Acta Math., 171 (1993), 263–297

A proof of Thurston's topological characterization of rational functions

by

ADRIEN DOUADY and JOHN H. HUBBARD[1]

Université Paris-Sud
Orsay, France

Cornell University
Ithaca, U.S.A.

The criterion proved in this paper was stated by Thurston in November 1982. Thurston lectured on its proof on several occasions, notably at the NSF summer conference in Duluth, 1983, where one of the authors (JHH) was present. Using the notes of various attendants at these lectures, we have reconstructed a proof that we have made as precise as we could. Since this required a certain amount of work on our part, we thought it might be of some use to present this proof to the reader.

We thank Dennis Sullivan for useful conversations, and Fritz von Haesseler and especially Ben Wittner for help with the writing and valuable suggestions.

After the first version was written, Clifford Earle pointed out that better estimates than what we had were to be found in [B].

Notations. $\#P$ = cardinality of P; $N=\{0,1,2,\dots\}$; $N^*=\{1,2,\dots\}$; P^1 = the Riemann sphere $\mathbf{C}\cup\{\infty\}$, i.e., the complex projective line.

1. Statement and definitions

Let $f\colon S^2\to S^2$ be an orientation-preserving branched covering map. We denote by $\deg_x f$ the local degree of f at x. We will call

$$\Omega_f=\{x\mid \deg_x f>1\}$$

the critical set of f, and

$$P_f=\bigcup_{n>0} f^n(\Omega_f)$$

(1) We thank the NSF for support under grant DMS 83-01564, and the Mittag-Leffler Institute for hospitality during the preparation of this paper

the post-critical set.

The mapping f will be called *critically finite* if P_f is a finite set. We will give in the appendix some examples of critically finite branched mappings, which bring out some of the difficulties in the proof of Thurston's Theorem.

We will assume throughout this paper that f is a critically finite branched mapping, of degree $d>1$, and we set $p=\#P_f$.

Remark. The critical set of f^n is usually larger than Ω_f for $n>1$. This is not true of P_f: we have $P_f=P_{f^n}$ for any $n\geqslant 1$.

Clearly there exists a smallest function ν_f among functions $\nu: S^2\to \mathbf{N}^*\cup\{\infty\}$ such that

(1) $\nu(x)=1$ when $x\notin P_f$, and

(2) $\nu(x)$ is a multiple of $\nu(y)\deg_y f$ for each $y\in f^{-1}(x)$.

We will say that the *orbifold* $O_f=(S^2,\nu_f)$ of f is hyperbolic if its *Euler characteristic*

$$\chi(O_f)=2-\sum_{x\in P_f}\left(1-\frac{1}{\nu_f(x)}\right)$$

satisfies $\chi(O_f)<0$.

Remark. We will see in Section 9 that $\chi(O_f)\leqslant 0$ for any critically finite branched mapping. Such orbifolds are usually hyperbolic: for instance, if $p\geqslant 5$, O_f will clearly be hyperbolic. We will completely classify branched mappings with non-hyperbolic orbifold orbifold in Section 9.

The theory of orbifolds is covered in [T1] and [T2]: we will not require any of this theory until Section 9. There is a natural definition of the universal covering space of an orbifold, and with this definition O_f is hyperbolic if for any complex structure on O_f (i.e., on S^2), the universal covering space $\widetilde{O}_f$ is isomorphic to the disc.

Two branched mappings $f,g: S^2\to S^2$ are *equivalent* iff there exist homeomorphisms $\theta,\theta': (S^2,P_f)\to(S^2,P_g)$ such that the diagram

$$\begin{tikzcd}
(S^2,P_f) \arrow[r, "\theta'"] \arrow[d, "f"'] & (S^2,P_g) \arrow[d, "g"] \\
(S^2,P_f) \arrow[r, "\theta"'] & (S^2,P_g)
\end{tikzcd}$$

commutes, and θ is isotopic to θ' rel P_f.

If γ is a simple closed curve on S^2-P_f, then the set $f^{-1}(\gamma)$ is a union of disjoint simple closed curves. If γ moves continuously, then so does each component of $f^{-1}(\gamma)$.

We will need to consider systems

$$\Gamma = \{\gamma_1, \ldots, \gamma_n\}$$

of simple, closed, disjoint, non-homotopic, non-peripheral curves on $S^2 - P_f$ (γ is non-peripheral if each component of $S^2 - \gamma$ contains at least 2 points of P_f). Such a system will be called a *multicurve* on $S^2 - P_f$.

A multicurve Γ will be called *f-stable* if for any $\gamma \in \Gamma$, all the non-peripheral components of $f^{-1}(\gamma)$ are homotopic in $S^2 - P_f$ to elements of Γ.

To each f-stable multi-curve Γ we can associate the *Thurston linear transformation*

$$f_\Gamma : \mathbf{R}^\Gamma \to \mathbf{R}^\Gamma$$

as follows: Let $\gamma_{i,j,\alpha}$ be the components of $f^{-1}(\gamma_j)$ homotopic to γ_i in $S^2 - P_f$. Define

$$f_\Gamma(\gamma_j) = \sum_{i,\alpha} \frac{1}{d_{i,j,\alpha}} \gamma_i,$$

where

$$d_{i,j,\alpha} = \deg f|_{\gamma_{i,j,\alpha}} : \gamma_{i,j,\alpha} \to \gamma_j.$$

LEMMA 1.1. *The Thurston transformation commutes with iteration*:

$$(f^n)_\Gamma = (f_\Gamma)^n.$$

The proof is left to the reader.

The following lemma, even though it is but a trivial remark, will be essential to the analysis in Section 8.

LEMMA 1.2. *There are only finitely many possible matrices of Thurston transformations for a given degree d of f and a given cardinal p of P_f.*

Proof. A multicurve Γ has at most $p-3$ elements, so the matrix has at most $(p-3)^2$ entries. Each entry is of the form

$$\sum_\alpha \frac{1}{d_{i,j,\alpha}},$$

where α runs through the components of $f^{-1}(\gamma_j)$ homotopic to γ_i. So there are at most d terms in the sum, each of which is of the form $1/d_\alpha$ with $d_\alpha \leqslant d$. □

Since f_Γ has a matrix with non-negative entries, there exists a largest eigenvalue $\lambda(\Gamma, f) \in \mathbf{R}_+$; the corresponding eigenvector has non-negative entries.

Thurston's criterion is the following:

THEOREM 1. *A critically finite branched map $f: S^2 \to S^2$ with hyperbolic orbifold is equivalent to a rational function if and only if for any f-stable multicurve Γ we have $\lambda(\Gamma, f) < 1$.*

In that case the rational function is unique up to conjugation by an automorphism of the Riemann sphere $\mathbf{P}^1$.

Remarks. (a) In principle, this reduces the problem of classifying critically finite rational functions to a purely topological problem.

In practice, it is not clear how to label branched mappings, or how to verify that the criterion is satisfied.

(b) It is not clear how to introduce parameters in the statement. Rational maps, even critically finite ones, can be "close". We know of no notion of "close" critically finite branched maps which would lead to close rational functions.

(c) One may hope that the theorem can be extended to branched mappings which are not critically finite by considering infinite dimensional Teichmüller spaces, laminations, etc.

Conventions. (a) The *Poincaré metric* on the unit disc D is given by

$$|dz|_D = \frac{|dz|}{1-|z|^2}.$$

For any Riemann surface X which admits a map $\pi: D \to X$ as a universal covering, define the Poincaré metric on X so that π is a local isometry.

For any closed curve γ on X, we denote $l_X(\gamma)$ the length of the geodesic homotopic to γ.

(b) *Modulus of an annulus.* Let

$$B_h = \{z = x+iy \mid 0 < y < h\}.$$

The modulus of the cylinder $B_h/\mathbf{Z}l$ is h/l. In particular,

$$\operatorname{mod}\{z \mid 1 < |z| < R\} = \frac{\log R}{2\pi}.$$

(c) *The measure induced by a quadratic form.* If $q(z) = u(z)\,dz^2$ then $|q|$ is the measure $|u(x+iy)|\,dx\,dy$.

2. The mapping σ_f

To prove the theorem, the basic construction is a mapping σ_f from an appropriate Teichmüller space to itself.

Definition. The Teichmüller space $\mathcal{T}_f$ is the Teichmüller space modelled on (S^2, P_f).

Remarks. (a) Of course, $\mathcal{T}_f$ could be identified with $\mathcal{T}_{0,p}$, but we will need functorial properties of $\mathcal{T}_f$, and $\mathcal{T}_{0,p}$ is only defined up to non-unique isomorphism.

(b) The space $\mathcal{T}_f$ can be constructed either as:

(i) The space of smooth almost-complex structures on S^2, two such structures μ_1 and μ_2 being identified if $\mu_1 = h^*\mu_2$ for some diffeomorphism $h: S^2 \to S^2$ with $h|_{P_f} = \mathrm{id}$ and h isotopic to the identity rel P_f,

or as:

(ii) The space of diffeomorphisms $\phi: (S^2, P_f) \to \mathbf{P}^1$, with ϕ_1 and ϕ_2 identified if and only if there exists an analytic isomorphism $h: \mathbf{P}^1 \to \mathbf{P}^1$ such that the diagram

$$\begin{tikzcd} & \mathbf{P}^1 \arrow[dd, "h"] \\ (S^2, P_f) \arrow[ur, "\phi_1"] \arrow[dr, "\phi_2"'] & \\ & \mathbf{P}^1 \end{tikzcd}$$

commutes on P_f, and commutes up to isotopy mod P_f.

The correspondence between these points of view is as follows:

(i) To ϕ one can associate $\phi^*\mu_0$, where μ_0 is the standard complex structure on $\mathbf{P}^1$;

(ii) Since any smooth almost-complex structure μ induces a complex structure, the sphere S^2 with the structure μ is a Riemann surface homeomorphic to S^2, hence isomorphic to $\mathbf{P}^1$; take ϕ to be such an isomorphism.

PROPOSITION 2.1 AND DEFINITION. (a) *The mapping $\mu \mapsto f^*\mu$ on almost complex structures induces an analytic mapping $\sigma_f: \mathcal{T}_f \to \mathcal{T}_f$.*

(b) *If f and g are equivalent and $\theta, \theta': (S^2, P_f) \to (S^2, P_g)$ realize an equivalence, then*

$$\theta^* = \theta'^*: \mathcal{T}_g \to \mathcal{T}_f$$

is an isomorphism such that $\theta^* \circ \sigma_g = \sigma_f \circ \theta^*$.

The proof is routine and left to the reader.

In terms of the second description of $\mathcal{T}_f$, this gives the following description of σ_f.

PROPOSITION 2.2. *If $\tau \in \mathcal{T}_f$ is represented by $\phi: (S^2, P_f) \to \mathbf{P}^1$, then $\tau' = \sigma_f(\tau)$ can be represented by $\phi': (S^2, P_f) \to \mathbf{P}^1$ such that*

$$f_\tau = \phi \circ f \circ (\phi')^{-1}: \mathbf{P}^1 \to \mathbf{P}^1$$

is analytic.

Proof. The point τ' is represented by $\mu' = f^*\phi^*\mu_0$, so take ϕ' to be an isomorphism of (S^2, μ') with $\mathbf{P}^1$.

PROPOSITION 2.3. *The mapping f is equivalent to a rational function if and only if σ_f has a fixed point.*

Proof. ($\Rightarrow$) If f is equivalent to a rational function g, then there exist ϕ, ϕ': $(S^2, P_f) \to (\mathbf{P}^1, P_g)$ isotopic rel P_f and such that the diagram

$$\begin{tikzcd}
(S^2, P_f) \arrow[r, "\phi'"] \arrow[d, "f"'] & (\mathbf{P}^1, P_g) \arrow[d, "g"] \\
(S^2, P_f) \arrow[r, "\phi"'] & (\mathbf{P}^1, P_g)
\end{tikzcd}$$

commutes. This means that if $\tau \in \mathcal{T}_f$ is the point represented by ϕ, then $\sigma_f(\tau)$ is the point represented by ϕ'.

