text{where} \quad t^- = \sum_{0<r/s<p/q} 2^{-s} \quad \text{and} \quad t^+ = \sum_{0<r/s\leq p/q} 2^{-s},$$

while for an irrational number θ we have

$$\rho^{-1}(\theta) = t = \sum_{0<r/s<\theta} 2^{-s}, \tag{4.2}$$

and the sums are taken over all rational numbers r/s (reduced or not). It is not hard to show that $t \mapsto \rho(t)$ is continuous and monotone; see [4] for further details.

Given an irrational θ, (4.2) gives an explicit description of the associated angle $t = \rho^{-1}(\theta)$. Of course you can then reverse this process and recover θ from (4.1). However, a simple observation allows you to put this reverse computation in a more elegant form: the orbit of t under the doubling map is contained in A_t when $\theta = \rho(t)$ is irrational. This means that the m_t- and m-orbits of t are identical, and have the same cyclic order as the orbit $\theta \mapsto 2\theta \mapsto 3\theta \mapsto \cdots$ of the angle θ under the rotation R_θ. If $0.t_1t_2t_3\cdots$ is the binary expansion of t, it follows that

$$m_t^{\circ n}(0.t_1t_2t_3\cdots) = m^{\circ n}(0.t_1t_2t_3\cdots) = 0.t_{n+1}t_{n+2}t_{n+3}\cdots$$

It is then easy to check that

$$\theta = \rho(t) = \lim_{n\to\infty} \frac{t_1 + t_2 + \cdots + t_n}{n}.$$

In other words, θ is the *frequency* of digit 1 in the binary expansion of t.

Based on such purely combinatorial observations, it is easy to verify the following:

Theorem 4.10. *Let $\theta \in \mathcal{B}$. If the external rays at angles $\omega/2$ and $(\omega+1)/2$ land at the critical point c of P_θ, then $\theta = \rho(\omega)$.*

Now suppose c is accessible and U is the component of $\mathbb{C} \smallsetminus (\mathcal{R}_{\omega/2} \cup \mathcal{R}_{(\omega+1)/2} \cup \{c\})$ containing the Siegel disk Δ. If there is a periodic orbit of P_θ in U other than the indifferent fixed point 0, it is necessarily a repelling orbit and hence there are finitely many external rays landing at each point of it [22]. The angles of these external rays form a finite invariant subset of the semicircle A_ω. This implies $\theta = \rho(\omega)$ is rational, which is a contradiction.

Corollary 4.11. *If the critical point c is accessible and off the boundary $\partial\Delta$, then $\partial\Delta$ cannot be accumulated by periodic orbits.*

Question 11. For what rotation numbers $\theta \in \mathcal{B}$ is the critical point c accessible? Does this happen precisely when $\theta \in \mathcal{H}$?

In view of Corollary 4.11, affirmative answers to Question 8 and Question 10 would imply that c is inaccessible when $\theta \notin \mathcal{H}$.

- "What can be said about the dynamics of the Siegel quadratic P_θ on $\partial\Delta$?"

Another interesting set of questions arise when one considers the action of P_θ on the boundary of the Siegel disk. They also shed light on some of the questions we asked before.

When $\partial\Delta$ is a Jordan curve, the theorem of Carathéodory shows that the linearizing map of Δ induces a homeomorphism $\partial\Delta \to \mathbb{S}^1$, so $P_\theta|_{\partial\Delta}$ is topologically conjugate to the rigid rotation R_θ. In particular, every boundary point has a dense orbit. Herman has proved a weaker statement without the assumption of $\partial\Delta$ being a Jordan curve [10]:

Theorem 4.12 (Herman)**.** *Fix $\theta \in \mathcal{B}$ and let μ be the harmonic measure on $\partial\Delta$ (obtained by pushing forward Lebesgue measure on the unit circle under the Riemann map of Δ). Then the support of μ is $\partial\Delta$, and $P_\theta|_{\partial\Delta}$ is ergodic with respect to μ. In particular, μ-almost every point in $\partial\Delta$ has a dense orbit.*

This theorem naturally leads to the following question:

Question 12. Is it true that for every $\theta \in \mathcal{B}$ and every $z \in \partial\Delta$ the orbit of z is dense in $\partial\Delta$?

A positive answer would be a step towards showing that $\partial\Delta$ is always a Jordan curve. At the other extreme, let me mention the following related question asked by Milnor:

Question 13. Can $P_\theta|_{\partial\Delta}$ ever have a periodic orbit?

Evidently the answer is negative in the case $\partial\Delta$ is a Jordan curve (see also Theorem 4.13(i) below). But the question seems to be very difficult in general.

Here is a rather surprising question, the importance of which has been shown in the work of Herman [10]:

Question 14. Is $P_\theta|_{\partial\Delta}$ always injective?

In fact, $P_\theta|_{\partial\Delta}$ is known to be injective in the following cases: (i) when $\partial\Delta$ is a Jordan curve; (ii) when the critical point c is accessible; (iii) when $c \notin \partial\Delta$ [38]. It is also easy to see using the Maximum Principle that $P_\theta|_{\partial\Delta}$ is injective on the set of points in $\partial\Delta$ which are accessible from within Δ.

Injectivity of $P_\theta|_{\partial\Delta}$ has nice implications:

Theorem 4.13. *Let $P_\theta|_{\partial\Delta}$ be injective. Then*

(i) $\mathbb{C} \smallsetminus \partial\Delta$ *has exactly two connected components.*
(ii) $P_\theta|_{\partial\Delta}$ *has no periodic orbits.*

The first statement holds since otherwise $\partial\Delta = J$ (see the comments after Question 3) and $P_\theta|_{\partial\Delta}$ would be 2-to-1. The second statement follows, for example, from Petersen [25]: A periodic point on $\partial\Delta$ must belong to the boundary of a preimage of Δ other than Δ. Each such point will eventually map to the critical point by injectivity of $P_\theta|_{\partial\Delta}$. But this is impossible since the critical point cannot be periodic.

