DYNAMICS IN ONE COMPLEX VARIABLE

Introductory Lectures

(Partially revised version of 9-5-91)

John Milnor

Institute for Mathematical Sciences, SUNY, Stony Brook NY

Table of Contents.

PREFACE.

These notes will study the dynamics of iterated holomorphic mappings from a Riemann surface to itself, concentrating on the classical case of rational maps of the Riemann sphere. They are based on introductory lectures given at Stony Brook during the Fall Term of 1989-90. I am grateful to the audience for a great deal of constructive criticism, and to Branner, Douady, Hubbard, and Shishikura who taught me most of what I know in this field. The surveys by Blanchard, Devaney, Douady, Keen, and Lyubich have been extremely useful, and are highly recommended to the reader. (Compare the list of references at the end.) Also, I want to thank A. Poirier for his criticisms of my first draft.

This subject is large and rapidly growing. These lectures are intended to introduce the reader to some key ideas in the field, and to form a basis for further study. The reader is assumed to be familiar with the rudiments of complex variable theory and of two-dimensional differential geometry.

John Milnor, Stony Brook, February 1990

CHRONOLOGICAL TABLE

Following is a list of some of the founders of the field of complex dynamics.

Hermann Amandus Schwarz	1843–1921
Henri Poincaré	1854–1912
Gabriel Koenigs	1858–1931
Léopold Leau	1868–1940 ±
Lucyan Emil Böttcher	1872– ?
Constantin Carathéodory	1873–1950
Samuel Lattès	1875 ±-1918
Paul Montel	1876–1975
Pierre Fatou	1878–1929
Paul Koebe	1882–1945
Gaston Julia	1893–1978
Carl Ludwig Siegel	1896–1981
Hubert Cremer	1897–1983
Charles Morrey	1907–1984

Among the many present day workers in the field, let me mention a few whose work is emphasized in these notes: Lars Ahlfors (1907), Lipman Bers (1914), Adrien Douady (1935), Vladimir I. Arnold (1937), Dennis P. Sullivan (1941), Michael R. Herman (1942), Bodil Branner (1943), John Hamal Hubbard (1945), William P. Thurston (1946), Mary Rees (1953), Jean-Christophe Yoccoz (1955), Mikhail Y. Lyubich (1959), and Mitsuhiro Shishikura (1960).

RIEMANN SURFACES

§1. Simply Connected Surfaces.

The first two sections will present an overview of well known material.

By a *Riemann surface* we mean a connected complex analytic manifold of complex dimension one. Two such surfaces S and S' are *conformally isomorphic* if there is a homeomorphism from S onto S' which is holomorphic, with holomorphic inverse. (Thus our conformal maps must always preserve orientation.) According to Poincaré and Koebe, there are only three simply connected Riemann surfaces, up to isomorphism.

1.1. Uniformization Theorem. *Any simply connected Riemann surface is conformally isomorphic either*

(a) *to the plane* $\mathbf{C}$ *consisting of all complex numbers* $z = x + iy$,

(b) *to the open unit disk* $D \subset \mathbf{C}$ *consisting of all* z *with* $|z|^2 = x^2 + y^2 < 1$, *or*

(c) *to the Riemann sphere* $\hat{\mathbf{C}}$ *consisting of* $\mathbf{C}$ *together with a point at infinity.*

The proof, which is quite difficult, may be found in Springer, or Farkas & Kra, or Ahlfors [A2], or in Beardon. □

For the rest of this section, we will discuss these three surfaces in more detail. We begin with a study of the unit disk D .

1.2. Schwarz Lemma. *If* $f : D \to D$ *is a holomorphic map with* $f(0) = 0$, *then the derivative at the origin satisfies* $|f'(0)| \le 1$. *If equality holds, then* f *is a rotation about the origin, that is* $f(z) = \lambda z$ *for some constant* $\lambda = f'(0)$ *on the unit circle. In particular, it follows that* f *is a conformal automorphism of* D . *On the other hand, if* $|f'(0)| < 1$, *then* $|f(z)| < |z|$ *for all* $z \ne 0$, *and* f *is not a conformal automorphism.*

Proof. We use the *maximum modulus principle*, which asserts that a non-constant holomorphic function cannot attain its maximum absolute value at any interior point of its region of definition. First note that the quotient function $g(z) = f(z)/z$ is well defined and holomorphic throughout the disk D . Since $|g(z)| < 1/r$ when $|z| = r < 1$, it follows by the maximum modulus principle that $|g(z)| < 1/r$ for all z in the disk $|z| \le r$. Taking the limit as $r \to 1$, we see that $|g(z)| \le 1$ for all $z \in D$. Again by the maximum modulus principle, we see that the case $|g(z)| = 1$, with z in the open disk, can occur only if the function $g(z)$ is constant. If we exclude this case $f(z)/z = c$, then it follows that $|g(z)| = |f(z)/z| < 1$ for all $z \ne 0$, and similarly that $|g(0)| = |f'(0)| < 1$. Evidently the composition of two such maps must also satisfy $|f_1(f_2(z))| < |z|$, and hence cannot be the identity map of D . □

1.3. Remarks. The Schwarz Lemma was first proved in this generality by Carathéodory. If we map the disk D_r of radius r into the disk D_s of radius s , with $f(0) = 0$, then a similar argument shows that $|f'(0)| \le s/r$. Even if we drop the condition that $f(0) = 0$, we certainly get the inequality

$$|f'(0)| \le 2s/r \qquad \text{whenever} \qquad f(D_r) \subset D_s$$

since the difference $f(z) - f(0)$ takes values in D_{2s}. (In fact the extra factor of 2 is unnecessary. Compare Problem 1-2.) One easy corollary is the *Theorem of Liouville*, which says that a bounded function which is defined and holomorphic everywhere on $\mathbf{C}$ must be constant. Another closely related statement is the following.

1.4. Theorem of Weierstrass. *If a sequence of holomorphic functions converges uniformly, then their derivatives also converge uniformly, and the limit function is itself holomorphic.*

In fact the convergence of first derivatives follows easily from the discussion above. For the proof of the full theorem, see for example [A1].