($\Leftarrow$) Consider the diagram

$$\begin{tikzcd}
(S^2, P_f) \arrow[r, "\phi'"] \arrow[d, "f"'] & \mathbf{P}^1 \arrow[d, "f_\tau"] \\
(S^2, P_f) \arrow[r, "\phi"'] & \mathbf{P}^1
\end{tikzcd}$$

of Proposition 2.2. If ϕ' represents the same point of $\mathcal{T}_f$ as ϕ, there exists an isomorphism $h: \mathbf{P}^1 \to \mathbf{P}^1$ such that the diagram

$$\begin{tikzcd}
 & \mathbf{P}^1 \arrow[dd, "h"] \\
(S^2, P_f) \arrow[ur, "\phi"] \arrow[dr, "\phi'"'] & \\
 & \mathbf{P}^1
\end{tikzcd}$$

commutes on P_f, and commutes up to isotopy rel P_f. Then $f_\tau \circ h$ is a rational map equivalent to f, as we see by considering the following diagram:

$$\begin{tikzcd}
(S^2, P_f) \arrow[r, "h^{-1} \circ \phi'"] \arrow[d, "f"'] & \mathbf{P}^1 \arrow[d, "f_\tau \circ h"] \\
(S^2, P_f) \arrow[r, "\phi"'] & \mathbf{P}^1
\end{tikzcd}$$

□

Remark. The above proof produces a map of the set of fixed points of σ_f onto the set of conjugacy classes of rational functions equivalent to f under $\operatorname{Aut} \mathbf{P}^1$.

COROLLARY 2.4. *If P_f has at most 3 elements, then f is equivalent to a rational mapping, unique up to conjugacy by* $\operatorname{Aut}\mathbf{P}^1$.

Proof. In that case $\mathcal{T}_f$ has one point.

Remark. According to Royden [R], all analytic mappings $\mathcal{T}_f \to \mathcal{T}_f$ are weakly contracting for the Teichmüller metric. We will not need this result, since we will compute the derivative of σ_f and verify it directly. Still, it does justify the feeling that something has been accomplished when a question has been reduced to whether a map $\mathcal{T}_f \to \mathcal{T}_f$ has a fixed point.

3. The derivative of σ_f

In addition to $\mathcal{T}_f$ we will need the moduli space $\mathcal{M}_f$.

Definition. The space $\mathcal{M}_f$ is the space of injections $i: P_f \to \mathbf{P}^1$, quotiented by the equivalence relation identifying i_1 and i_2 if $i_1 = h \circ i_2$ for some automorphism h of $\mathbf{P}^1$.

Especially using the second description of $\mathcal{T}_f$, there is an obvious forgetful map $\pi: \mathcal{T}_f \to \mathcal{M}_f$, which is in fact a universal covering space. So the tangent $T_\tau \mathcal{T}_f$ is the same as $T_{\pi(\tau)}\mathcal{M}_f$.

Let $i: P_f \to \mathbf{P}^1$ represent a point of $\mathcal{M}_f$, and set $P = i(P_f)$.

Define $Q(P)$ to be the space of holomorphic quadratic forms on $\mathbf{P}^1 - P$ with at most simple poles on P.

PROPOSITION 3.1. *The cotangent space $T_i^*\mathcal{M}_f$ is canonically isomorphic to $Q(P)$.*

Proof. This result is standard, using Kodaira–Spencer deformation theory and Serre duality. The precise statement we require is in [H], [A], so we will just sketch the proof.

An infinitesimal variation of the complex structure on $\mathbf{P}^1$ is a Beltrami form $\mu \in A^{1,-1}$, i.e., an object which in a local coordinate z can be written $\mu(z)\, d\bar{z}/dz$. In fact, the space of complex structures is the unit ball in $A^{1,-1}$ for the sup norm.

We will use smooth Beltrami forms, which are sufficient for our purposes. The traditional treatment as in [A] uses L^∞ Beltrami forms, and Gunning [G] uses an appropriate Sobolev space. All these methods lead to the same results when the Teichmüller spaces involved are finite dimensional.

An infinitesimal diffeomorphism which is the identity on P is a vector-field which vanishes on P; we will denote the space of such vector-fields $A^{0,-1}(-P)$. If ξ is such a vector-field, the Lie derivative $L_\xi(\mu_0)$ of the standard complex structure is $\bar{\partial}_\xi$. Thus, the tangent space to Teichmüller space is

$$A^{1,-1}/\bar{\partial}(A^{0,-1}(-P)).$$

Now the dual of the space of C^∞ Beltrami forms is the space $\mathcal{D}(Q)$ of distribution quadratic forms, since the product

$$\mu(z)\,\frac{d\bar{z}}{dz}\,q(z)\,dz^2 = \mu(z)q(z)\,d\bar{z}\,dz$$

is naturally a measure.

The cotangent space is the subspace of $\mathcal{D}(Q)$ orthogonal to $\bar{\partial}(A^{0,-1}(-P))$. At points not in P it is easy to show by a bump function argument and Weyl's lemma that in order for q to satisfy

$$\int q\,\bar{\partial}\xi = 0$$

for all $\xi \in A^{0,-1}(-P)$ it is necessary that q be holomorphic on $\mathbf{P}^1 - P$.

For $p \in P$, consider a coordinate z defined on a domain U with $z(p)=0$. We must have $\int zq\,\bar{\partial}\xi = \int q\,\bar{\partial}(z\xi) = 0$ for ξ with support in U, so that zq is holomorphic, and q can have at worst a simple pole at p. □

If U and V are Riemann surfaces, $g\colon U \to V$ is a proper analytic mapping and q is a quadratic form on U, then let $g_* q$ be the quadratic form on V defined by

$$(g_* q)_v(\xi) = \sum_{\mu \in g^{-1}(v)} q((d_\mu g)^{-1}\xi)$$

for all $v \in V$ and $\xi \in T_v V$.

Note that this definition does not require that q be analytic; in fact, even if q is analytic on U, $g_* q$ may acquire poles on V at the critical values of g. However, if q is integrable on U then $g_* q$ is integrable on V.

To see exactly how this may occur, let $U=V=D$ and suppose $w=g(z)=z^k$. Then we have the formula

$$g_*\left(\sum a_i z^i\,dz^2\right) = \sum b_j w^j\,dw^2,$$

with

$$b_j = \frac{1}{k} a_{kj+2k-2}.$$

In particular, if $\sum a_i z^i dz^2$ has at worst a simple pole at 0 then so does $\sum b_j w^j dw^2$.

Let $\tau \in \mathcal{T}_f$, $\tau' = \sigma_f(\tau)$, and let ϕ, ϕ' and f_τ be as in Proposition 2.2. Set $P=\phi(P_f)$, $P'=\phi'(P_f)$. Then $(f_\tau)_* Q(P') \subset Q(P)$.

PROPOSITION 3.2. *The transpose* $(d_\tau \sigma_f)^*\colon Q(P') \to Q(P)$ *is* $(f_\tau)_*$.

Proof. Recall that σ_f was induced by $\mu \to f^*\mu$, for μ a complex structure on S^2. Clearly then if $\mu \in A^{1,-1}$ is an infinitesimal deformation of a complex structure at τ, then

f^* is the corresponding deformation of τ'. The proposition follows from the observation that

$$(f_\tau)^*: A_\tau^{1,-1} \to A_\tau^{1,-1}$$

is the transpose of

$$(f_\tau)_*: \mathcal{D}(Q) \to \mathcal{D}(Q). \qquad \square$$

The space $Q(P)$ carries the natural norm

$$\|q\| = 2\int_{\mathbf{P}^1} |q|.$$

The metric on Teichmüller space induced by the dual norm on each tangent space is called the Teichmüller metric, [A], [H]; it has the following two properties which we will use in an essential way:

(i) *The space $\mathcal{T}_f$ equipped with the Teichmüller metric is a complete metric space.*

(ii) *If $d(\tau,\tau')=\delta$, then there exists a K-quasi-conformal mapping h such that the diagram*

$$\begin{tikzcd} & \mathbf{P}^1 \arrow[dd, "h"] \\ (S^2, P_f) \arrow[ur, "\phi"] \arrow[dr, "\phi'"'] & \\ & \mathbf{P}^1 \end{tikzcd}$$

commutes on P_f, and commutes up to isotopy mod *P_f, if and only if $K \geqslant e^{2\delta}$.*

Proposition 3.3. (a) $\|(f_\tau)_*\| \leqslant 1$.

(b) *If O_f is hyperbolic, then $\|(f_\tau^2)_*\| < 1$.*

Comment. Part (b) is concerned with $f_\tau^2 = f_\tau \circ f_\tau$, in a diagram

$$\begin{tikzcd} (S^2, P_f) \arrow[r, "\phi''"] \arrow[d, "f"] & (\mathbf{P}^1, P'') \arrow[d, "f_{\tau'}"] \\ (S^2, P_f) \arrow[r, "\phi'"] \arrow[d, "f"] & (\mathbf{P}^1, P') \arrow[d, "f_\tau"] \\ (S^2, P_f) \arrow[r, "\phi"'] & (\mathbf{P}^1, P) \end{tikzcd}$$

where the pairs (ϕ,ϕ') and (ϕ',ϕ'') are as in Proposition 2.2. The map considered is

$$(f_\tau^2)_* = (f_\tau)_* \circ (f_{\tau'})_*: Q(P'') \to Q(P).$$

Part (a) of the proposition is obvious. The proof of part (b) uses the following two lemmas.

LEMMA 1. *Let $F: \mathbf{P}^1 \to \mathbf{P}^1$ be a rational map of degree d and q a meromorphic quadratic form with simple poles on $\mathbf{P}^1$. Let Z be the set of poles of q. Suppose $\|F_* q\| = \|q\| \neq 0$. Then*

(a) $q = (1/d) F^* F_* q$,

(b) $F^{-1}(F(Z)) \subset Z \cup \Omega_F$.

Proof. At the neighborhood of a non-critical value, the terms in $F_* q$ coming from the different sheets of the covering must have the same argument. $F^* F_* q$ is a multiple of q by a function which is meromorphic and real, hence constant, and its value must be d. This proves part (a). Part (b) follows. □

LEMMA 2. *Let $f: S^2 \to S^2$ be a critically finite branched mapping, and suppose that $Z \subset P_f$ satisfies $f^{-1}(Z) \subset P_f \cup \Omega_f$.*

(a) *We have that $\# Z \leqslant 4$.*

(b) *If $\# Z = 4$, then all critical points are ordinary, Z contains the set of critical values, and $Z \cap \Omega_f = \varnothing$.*

(c) *Case* (b) *above can occur in two ways: either $f(Z) \subset Z$ in which case $Z = P_f$ and O_f is not hyperbolic, or $Z' = f^{-1}(Z) - \Omega_f$ does not satisfy $f^{-1}(Z') \subset P_f \cup \Omega_f$.*

Proof. Write $f^{-1}(Z) = X_1 \cup X_2$, where

$$X_1 = \{x \in f^{-1}(Z) \mid \exists k \geqslant 0 \text{ and } \omega \in \Omega_f \text{ with } f^{\circ k}(\omega) = x \text{ and } f^{\circ m}(\omega) \text{ not in } Z \text{ for } m \leqslant k\}$$

and $X_2 = f^{-1}(Z) - X_1$. In words, X_1 is the set of points in $f^{-1}(Z)$ which can be reached from Ω_f without passing through Z, and X_2 is the set where you must pass through Z.

Associate to each $x \in X_1$ the subset $\Omega_x \subset \Omega_f$ defined by

$$\Omega_x = \{\omega \in \Omega_f \mid \exists k \geqslant 0 \text{ with } f^{\circ k}(\omega) = x \text{ and } f^{\circ m}(\omega) \text{ not in } Z \text{ for } m \leqslant k\}.$$

Clearly the Ω_x are disjoint or identical.

Similarly, associate to each $x \in X_2$ the subset $Z_x \subset Z$ defined by

$$Z_x = \{z \in Z \mid k \geqslant 0 \text{ such that } f^{\circ k}(z) = x \text{ and } k \text{ is minimal for this property}\}.$$

Again the Z_x are disjoint or identical.

Putting these decompositions together, we find

$$\# f^{-1}(Z) = \# X_1 + \# X_2 \leqslant \# \Omega_f + \# Z \leqslant (2d-2) + \# Z. \tag{1}$$

On the other hand, Z has $d \# Z$ elements in its inverse, counted with multiplicity, where the multiplicity at an inverse image is the local degree there. Since there are precisely $2d-2$ critical points, counting each with multiplicity the local degree minus 1, we see that

$$\begin{aligned} d(\#Z) &= \sum_{x\in f^{-1}(Z)} \deg_x f = \#f^{-1}(Z) + \sum_{x\in f^{-1}(Z)} (\deg_x f - 1) \\ &\leqslant \#f^{-1}(Z) + 2d - 2. \end{aligned} \tag{2}$$

Putting (1) and (2) together, we find

$$(\#Z)(d-1) \leqslant 4(d-1)$$

and since $d>1$ this proves (a).