5. A new development

In [28], Petersen and the author introduced the new arithmetical class $\mathcal{E}$ consisting of all irrational numbers $\theta = [a_1, a_2, a_3, \ldots]$ in $\mathbb{T}$ for which

$$\log a_n = O(\sqrt{n}) \qquad \text{as } n \to +\infty.$$

It is not hard to check that $\mathcal{D}_2 \subsetneq \mathcal{E} \subsetneq \mathcal{D}_\nu$ for every $\nu > 2$. Moreover, it follows from a theorem of Khinchin [19] that $\mathcal{E}$ has full measure in $\mathbb{T}$.

Theorem 5.1 (Petersen-Zakeri)**.** *If $\theta \in \mathcal{E}$, then the Julia set of P_θ is locally-connected and has measure zero. In particular, $\partial\Delta$ is a Jordan curve passing through the critical point.*

If θ belongs to the full measure set $\mathcal{E} \smallsetminus \mathcal{D}_2$, it follows from Theorem 4.3 that $\partial\Delta$ is a Jordan curve but not a quasicircle (compare Fig. 3).

The theorem gives a complete description of the combinatorics and topology of the Julia set and the dynamics of P_θ on it for *almost every* rotation number θ. In fact, it follows from the discussion at the end of §3 that when $\theta \in \mathcal{E}$ the Julia set J is homeomorphic to the quotient of the circle by the ray equivalence relation, and the action of P_θ on J is topologically conjugate to the action of the doubling map on this quotient. But by Theorem 4.9 and Theorem 4.10, the ray equivalence is generated by the pair $(\omega/2, (\omega+1)/2)$ and all its preimages under the doubling map. Since ω is uniquely determined by the rotation number θ, it follows that for $\theta \in \mathcal{E}$, the Julia set J can be constructed up to homeomorphism in a combinatorial way, only with knowledge of θ.

The rest of this paper will be devoted to a rather long sketch the proof of Theorem 5.1; see [28] for more technical details. The argument contains, as a special case, the proof of Theorem 4.2(i)-(ii). I will assume some familiarity with the theory of quasiconformal mappings, as in [1].

Fix an irrational number $\theta \in \mathbb{T}$. Following Douady [8], we consider the degree 3 Blaschke product

$$f_\theta : z \mapsto e^{2\pi i t}\, z^2 \left(\frac{z-3}{1-3z} \right),$$

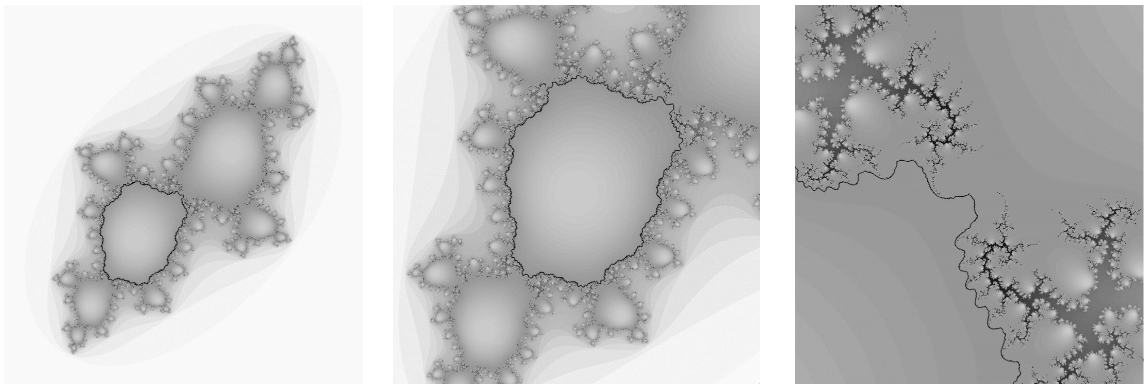

FIGURE 3. *The picture on the left is the filled Julia set of the quadratic* P_θ *for* $\theta = [a_1, a_2, a_3, \ldots] \approx 0.446746$, *where* $a_n = \lfloor e^{\sqrt{n}} \rfloor$. *The boundary of the Siegel disk, emphasized in black, is a Jordan curve but not a quasicircle. The picture in the middle is a close up of this boundary and the one on the right is a further magnification near the critical point. Notice how the geometry of the boundary differs from the bounded type case quasicircle in Fig. 2.*

which has a double critical point at 1, a zero at 3, a pole at 1/3, and superattracting fixed points at 0 and ∞. The restriction $f_\theta|_{\mathbb{S}^1}$ is a *critical circle map*, that is a real-analytic homeomorphism with a critical point at 1. We choose the (unique) parameter $t = t(\theta) \in \mathbb{T}$ so that $f_\theta|_{\mathbb{S}^1} : \mathbb{S}^1 \to \mathbb{S}^1$ has rotation number θ. Fig. 4 left shows the Julia set of f_θ.

By a theorem of Yoccoz [33], there exists a unique homeomorphism $h : \mathbb{S}^1 \to \mathbb{S}^1$ with $h(1) = 1$ such that $h \circ f_\theta|_{\mathbb{S}^1} = R_\theta \circ h$. Let $H : \overline{\mathbb{D}} \to \overline{\mathbb{D}}$ be any homeomorphic extension of h and define

$$F_\theta(z) = F_{\theta,H}(z) = \begin{cases} f_\theta(z) & \text{if } |z| \geq 1 \\ (H^{-1} \circ R_\theta \circ H)(z) & \text{if } |z| < 1 \end{cases}$$

Then F_θ is a degree 2 topological branched covering of the sphere. It is holomorphic outside of $\overline{\mathbb{D}}$ and topologically conjugate to the rigid rotation R_θ on $\overline{\mathbb{D}}$. This is a *synthetic* model for the Siegel quadratic P_θ.

By way of comparison, if there is any correspondence between P_θ and F_θ, the Siegel disk Δ for P_θ should correspond to the unit disk for F_θ, so the preimages of Δ should correspond to the iterated F_θ-preimages of the unit disk, which we call *drops* (these are the teardrop shaped regions in Fig. 4 right). The basin of attraction of infinity for P_θ should correspond to a similar basin $A(\infty)$ for F_θ (which is the immediate basin of infinity for f_θ). By imitating the case of polynomials, we define the "filled Julia set" $K(F_\theta)$ as $\mathbb{C} \smallsetminus A(\infty)$ and the "Julia set" $J(F_\theta)$ as the topological boundary of $K(F_\theta)$, both of which are independent of the homeomorphism H.