It follows from this discussion that our three model surfaces really are distinct. For there are natural inclusion maps $D \to \mathbf{C} \to \hat{\mathbf{C}}$; yet it follows from the maximum modulus principle and Liouville's Theorem that every holomorphic map $\hat{\mathbf{C}} \to \mathbf{C}$ or $\mathbf{C} \to D$ must be constant.

It is often more convenient to work with the *upper half-plane* H, consisting of all complex numbers $w = u + iv$ with $v > 0$.

1.5. Lemma. *The half-plane H is conformally isomorphic to the disk D under the holomorphic mapping $z = (i-w)/(i+w)$, with inverse $w = i(1-z)/(1+z)$.*

Proof. We have $|z|^2 < 1$ if and only if $|i - w|^2 = u^2 + (1 - 2v + v^2)$ is less than $|i + w|^2 = u^2 + (1 + 2v + v^2)$, or in other words if and only if $0 < v$. $\square$

1.6. Lemma. *Given any point z_0 of D, there exists a conformal automorphism f of D which maps z_0 to the origin. Furthermore, f is uniquely determined up to composition with a rotation which fixes the origin.*

Proof. Given any two points w_1 and w_2 of the upper half-plane H, it is easy to check that there exists a automorphism of the form $w \mapsto a + bw$ which carries w_1 to w_2. Here the coefficients a and b are to be real, with $b > 0$. Since H is conformally isomorphic to D, it follows that there exists a conformal automorphism of D carrying any given point to zero. As noted above, the Schwarz Lemma implies that any automorphism of D which fixes the origin is a rotation. $\square$

1.7. Theorem. *The group $G(H)$ consisting of all conformal automorphisms of the upper half-plane H can be identified with the group of all fractional linear transformations $w \mapsto (aw + b)/(cw + d)$ with real coefficients and with determinant $ad - bc > 0$.*

If we normalize so that $ad - bc = 1$, then these coefficients are well defined up to a simultaneous change of sign. *Thus $G(H)$ is isomorphic to the group* $\mathrm{PSL}(2, \mathbf{R})$, *consisting of all 2×2 real matrices with determinant $+1$ modulo the subgroup $\{\pm I\}$.* To prove 1.7, it is only necessary to note that this group acts transitively on H, and that it contains the group of "rotations"

$$g(w) = (w \cos\theta + \sin\theta)/(-w \sin\theta + \cos\theta)\,, \tag{1}$$

which fix the point $g(i) = i$ with $g'(i) = e^{2i\theta}$. By 1.6, there can be no further automorphisms. (Compare Problem 1-2.) □

Next we introduce the *Poincaré metric* on the half-plane H .

1.8. Theorem. *There exists one and, up to multiplication by a constant, only one Riemannian metric on the half-plane H which is invariant under every conformal automorphism of H .*

It follows immediately that the same statement is true for the disk D , or for any other Riemann surface which is conformally isomorphic to H .

Proof of 1.8. A Riemannian metric $ds^2 = g_{11}dx^2 + 2g_{12}dxdy + g_{22}dy^2$ is said to be *conformal* if $g_{11} = g_{22}$ and $g_{12} = 0$, so that the matrix $[g_{ij}]$, evaluated at any point z , is some multiple of the identity matrix. Such a metric can be written as $ds^2 = \gamma(x+iy)^2(dx^2+dy^2)$, or briefly as $ds = \gamma(z)|dz|$, where the function $\gamma(z)$ is smooth and strictly positive. By definition, such a metric is *invariant* under a conformal automorphism $z' = f(z)$ if and only if it satisfies the identity $\gamma(z')|dz'| = \gamma(z)|dz|$, or in other words.

$$\gamma(f(z)) = \gamma(z)/|f'(z)| . \tag{2}$$

Equivalently, we may say that f is an *isometry* with respect to the metric.

As an example, suppose that a conformal metric $\gamma(w)|dw|$ on the upper half-plane is invariant under every linear automorphism $f(w) = a + bw$. Then we must have $\gamma(a+bi) = \gamma(i)/b$. If we normalize by setting $\gamma(i) = 1$, then we are led to the formula $\gamma(u+iv) = 1/v$, or in other words

$$ds = |dw|/v \qquad \text{for} \quad w = u + iv \in H . \tag{3}$$

In fact, the metric defined by this formula is invariant under every conformal automorphism g of H . For, if we select some arbitrary point $w_1 \in H$ and set $g(w_1) = w_2$, then g can be expressed as the composition of a linear automorphism of the form $g_1(w) = a+bw$ which maps w_1 to w_2 and an automorphism g_2 which fixes w_2 . We have constructed the metric (3) so that g_1 is an isometry, and it follows from 1.2 and 1.5 that $|g_2'(w_2)| = 1$, so that g_2 is an isometry at w_2 . Thus our metric is invariant at an arbitrarily chosen point under an arbitrary automorphism.

To complete the proof of 1.8, we must show that a metric which is invariant under all automorphisms of H is necessarily conformal. For this purpose, it is convenient to pass to the equivalent problem on D . In fact a brief computation shows that any metric on D which is preserved by all rotations about the origin must be conformal at the origin. Further details will be left to the reader. □

Definition. This metric $ds = |dw|/v$ is called the *Poincaré metric* on the upper half-plane H . It can be shown that this is the unique conformal metric on H which is complete, with constant Gaussian curvature equal to -1 . (Compare Problems 1-9 and 2-2.) The corresponding expression on D is

$$ds = 2|dz|/(1-|z|^2) \qquad \text{for} \quad z = x + iy \in D , \tag{4}$$

as can be verified using 1.5 and (2).

Caution: Some authors call $|dz|/(1-|z|^2)$ the Poincaré metric on D, and correspondingly call $\frac{1}{2}|dw|/v$ the Poincaré metric on H. These modified metrics have constant Gaussian curvature equal to -4.

Thus there is a preferred Riemannian metric ds on D or on H. More generally, if S is any Riemann surface which is conformally isomorphic to D, then there is a corresponding Poincaré metric ds on S. Hence we can define the *Poincaré distance* $\rho(z_1, z_2) = \rho_S(z_1, z_2)$ between two points of S as the minimum, over all paths P joining z_1 to z_2, of the integral $\int_P ds$.