If $\#Z=4$, then all the inequalities above must be equalities. In particular, $\#\Omega_f = 2d-2$, so that sll the critical points are ordinary. If a point of Z is critical, the first inequality in (1) cannot be an equality. Moreover, in order for the inequality in (2) to be an equality, all the critical points must be in $f^{-1}(Z)$, so that Z contains the critical values. This proves (b).

In this case, moreover, we have, by (2), $4d=\#f^{-1}(Z)+2d-2=\#Z'+4d-4$, hence $\#Z'=4=\#Z$. Set $Y_1=Z-Z'$, $Y_2=Z'-Z$ and $Y_2'=F^{-1}(Y_2)$. We have $\#Y_1=\#Y_2$, and since Y_2 contains no critical value $\#Y_2'-d\#Y_2$.

Suppose now that $f^{-1}(Z')\subset P_f\cup\Omega_f$. Then $Y_2'\subset P_f$. For each $y\in Y_2'$, one can choose an x in $f(\Omega_f)$ and a $k\geqslant 0$ such that $f^k(x)=y$. Take the last j in $\{0,\dots,k\}$ such that $f^j(x)\in Z$, set $y'=f^j(x)$ and $i=k-j$. Then $y'\in Y_1$, $f^i(y')\in Y_2'$, and $f^{i'}(y')\notin Y_2'$ for $i'<i$. It follows that the assignment $y\mapsto y'$ is injective, and $\#Y_1\geqslant\#Y_2'=d\#Y_2$. This implies $\#Y_1=\#Y_2=0$, $Z=Z'=P_f$, $\#P_f=4$ and thus f is not hyperbolic. □

Proof of Proposition 3.3, *part* (b). Let $q''\in Q(P'')$ satisfy $\|(f_\tau)^2_* q''\|=\|q''\|\neq 0$, and denote by Z'', Z' and Z the set of poles of q'', q' and q where $q'=(f_\tau)_* q''$ and $q=(f_\tau^2)_* q''=(f_{\tau'})_* q'$.

Then by Lemma 1 the subsets $\phi^{-1}(Z)$ and $(\phi')^{-1}(Z')$ of P_f satisfy the hypothesis of Lemma 2.

By part (c) of Lemma 2, this is impossible if O_f is hyperbolic. □

Remark. The above proof can be simplified when $d\neq 2,4$. Indeed, in this case, if O_f is hyperbolic, $f(\Omega_f)$ has at least 4 elements.

COROLLARY 3.4. *Suppose the orbifold O_f is hyperbolic, then:*

(a) *σ_f^2 is strictly contracting, i.e., for all $\tau,\tau'\in\mathcal{T}_f$, we have*

$$d(\sigma_f^2(\tau),\sigma_f^2(\tau')) < d(\tau,\tau').$$

(b) *If f and g are equivalent rational functions, then they are conjugate by an automorphism of $\mathbf{P}^1$.*

Remarks. (a) Even though the Teichmüller space is complete, part (a) does not imply the existence of a fixed point.

(b) The case $\#P_f=4$, $\nu_f(x)=2$ if $x\in P_f$ does in fact occur; we will examine it in detail in Section 9.

4. The necessity of the criterion

THEOREM 4.1. *Let f be a critically finite rational function, P its post-critical set, and $\Gamma=\{\gamma_1,\ldots,\gamma_n\}$ an f-stable multicurve. Then $\lambda(\Gamma,f)\leqslant 1$, and if O_f is hyperbolic then $\lambda(\Gamma,f)<1$.*

The proof will require a theorem of Jenkins and Strebel [J], [S], [H-M], and an inequality analogous to one due to Grotzsch [A]; this precise form can be found in [S], [H-M]. These results are stated as Propositions 4.2 and 4.3 below.

PROPOSITION 4.2. *For any vector $v\in\mathbf{R}^\Gamma$ with positive entries, there exists a unique $q\in Q(P)$ with $\int_{\mathbf{P}_1}|q|=1$, having closed trajectories, with annuli $A_1,\ldots,A_n$ homotopic to $\gamma_1,\ldots,\gamma_n$ and with vector of moduli*

$$(\operatorname{mod}(A_1),\ldots,\operatorname{mod}(A_n))$$

a multiple of v.

PROPOSITION 4.3. *Let $q\in Q(P)$ be a quadratic form with closed horizontal trajectories, A an annulus of q with equator γ, with height h and circumference c. Let A' be a straight cylinder with height h' and circumference c', and $g\colon A'\to\mathbf{P}^1-P$ an analytic injection with $g(A')$ homotopic to γ. Then*

$$\int_{g(A')}|q|\geqslant\frac{h'}{c'}c^2.$$

Equality is realized only if g is the inclusion of a straight subcylinder.

Proof of Theorem 4.1. Without loss of generality we may assume that Γ is minimal, so that every $\gamma\in\Gamma$ is homotopic to some component of $f^{-1}(\gamma_i)$, for some $\gamma_i\in\Gamma$.

Let q be the quadratic form given by Proposition 4.2, with the vector of moduli

$$(m_1,\ldots,m_n)$$

an eigenvector for f_Γ with eigenvalue $\lambda(\Gamma,f)$. Denote by h_i, c_i the height and the circumference of A_i, so that $m_i=h_i/c_i$, and let $q'=f^*q$.

Since Γ is f-stable, we may label $A_{i,j,\alpha}$ the cylinders of q' which are inverse images of A_j, homotopic in $\mathbf{P}^1-P$ to A_i; set $d_\alpha=\deg f|_{A_{i,j,\alpha}}$. The the height of $A_{i,j,\alpha}$ is h_j, and its circumference is $d_\alpha c_j$.

Now apply Proposition 4.3. We find

$$\begin{aligned} 1 &= \int_{\mathbf{P}^1} |q| \geqslant \sum_{i,j,\alpha} \int_{A_{i,j,\alpha}} |q| \geqslant \sum_i \sum_j \sum_\alpha \frac{h_j}{d_\alpha c_j} c_i^2 \\ &= \sum_i \Big(\sum_j (f_\Gamma)_{i,j} m_j \Big) c_i^2 = \sum_i \lambda(\Gamma, f) \frac{h_i}{c_i} c_i^2 = \lambda(\Gamma, f). \end{aligned}$$

So we see that $\lambda(\Gamma, f)\leqslant 1$, and that equality is realized only if the cylinders of q' are straight subcylinders of those of q. This can happen only if q' is a real multiple of q, so $f_*q=\pm q$. Then $f_*^2 q=q$ and $\|f_*^2 q\|=\|q\|$. In Proposition 3.3 we see that this cannot happen if O_f is hyperbolic. □

5. Convergence in $\mathcal{T}_f$ and $\mathcal{M}_f$

Generally speaking, given a sequence (τ_i) in Teichmüller space, it is much easier for the images $\pi(\tau_i)$ to converge in $\mathcal{M}_f$ than for the original sequence to converge in $\mathcal{T}_f$.

Pick $\tau_0\in\mathcal{T}_f$ and define $\tau_{i+1}=\sigma_f(\tau_i)$. In this section we will see that it is equivalent for (τ_i) and for $\pi(\tau_i)$ to converge and even for the set $\{\pi(\tau_i)\}$ to have compact closure in $\mathcal{M}_f$.

PROPOSITION 5.1. *If the orbifold O_f is hyperbolic, then (τ_i) converges if and only if the closure of the sequence $\{\pi(\tau_i)\}$ in $\mathcal{M}_f$ is compact.*

In that case, $\tau=\lim_{i\to\infty}\tau_i$ is the unique fixed point of σ_f.

Proof. We will show that the amount by which σ_f contracts at τ depends only on $\pi(\tau)$ and a finite amount of extra information.

LEMMA 5.2. *There exists a tower $\mathcal{T}_f \xrightarrow{\tilde\pi} \widetilde{\mathcal{M}}_f \xrightarrow{\bar\pi} \mathcal{M}_f$ of covering spaces with $\bar\pi$ finite and a map $\bar\sigma_f\colon \widetilde{\mathcal{M}}_f\to\mathcal{M}_f$ such that the diagram*

$$\begin{tikzcd} \mathcal{T}_f \arrow[r, "\sigma_f"] \arrow[d, "\tilde\pi"'] & \mathcal{T}_f \arrow[d, "\pi"] \\ \widetilde{\mathcal{M}}_f \arrow[r, "\bar\sigma_f"'] & \mathcal{M}_f \end{tikzcd}$$

commutes.

Proof of Lemma 5.2. Given a point in $\mathcal{M}_f$ represented by an inclusion $i\colon P_f\to\mathbf{P}^1$, there exist only finitely many isomorphism classes of covering maps $g\colon X'\to\mathbf{P}^1$ of degree d, ramified only over $i(P_f)$. Indeed, pick $x \in \mathbf{P}^1 - i(P_f)$; such a class is determined by the action of the generators of $\pi_1(\mathbf{P}^1 - i(P_f), x)$ on the fiber $g^{-1}(x)$. For each such covering, there are finitely many injections $i': P_f \to g^{-1}(i(P_f))$.

The pairs (g, i') for which there exists homeomorphisms ϕ and ϕ' such that the diagram

$$\begin{tikzcd} S^2 \arrow[r, "\phi'"] \arrow[d, "f"'] & X' \arrow[d, "g"] \\ S^2 \arrow[r, "\phi"'] & \mathbf{P}^1 \end{tikzcd}$$

commutes and $\phi|_{P_f} = i$, $\phi'|_{P_f} = i'$ form a finite set. This is the fiber of $\bar{\pi}$ over i and we can define $\bar{\sigma}_f$ by

$$\bar{\sigma}_f((g, i')) = i'. \qquad \square$$

Proof of Proposition 5.1. By Proposition 3.2, the norm of $d_\tau \sigma_f$ or $d_\tau \sigma_f^2$ depends only on $\tilde{\pi}(\tau)$.

Let δ_0 be a C^1-curve from τ_0 to τ_1, with length l_0; let $\delta_i = \sigma_f^i(\delta_0)$, and set $\delta = \bigcup_{i \geqslant 0} \delta_i$. If the $\pi(\tau_i)$ have compact closure in $\mathcal{M}_f$, then $\tilde{\pi}(\delta)$ has compact closure in $\widetilde{\mathcal{M}}_f$.

By Proposition 3.3, we see that $K = \sup_{\tau \in \delta} |d_\tau \sigma_f^2| < 1$, and since

$$\operatorname{length}(\delta_i) \leqslant K \cdot \operatorname{length}(\delta_{i-2}),$$

the τ_i form a Cauchy sequence, and converge since Teichmüller space is complete.

Clearly $\tau = \lim_{i \to \infty} \tau_i$ is a fixed point of σ_f. $\square$

6. Annuli in Riemann surfaces

Let X be a Reimann surface with its Poincaré metric. If some curves on X are very short, then in some sense the geometry of X breaks up into "thin parts" which are annuli isomorphic to a standard model, and "thick parts" whose geometry remains bounded.

Theorem 6.3 makes this idea precise; Proposition 6.1 is a study of the standard model. Our proof of Theorem 6.3 is borrowed from Beardon [B, Theorem 11.7.1].

Let A_l be an annulus of modulus $\pi/2l$, so that in its Poincaré metric the length of the unique simple closed geodesic γ is l. For $\eta > 0$, set

$$A_l(\eta) = \{z \in A_l \mid d_{A_l}(z, \gamma) \leqslant \eta\}.$$

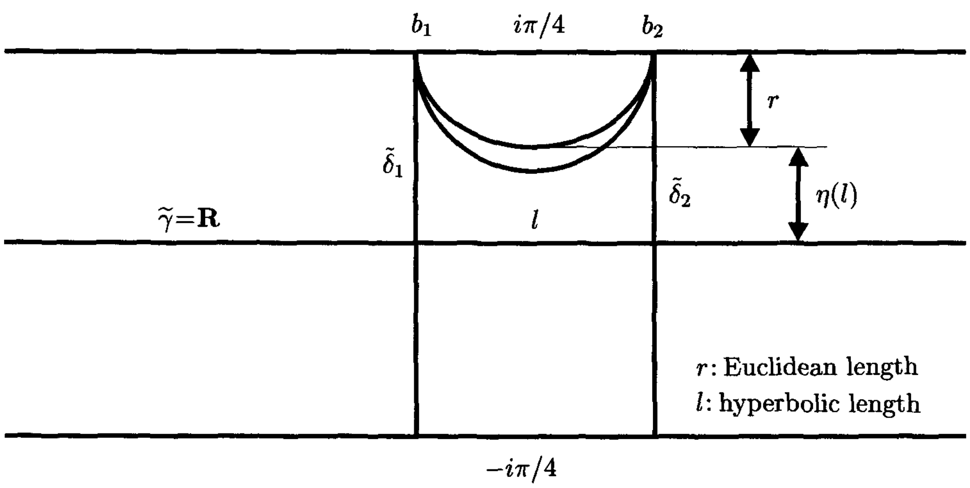


Fig. 1

PROPOSITION 6.1. (a) *There is a largest number* $\eta(l)>0$ *such that no geodesic* γ' *on* A_l *with* $\gamma'\cap\gamma=\varnothing$ *and* $d(\gamma,\gamma')<\eta(l)$ *is simple.*

(b) *The function* $\eta(l)$ *is strictly decreasing.*

(c) *Set* $\bar{A}_l=A_l(\eta(l))$, *and* $m(l)=\operatorname{mod}\bar{A}_l$. *Then*

$$\frac{\pi}{2l}-1<m(l)<\frac{\pi}{2l}.$$

Proof. Let $B=\{z \mid |\operatorname{Im} z|<\pi/4\}$. Since

$$z\mapsto \tanh^{-1}(z)=\frac{1}{2}\log\frac{1+z}{1-z}$$

is an isomorphism of D onto B, the Poincaré metric of B is $|dz|/\cos 2y$ and so if we choose an isomorphism $\tilde{A}_l\to B$ sending $\tilde{\gamma}$ to $\mathbf{R}$, we find

$$A_l=B/l\mathbf{Z}.$$

Let δ be a geodesic in A_l perpendicular to γ, and $\tilde{\delta}_1,\tilde{\delta}_2$ be two successive lifts of δ in B intersecting the line $\operatorname{Im}(z)=\pi/4$ in b_1 and b_2. Consider the geodesic α joining b_1 to b_2; $\alpha\cap\mathbf{R}=\varnothing$ since geodesics can intersect at most in one point, and the image of α in A_l is simple.