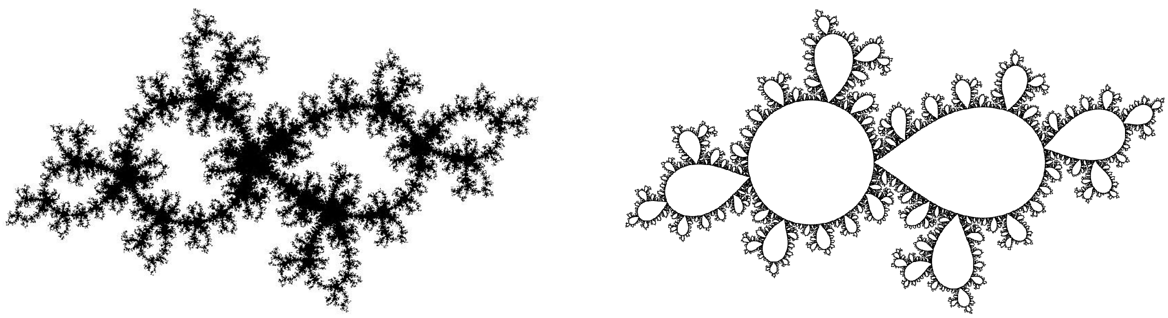

FIGURE 4. *Left: The Julia set of the Blaschke product f_θ for $\theta = [a_1, a_2, a_3, \ldots] \approx 0.446746$, where $a_n = \lfloor e^{\sqrt{n}} \rfloor$. Computation gives $t(\theta) \approx 0.441759$ for this value of θ. Right: The "Julia set" of the synthetic model F_θ. After a David surgery, this synthetic Julia set maps to the Julia set of the quadratic P_θ shown in Fig. 3 left.*

By the work of Petersen in [26] which utilizes the so-called "complex a priori bounds," we know that $J(F_\theta)$ is locally-connected and has measure zero for *all* irrational numbers θ. Thus, the local-connectivity statement in Theorem 5.1 will follow once we prove that for $\theta \in \mathcal{E}$ there exists a homeomorphism $\varphi : \mathbb{C} \to \mathbb{C}$ such that $\varphi \circ F_\theta \circ \varphi^{-1} = P_\theta$. The measure zero statement in Theorem 5.1 will follow if we show that φ is absolutely continuous.

The basic idea described by Douady in [8] is to choose the homeomorphic extension H in the definition of F_θ to be quasiconformal. This turns out to be possible precisely when θ is bounded type. In fact, by the theorem of Beurling-Ahlfors [1], an orientation-preserving circle homeomorphism h has a quasiconformal extension to the disk if and only if it is *quasisymmetric* in the sense that its lift $\widetilde{h}$ to $\mathbb{R}$ satisfies

$$\sup_{x \in \mathbb{R}} \sup_{t>0} \frac{\widetilde{h}(x+t) - \widetilde{h}(x)}{\widetilde{h}(x) - \widetilde{h}(x-t)} < +\infty.$$

On the other hand, the existence of "real a priori bounds" for critical circle maps (see [12] and [32]) implies

Theorem 5.2 (Herman-Światek)**.** *The linearizing map $h : \mathbb{S}^1 \to \mathbb{S}^1$ is quasisymmetric if and only if $\theta \in \mathcal{D}_2$.*

Thus, assuming $\theta \in \mathcal{D}_2$, one can choose a quasiconformal extension $H : \mathbb{D} \to \mathbb{D}$ of h. Let $\mu_H = (\overline{\partial} H / \partial H)\, d\overline{z}/dz$ be the Beltrami differential of H in $\mathbb{D}$. We can make μ_H into an F_θ-invariant Beltrami differential $\widetilde{\mu}_H$ in the plane by saturating it as follows:

Let U be any drop and $F_\theta^{\circ k}$ be the first iterate which maps U conformally to $\mathbb{D}$. Define $\widetilde{\mu}_H$ in U as the pull-back $(F_\theta^{\circ k})^* \mu_H$. On the unit disk set $\widetilde{\mu}_H = \mu_H$. This defines $\widetilde{\mu}_H$ in the interior of $K(F_\theta)$. To complete the definition, simply set $\widetilde{\mu}_H = 0$ anywhere else. The Beltrami differential $\widetilde{\mu}_H$ is called the *saturation* of μ_H; that $\widetilde{\mu}_H$ is F_θ-invariant is clear from its definition. Since the iterated inverse branches of F_θ used for pulling back are all conformal, they do not increase the dilatation, so $\widetilde{\mu}_H$ has bounded dilatation. The measurable Riemann mapping theorem [2] then shows that $\widetilde{\mu}_H$ can be integrated by a quasiconformal homeomorphism $\varphi : \mathbb{C} \to \mathbb{C}$, which means its Beltrami differential μ_φ is equal to $\widetilde{\mu}_H$ almost everywhere. It follows that $P = \varphi \circ F_\theta \circ \varphi^{-1}$ is a quasiregular degree 2 branched covering of $\mathbb{C}$ which preserves the zero Beltrami differential, hence it is holomorphic. Since $P^{-1}(\infty) = \infty$, the map P must be a quadratic polynomial. With appropriate normalization, we obtain $P = P_\theta$ and φ is the desired conjugacy between F_θ and P_θ. This proves Theorem 4.2(i)-(ii).

It is clear from Theorem 5.2 that to go beyond the bounded type rotation numbers, we have no choice but to give up the above quasiconformal surgery. The main idea of [28] is to use extensions H which are no longer quasiconformal, but their dilatation grows in a controlled fashion. What gives this approach a chance to succeed is David's theorem on integrability of certain Beltrami differentials with unbounded dilatation. A measurable $(-1, 1)$ form μ in a domain U is called a *David-Beltrami differential* if there are positive constants C, α, ε_0 such that

$$\text{area}\{z \in U : |\mu|(z) > 1 - \varepsilon\} \leq C\, e^{-\alpha/\varepsilon} \quad \text{for all} \quad \varepsilon < \varepsilon_0. \tag{5.1}$$

In other words, the area of the set of points where the dilatation of μ is large must be exponentially small. In [7], David showed that the measurable Riemann mapping theorem holds for the class of David-Beltrami differentials:

Theorem 5.3 (David). *Every David-Beltrami differential μ in a domain U is integrable, that is there exists an orientation-preserving homeomorphism $\varphi : U \xrightarrow{\cong} \varphi(U)$ in $W^{1,1}_{\text{loc}}(U)$ whose Beltrami differential μ_φ coincides with μ almost everywhere in U. This φ is unique up to postcomposition with a conformal map, and is absolutely continuous in the sense that* $\text{area}\, E = 0$ *if and only if* $\text{area}\, \varphi(E) = 0$.