1.9. Lemma. *The disk D with this Poincaré metric is complete. That is:*

(a) every Cauchy sequence with respect to the metric ρ_D converges,

(b) every closed neighborhood $N_r(z_0, \rho_D) = \{z \in D : \rho_D(z, z_0) \le r\}$ is a compact subset of D, and

(c) any path leading from z_0 to a point of the boundary $\partial D = \bar{D} \smallsetminus D \subset \hat{\mathbf{C}}$ has infinite Poincaré length.

Furthermore, any two points of D are joined by a unique minimal geodesic.

(Compare Willmore.) Equivalently, we have exactly the same statements for the half-plane H.

Proof. Given any two points of D we can first choose a conformal automorphism which moves the first point to the origin and the second to some point ξ on the positive real axis. For any path P between 0 and ξ within D we have

$$\int_P ds = \int_P \frac{2|dz|}{1-|z|^2} \ge \int_P \frac{2|dx|}{1-x^2} \ge \int_{[0,\xi]} \frac{2dx}{1-x^2} = \log\frac{1+\xi}{1-\xi},$$

with equality if and only if P is the straight line segment $[0,\xi]$. For any $z \in D$, it follows easily that the Poincaré distance $\rho = \rho_D(0, z)$ is equal to $\log\big((1+|z|)/(1-|z|)\big)$, and that the straight line segment from 0 to z is the unique minimal Poincaré geodesic. Clearly $\rho \to \infty$ as $|z| \to 1$, which proves (c). The proof of (b), and hence of (a) is now straightforward. □

The Poincaré metric has the marvelous property of never increasing under holomorphic maps.

1.10. Theorem of Pick. *If $f : S \to T$ is a holomorphic map between Riemann surfaces, both of which are conformally isomorphic to D, then*

$$\rho_T(f(z_1), f(z_2)) \le \rho_S(z_1, z_2).$$

Furthermore, if equality holds for some $z_1 \ne z_2$ in S, then f must be a conformal isomorphism from S onto T.

Proof. Join z_1 to z_2 by a geodesic of length equal to the distance $\rho_S(z_1, z_2)$. Let ds be the element of Poincaré length along this geodesic, and let ds' be the element of length along the image curve in T. It follows from the Schwarz Lemma and the definition of the Poincaré metric that $|ds'/ds| \le 1$, with equality if and only if f is a conformal isomorphism; and the conclusion follows. □

Note that this distance depends sharply on the choice of S. As an example, suppose that $U \subset D$ is a simply connected open subset with $U \neq D$. Then U is conformally isomorphic to D by the Riemann Mapping Theorem. Applying 1.10 to the inclusion $U \to D$, we see that

$$\rho_D(z_1\,,\,z_2) \;<\; \rho_U(z_1\,,\,z_2)$$

for any two distinct points z_1 and z_2 in U.

1.11. Remark. In the special case of a holomorphic map from S to itself, if the sharper inequality

$$\rho(f(z_1), f(z_2)) \le c\rho(z_1\,,\,z_2)$$

is satisfied for all z_1 and z_2, where $0 < c < 1$ is constant, then f necessarily has a unique fixed point. However, the example $z \mapsto z^2$ on the unit disk shows that a map with a unique fixed point need not satisfy this sharper inequality; and the example $w \mapsto w + i$ on the upper half-plane shows that map which simply reduces Poincaré distance need not have any fixed point. (See also Problem 1-1.)

Next let us consider the *Riemann sphere* $\hat{\mathbf{C}}$, that is the compact Riemann surface consisting of the complex numbers $\mathbf{C}$ together with a single point at infinity. The conformal structure on this complex manifold can be specified by using the usual coordinate z as uniformizing parameter in the finite plane $\mathbf{C}$, and using $\zeta = 1/z$ as uniformizing parameter in $\hat{\mathbf{C}} \smallsetminus \{0\}$.

When studying $\hat{\mathbf{C}}$, we need some *spherical metric* σ which is adapted to the topology, so that the point at infinity has finite distance from other points of $\hat{\mathbf{C}}$. The precise choice of such a metric does not matter for our purposes. However, one good choice would be the distance function $\sigma(z_1\,,\,z_2)$ which is associated with the Riemannian metric

$$ds = 2|dz|/(1 + |z|^2)\,. \tag{6}$$

This metric is smooth and well behaved, even in a neighborhood of the point at infinity, and has constant Gaussian curvature $+1$. It corresponds to the standard metric on the unit sphere in $\mathbf{R}^3$ under stereographic projection. Note that the map $z \mapsto 1/z$ is an isometry for this metric. (However, it is certainly not true that every conformal self-map of $\hat{\mathbf{C}}$ is an isometry.)

The group $G(\hat{\mathbf{C}})$ consisting of all conformal automorphisms of $\hat{\mathbf{C}}$ can be described as follows. By definition, an automorphism $g \in G(\hat{\mathbf{C}})$ is called an *involution* if the composition $g \circ g$ is the identity map of $\hat{\mathbf{C}}$.

1.12. Theorem. *Every conformal automorphism g of $\hat{\mathbf{C}}$ can be expressed as a fractional linear transformation or Möbius transformation*

$$g(z) \;=\; (az + b)/(cz + d)\,,$$

where the coefficients are complex numbers with determinant $ad - bc \neq 0$. Every non-identity automorphism of $\hat{\mathbf{C}}$ either has two distinct fixed points or one double fixed point in $\hat{\mathbf{C}}$. In general, two non-identity elements of $G(\hat{\mathbf{C}})$ commute if *and only if they have precisely the same fixed points: the only exceptions to this statement are provided by pairs of commuting involutions.*

Here by a "double" fixed point we mean one at which the derivative $g'(z)$ is equal to $+1$. If we normalize so that $ad - bc = 1$, then the coefficients are well defined up to a simultaneous change of sign. *Thus the group* $G(\hat{\mathbf{C}})$ *of conformal automorphisms can be identified with the group* $\mathrm{PSL}(2,\mathbf{C})$ *consisting of all complex matrices with determinant* $+1$ *modulo the subgroup* $\{\pm I\}$.