To prove (a), we claim $\eta(l)=d(\alpha,\tilde{\gamma})$. Indeed, if β is a geodesic of B coming closer to $\mathbf{R}$ than η and disjoint from $\mathbf{R}$, then its endpoints are a Euclidean distance $>l$ apart, so it cannot be disjoint from its translate by l, and its image in A_l is not simple.

Clearly as l increases, $\eta(l)$ decreases.

Let r be the Euclidean length indicated on Figure 1.

LEMMA 6.2. *We have* $r<\frac{1}{2}l$.

Proof of Lemma 6.2. Consider the bounded harmonic functions h_0 on

$$U=\{z \mid \operatorname{Im} z<\pi/4\}$$

with boundary value

$$\begin{aligned} &0 \ \text{ on } [b_1,b_2]\\ &1 \ \text{ on } \partial U-[b_1,b_2], \end{aligned}$$

and h on B with boundary value

$$\begin{aligned} &0 \ \text{ on } [b_1,b_2]\\ &1 \ \text{ on } \partial B-[b_1,b_2]. \end{aligned}$$

We have $h\geqslant h_0$ on ∂B, thus $h>h_0$ in B. Now $\alpha=h^{-1}\left(\frac{1}{2}\right)$, and $h_0^{-1}\left(\frac{1}{2}\right)$ is the geodesic of U joining b_1 to b_2, i.e., the semi-circle of radius $\frac{1}{2}l$ centered at $\frac{1}{2}(b_1+b_2)$. So α is within this circle. □

End of proof of Proposition 6.1. (c) We have

$$\operatorname{mod}(A_l(\eta(l)))=\frac{\pi}{2l}-\frac{2r}{l}>\frac{\pi}{2l}-1. \qquad \square$$

THEOREM 6.3. *Let X be a Riemann surface with its Poincaré metric, and $\gamma_1,\ldots,\gamma_n$ disjoint simple closed geodesics of length $l_1,\ldots,l_n$. Then there exist in X disjoint annuli $C_1,\ldots,C_n$, isometric to $A_{l_i}(\eta(l_i))$ with equators the γ_i.*

Proof. The annulus A_{l_i} is isomorphic to the covering space $\widetilde{X}_{\gamma_i}$ in which a lift $\widetilde{\gamma}_i$ of γ_i is the only closed curve. The restrictions of the projections

$$\pi_{\gamma_i}\colon A_{l_i}=\widetilde{X}_{\gamma_i}\to X$$

to the $A_{l_i}(\eta(l_i))$ give a map $\pi\colon \amalg_i A_{l_i}(\eta(l_i))\to X$; we need to show that π is injective.

By contradiction, let $x\in X$ be a point which has two distinct inverse images y,y' in $\amalg_i A_{l_i}(\eta(l_i))$, say $y\in A(\eta(l))$ and $y'\in A'(\eta(l'))$. The case $A=A'$ corresponds to that annulus injecting into X and the case $A\neq A'$ corresponds to the two annuli being disjoint.

Let δ and δ' be the geodesics joining y and y' to their respective equators; then $l_A(\delta)<\eta(l)$ and $l_{A'}(\delta')<\eta(l')$.

Choose an isomorphism of the universal covering surface $\widetilde{X}$ with the unit disc and let $\tilde{x}$ be an inverse image of x. The lifts of $\pi(\delta)$ and $\pi(\delta')$ starting at $\tilde{x}$ lead to lifts $\widetilde{\gamma}$ and $\widetilde{\gamma}'$ of the equators γ and γ' of A and A'. The distance between $\widetilde{\gamma}$ and $\widetilde{\gamma}'$ is less than $\eta(l)+\eta(l')$. Lemma 6.4 below says that this is impossible.

Since $\widetilde{\gamma}$ and $\widetilde{\gamma}'$ are lifts of disjoint curves or 2 lifts of 1 simple curve, they are disjoint. Let α be their common perpendicular. Let β_1, β_2 be the perpendicular to $\widetilde{\gamma}$ at distance $\frac{1}{2}l$ from α, and β_1', β_2' the perpendicular to $\widetilde{\gamma}'$ at distance $\frac{1}{2}l'$ from α, labelled so that β_1 and β_1' are on the same side of α. In view of the symmetry with respect to α, there are a priori 4 possible configurations as shown in Figure 2.

LEMMA 6.4. *Only Configuration* I *can occur.*

Proof. For any geodesic θ, call ϱ_θ the reflexion with respect to θ. The automorphisms $g=\varrho_{\beta_1}\circ\varrho_\alpha$ and $g'=\varrho_{\beta_1'}\circ\varrho_\alpha$ of D are replaced by elements of $\pi_1(X)$, and so is $h=g'\circ g^{-1}=\varrho_{\beta_1'}\circ\varrho_{\beta_1}$. In Case II, h has a fixed point, which is impossible. In Case III, $g(\widetilde{\gamma}')$ is a geodesic which intersects $\widetilde{\gamma}'$ transversally, which is impossible since $\widetilde{\gamma}'$ is a lift of a simple geodesic. Case IV is excluded similarly. □

Theorem 6.3 follows. □

COROLLARY 6.5. *Let X be a hyperbolic Riemann surface and γ_1, γ_2 simple closed geodesics of lengths l_1 and l_2. If $l_2<2\eta(l_1)$, then either $\gamma_1=\gamma_2$ or $\gamma_1\cap\gamma_2=\varnothing$.*

Proof. If $\gamma_1\neq\gamma_2$ and $\gamma_1\cap\gamma_2\neq\varnothing$, then in $\widetilde{X}_{\gamma_1}$ a lift $\widetilde{\gamma}_2$ of γ_2 intersects the equator. Since the projection $\widetilde{X}_{\gamma_1}\to X$ is injective on the part of $\widetilde{\gamma}_2$ which is within $\eta(l_1)$ of the equator, we see that $l_2\geqslant 2\eta(l_1)$. □

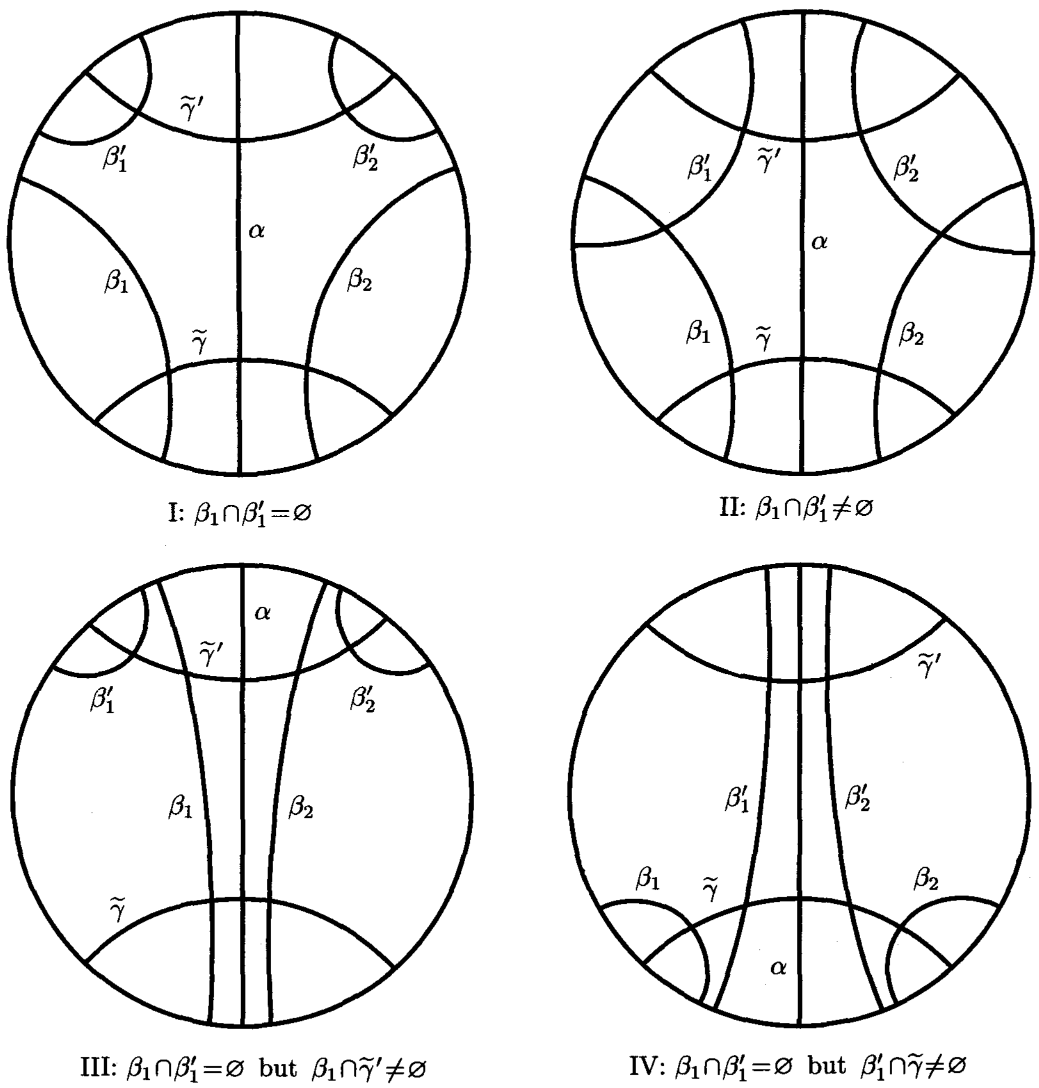
Remark. This bound is sharp, in the sense that for any $l>0$, there exists a Riemann surface X and two geodesics γ_1 and γ_2 on X which intersect, with lengths l and $2\eta(l)$. In fact, take X to be the once punctured torus quotient of D by hyperbolic translations by l and $2\eta(l)$ with perpendicular axes. A fundamental domain is the ideal quadrilateral in Figure 3.

In higher genera, you probably cannot realize the bound exactly, but you can approximate it as closely as you like by squeezing off a handle.

COROLLARY 6.6. *Let X be a Riemann surface and γ_1, γ_2 be two geodesics of length $<\log(\sqrt{2}+1)$. Then either $\gamma_1=\gamma_2$ or $\gamma_1\cap\gamma_2=\varnothing$. Moreover,* $\log(\sqrt{2}+1)$ *is the largest constant for which this is true.*

Proof. First we need to solve $l=2\eta(l)$. Clearly, the length of the common perpendiculars in the regular ideal quadrilateral solves this equation (see Figure 4).

An easy integral shows that this length is $\log(\sqrt{2}+1)$.



I: $\beta_1 \cap \beta_1' = \varnothing$

II: $\beta_1 \cap \beta_1' \neq \varnothing$

III: $\beta_1 \cap \beta_1' = \varnothing$ but $\beta_1 \cap \tilde{\gamma}' \neq \varnothing$

IV: $\beta_1 \cap \beta_1' = \varnothing$ but $\beta_1' \cap \tilde{\gamma} \neq \varnothing$

Fig. 2

If γ_1 is not longer than γ_2, then since η is decreasing, we have

$$l_X(\gamma_1) \leqslant l_X(\gamma_2) < 2\eta(l(\gamma_2)),$$

so by Corollary 6.5, γ_1 and γ_2 are equal or disjoint.