A homeomorphism φ in the Sobolev class $W^{1,1}_{\text{loc}}$ for which μ_φ is a David-Beltrami differential is called a *David homeomorphism.*

To carry out a non-quasiconformal surgery in the unbounded type case, let us consider the following statements:

- **A.** The rotation number θ of $F_\theta|_{\mathbb{S}^1}$ belongs to the arithmetical class $\mathcal{E}$.
- **B.** The homeomorphism $h : \mathbb{S}^1 \to \mathbb{S}^1$ which linearizes $F_\theta|_{\mathbb{S}^1}$ has a David extension $H : \mathbb{D} \to \mathbb{D}$.
- **C.** The saturation of every David Beltrami-Differential in $\mathbb{D}$ by F_θ is a David-Beltrami differential in $\mathbb{C}$.

The main ingredient of the proof of Theorem 5.1 is then to show the implications

$$\mathbf{A} \Rightarrow \mathbf{B} \Rightarrow \mathbf{C}$$

In fact, given a $\theta \in \mathcal{E}$, the implication $\mathbf{A} \Rightarrow \mathbf{B}$ tells us that h has a David extension H to $\mathbb{D}$, and the implication $\mathbf{B} \Rightarrow \mathbf{C}$ guarantees that the saturation of the Beltrami differential of H in $\mathbb{D}$ is an F_θ-invariant David-Beltrami differential in $\mathbb{C}$. Integrating this differential by Theorem 5.3 gives a David homeomorphism φ for which $\varphi \circ F_\theta \circ \varphi^{-1}$ is a quadratic polynomial with a Siegel disk of rotation number θ. Theorem 5.1 follows since φ, being David, is an absolutely continuous homeomorphism.

So let me address these implications separately. A piece of notation will be useful in what follows. For positive quantities a and b, the notation

$$a \preccurlyeq b$$

means that there exists a universal constant $C > 0$ such that $a \leq C\, b$. The notation

$$a \asymp b$$

means that $a \preccurlyeq b$ and $b \preccurlyeq a$, that is there exists a universal constant $C > 0$ such that $C^{-1}\, b \leq a \leq C\, b$. In this case, we say that a and b are *comparable.*

• Outline of the proof of $\mathbf{A} \Rightarrow \mathbf{B}$

For convenience, think of the circle maps F_θ, R_θ and h as homeomorphisms on the real line which commute with the translation $x \mapsto x+1$ (this can be achieved by lifting them via the exponential map $x \mapsto \exp(2\pi i x)$). The goal is then to construct an extension of h to the upper half-plane $\mathbb{H}$, which commutes with the translation $z \mapsto z+1$, whose Beltrami differential satisfies the David's condition (5.1) on the vertical strip $0 < \Re z < 1$. Following Yoccoz, we construct two combinatorially equivalent, affine cell decompositions Γ_F and Γ_R of $\mathbb{H}$, one using F_θ and the other one using R_θ. For each n, consider the finite segment $\{F_\theta^{-j}(0)\}_{0 \leq j < q_n}$ of the backward orbit of 0 and translate it by $\mathbb{Z}$ to obtain a "lattice" in $\mathbb{R}$. Above each point x in this lattice, mark the point $x + iy \in \mathbb{H}$, where y is the average distance of x to its two neighbors in the lattice. Repeat this for every n, and mark all the points obtained this way in $\mathbb{H}$. These points form the vertices of the cells in Γ_F. A similar construction with R_θ defines the cells in Γ_R, and the conjugacy h between F_θ and R_θ gives a correspondence between cells in Γ_F and Γ_R. The cells are arranged in horizontal layers labeled by their *level* n. The closer a cell is to the boundary of $\mathbb{H}$, the larger its level will be. The Euclidean diameter of an n-cell, that is a cell of level n, decays exponentially in n. Each n-cell is bounded by an "upper edge," two "sides edges" and k "lower edges" with $k = a_{n+1}$ or $a_{n+1}+1$ (see Fig. 5).

The cells in both Γ_F and Γ_R have bounded geometry independent of the rotation number θ, that is they do not look too stretched out either horizontally or vertically. For Γ_R this statement is easy to verify but for Γ_F it follows from Herman-Światek's real a priori bounds. What makes the cells in Γ_F and Γ_R different is the relative size

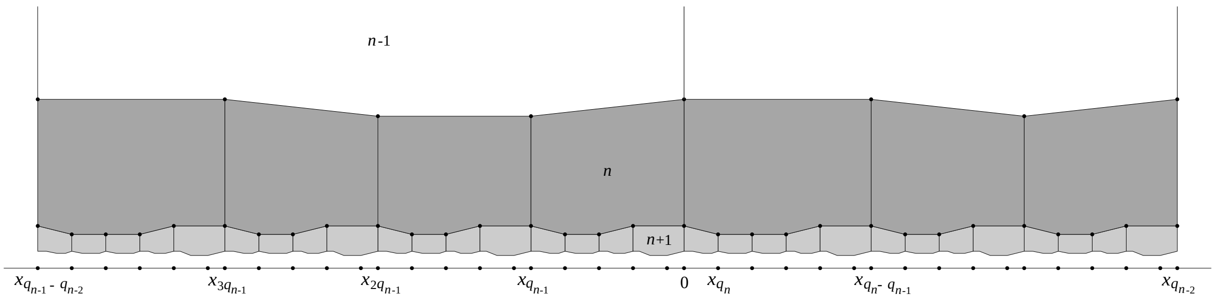

FIGURE 5. *Part of the cell decomposition Γ_R constructed using the rigid rotation R_θ. Some cells of level $n-1$, n, and $n+1$ are shown. In this example $a_n = 3$ and $a_{n+1} = 4$. Each marked point x_j is the lift of the preimage $F_\theta^{-j}(1)$ of the critical point.*

of their lower edges: for an n-cell in Γ_R all the lower edges have roughly the same length, but for an n-cell γ in Γ_F, the j-th lower edge (counted from left or right) has length comparable to

$$\frac{\mathrm{diam}(\gamma)}{\min\{j, k+1-j\}^2}. \tag{5.2}$$

In other words, the relative size of the lower edges of γ decreases quadratically from both sides (see Fig. 6). Of course, this phenomenon becomes relevant only when a_{n+1} (hence k) is very large.