1.13. Remark. The group $G(\mathbf{C})$ of all conformal automorphisms of the complex plane can be identified with the subgroup of $G(\hat{\mathbf{C}})$ consisting of automorphisms g which fix the point ∞ , since every conformal automorphism of $\mathbf{C}$ extends uniquely to a conformal automorphism of $\hat{\mathbf{C}}$. (Compare Ahlfors [A1, p.124] .) It follows easily that $G(\mathbf{C})$ consists of all affine transformations

$$g(z) \;=\; \lambda z + c$$

with complex coefficients $\lambda \neq 0$ and c .

Proof of 1.12. Clearly $G(\hat{\mathbf{C}})$ contains this group of fractional linear transformations as a subgroup. After composing the given $g \in G(\hat{\mathbf{C}})$ with a suitable element of this subgroup, we may assume that $g(0) = 0$ and $g(\infty) = \infty$. But then the quotient $g(z)/z$ is a bounded holomorphic function, hence constant by Liouville's Theorem (§1.3). Thus g is linear, and hence itself is an element of $\mathrm{PSL}(2,\mathbf{C})$.

The fixed points of a fractional linear transformation can be determined by solving a quadratic equation, so it is easy to check that there must be at least one and at most two distinct solutions in the extended plane $\hat{\mathbf{C}}$. In particular, if an automorphism of $\hat{\mathbf{C}}$ fixes three distinct points, then it must be the identity map.

In general, if two automorphisms g and h commute, we can say that g maps every fixed point of h to a fixed point of h , and that h maps every fixed point of g to a fixed point of g . However, this leaves open the possibility that g interchanges the two fixed of h and that h interchanges the two fixed points of g . If this is indeed the case, then both $g \circ g$ and $h \circ h$ have at least four fixed points, hence both must equal the identity map. (An example of this phenomenon is provided by the two commuting involutions $g(z) = -z$ with fixed points $\{0, \infty\}$, and $h(z) = 1/z$ with fixed points $\{\pm 1\}$.) If we exclude this possibility, then elements which commute must have exactly the same fixed points.

Conversely, let us consider the subgroup consisting of all elements of $G(\hat{\mathbf{C}})$ which fix two specified points $z_0 \neq z_1$. It is convenient to conjugate by an automorphism which carries z_0 to zero and z_1 to infinity. The argument above shows that an automorphism g fixes both zero and infinity if and only if it has the form $g(z) = \lambda z$ for some $\lambda \neq 0$. Thus the subgroup consisting of all such elements is isomorphic to the multiplicative group $\mathbf{C} \smallsetminus \{0\}$, and hence is commutative as required.

Finally, consider automorphisms g which fix only the point at infinity. A similar argument shows that $g(z)$ must be equal to $z + c$ for some constant $c \neq 0$. (Compare 1.13.) These automorphisms, together with the identity map, form a subgroup of $G(\hat{\mathbf{C}})$

which is isomorphic to the additive group of $\mathbf{C}$; which again is commutative as required. Further details will be left to the reader. $\square$

The corresponding discussion for the group $G(H)$ (or $G(D)$) will be important in §3. To fix our ideas, let us concentrate on the case of the half-plane. It follows from 1.7 that every automorphism of H extends uniquely to an automorphism of $\hat{\mathbf{C}}$. Hence 1.12 applies immediately. However, we must consider not only fixed points inside of H but also fixed points which lie on the boundary $\partial H = \mathbf{R} \cup \infty$. A priori, we should also consider fixed points which lie in the lower half-plane, completely outside of the closure $\bar{H}$. However, it is easy to check that w is a fixed point of a fractional linear transformation with real coefficients if and only if the complex conjugate $\bar{w}$ is also a fixed point. Thus each fixed point in the lower half-plane, outside of $\bar{H}$, is paired with a fixed point which is inside the upper half-plane H .

1.14. Definition. The non-identity automorphisms of H fall into three classes, as follows:

An automorphism $g \in G(H)$ is said to be *elliptic* if it has a fixed point $w_0 \in H$. We may also describe g as a *rotation* around w_0 . (Compare (1) above.)

The automorphism $g \in G(H)$ is *hyperbolic* if it has two distinct fixed points on the boundary ∂H . As an example, g fixes the two points 0 and ∞ if and only if it has the form $g(w) = \lambda w$ with $\lambda > 0$.

The automorphism $g \in G(H)$ is *parabolic* if it has just one double fixed point, which necessarily belongs to the boundary $\partial H = \mathbf{R} \cup \infty$. For example, if this double fixed point is the point at infinity, then g is necessarily a translation: $g(w) = w + c$ for some constant $c \neq 0$ in $\mathbf{R}$.

1.15. Lemma. *Two non-identity elements of $G(H)$ commute if and only if they have exactly the same fixed points in $\bar{H} = H \cup \mathbf{R} \cup \infty$. The set of all group elements with some specified fixed point set, together with the identity element, forms a commutative subgroup, which is isomorphic to a circle in the elliptic case and to a real line in the parabolic or hyperbolic cases.*

The proof is easily supplied. $\square$

Again, we could equally well work with D in place of H . One defect of this exposition is that it requires going outside of the half-plane H in order to distinguish between parabolic and hyperbolic automorphisms. For a more intrinsic description of the difference between these cases, see Problem 1-5.

We conclude this section with a number of problems for the reader.

Problem 1-1. If a holomorphic map $f : D \to D$ fixes the origin, and is not a rotation, prove that the successive images $f^{\circ n}(z)$ converge to zero for all z in the open disk D , and prove that this convergence is uniform on compact subsets of D . (Here $f^{\circ n}$ stands for the n-fold iterate $f \circ \cdots \circ f$. The example $f(z) = z^2$ shows that convergence need not be uniform on all of D .)

Problem 1-2. Show that the group $G(D)$ of conformal automorphisms of the unit disk D consists of all maps

$$g(z) \;=\; e^{i\theta}(z-a)/(1-\bar{a}z)$$

with $|e^{i\theta}|=1$, where $a\in D$ is the unique point which maps to zero. (Compare 1.7.) Check that $|g'(0)|\le 1$, and conclude using 1.2 that any holomorphic map $f:D\to D$ must satisfy $|f'(0)|\le 1$.