The same example as in the remark above, in the case $l = 2\eta(l)$ shows that on the appropriate punctured torus, there exist intersecting geodesics both with length $\log(\sqrt{2}+1)$. $\square$

We will need one more result from hyperbolic geometry.

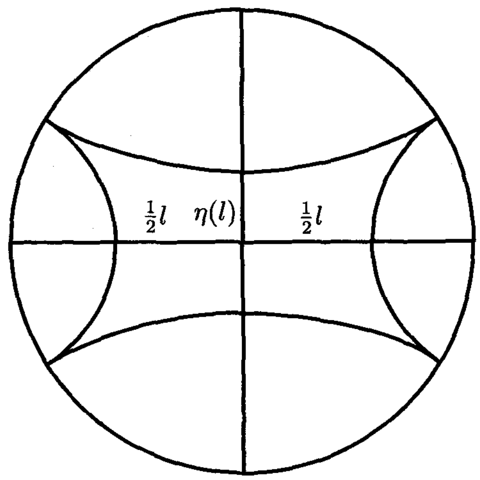


Fig. 3

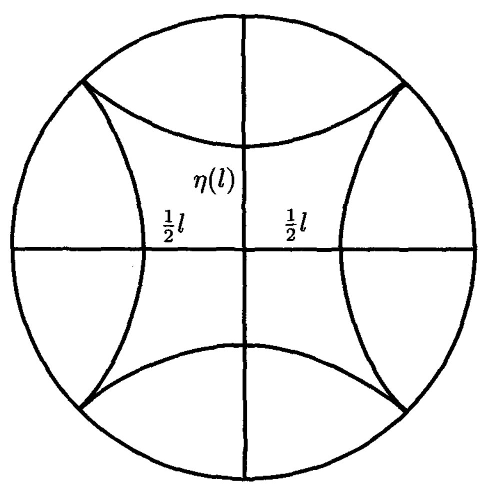


Fig. 4

PROPOSITION 6.7. *Let X be a hyperbolic Riemann surface and γ be a geodesic on X which intersects itself tranversally at least once. Then* $l_X(\gamma) \geqslant 2\log(\sqrt{2}+1)$.

Again the bound is sharp.

Proof. We can suppose without loss of generality that γ has a unique point of self-intersection x and thus consists of two loops γ_1 and γ_2 of lengths l_1 and l_2 respectively. Since γ_1 and γ_2 are simple closed curves, their lifts to the universal covering space D through a lift $\tilde{x}$ of x look like Figure 5.

Let γ be the indicated bisector of the angle at $\tilde{x}$, and let β_1, β_2 be the indicated

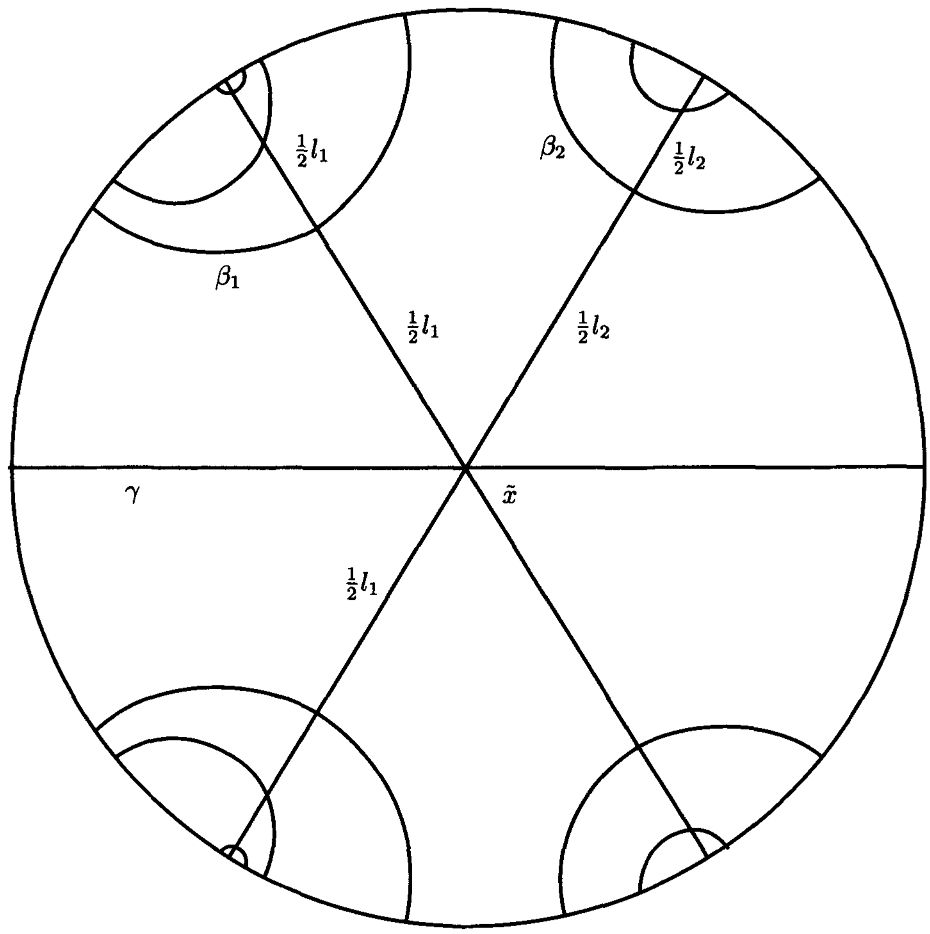


Fig. 5

perpendiculars. Then as in Theorem 6.3, the products of reflections

$$\varrho_{\beta_1} \circ \varrho_\gamma \quad \text{and} \quad \varrho_{\beta_2} \circ \varrho_\gamma$$

are both in $\pi_1(X)$, and so β_1 and β_2 do not intersect.

We see the following configuration in D: three disjoint geodesics γ, β_1 and β_2 with β_1 and β_2 on the same side of γ, a point $\tilde{x}$ on γ and geodesics from $\tilde{x}$ to β_1 and β_2 of lengths $\frac{1}{2}l_1$ and $\frac{1}{2}l_2$ respectively. It is easy to show that the minimum of l_1+l_2 in this situation is realized when γ, β_1 and β_2 bound an ideal triangle, and $\tilde{x}$ is the projection of the point at infinity where β_1 and β_2 meet onto γ.

This minimum is realized on the sphere with three points removed, say $-1, 1$ and ∞, by the figure eight curve as in Figure 6(a), whose length can be computed to be $2\log(\sqrt{2}+1)$ from Figure 6(b).

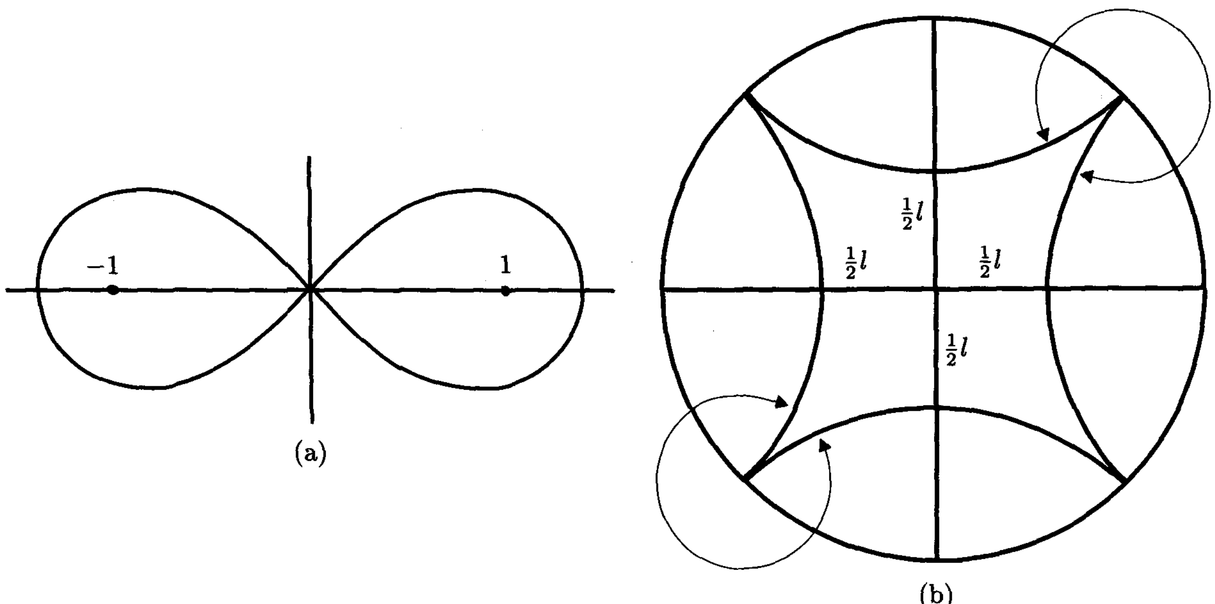


Fig. 6

7. Asymptotic geometry of Riemann surfaces

If $X'\subset X$ and γ is a curve on X', then

$$l_{X'}(\gamma)\geqslant l_X(\gamma),$$

since the injection $X'\to X$ is analytic, hence length decreasing for the Poincaré metric.

If γ is a short curve on X, and X' is obtained from X by deleting a finite number of points, then this inequality can be sharpened.

THEOREM 7.1. *Let X be a Riemann surface, $P\subset X$ a finite set, with $\#P=p>0$. Set $X'=X-P$, and choose $L<\log(\sqrt{2}+1)$. Let γ be a simple closed curved geodesic on X, and $\{\gamma_1',...,\gamma_s'\}$ be the closed geodesics of X^i homotopic to γ in X and of length $<L$. Set $l=l_X(\gamma)$, $l_i'=l_{X'}(\gamma_i')$. Then*

(a) $s\leqslant p+1$;

(b) *for all* i, $l_i'>l$;

(c) $1/l-2/\pi-(p+1)/L<\sum_{i=1}^s 1/l_i'<1/l+2(p+1)/\pi$.

Proof. (a) By Corollary 6.6, the γ_i' are disjoint since $L<\log(\sqrt{2}+1)$. Then $s-1$ of the components of $X-\{\bigcup_{i=1}^s\gamma_i'\}$ are annuli, and at least one point of P must belong to each, so $p\geqslant s-1$.

(b) The inclusion $X'\to X$ is analytic hence length decreasing.

(c) First let us verify the right-hand inequality. According to Theorem 6.3 and Proposition 6.1, there exist disjoint cylinders $C_i' \subset X'$ with equators γ_i' and moduli

$$\operatorname{mod}(C_i') \geqslant \frac{\pi}{2l_i'} - 1.$$

These cylinders lift to disjoint cylinders in $\widetilde{X}_\gamma$ homotopic to the equator $\widetilde{\gamma}$. Whenever an annulus A contains disjoint annuli A_i homotopic to the equator of A, we have, [O, Theorem 2.44], $\operatorname{mod}(A) \geqslant \sum \operatorname{mod}(A_i)$. Therefore

$$\frac{\pi}{2l} = \operatorname{mod}(\widetilde{X}_\gamma) \geqslant \sum \operatorname{mod}(C_i') > \frac{\pi}{2}\sum \frac{1}{l_i'} - s > \frac{\pi}{2}\left(\sum \frac{1}{l_i'} - \frac{2(p+1)}{\pi}\right).$$

Now for the left-hand inequality. According to Theorem 6.3 and Proposition 6.1, there is a cylinder $C \subset X$ with equator γ and

$$\operatorname{mod} C > \frac{\pi}{2l} - 1.$$

The parallels (curves at constant distance from the equator) of C passing through the points of $C \cap P$ cut C into s' annuli C_j, $j=1, \ldots, s'$ with $s' \leqslant p+1$.