Now extend h to a homeomorphism $H : \mathbb{H} \to \mathbb{H}$ as follows: Let H map the boundary of each cell $\gamma \in \Gamma_F$ affinely to the boundary of the corresponding cell $\gamma' \in \Gamma_R$. To define H in the interior of these cells, it will be more convenient to map γ and γ' quasiconformally to the unit disk, with the upper edges mapped to the arc $[\zeta, \zeta^3]$, the left edges mapped to $[\zeta^3, \zeta^5]$, and the right edges mapped to $[\zeta^7, \zeta]$; here $\zeta = e^{i\pi/4}$. The lower edges of γ and γ' are mapped into subarcs of $[\zeta^5, \zeta^7]$, which in the case of γ have quadratically decreasing lengths and in the case of γ' have comparable lengths (see Fig. 6). This normalization process can be done with a dilatation independent of γ and γ' since the cells have bounded geometry. The induced boundary map $\mathbb{S}^1 \to \mathbb{S}^1$ can be arranged to be the identity on $\mathbb{S}^1 \smallsetminus [\zeta^5, \zeta^7]$. Its action on each of the arcs $[\zeta^5, -i]$ and $[-i, \zeta^7]$ is by (5.2) reminiscent of a hyperbolic Möbius transformation which fixes the endpoints of the arc and is repelling at $-i$. Using a construction of Strebel, it is not hard to show that this boundary map has a quasiconformal extension to a map $\mathbb{D} \to \mathbb{D}$ whose real dilatation is $\preccurlyeq (\log k)^2 \asymp (\log a_{n+1})^2$. But since $\theta \in \mathcal{E}$, we have $\log a_n \preccurlyeq \sqrt{n}$, and so this real dilatation is $\preccurlyeq n$. Back to the original extension problem, it follows that H extends to the upper half-plane, and the real dilatation of H is $\preccurlyeq n$ on the union of n-cells of Γ_F whose area in the strip $0 < \Re z < 1$ is known to be exponentially small in n. This is precisely the David's condition (5.1).

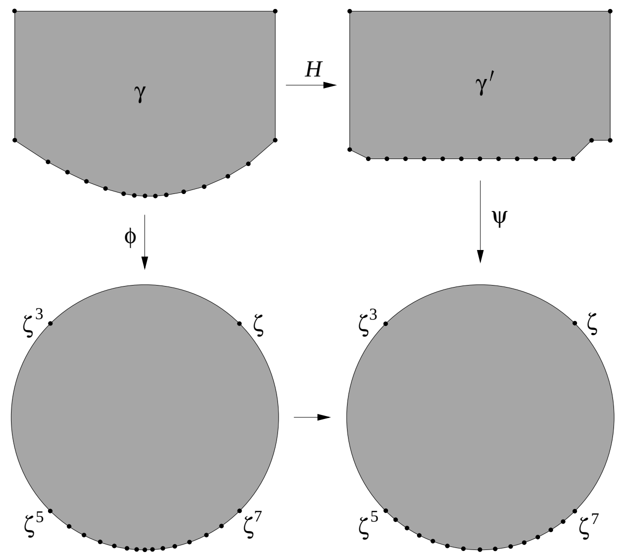

FIGURE 6. *A typical cell $\gamma \in \Gamma_F$ and its corresponding cell $\gamma' \in \Gamma_R$. The number k of lower edges is 14 in this example. They shrink quadratically in size for γ and have roughly the same length for γ'. The normalizing maps ϕ and ψ are quasiconformal with dilatation independent of γ and γ'. The induced boundary map $\mathbb{S}^1 \to \mathbb{S}^1$ has a quasiconformal extension $\mathbb{D} \to \mathbb{D}$ with real dilatation $\preccurlyeq (\log k)^2$.*

• Outline of the proof of $\mathbf{B} \Rightarrow \mathbf{C}$

This is the most non-trivial step in the proof of Theorem 5.1. It amounts to showing that if μ is a David-Beltrami differential in $\mathbb{D}$, then the saturation $\widetilde{\mu}$ will be a David-Beltrami differential in $\mathbb{C}$. For $\varepsilon > 0$ small, let $A_{\mathbb{D}}(\varepsilon) = \{z \in \mathbb{D} : |\mu|(z) > 1 - \varepsilon\}$. More generally, for a drop U, define

$$A_U(\varepsilon) = \{z \in U : |\widetilde{\mu}|(z) > 1 - \varepsilon\}.$$

Think of the regions $A_{\mathbb{D}}(\varepsilon)$ and $A_U(\varepsilon)$ as "red zones" in $\mathbb{D}$ and U respectively, where the dilatation is large. Our goal is to show that if the area of red in $\mathbb{D}$ is exponentially small in ε, so is the total area of red in the plane (remember that there is no red outside the union of drops).

Let U_0 denote the unique preimage of $\mathbb{D}$ other than $\mathbb{D}$ (in Fig. 4 right, U_0 is the prominently visible drop attached to $\mathbb{D}$ at the critical point $z = 1$). Let g denote the univalent branch of F_θ^{-1} mapping $\mathbb{D}$ to U_0. Since F_θ has a critical point at $1 \in \partial U_0$, this g distorts areas a great deal in the vicinity of the critical value $F_\theta(1)$. However, if area $A_{\mathbb{D}}(\varepsilon)$ is exponentially small in ε, so is area $A_{U_0}(\varepsilon) = \text{area}\, g(A_{\mathbb{D}}(\varepsilon))$.