Problem 1-3. Show that the action of the group $G(\hat{\mathbf{C}})$ on $\hat{\mathbf{C}}$ is *simply 3-transitive*. That is, there is one and only one automorphism which carries three distinct specified points of $\hat{\mathbf{C}}$ into three other specified points. Similarly, show that the action of of $G(\mathbf{C})$ on $\mathbf{C}$ is simply 2-transitive. Show that the action of $G(D)$ on the boundary circle ∂D carries three specified points into three other specified points if and only if they have the same cyclic order; and show that the action of $G(D)$ carries two points of D into two other specified points if and only if they have the same Poincaré distance.

Problem 1-4. By an *anti-holomorphic* automorphism of $\hat{\mathbf{C}}$ we mean an orientation reversing self-homeomorphism of the form $z\mapsto\eta(\bar{z})$, where η is holomorphic. If L is a straight line or circle in $\hat{\mathbf{C}}$, show that there is one and only one anti-holomorphic involution of $\hat{\mathbf{C}}$ having L as fixed point set, and show that no other non-vacuous fixed point sets can occur. Show that the automorphism group $G(\hat{\mathbf{C}})$ acts transitively on the set of straight lines and circles in $\hat{\mathbf{C}}$. Similarly, show that any anti-holomorphic involution of D is the reflection in some Poincaré geodesic; and show that $G(D)$ acts transitively on the set of such geodesics.

Problem 1-5. Let g be an automorphism of D with $g\circ g$ not the identity map. Show that g is hyperbolic if and only if there exists an automorphism h which satisfies

$$h\circ g\circ h^{-1} \;=\; g^{-1}\,,$$

or if and only if there exists some necessarily unique geodesic L with respect to the Poincaré metric which is mapped onto itself by g, or if and only if g commutes with some anti-holomorphic involution. (The possible choices for h are just the $180°$ rotations about the points of L.)

Problem 1-6. Define the *infinite band* $B\subset\mathbf{C}$ of width π to be the set of all $z=x+iy$ with $|y|<\pi/2$. Show that the exponential map carries B isomorphically onto the right half-plane $\{u+iv : u>0\}$. Show that the Poincaré metric on B takes the form

$$ds = |dz|/cos\,y\,. \tag{7}$$

Show that the real axis is a geodesic whose Poincaré arclength coincides with its usual Euclidean arclength, and show that each real translation $z\mapsto z+c$ is a hyperbolic automorphism of B having the real axis as its unique invariant geodesic.

Problem 1-7. Given four distinct points z_j in $\hat{\mathbf{C}}$ show that the *cross ratio*

$$\chi(z_1\,,\,z_2\,;\,z_3\,,\,z_4) \quad = \quad \frac{(z_1-z_3)(z_2-z_4)}{(z_1-z_4)(z_2-z_3)} \quad \in \quad \mathbf{C}\smallsetminus\{0,1\}$$

is invariant under fractional linear transformations, and show that χ is real if and only if the four points lie on a straight line or circle.

Problem 1-8. Show that each Poincaré neighborhood $N_r(w_0\,,\,\rho_H)$ in the upper half-plane is bounded by a Euclidean circle, but that w_0 is not its Euclidean center. Show that each Poincaré geodesic in the upper half-plane is a straight line or semi-circle which meets the real axis orthogonally. If the geodesic through w_1 and w_2 meets $\partial H = \mathbf{R}\cup\infty$ at the points α and β , show that the Poincaré distance between w_1 and w_2 is equal to the logarithm of the cross ratio $\chi(w_1\,,\,w_2\,;\,\alpha\,,\,\beta)$. Prove corresponding statements for the unit disk.

Problem 1-9. The Gaussian curvature of a conformal metric $ds = \gamma(w)|dw|$ with $w = u + iv$ is given by the formula

$$K \quad = \quad \frac{\gamma_u^2 + \gamma_v^2 - \gamma(\gamma_{uu} + \gamma_{vv})}{\gamma^4}$$

where the subscripts stand for partial derivatives. (Compare Willmore, p. 79.) Check that the Poincaré metrics (3), (4) and (7) have curvature $K \equiv -1$, and more generally that the metric $ds = c|dw|/v$ has curvature $K \equiv -1/c^2$. Check that the spherical metric (6) has curvature $K \equiv +1$.

Problem 1-10. Classify conjugacy classes in the groups $G(H) \cong \mathrm{PSL}(2,\mathbf{R})$ as follows. Show that every automorphism of H without fixed point is conjugate to a unique transformation of the form $w \mapsto w+1$ or $w \mapsto w-1$ or $w \mapsto \lambda w$ with $\lambda > 1$; and show that the conjugacy class of an automorphism g with fixed point $w_0 \in H$ is uniquely determined by the derivative $\lambda = g'(w_0)$, where $|\lambda| = 1$. Show also that each non-identity element of $\mathrm{PSL}_2(\mathbf{R})$ belongs to one and only one "one-parameter subgroup", and that each one-parameter subgroup is conjugate to either

$$t \;\mapsto\; \begin{bmatrix} 1 & t \\ 0 & 1 \end{bmatrix} \quad \text{or} \quad \begin{bmatrix} e^t & 0 \\ 0 & e^{-t} \end{bmatrix} \quad \text{or} \quad \begin{bmatrix} \cos t & \sin t \\ -\sin t & \cos t \end{bmatrix}$$

according as its elements are parabolic or hyperbolic or elliptic. Here t ranges over the additive group of real numbers.

Problem 1-11. For a non-identity automorphism $g \in G(\hat{\mathbf{C}})$, show that the derivatives $g'(z)$ at the two fixed points are reciprocals, say λ and λ^{-1} , and show that the sum $\lambda + \lambda^{-1}$ is a complete conjugacy class invariant. (Here $\lambda = 1$ if and only if the two fixed points coincide. In the special case of a fixed point at infinity, one must evaluate the derivative using the local coordinate $\zeta = 1/z$.)

Problem 1-12. Show that the conjugacy class of a non-identity automorphism $g(z) = \lambda z + c$ in the group $G(\mathbf{C})$ is uniquely determined by its image under the homomorphism $g \mapsto \lambda \in \mathbf{C}\smallsetminus\{0\}$.