For each j let γ_j be the geodesic of X' homotopic to C_j and $l_j = l_{X'}(\gamma_j)$. Then

$$\frac{\pi}{2}\left(\frac{1}{l} - \frac{2}{\pi}\right) < \operatorname{mod} C = \sum_{j=1}^{s'} \operatorname{mod} C_j < \sum_{j=1}^{s'} \frac{\pi}{2l_j},$$

so

$$\frac{1}{l} - \frac{2}{\pi} < \sum_{j=1}^{s'} \frac{1}{l_j}.$$

Let $J_- = \{j \mid l_j < L\}$ and $J_+ = \{j \mid l_j \geqslant L\}$. The γ_j for $j \in J_-$ are among the γ_i' so

$$\sum_{j \in J_-} \frac{1}{l_j} \leqslant \sum \frac{1}{l_i'}.$$

On the other hand, $\sum_{j \in J_+} 1/l_j \leqslant (p+1)/L$. So

$$\frac{1}{l} - \frac{2}{\pi} \leqslant \sum \frac{1}{l_i'} + \frac{p+1}{L}. \qquad \square$$

Let $P \subset S^2$ be a finite set. For any closed curve $\gamma \subset S^2 - P$, and any $\tau \in \mathcal{T}_{(S^2,P)} = \mathcal{T}(P)$, represented by

$$\pi\colon (S^2, P) \to \mathbf{P}^1,$$

we can define $l_\tau(\gamma)$ to be the length of the geodesic homotopic to $\phi(\gamma)$ on $\mathbf{P}^1 - \phi(P)$; define $w(\gamma, \tau) = -\log l_\tau(\gamma)$.

PROPOSITION 7.2. *The function $\mathcal{T}(P)\to\mathbf{R}$ given by $\tau\mapsto w(\gamma,\tau)$ is Lipshitz, with Lipshitz constant* 2.

Proof. Let τ and τ' be represented respectively by $\phi,\phi'\colon(S^2,P)\to\mathbf{P}^1$, and set $P_\tau=\phi(P)$, $P_{\tau'}=\phi'(P)$.

If $d(\tau,\tau')=\delta$, then there exists a K-quasiconformal map $\psi\colon\mathbf{P}^1\to\mathbf{P}^1$ with $\psi(P_\tau)=P_{\tau'}$ and $K=e^{2\delta}$.

The mapping ψ lifts to the covering spaces:

$$\tilde{\psi}\colon(\widetilde{\mathbf{P}^1-P_\tau})_\gamma\to(\widetilde{\mathbf{P}^1-P_{\tau'}})_\gamma,$$

which are annuli of moduli $m=\pi/2l_\tau(\gamma)$ and $m'=\pi/2l_{\tau'}(\gamma)$ respectively, and this is possible only if

$$\frac{1}{K}\leqslant\frac{m}{m'}\leqslant K,\quad\text{i.e.,}$$

$$|w(\gamma,\tau)-w(\gamma,\tau')|\leqslant\log K=2\delta.\qquad\square$$

For $\tau\in\mathcal{T}(P)$ set $w(\tau)=\sup_\gamma w(\gamma,\tau)$; this sup is finite since there is only a finite number of curves of length $<\log(\sqrt{2}+1)$.

PROPOSITION 7.3. (a) *The function $\tau\mapsto w(\tau)$ has Lipshitz constant* 2.

(b) *For any $M\in\mathbf{R}$,*

$$\{\tau\in\mathcal{T}(P)\mid w(\tau)\leqslant M\}$$

is the inverse image of a compact subset of $\mathcal{M}(P)$.

Proof. (a) Comes from Proposition 7.2.

(b) Let (τ_n) be a sequence in $\mathcal{T}(P)$, and suppose that the images $\pi(\tau_n)$ in $\mathcal{M}(P)$ are represented by injections $i_n\colon P\to\mathbf{P}^1$, normalized so that for some 3 points x_1,x_2,x_3 of p we have $i_n(x_1)=0$, $i_n(x_2)=1$ and $i_n(x_3)=\infty$ for all n. Since $\mathbf{P}^1$ is compact, we can extract a subsequence, say j_n such that $j=\lim_{n\to\infty}j_n$ exists.

If j is injective, the subsequence converges in $\mathcal{M}(P)$.

If j is not injective, there exists $y_1\neq y_2$ in P with $j(y_1)=j(y_2)=Y$; we may assume $Y\neq\infty$. Let

$$R=\inf_{j(x)\neq Y}|j(x)-Y|<\infty.$$

Then for any ε, there exists N such that for $n>N$, there are no points of $j_n(P)$ in

$$\{z\mid\varepsilon<|z-Y|<R-\varepsilon\},$$

but at least two points inside and outside.

Then the curve $|z-Y|=\frac{1}{2}R$ has length less than $\pi^2/(\log R-\log\varepsilon)$, which goes to 0 with ε. $\square$

8. Sufficiency of the criterion

For any $\tau\in\mathcal{T}_f$, let

$$L_\tau=\{w(\gamma,\tau)\mid\gamma \text{ a closed curve on } S^2-P_f\},$$

so that $w(\tau)=\sup(L_\tau)$. Also, if Γ is a multicurve, let

$$w(\Gamma,\tau)=\sup_{\gamma\in\Gamma} w(\gamma,\tau).$$

For any $J>0$, let $]a,b[$ be the lowest interval in $\mathbf{R}-L_\tau$ of length J, with

$$a\geqslant -\log\log\left(\sqrt{2}+1\right)=A,$$

and

$$\Gamma_{J,\tau}=\{\gamma\mid w(\gamma,\tau)\geqslant b\}.$$

Let $\tau'=\sigma_f(\tau)$, and ϕ,ϕ' and f_τ be as in Proposition 2.2. Let $P=\phi(P_f)$, $P'=\phi'(P_f)$ and $P''=f_\tau^{-1}(P)$.

PROPOSITION 8.1. (a) *If $J\geqslant\log d+2d(\tau,\tau')$ and if $\Gamma_{J,\tau}\neq\varnothing$, then $\Gamma_{J,\tau}$ is an f-stable multicurve.*

(b) *The simple closed geodesics on $\mathbf{P}^1-P''$ of length less than de^{-b} are the components of $f_\tau^{-1}(\gamma)$ for $\gamma\in\Gamma_{J,\tau}$.*

Proof. Since $e^{-b}<\log(\sqrt{2}+1)$, all the curves of $\Gamma_{J,\tau}$ are disjoint by Corollary 6.6.

If $\gamma\in\Gamma_{J,\tau}$ and γ' is a component of $f^{-1}(\gamma)$ then

$$l_{\mathbf{P}^1-P''}(\gamma')-d_\alpha l_\tau(\gamma),$$

where $d_\alpha=\deg f_\tau|_{\gamma'}:\gamma'\to\gamma$ and so $d_\alpha\leqslant d$, so

$$w(\tau',\gamma')>w(\tau,\gamma)-\log d\geqslant b-\log d.$$

On the other hand, if $\gamma''\in\Gamma_{J,\tau}$, then $w(\gamma'',\tau)\leqslant a$ and so $w(\gamma'',\tau)\leqslant a+2d(\tau,\tau')$, by Proposition 7.2.

Since $b-a=J>\log d+2d(\tau,\tau')$, we see that $\gamma'\neq\gamma''$, so $\gamma'\in\Gamma_{J,\tau}$. This proves (a) and half of (b).

For the other half of (b), let γ' be any simple closed geodesic on $\mathbf{P}^1=P''$ of length $<de^{-b}$. Then $f_\tau(\gamma')$ is a geodesic on $\mathbf{P}^1-P$ of the same length, which may fail to be simple. It cannot have any transverse self-intersections by Proposition 6.7, since $de^{-b}<2\log(\sqrt{2}+1)$. So it must cover some simple closed geodesic γ on $\mathbf{P}^1-P$ with degree $\leqslant d$, so $l_\tau(\gamma)\leqslant l_{\mathbf{P}^1-P''}(\gamma')$, i.e., $w(\tau,\gamma)\geqslant b-\log d$.

Since there is a gap of length J in L_τ, this shows that $\gamma\in\Gamma_{J,\tau}$. □

The theorem will now follow easily from the following proposition.

PROPOSITION 8.2. *There exists an integer $m \geqslant 1$ depending only on the degree d of f and the cardinality p of P_f, and a constant C depending only on p, d, and $D=d(\tau, \sigma_f(\tau))$, such that if $J=m(\log d+2D)$ and $\Gamma=\Gamma_{J,\tau}$ then whenever $w(\tau)>C$, Γ is non-empty and $w(\Gamma, \sigma_f^m(\tau))<w(\Gamma, \tau)$.*

Proof. It is here that we use Lemma 1.2 in an essential way. Let Γ be a multicurve with $\lambda(\Gamma, f)<1$. Give $\mathbf{R}^\Gamma$ the sup norm, and $\mathrm{Hom}(\mathbf{R}^\Gamma, \mathbf{R}^\Gamma)$ the corresponding norm.

Since there are only finitely many possible matrices for f_Γ, we can choose m such that

$$\|f_\Gamma^m\| < \tfrac{1}{2},$$

independent of the multicurve Γ.

Now let $]a, b[$ be the lowest gap in L_τ of length J with $a \geqslant A$ as before. Since there are at most $p-3$ elements of L_τ greater than A, we see that if $w(\tau)>(p-3)J+A=B$ then $b<B$. Let $L_0=d^m e^{-B}$; note that L_0 depends only on p, d and D.

Clearly if $w(\tau)>B$ then $\Gamma=\Gamma_{J,\tau}\neq\varnothing$, and Γ is an f-stable multicurve by Proposition 8.1(a).

Let $\tau'=\sigma_f^m(\tau)$, and let ϕ, ϕ' and f_τ^m be as in Proposition 2.2 applied to f^m, so that f_τ^m is analytic and the diagram

$$\begin{array}{ccc} (S^2, P_f) & \xrightarrow{\phi'} & \mathbf{P}^1 \\ {\scriptstyle f^m}\downarrow & & \downarrow{\scriptstyle f_\tau^m} \\ (S^2, P_f) & \xrightarrow[\phi]{} & \mathbf{P}^1 \end{array}$$

commutes.

Let $P=\phi(P_f)$, $P'=\phi'(P_f)$ and $P''=(f_\tau^m)^{-1}(P)$. Define $v, v' \in \mathbf{R}^\Gamma$ by

$$v=\begin{bmatrix} 1/l_\tau(\gamma_1) \\ \vdots \\ 1/l_\tau(\gamma_n) \end{bmatrix}, \quad v'=\begin{bmatrix} 1/l_{\tau'}(\gamma_1) \\ \vdots \\ 1/l_{\tau'}(\gamma_n) \end{bmatrix}.$$

Consider a curve $\gamma_i \in \Gamma$, and the components $\gamma_{i,j,\alpha}$ of $(f_\tau^m)^{-1}(\gamma_j)$ homotopic to γ_i in $\mathbf{P}^1-P'$. We wish to apply the left-hand inequality of Theorem 7.1(c) to the geodesic on $\mathbf{P}^1-P'$ (the X of Proposition 7.1) in the homotopy class of γ_i, and the geodesics in the classes of $\gamma_{i,j,\alpha}$ on $X''=\mathbf{P}^1-P''$ (the X' of Theorem 7.1). Using Proposition 8.1(b), we see that the hypothesis of Theorem 7.1 is satisfied with $L=d^m e^{-b}>d^m e^{-B}=L_0$.

Moreover, since $f_\tau^m|_{\mathbf{P}^1-P''}$ is a covering map, we have

$$\sum_{\alpha, j} \frac{1}{l_{X''}(\gamma_{i,j,\alpha})}=[(f_\Gamma)^m v]_i.$$

Define r by $v'=(f_\Gamma)^m v+r$; by Theorem 7.1 we have

$$r_i<\frac{2}{\pi}+\frac{pd^m}{L}\leqslant\frac{2}{\pi}+\frac{pd^m}{L_0},$$

so

$$|v'|\leqslant\tfrac{1}{2}|v|+\frac{2}{\pi}+\frac{pd^m}{L_0}.$$

Now if x and y are any two numbers such that

$$x\leqslant\tfrac{1}{2}y+K\quad\text{and}\quad y>2K,$$

we have $x<y$.

So if

$$|v|\geqslant 2\left(\frac{2}{\pi}+\frac{pd^m}{L_0}\right),$$

we have

$$|v'|<|v|.$$

We see that if we choose

$$C=\sup\left(\log 2\left(\frac{2}{\pi}+\frac{pd^m}{L_0}\right),B\right),$$

the proposition is proved. □

Proof of the theorem. Suppose that f is not equivalent to a rational function. Choose $\tau_0\in\mathcal{T}_f$, set $\tau_i=\sigma_f(\tau_{i-1})$ and let C and m be as in Proposition 8.2 with $D=d(\tau_0,\tau_1)$. Raise C if necessary so that $w(\tau_0)<C$.

By Propositions 2.3, 5.1 and 7.3, the sequence $w(\tau_i)$ is unbounded. Consider the first i for which $w(\tau_i)$ is unbounded. Consider the first i for which $w(\tau_i)>C+2mD$, and let $\Gamma=\Gamma_{j,\tau_i}$ as in Proposition 8.2.