For any drop U, let $g_U : U_0 \to U$ be the inverse of the iterate $F_\theta^{\circ k}$ which maps U to U_0 conformally (in particular, $g_{U_0} = \text{id}$). It easily follows from the definition of saturation that

$$A_U(\varepsilon) = g_U(A_{U_0}(\varepsilon)).$$

This, in particular, shows that the red zone is backward invariant. Now, each g_U has a univalent extension to a neighborhood of $\overline{U_0}$ whose size depends on U. If we knew that all the g_U had bounded distortion *independent of* U, we could conclude that the proportion of red in U is comparable to the proportion of red in U_0:

$$\frac{\operatorname{area} A_U(\varepsilon)}{\operatorname{area} U} = \frac{\operatorname{area} g_U(A_{U_0}(\varepsilon))}{\operatorname{area} g_U(U_0)} \asymp \frac{\operatorname{area} A_{U_0}(\varepsilon)}{\operatorname{area} U_0} \asymp \operatorname{area} A_{U_0}(\varepsilon).$$

This would imply

$$\begin{aligned} \operatorname{area}\{z \in \mathbb{C} : |\widetilde{\mu}|(z) > 1-\varepsilon\} &= \operatorname{area} A_{\mathbb{D}}(\varepsilon) + \sum_U \operatorname{area} A_U(\varepsilon) \\ &\asymp \operatorname{area} A_{\mathbb{D}}(\varepsilon) + \operatorname{area} A_{U_0}(\varepsilon) \ \sum_U \operatorname{area} U \\ &\preccurlyeq \operatorname{area} A_{\mathbb{D}}(\varepsilon) + \operatorname{area} A_{U_0}(\varepsilon) \ \operatorname{area} K(F_\theta). \end{aligned}$$

Since $\operatorname{area} A_{\mathbb{D}}(\varepsilon)$ and $\operatorname{area} A_{U_0}(\varepsilon)$ are both exponentially small in ε, so is the area on the left. This would show that $\widetilde{\mu}$ is a David-Beltrami differential in $\mathbb{C}$.

Unfortunately, that the g_U have uniformly bounded distortion is wishful thinking. In fact, the recurrence of the critical point $z = 1$ forces the distortion of $g_U = F_\theta^{-k}$ to grow arbitrarily large as $k \to +\infty$ at least when U has its root on the unit circle. The red zone occupies almost all of such a drop U (compare Fig. 7).

To circumvent this difficulty, we change the strategy: Starting with U_0 and taking backward iterates, we can guarantee bounded distortion for g_U for some time and the red zone will be small. When the distortion of g_U begins to get so big to interfere with the area bounds, the drop U is considerably red, but by this time U is extremely small, so it would not hurt the area bound much to ignore the distortion issue and assume *all* of U is red. This is the philosophy behind the argument, but one needs a great deal of technicalities to make it work.

Formally, we define a Borel measure ν on $\mathbb{D}$ by

$$\nu(E) = \operatorname{area} E + \sum_U \operatorname{area}\,(g_U \circ g(E)),$$

Clearly ν is absolutely continuous with respect to Lebesgue measure on $\mathbb{D}$, but we prove the much sharper estimate

$$\nu(E) \preccurlyeq (\operatorname{area} E)^\delta \quad \text{for some } \delta > 0. \tag{5.3}$$

This key inequality immediately proves the exponential estimate we are after, since it shows

$$\operatorname{area}\{z \in \mathbb{C} : |\widetilde{\mu}|(z) > 1-\varepsilon\} = \nu(A_{\mathbb{D}}(\varepsilon)) \preccurlyeq (\operatorname{area} A_{\mathbb{D}}(\varepsilon))^\delta \preccurlyeq e^{-\delta\,\alpha/\varepsilon}$$

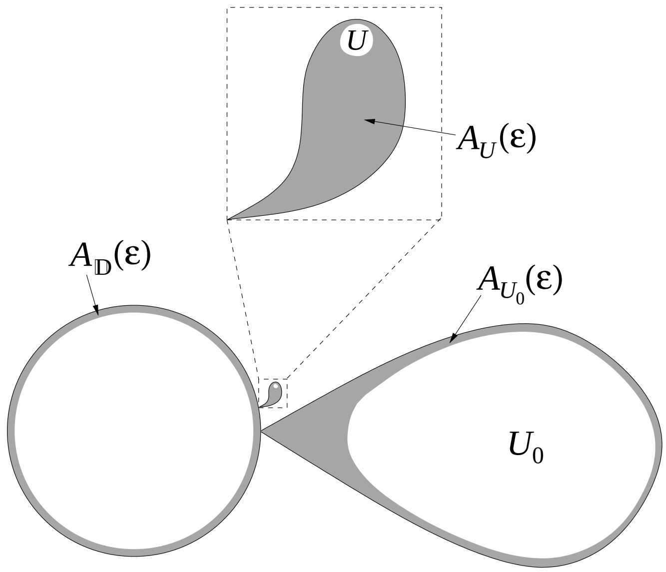

FIGURE 7. *The shaded region in each drop is the "red zone" where the dilatation of $\widetilde{\mu}$ is larger than $1-\varepsilon$. The area of this region in $\mathbb{D}$ is exponentially small. The first pull-back to the drop U_0 introduces some distortion near $z = 1$ (the red zone in U_0 is much thicker there) but still keeps the area exponentially small. However, a long pull-back along a drop U with the root on $\mathbb{S}^1$ introduces so much distortion that almost all of U becomes red.*

whenever $\varepsilon < \varepsilon_0$. The proof of (5.3) is based upon constructing a sequence of dynamically defined "puzzle pieces"

$$\overline{\mathbb{D}} \supset P_1 \supset P_2 \supset \cdots$$

nesting down to the critical value $F_\theta(1)$. Each P_n is a closed topological disk which intersects $\mathbb{S}^1$ along a non-degenerate interval having $F_\theta(1)$ almost in its middle (see Fig. 8). We also define the difference sets

$$D_0 = \overline{\mathbb{D}} \smallsetminus P_1 \quad \text{and} \quad D_n = P_n \smallsetminus P_{n+1} \quad \text{for} \quad n \geq 1.$$