§2. The Universal Covering, Montel's Theorem.

If S is a completely arbitrary Riemann surface, then the universal covering $\tilde{S}$ is a well defined simply connected Riemann surface, with a canonical projection map $p:\tilde{S}\to S$. (Compare Munkres, and also Appendix E.) According to the Uniformization Theorem, this universal covering $\tilde{S}$ must be conformally isomorphic to one of the three model surfaces (§1.1). Thus we have the following.

2.1. Lemma. *Every Riemann surface S is conformally isomorphic to a quotient of the form $\tilde{S}/\Gamma$, where $\tilde{S}$ is a simply connected surface which is isomorphic to either D, $\mathbf{C}$, or $\hat{\mathbf{C}}$. Here Γ is a discrete group of conformal automorphisms of $\tilde{S}$, such that every non-identity element of Γ acts without fixed points on $\tilde{S}$.*

This discrete group $\Gamma\subset G(\tilde{S})$ can be identified with the fundamental group $\pi_1(S)$. The elements of Γ are called *deck transformations*. They can be characterized as maps $\gamma:\tilde{S}\to\tilde{S}$ which satisfy the identity $\gamma=p\circ\gamma$, that is maps for which the diagram

$$\begin{array}{ccc}\tilde{S} & \xrightarrow{\gamma} & \tilde{S}\\ \downarrow p & & \downarrow p\\ S & \xrightarrow{=} & S\end{array}$$

is commutative. Conversely, if we are given a group Γ of conformal automorphisms of a simply connected surface $\tilde{S}$ which is discrete as a subgroup of $G(\tilde{S})$, and such that every non-identity element of Γ acts without fixed points, then it is not difficult to check that Γ is the group of deck transformations for a universal covering map $\tilde{S}\to\tilde{S}/\Gamma$. (Compare Problem 2-1.)

We can analyze the three possibilities for $\tilde{S}$ as follows.

Spherical Case. According to 1.12, every conformal automorphism of the Riemann sphere $\hat{\mathbf{C}}$ has at least one fixed point. Therefore, if $S\cong\hat{\mathbf{C}}/\Gamma$ is a Riemann surface with universal covering $\tilde{S}\cong\hat{\mathbf{C}}$, then the group $\Gamma\subset G(\hat{\mathbf{C}})$ must be trivial, hence S itself must be isomorphic to $\hat{\mathbf{C}}$.

Euclidean Case. By 1.13, the group $G(\mathbf{C})$ of conformal automorphisms of the complex plane consists of all affine transformations $z\mapsto\lambda z+c$ with $\lambda\neq 0$. Every such transformation with $\lambda\neq 1$ has a fixed point. Hence, if $S\cong\mathbf{C}/\Gamma$ is a surface with universal covering $\tilde{S}\cong\mathbf{C}$, then Γ must be a discrete group of translations $z\mapsto z+c$ of the complex plane $\mathbf{C}$. There are three subcases:

If Γ is trivial, then S itself is isomorphic to $\mathbf{C}$.

If Γ has just one generator, then S is isomorphic to the cylinder $\mathbf{C}/\mathbf{Z}$, which in turn is isomorphic under the exponential map to the *punctured plane* $\mathbf{C}\smallsetminus\{0\}$.

If Γ has two generators, then it can be described as a two-dimensional *lattice* $\Lambda\subset\mathbf{C}$, that is an additive group generated by two complex numbers which are linearly independent over $\mathbf{R}$. The quotient $\mathbf{T}=\mathbf{C}/\Lambda$ is called a *torus*.

In all three of these cases, note that our surface inherits a locally Euclidean geometry from the Euclidean metric $|dz|$ on its universal covering surface. For example the punctured plane $\mathbf{C}\smallsetminus\{0\}$, consisting of points $\exp(z)=w$, has a complete locally Euclidean metric $|dz| = |dw/w|$. It is easy to check that such a metric is unique up to a multiplicative constant. (Compare Problem 2-2.) We will refer to these Riemann surfaces as [complete locally] *Euclidean surfaces*. The term "*parabolic surface*" is also commonly used in the literature.

Hyperbolic Case. In all other cases, the universal covering $\tilde{S}$ must be conformally isomorphic to the unit disk. Such Riemann surfaces are said to be *Hyperbolic*. *As an example, any Riemann surface with non-abelian fundamental group, and in particular any surface of higher genus, is necessarily Hyperbolic.*

Remark. Here the word "Hyperbolic" is a reference to Hyperbolic Geometry, that is the geometry of Lobachevsky. Unfortunately the term "hyperbolic" has at least three different well established meanings in conformal dynamics. (Compare §1.14 and §14.) In a crude attempt to avoid confusion, I will alway capitalize the word when it is used with this geometric meaning.

Every Hyperbolic surface S possesses a unique *Poincaré metric*, which is complete, with Gaussian curvature identically equal to -1. To construct this metric, we note that the Poincaré metric on $\tilde{S}$ is invariant under the action of Γ. Hence there is one and only one metric on S so that the projection $\tilde{S} \cong D \to S$ is a local isometry. Hence, just as in §1.9, there is an associated *Poincaré distance function* $\rho(z, z') = \rho_S(z, z')$ which is equal to the length of a shortest geodesic from z to z'. As in 1.10 we have:

2.2. Lemma. *If $f : S \to T$ is a holomorphic map between Hyperbolic surfaces, then*

$$\rho_T(f(z),\, f(z')) \;\le\; \rho_S(z, z')\,.$$

Furthermore, if equality holds for some $z \neq z'$, then it follows that f is a local isometry. That is, f preserves the infinitesimal distance element ds, and hence preserves the distances between nearby points.

Proof. This follows, just as in 1.10, since a minimal geodesic joining z to z' must be mapped isometrically. □

Caution: It no longer follows that f must be a conformal isomorphism from S to T. However, if f is a local isometry, then we can at least assert that f induces an isometry $\tilde{S} \xrightarrow{\cong} \tilde{T}$ between the universal covering surfaces, and hence that f is a covering map from S onto T.