By Proposition 7.2, $w(\Gamma,\tau_{i-m})>C$, so that if $\lambda(f,\Gamma)<1$, we find that

$$w(\Gamma,\tau_1)<w(\Gamma,\tau_{i-m})<C+2mD,$$

a contradiction. So $\lambda(f,\Gamma)\geqslant 1$. □

9. The non-hyperbolic case

PROPOSITION 9.1. (a) *If $f\colon S^2\to S^2$ is a critically finite branched mapping, then $\chi(O_f)\leqslant 0$.*

(b) *If $\chi(O_f)=0$, then $f\colon O_f\to O_f$ is a covering map of orbifolds.*

Proof. Let $O'_f=(S^2,\nu'_f)$ where

$$\nu'_f(x)=\frac{\nu_f(f(x))}{\deg_x f}.$$

Then $\nu'_f \geqslant \nu_f$ everywhere (recall that $\nu_f(x)=1$ if $x \notin P_f$), so

$$\chi(O'_f) \leqslant \chi(O_f).$$

However, $f\colon O'_f \to O_f$ is a covering map of orbifolds of degree d, so

$$\chi(O'_f)=d\chi(O_f).$$

Thus $(d-1)\chi(O_f)\leqslant 0$ and (a) follows since $d>1$.

Moreover, to get equality we must have $O'_f=O_f$. □

There are precisely six orbifolds homeomorphic to S^2 with Euler characteristic 0. They are given by the following weights at the weighted points:

(1) (∞,∞),
(2) $(2,2,\infty)$,
(3) $(2,4,4)$,
(4) $(2,3,6)$,
(5) $(3,3,3)$,
(6) $(2,2,2,2)$.

In cases (1)–(5), the orbifolds have a unique complex structure, since there are at most 3 marked points, and any three distinct points can be moved to any other by an automorphism of $\mathbf{P}^1$. They can be realized as $\mathbf{C}/\Gamma$ where Γ is a discrete subgroup of $\mathrm{Aut}(\mathbf{C})$, as follows.

(1) $\Gamma=\mathbf{Z}$, acting by translations;
(2) Γ generated by $\mathbf{Z}$ as above, and $z\mapsto -z$;
(3) Γ generated by $z\mapsto z+a$, $a\in\mathbf{Z}[i]$; $z\mapsto iz$;
(4) Γ generated by $z\mapsto z+a$, $a\in\mathbf{Z}[\omega]$, $\omega=e^{i\pi/3}$; $z\mapsto \omega z$;
(5) Γ generated by $z\mapsto z+a$, $a\in\mathbf{Z}[\omega]$; $z\mapsto \omega^2 z$.

By Corollary 2.4, any branched map f with $P_f=3$ is equivalent to one which preserves the unique complex structure of O_f, so using the identifications (1) through (5) above, we see that $\tilde f\colon \widetilde{O}_f\to\widetilde{O}_f$ can be taken to be an automorphism $z\mapsto az+b$ of $\mathbf{C}$ with $\deg f=|a|^2$.

It is now routine (rather tedious) to write down the maps $\tilde f$ which induce a map on $\mathbf{C}/\Gamma$.

PROPOSITION 9.2. *The critically finite branched maps with* $\chi(O_f)=0$ *and* $\#P_f\leqslant 3$ *are all equivalent to one of the maps* f *induced on* $\mathbf{C}/\Gamma$ *by*:

(1) $z\mapsto nz$, $n\in\mathbf{Z}$, $|n|>1$;

(2) $z\mapsto nz$, *n as above*; $z\mapsto nz+\frac{1}{2}$, *n as above*;

(3) $z\mapsto az$, $a\in\mathbf{Z}[i]$, $|a|\geqslant 2$; $z\mapsto az+\frac{1}{2}(1+i)$, *a as above*;

(4) $z\mapsto az$, $a\in\mathbf{Z}[\omega]$, $|a|\geqslant 3$;

(5) $z\mapsto az$, *a as above*; $z\mapsto az+\frac{1}{3}(\omega+1)$, *a as above*; $z\mapsto az+\frac{1}{3}i\sqrt{3}$, *a as above.*

Proof. The verification is left to the assiduous reader.

Remarks. (i) The maps above are not all inequivalent.

(ii) In case (1) above, the associated rational functions are

$$z\mapsto z^n,\quad |n|>1,$$

since $z\mapsto e^{2\pi iz}$ is the universal covering map of O_f.

In case (2), they are (up to sign) the Tchebycheff polynomials $P_n(z)$, i.e., the polynomials such that

$$P_n(\cos z)=\cos nz,$$

since $z\mapsto\cos 2\pi z$ is the universal covering map of O_f.

Note that in both of these cases, the rational functions are related to the addition formula for the exponential function. In cases (3)–(5) (and (6) below), the rational functions are related to the addition formulae for elliptic functions.

(iii) Cases (1) and (2) are precisely the rational functions we know for which the Julia set is not $\mathbf{P}^1$ and is not "fractal". Is this indeed the complete list?

Finally we come to (6), the most interesting case. In this case, the possible complex structures on O_f are given by $\mathbf{C}/\Gamma_\tau$, where τ is in the upper half plane H and Γ_τ is the subgroup of $\mathrm{Aut}(\mathbf{C})$ generated by $z\mapsto z+1$, $z\mapsto z+\tau$ and $z\mapsto -z$.

PROPOSITION 9.3. *The rational maps* $f\colon S^2\to S^2$ *with orbifold* $(2,2,2,2)$ *are induced on* $\mathbf{P}^1$ *by an isomorphism* $\mathbf{P}^1\to\mathbf{C}/\Gamma_\tau$ *and a map* $\tilde f\colon\mathbf{C}\to\mathbf{C}$, $z\mapsto z\alpha+\beta$, *where*

(a) α *is an integer in an imaginary quadratic field* K;

(b) $2\beta\in\Gamma_\tau$;

(c) *if* α *is not real, then* Γ_τ *is a module over the subring of* K *generated by* 1 *and* α, *two such modules giving the same mapping if they are isomorphic.*

Proof. Let $\pi\colon\mathbf{C}\to\mathbf{P}^1$ be the universal covering of O_f. Such a rational mapping must lift to an automorphism $\tilde f\colon\mathbf{C}\to\mathbf{C}$ of the form $z\mapsto\alpha z+\beta$.

Since $\pi(0)\in P_f$, we must have $\pi(\tilde f(0))\in P_f$ so

$$\beta=\tilde f(0)\in\tfrac{1}{2}\Gamma_\tau.$$

Moreover, $(\tilde{f}-\tilde{f}(0))\Gamma_\tau \subset \Gamma_\tau$, so $\alpha\Gamma_\tau \subset \Gamma_\tau$. Therefore there exist integers a, b, c, d, such that

$$\alpha = a + b\tau, \quad \alpha\tau = c + d\tau,$$

which gives

$$\alpha^2 - (a+d)\alpha + ad - bc = 0.$$

So α is a quadratic integer, necessarily either a rational integer or an integer in an imaginary quadratic field.

Part (2) was done above, and (3) is obvious. □

Remark. This proposition implies that there are only finitely many rational functions of given degree (up to conjugation by automorphisms of $\mathbf{P}^1$) with orbifold $(2,2,2,2)$ and induced by multiplication by a non-real quadratic integer α. Indeed, there are only finitely many such α with given $|a|^2 = \deg f$, and the class group of α is finite.

On the other hand, in degrees which are squares, there are one-parameter families of critically finite rational functions all of which are equivalent as branched mappings, but which are not conjugate by automorphisms of $\mathbf{P}^1$.

Proposition 9.3 does not solve our problem; we still need a topological criterion to decide if a branched map $f: S^2 \to S^2$ with orbifold $(2,2,2,2)$ is equivalent to a rational function. We will do this by finding an isomorphism of $\mathcal{T}_f$ with the upper half plane H, and identify σ_f as a fractional linear transformation.

The (differentiable) orbifold O_f can be identified with $\mathbf{R}^2/\Gamma$, where Γ is the group of isometries of $\mathbf{R}^2$ generated by

$$x \mapsto x + a, \ a \in \mathbf{Z}^2, \quad \text{and} \quad x \mapsto -x.$$

Let $T_f = \mathbf{R}^2/\mathbf{Z}^2$; T_f is a torus and the canonical map $\pi: T_f \to S^2$ is a double cover ramified above P_f.

LEMMA 9.4. *The map f lifts to a covering map $\tilde{f}: T_f \to T_f$.*

Proof. This is a straightforward application of the lifting criterion for covering spaces. Since $f \circ \pi$ is a covering map, it induces an injection on fundamental groups, so the image is a subgroup of $\pi_1(O_f)$ isomorphic to $\mathbf{Z}^2$.

However, an element of Γ is either a translation or of order 2. So

$$(f \circ \pi)_*(\pi_1(T_f)) \subset \pi_*(\pi_1(T_f)).$$ □

Let γ_1, γ_2 be the curves on T_f images of the segments joining $(0,0)$ to $(1,0)$ and $(0,1)$ in $\mathbf{R}^2$.

Any complex structure μ on (S^2, P_f) induces a complex structure $\pi^*\mu$ on T_f. The space of 1-forms ϕ holomorphic for $\pi^*\mu$ is 1-dimensional, so the number

$$\tilde{\varrho}(\mu) = \frac{\int_{\gamma_2}\phi}{\int_{\gamma_1}\phi}$$

does not depend on the choice of ϕ.

LEMMA 9.5. *The map $\tilde{\varrho}$ induces an isomorphism*

$$\varrho\colon T_f \to H.$$

Proof. Clearly ϱ is well-defined. To show that it is an isomorphism, we will construct an inverse mapping.

For any $\tau \in H$, let $\psi_\tau\colon \mathbf{R}^2 \to \mathbf{C}$ be given by

$$\psi_\tau(x, y) = x + \tau y.$$

If μ_0 is the standard complex structure on $\mathbf{C}$ and $\mu_\tau = \psi_\tau^* \mu_0$, then the complex structure $\bar{\mu}_\tau$ on $\mathbf{R}^2/\Gamma = O_f$ satisfies $\varrho(\bar{\mu}_\tau) = \tau$. □

Let A_f be the matrix of $f_*\colon H_1(T_f) \to H_1(T_f)$ in the basis γ_1, γ_2; clearly $\det A_f = \deg f \geqslant 2$, and A is determined by f up to sign. Conversely, any matrix A with integer entries and $\det A \geqslant 2$ arises as A_f, namely for the map f induced on $\mathbf{R}^2/\Gamma$ by $A\colon \mathbf{R}^2 \to \mathbf{R}^2$.

LEMMA 9.6. *If*

$$A_f = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

then $\varrho \circ \sigma_f \circ \varrho^{-1}$ is the fractional linear transformation

$$z \mapsto \frac{dz + b}{cz + a}.$$

Proof. By Stokes' theorem,

$$\int_{f_*\gamma} \phi = \int_\gamma f^*\phi,$$

so

$$(\varrho \circ \sigma_f \circ \varrho^{-1}) \frac{\int_{\gamma_2} f^*\phi}{\int_{\gamma_1}\phi} = \frac{\int_{\gamma_2} f^*\phi}{\int_{\gamma_1} f^*\phi} = \frac{b\int_{\gamma_1}\phi + d\int_{\gamma_2}\phi}{a\int_{\gamma_1}\phi + c\int_{\gamma_2}\phi} = \frac{b + d\int_{\gamma_2}\phi / \int_{\gamma_1}\phi}{a + c\int_{\gamma_2}\phi / \int_{\gamma_1}\phi}.$$ □

PROPOSITION 9.7. *A branched mapping $f: S^2 \to S^2$ with orbifold $(2,2,2,2)$ is equivalent to a rational function if and only if the eigenvalues of A_f are not real, or if A_f is a multiple of the identity.*

Proof. This follows immediately from Lemma 9.6 and Proposition 2.3. Indeed, the fractional linear transformation $z \mapsto (dz+b)/(cz+a)$ has a unique fixed point in H if and only if the eigenvalues of A_f are not real, and otherwise has no fixed point in H unless it is the identity. □

Remarks. (i) The eigenvalues of A_f are not real when the number α of Proposition 9.3 is not real, and A_f is a multiple of the identity if α is a rational integer.

(ii) Lemma 9.6 gives examples of branched mappings f where σ_f is the identity, or an elliptic, parabolic or hyperbolic automorphism of $\mathcal{T}_f = H$. These examples are however misleading; in general, σ_f is neither injective, surjective or proper.

Appendix: Examples of Thurston mappings

Following Milnor's suggestion, we will call a critically finite branched mapping a *Thurston mapping*.