The pieces P_n and D_n have nice geometry which help establish (5.3). First, using real a priori bounds, it is not hard to show that there is a $0 < \sigma_1 < 1$ such that

$$\sigma_1^n \preccurlyeq \operatorname{area} P_n \asymp \operatorname{area} D_n. \tag{5.4}$$

On the other hand, a very careful inductive analysis of the pull-backs of P_n, which must be the most non-trivial estimate in the entire proof, shows that there is a $0 < \sigma_2 < 1$ such that

$$\nu(D_n) \leq \nu(P_n) \preccurlyeq \sigma_2^n. \tag{5.5}$$

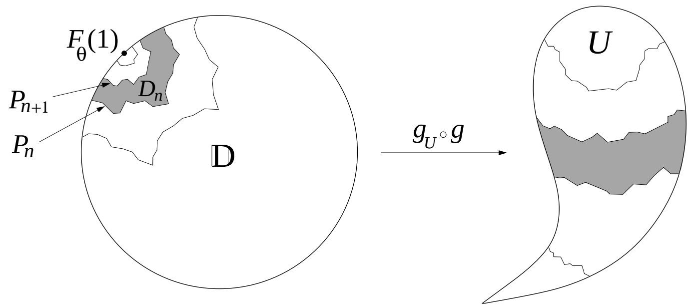

FIGURE 8. *The nested pieces P_n around the critical value $F_\theta(1)$ and the difference sets D_n. Although the pull-backs $g_U \circ g$ do not have bounded distortion on $\mathbb{D}$, their distortion restricted to each D_n is bounded.*

The estimates (5.4) and (5.5) together imply

$$\nu(D_n) \preccurlyeq \operatorname{area}(D_n)^{\delta_1} \tag{5.6}$$

for some $0 < \delta_1 < 1$ (any $0 < \delta_1 < \log \sigma_2 / \log \sigma_1$ will do). This establishes (5.3) for the pieces D_n. To deduce the general case, one needs a further estimate: The Euclidean diameter of D_n is shown to be comparable to the Euclidean distance of D_n to the critical value $F_\theta(1)$. Thus, by Köebe distortion theorem, the univalent maps $g_U \circ g$ have uniformly bounded distortion on D_n, so that

$$\frac{\operatorname{area}(g_U \circ g(E))}{\operatorname{area}(g_U \circ g(D_n))} \asymp \frac{\operatorname{area} E}{\operatorname{area} D_n} \quad \text{whenever } E \subset D_n. \tag{5.7}$$

Now is the endgame: take any measurable set $E \subset \mathbb{D}$ and decompose it into the disjoint union of $E_n = E \cap D_n$. Then, by (5.6) and (5.7),

$$\begin{aligned}
\nu(E_n) = {} & \operatorname{area} E_n + \sum_U \operatorname{area}(g_U \circ g(E_n)) \\
\preccurlyeq {} & \frac{\operatorname{area} E_n}{\operatorname{area} D_n} \left(\operatorname{area} D_n + \sum_U \operatorname{area}(g_U \circ g(D_n)) \right) \\
\preccurlyeq {} & \frac{\operatorname{area} E_n}{\operatorname{area} D_n} \; \nu(D_n) \\
\preccurlyeq {} & \frac{\nu(D_n)}{(\operatorname{area} D_n)^{\delta_1}} \left(\frac{\operatorname{area} E_n}{\operatorname{area} D_n} \right)^{1-\delta_1} (\operatorname{area} E_n)^{\delta_1},
\end{aligned}$$

which gives

$$\nu(E_n) \preccurlyeq (\operatorname{area} E_n)^{\delta_1}. \tag{5.8}$$

Choose any $0 < \delta < \delta_1$ and set $\delta_2 = \delta_1 - \delta$. Then, by (5.8) and Hölder inequality, we obtain

$$\begin{aligned}
\nu(E) &\preccurlyeq \sum_{n=0}^{\infty} (\operatorname{area} E_n)^{\delta+\delta_2} \\
&\preccurlyeq \left(\sum_{n=0}^{\infty} (\operatorname{area} E_n)^{\delta_2/(1-\delta)} \right)^{1-\delta} \left(\sum_{n=0}^{\infty} \operatorname{area} E_n \right)^{\delta} \\
&\preccurlyeq \left(\sum_{n=0}^{\infty} \sigma_2^{n\delta_2/(1-\delta)} \right)^{1-\delta} (\operatorname{area} E)^{\delta} \\
&\preccurlyeq (\operatorname{area} E)^{\delta}.
\end{aligned}$$

This finishes the proof of (5.3), hence the implication $\mathbf{B} \Rightarrow \mathbf{C}$, hence Theorem 5.1.

The idea of David surgery and the proof of Theorem 5.1 generates a series of new questions. For example,

Question 15. Let $f : \mathbb{S}^1 \to \mathbb{S}^1$ be a critical circle map with irrational rotation number θ, and let $h : \mathbb{S}^1 \to \mathbb{S}^1$ be the unique normalized homeomorphism which conjugates f to R_θ. What is the optimal arithmetical condition on θ which guarantees h has a David extension to the unit disk?

The proof of Theorem 5.1 shows that this optimal class contains $\mathcal{E}$; there is some heuristic evidence that shows $\mathcal{E}$ might indeed be the optimal condition.

One can go farther as to ask, in the spirit of Beurling-Ahlfors theory, what conditions guarantee the existence of David extensions for general circle homeomorphisms:

Question 16. Find a necessary and sufficient condition for a circle homeomorphism to have a David extension to the disk.

While it is not hard to find some sufficient conditions, I am not aware of any concrete necessary condition for this extension problem.

The boundary $\partial\Delta$ of the Siegel disk of a quadratic P_θ for a bounded type $\theta \in \mathcal{D}_2$ is a quasicircle, so it has Hausdorff dimension < 2. By Theorem 5.1 and Theorem 4.3, when $\theta \in \mathcal{E} \smallsetminus \mathcal{D}_2$, this boundary is a *David circle* (the image of the round circle under a David homeomorphism) but not a quasicircle. In this case, it is natural to ask how large the dimension of $\partial\Delta$ can get. It can be shown that, unlike quasicircles, there are David circles of Hausdorff dimension 2 [39]. Inspired by this, we ask:

Question 17. Are there rotation numbers $\theta \in \mathcal{E} \smallsetminus \mathcal{D}_2$ for which $\partial\Delta$ has Hausdorff dimension 2?