Here is an example. The map $f(z) = z^2$ on the *punctured disk* $D \smallsetminus \{0\}$ is certainly not an automorphism, since it is two-to-one. However, the universal covering of $D \smallsetminus \{0\}$ can be identified with the left half-plane, mapped to $D \smallsetminus \{0\}$ by the exponential map, and f lifts to the automorphism $w \mapsto 2w$ of this left half-plane. Therefore, f is a covering map, and preserves the Poincaré metric locally.

Note that the punctured disk has abelian fundamental group $\pi_1(D \smallsetminus \{0\}) \cong \mathbf{Z}$. Here is another such example. For any $r > 1$, the *annulus*

$$\mathcal{A}_r \;=\; \{\, z \;:\; 1 < |z| < r \,\}$$

is a Hyperbolic surface, since it admits a holomorphic map to the unit disk. The fundamental group $\pi_1(\mathcal{A}_r)$ is evidently also free cyclic. In fact annuli and the punctured disk are the only Hyperbolic surfaces with abelian fundamental group, other than the disk itself. (Compare Problem 2-3.)

Maximal Hyperbolic Example. If a_1, a_2, a_3 are three distinct points of $\hat{\mathbf{C}}$, then the complement $\Sigma_3 = \hat{\mathbf{C}} \smallsetminus \{a_1, a_2, a_3\} \cong \mathbf{C} \smallsetminus \{0, 1\}$ is called a *thrice punctured sphere.* This is evidently a Hyperbolic surface, for example since its fundamental group is not abelian. One immediate corollary of this observation is the following.

2.3. Picard's Theorem. *Any holomorphic map from* $\mathbf{C}$ *to* $\hat{\mathbf{C}}$ *which omits three different values must be constant. More generally, if the Riemann surface* S *admits some non-constant holomorphic map to the thrice punctured sphere* Σ_3, *then* S *must be Hyperbolic.*

For this map $f: S \to \Sigma_3$ can be lifted to a holomorphic map from the universal covering $\tilde{S}$ to the universal covering $\tilde{\Sigma}_3 \cong D$. By Liouville's Theorem (§1.3) it follows that $\tilde{S} \cong D$. □

Let $U \subset \hat{\mathbf{C}}$ be any connected open set which omits at least three points, and hence is Hyperbolic. It is often useful to compare the Poincaré distance between two points of U with the spherical distance between the same two points within $\hat{\mathbf{C}}$. (See §1(6).) Here is a crude estimate. Again let $N_r(z, \rho_U)$ be the closed neighborhood of some fixed radius r with respect to the Poincaré metric about the point z of U.

2.4. Lemma. *As* z *converges towards the boundary* ∂U *in the spherical metric, the spherical diameter of the neighborhood* $N_r(z, \rho_U)$ *tends to zero.*

(For a sharper statement in the simply connected case, see A.8 in the Appendix.)

Proof of 2.4. First consider the special case $U = \Sigma_3 = \hat{\mathbf{C}} \smallsetminus \{a_1, a_2, a_3\}$. Choose some fixed base point z_0 in $U = \Sigma_3$. For each fixed $r > 0$, the neighborhood $N_r(z_0, \rho_U)$ is the image under projection of a corresponding neighborhood in the universal covering surface, and hence is compact and connected. (Compare 1.9.) Now, as z tends to one of the three boundary points a_1 of U, it must eventually leave any compact subset of U, hence the distance $\rho_U(z, z_0)$ must tend to infinity. For fixed r, it follows that the neighborhood $N_r(z, \rho_U)$ will eventually be disjoint from any given compact subset of Σ_3. In fact, since the set $N_r(z, \rho_U)$ is connected, this entire set must tend to just one boundary point a_1 with respect to the spherical metric.

Now consider an arbitrary Hyperbolic open set $U \subset \hat{\mathbf{C}}$. Given a sequence of points of U tending to the boundary, we can choose a subsequence $\{z_j\}$ which converges to a single boundary point a_1. Choose two other boundary points a_2 and a_3, and consider the inclusion map from U to $\Sigma_3 = \hat{\mathbf{C}} \smallsetminus \{a_1, a_2, a_3\}$. Applying the Pick inequality 2.2, we have $N_r(z_j, \rho_U) \subset N_r(z_j, \rho_{\Sigma_3})$, and it follows from the discussion above that this entire neighborhood converges to the boundary point a_1 as $j \to \infty$. □

Using the Poincaré metric, we will develop another important tool. A sequence of maps $f_n : S \to \hat{\mathbf{C}}$ on a Riemann surface S is said to *converge locally uniformly* (or *uniformly on compact sets*) to the limit $g : S \to \hat{\mathbf{C}}$ if for every compact subset $K \subset S$ the sequence $\{f_n|K\}$ of maps f_n restricted to K converges uniformly to $g|K$. Here it is to be understood that we use the spherical metric $\sigma(z_1, z_2)$ on the target space $\hat{\mathbf{C}}$.

Definition. Let S and S' be Riemann surfaces, with S' compact. A collection $\mathcal{F}$ of holomorphic maps $f_\alpha : S \to S'$ is *normal* if every infinite sequence of maps from $\mathcal{F}$ contains a subsequence which converges locally uniformly to a limit.

(For the case of a non-compact target space S', see 2.6 below.) Note that the limit function g must itself be holomorphic, by the Theorem of Weierstrass. However, this limit g need not belong to the given family. Roughly speaking, a family of maps is normal if and only if its closure, in the space of all holomorphic maps from S to S', is a compact set. (Ahlfors [A1, p.213].)

2.5. Montel's Theorem. *Let S be any Riemann surface, which we may as well suppose to be Hyperbolic. If a collection $\mathcal{F}$ of holomorphic maps from S to the Riemann sphere $\hat{\mathbf{C}}$ takes values in some Hyperbolic open subset $U \subset \hat{\mathbf{C}}$, or equivalently if there are three distinct points of $\hat{\mathbf{C}}$ which never occur as values, then this collection $\mathcal{F}$ is normal.*

More explicitly, any sequence of holomorphic maps $f_n : S \to U$ contains a subsequence which converges, uniformly on compact subsets of S, to some holomorphic map $g : S \to \bar{U}$. Here it is essential that g be allowed to take values in the closure $\bar{U}$. However, the proof will show that the image $g(S)$ is either contained in U, or is a single point belonging to the boundary of U.