Example 1. The easiest examples of Thurston mappings are simply postcritically finite rational functions, such as

$$\begin{array}{lll} f(z)=z^k,\ |k|>1, & \Omega_f=\{0,\infty\}, & P_f=\{0,\infty\}, \\ f(z)=z^2-2, & \Omega_f=\{0,\infty\}, & P_f=\{\infty,-2,2\}, \\ f(z)=z^2+i, & \Omega_f=\{0,\infty\}, & P_f=\{\infty,i,-1+i,-i\}, \\ f(z)=\frac{1}{2}i(z+1/z), & \Omega_f=\{1,-1\}, & P_f=\{i,-i,0,\infty\}. \end{array}$$

These examples are a bit misleading; one should not think of a Thurston mapping as a rigid, analytic object, but as something topological, "defined up to homotopy". It is not hard to construct such things: for instance, take the third example above, and compose it with the Dehn twist around a curve on $\mathbf{P}-P_f$. Such examples are quite mysterious; we do not know if they admit Thurston obstructions, nor if they do not, what polynomial they are equivalent to (even though there is not much choice).

The next family of examples is a slight modification of the spiders considered in [BFH]; the reader is invited to read the general treatment there. We will need the construction for Example 3 below, which brings out some of the difficulties in the proof of Thurston's theorem, and justifies the repetition.

Example 2. Choose an angle

$$\theta=\theta_1=\frac{p}{2^k(2^l-1)}$$

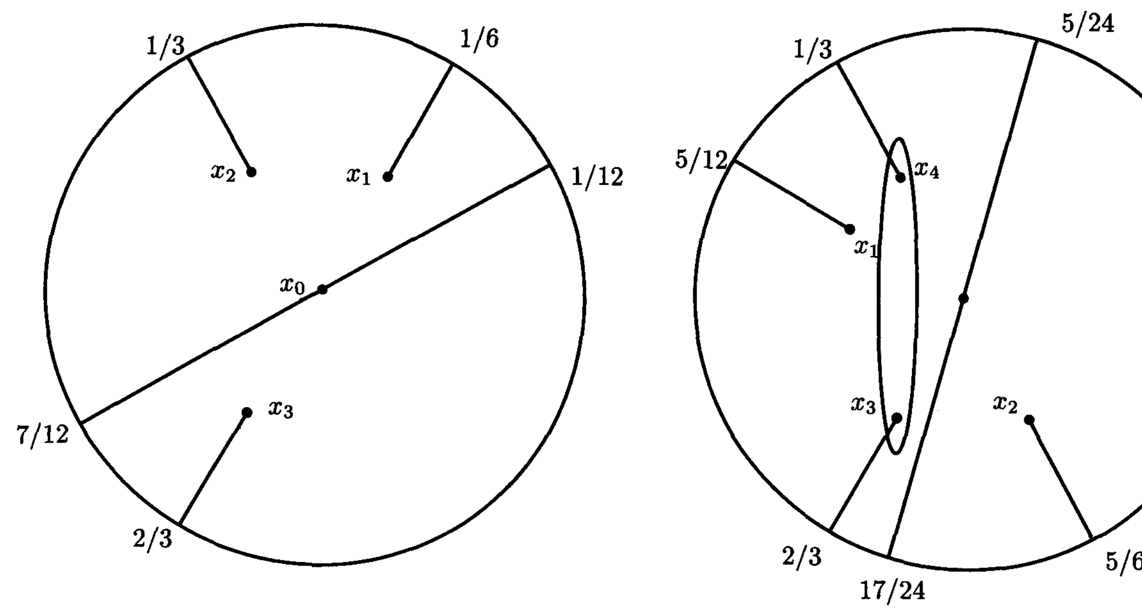


Fig. 7. The spiders for $\theta=1/6$ and $\theta=5/12$

with $k>0$, and set $\theta_n=2\theta_{n-1}=2^{n-1}\theta_1$, so that $\theta_{k+l+1}=\theta_{k+1}$. In the unit disc, draw the segments γ_n joining the points

$$\tfrac{1}{2}e^{2\pi i\theta_n} \quad \text{and} \quad e^{2\pi i\theta_n}$$

and the diagonal γ_0 joining

$$e^{\pi i\theta_1} \quad \text{and} \quad -e^{\pi i\theta_1}.$$

Then there exists a branched mapping f_θ which:

(1) Outside of the unit disc is $z\mapsto z^2$;

(2) Folds γ_0 at the origin and maps it to γ_1;

(3) Maps each γ_n homeomorphically to γ_{n+1};

(4) And except for the folding of γ_0 is a homeomorphism mapping each half of the unit disc cut along γ_0 to the whole disc.

Here are two examples of this construction, one for $\theta=1/6$ and one for $\theta=5/12$. (See Figure 7.)

The branched mapping $f_{1/6}$ is equivalent to a rational function, in fact to the polynomial $z\mapsto z^2+i$, and therefore has no Thurston obstruction. On the other hand, for $f_{5/12}$, the curve surrounding x_3 and x_4 is a Thurston obstruction by itself. Its inverse image consists of a curve homotopic to itself, and another which surrounds only x_2 and hence is peripheral. Clearly the Thurston matrix is in this case simply the number 1.

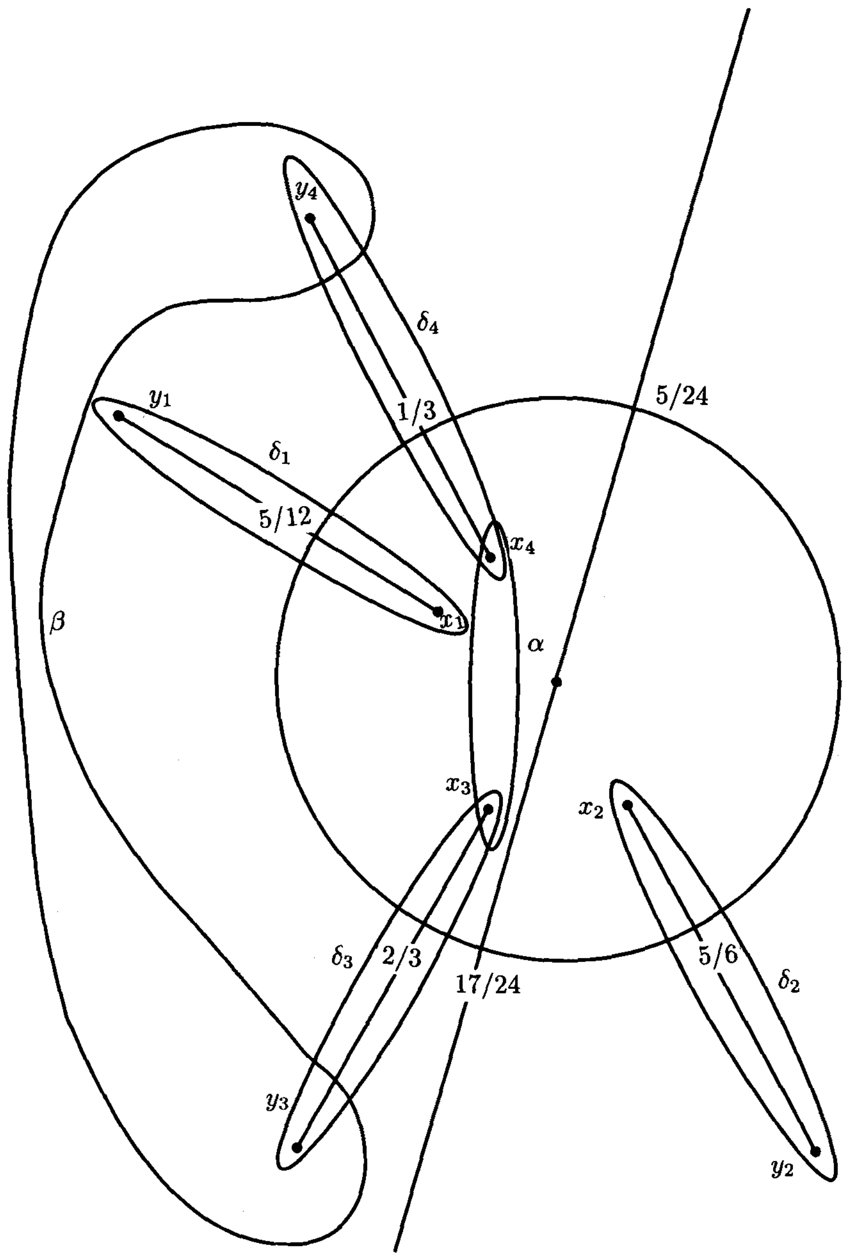


Fig. 8. The mating of $f_{5/12}$ with its conjugate

Example 3. Now for a really complicated example (see Figure 8): let us keep the mapping above inside the unit disc, and put its symmetric on the outside of the unit disc. We still have a Thurston mapping, with $\Omega_f=\{0,\infty\}$, and $P_f=\{x_1,...,x_4,y_1,...,y_4\}$.

In this case, there are four Thurston obstructions:

(1) The curve $\Gamma_1=\{\alpha\}$, with matrix 1;

(2) The curve $\Gamma_2=\{\beta\}$, with matrix 1;

(3) The multicurve $\Gamma_3=\{\delta_1,...,\delta_4\}$, with matrix

$$f_{\Gamma_3}=\begin{bmatrix}0&1&0&0\\2&0&1&0\\0&0&0&1\\0&0&1&0\end{bmatrix}\quad\text{and}\quad\lambda(\Gamma_3)=\sqrt{2};$$

(4) The multicurve $\Gamma_4=\{\delta_1,\delta_2\}$, with matrix

$$f_{\Gamma_4}=\begin{bmatrix}0&1\\2&0\end{bmatrix}\quad\text{and}\quad\lambda(\Gamma_4)=\sqrt{2}.$$

The first two are fairly obvious, but the third and fourth require a bit of checking: the inverse image of δ_1 consists of two curves, one on each side of the diameter. But recall that the homotopies are relative to the post-critical set, and that the critical points are not in the post-critical set in this case. Therefore these curves are both homotopic to δ_2.

The surprizing thing about this mapping is that as the Thurston transformation σ_f is iterated, starting from some τ_0, the lengths of α,β, and the supremum of the lengths of δ_3 and δ_4 all have to strictly decrease. But this prevents any of these from tending to 0, because they intersect, and two short geodesics can never intersect. By Thurston's theorem some curves must have lengths shrinking to 0, and it is not too hard to see that it is the curves of Γ_4. This also follows from the proof of Thurston's theorem, since Γ_4 is a minimal invariant multicurve, with leading eigenvalue $\sqrt{2}>1$.

References

[A] AHLFORS, L., *Lectures on Quasiconformal Mappings.* Van Nostrand, 1966.

[B] BEARDON, A. F., *The Geometry of Discrete Groups.* Springer-Verlag, 1983.

[BFH] BIELEFELD, B., FISCHER, Y. & HUBBARD, J. H., The classification of critically preperiodic polynomials as dynamical systems. *J. Amer. Math. Soc.*, 5 (1992), 721–762.

[G] GUNNING, R., *Lectures on Riemann Surfaces.* Mathematical Notes, Princeton University Press, 1966.

[H] HUBBARD, J. H., *Sur les sections analytiques de la courbe universelle de Teichmüller.* Mem. Amer. Math. Soc., 166, 1976.

[H-M] HUBBARD, J. H. & MASUR, H., Quadratic differentials and foliations. *Acta Math.*, 142 (1979), 221–274.

[J] JENKINS, J. A., On the existence of certain general extremal metrics. *Ann. of Math.*, 66 (1957), 440–453.

[O] OHTSUKA, M., *Dirichlet Problem, Extremal Length and Prime Ends.* Van Nostrand Reinhold Math. Stud., 22, 1970.

[R] ROYDEN, H., Automorphisms and isometries of Teichmüller space, in *Advances in the Theory of Riemann Surfaces*, pp. 369–383. Ann. of Math. Stud., 66. Princeton University Press, 1971.

[S] STREBEL, K., *On Quadratic Differentials and Extremal Quasiconformal Mappings.* Lecture Notes, University of Minnesota, 1967.

[T1] THURSTON, W., *Lecture Notes.* Princeton University.

[T2] — *Lecture Notes.* CBMS Conference, University of Minnesota at Duluth, 1983.

ADRIEN DOUADY
Département Mathématiques
Université Paris-Sud
Bâtiment 425
F-91405 Orsay Cedex
France

JOHN H. HUBBARD
Department of Mathematics
Cornell University
Ithaca, NY 14853-7801
U.S.A.

Received February 28, 1990
Received in revised form March 4, 1993