I suspect the answer is yes, but I do not yet have a proof.

References

[1] L. Ahlfors, *Lectures on quasiconformal mappings*, Van Nostrand, 1966.

[2] L. Ahlfors and L. Bers, *Riemann mapping's theorem for variable metrics*, Ann. of Math. **72** (1960) 385-404.

[3] A. Brjuno, *Analytical form of differential equations*, Trans. Moscow Math. Soc. **25** (1971) 131-288 and **26** (1972) 199-239.

[4] S. Bullett and P. Sentenac, *Ordered orbits of the shift, square roots, and the devil's staircase*, Math. Proc. Cambridge Philos. Soc. **115** (1994) 451-481.

[5] L. Carleson and T. Gamelin, *Complex dynamics*, Springer-Verlag, New York, 1993.

[6] H. Cremer, *Zum Zentrumproblem*, Math. Ann. **98** (1927) 151-163.

[7] G. David, *Solutions de l'équation de Beltrami avec* $\|\mu\|_\infty = 1$, Ann. Acad. Sci. Fenn. Ser. A I Math. **13** (1988) 25-70.

[8] A. Douady, *Disques de Siegel at aneaux de Herman*, Seminar Bourbaki, Astérisque **152-153** (1987) 151-172.

[9] J. Graczyk and P. Jones, *Geometry of Siegel disks*, manuscript, 1997.

[10] M. Herman, *Are there critical points on the boundaries of singular domains?* Comm. Math. Phys. **99** (1985) 593-612.

[11] M. Herman, *Construction of some curious diffeomorphisms of the Riemann sphere*, J. London Math. Soc. (2) **34** (1986) 375-384.

[12] M. Herman, *Conjugaison quasi-symetrique des homeomorphismes analytique du cercle a des rotations*, manuscript, 1986.

[13] M. Herman, *Conjugaison quasi-symetrique des diffeomorphismes du cercle a des rotations et applications aux disques singuliers de Siegel*, manuscript, 1986.

[14] M. Herman, *Some open problems in dynamical systems*, Proceedings of the International Congress of Mathematicians, Vol. II (Berlin, 1998). Doc. Math. 1998, Extra Vol. II, 797-808.

[15] J. Hocking and G. Young, *Topology*, 2nd edition, Dover Publications, Inc., New York, 1988.

[16] J. Hubbard, *The forced damped pendulum: chaos, complication and control*, Amer. Math. Monthly **106** (1999) 741-758.

[17] B. Jones, *A note on homogeneous plane continua*, Bull. Amer. Math. Soc. **55** (1949) 113-114.

[18] G. Julia, *Œuvres de Gaston Julia, vol. III*, Gauthier-Villars, Paris, 1969.

[19] A. Khinchin, *Continued fractions*, Dover Publications, Inc., New York, 1997.

[20] R. Mañé, *On a theorem of Fatou*, Bol. Soc. Brasil. Mat. (N.S.) **24** (1993) 1-11.

[21] C. McMullen, *Self-similarity of Siegel disks and Hausdorff dimension of Julia sets*, Acta Math. **180** (1998) 247-292.

[22] J. Milnor, *Dynamics in one complex variable: Introductory lectures*, Friedr. Vieweg & Sohn, Braunschweig, 1999.

[23] R. Perez-Marco, *Fixed points and circle maps*, Acta Math. **179** (1997) 243-294.

[24] R. Perez-Marco, *Siegel disks with quasi-analytic boundary*, manuscript, 1997.

[25] C. L. Petersen, *On the Pommerenke-Levin-Yoccoz inequality*, Ergodic Theory Dynam. Systems **13** (1993) 785-806.

[26] C. L. Petersen, *Local connectivity of some Julia sets containing a circle with an irrational rotation*, Acta Math. **177** (1996) 163-224.

[27] C. L. Petersen, *On critical quasicircle maps*, manuscript, 1999.

[28] C. L. Petersen and S. Zakeri, On the Julia set of a typical quadratic polynomial with a Siegel disk, available at `www.math.upenn.edu/~zakeri`, to appear in Ann. of Math.

[29] G. Pfeifer, *On the conformal mapping of curvilinear angles; the functional equation $\phi[f(x)] = a_1\phi(x)$*, Trans. Amer. Math. Soc. **18** (1917) 185-198.

[30] J. Rogers, *Singularities in the boundaries of local Siegel disks*, Ergodic Theory Dynam. Systems **12** (1992) 803-821.

[31] C. L. Siegel, *Iteration of analytic functions*, Ann. of Math.(2) **43** (1942) 607-612.

[32] G. Światek, *On critical circle mappings*, Bol. Soc. Brasil. (N.S.), **29** (1998) 329-351.

[33] J. C. Yoccoz, *Il n'y a pas de contre-exemple de Denjoy analytique*, C. R. Acad. Sci. Paris Sér. I Math. **298** (1984) 141-144.

[34] J. C. Yoccoz, *Linéarisation des germes de difféomorphismes holomorphes de* $(\mathbb{C}, 0)$, C. R. Acad. Sci. Paris Sér. I Math. **306** (1988) 55-58.

[35] J. C. Yoccoz, *Recent developments in dynamics*, Proceedings of the International Congress of Mathematicians in Zürich, Birkhäuser Verlag, 1994.

[36] J. C. Yoccoz, *Petits diviseurs en dimension 1: Théorème de Siegel, nombres de Bruno et polynômes quadratiques*, Astérisque **231** (1995) 3-88.

[37] S. Zakeri, *Dynamics of cubic Siegel polynomials*, Comm. Math. Phys., **206** (1999) 185-233.

[38] S. Zakeri, *Biaccessibility in quadratic Julia sets*, Ergodic Theory Dynam. Systems **20** (2000) 1859-1883.

[39] S. Zakeri, *Hausdorff dimension and David homeomorphisms*, manuscript, 2001.

S. Zakeri, Department of Mathematics, University of Pennsylvania, Philadelphia 19104-6395, USA

E-mail address: zakeri@math.upenn.edu