Proof of 2.5. First note by 2.3 that the surface S must be Hyperbolic, unless all of our maps are constant. Hence S also has a Poincaré metric. Choose a countable dense subset of points $z_j \in S$, $j \geq 1$. (It follows easily from §1.1 that every Riemann surface possesses a countable dense subset. This statement was first proved by Radó. Compare Ahlfors & Sario.) Given any sequence of holomorphic maps $f_n : S \to U \subset \hat{\mathbf{C}}$, we can first choose an infinite subsequence $\{f_{n(p)}\}$ of the f_n so that the images $f_{n(p)}(z_1)$ converge to a limit within the closure of U. Then choose a sub-sub-sequence $f_{n(p(q))}$ so that the $f_{n(p(q))}(z_2)$ also converge to a limit, and continue inductively. By a diagonal procedure, taking the first element of the first subsequence, the second element of the second subsequence, and so on, we construct a new infinite sequence of maps $g_m = f_{n_m}$ so that $\lim_{m\to\infty} g_m(z_j) \in \bar{U}$ exists for every choice of z_j.

Case 1. Suppose that every one of these limit points in $\bar{U}$ actually belongs to the set U itself. Given any compact set $K \subset S$ and any $\epsilon > 0$, we can choose finitely many z_j so that every point $z \in K$ has Poincaré distance $\rho_S(z, z_j) < \epsilon$ from one of these z_j. Further, we can choose m_0 so that $\rho_U(g_m(z_j), g_n(z_j)) < \epsilon$ for each of these finitely many z_j, whenever $m, n > m_0$. For any $z \in K$ it then follows using 2.2 that $\rho_U(g_m(z), g_n(z)) < 3\epsilon$ whenever $m, n > m_0$. Thus the $g_m(z)$ form a Cauchy sequence. It follows that the sequence of functions $\{g_m\}$ converges to a limit, and that this convergence is uniform on compact subsets of S.

Case 2. Suppose on the other hand that $\lim_{m\to\infty} g_m(z_j)$ is actually a boundary point $a_0 \in \partial U$ for some z_j. Then it follows from 2.4 that $g_m(z)$ converges to a_0 for every $z \in S$, and that this convergence is uniform on compact subsets of S. □

2.6. Remark. If we consider maps $S \to S'$ where the target space S' is not compact, then the definition should be modified as follows. We continue to assume that S is connected. A collection $\mathcal{F}$ of maps $f_\alpha : S \to S'$ is *normal* if every infinite sequence of maps from $\mathcal{F}$ contains a subsequence which either

(1) converges locally uniformly to a holomorphic map from S to S', or

(2) diverges locally uniformly to infinity, in the sense that the successive images of any compact subset of S eventually miss any given compact subset of S'.

Concluding Problems.

Problem 2-1. Let S be a simply connected Riemann surface, and let $\Gamma \subset G(S)$ be a discrete subgroup; that is, suppose that the identity element is an isolated point of Γ. If every non-identity element of Γ acts on S without fixed points, show that the action of Γ is *properly discontinuous*. That is, for every compact $K \subset S$ show that only finitely many group elements satisfy $K \cap \gamma(K) \neq \emptyset$. Show that each $z \in S$ has a neighborhood U whose translates $\gamma(U)$ are pairwise disjoint. Conclude that S/Γ is a well defined Riemann surface with S as its universal covering.

Problem 2-2. Show that any Riemann surface which is not conformally isomorphic to $\hat{\mathbf{C}}$, has one and up to a multiplicative constant only one conformal Riemann metric which is complete, with constant Gaussian curvature. (Make use of Hopf's Theorem which asserts that, for each real number K, there is one and only one complete simply connected surface of constant curvature K up to isometry. See Willmore p. 162.) On the other hand, show that $\hat{\mathbf{C}}$ has a 3-dimensional family of conformal metrics with curvature $+1$.

Problem 2-3. Using Problems 1-10 and 1-6, show that every Hyperbolic surface with abelian fundamental group is conformally isomorphic either to the disk D, or the punctured disk $D \smallsetminus \{0\}$, or to the annulus

$$\mathcal{A}_r = \{z \in \mathbf{C} : 1 < |z| < r\}$$

for some uniquely defined $r > 1$. Define the *modulus* of this annulus to be the number $\mathrm{mod}(\mathcal{A}_r) = \log r / 2\pi > 0$. Show that this annulus has a unique simple closed geodesic, which has length $\ell = \pi/\mathrm{mod}(\mathcal{A}_r) = 2\pi^2/\log r$. On the other hand, show that the punctured disk $D \smallsetminus \{0\}$ has no closed geodesic. This punctured disk, which is conformally isomorphic to the set $\mathcal{A}_\infty = \{z \in \mathbf{C} : 1 < |z| < \infty\}$, is sometimes described as an annulus of infinite modulus. (However this designation is ambiguous, since the Euclidean surface $\mathbf{C} - \{0\}$ might also be described as an annulus of infinite modulus.)

Problem 2-4. If S and S' are Hyperbolic Riemann surfaces (not necessarily compact), show that *every* family of maps from S to S' is normal.

Problem 2-5. Show that normality is a local property. More precisely, let S and S' be any Riemann surfaces, and let $\{f_\alpha\}$ be a family of holomorphic maps from S to S'. If every point of S has a neighborhood U such that the collection $\{f_\alpha|U\}$ of restricted maps is normal, show by a diagonal argument as in the proof of 2.5 that $\{f_\alpha\}$ is normal.

THE JULIA SET

§3. Fatou and Julia: Dynamics on the Riemann Sphere.

The local study of iterated holomorphic mappings, in a neighborhood of a fixed point, was quite well developed in the late 19^{th} century. (Compare §§6,7.) However, except for one very simple case studied by Cayley, nothing was known about the global behavior of iterated holomorphic maps until 1906, when Pierre Fatou described the following startling example. For the map $z \mapsto z^2/(z^2+2)$, he showed that almost every orbit under iteration converges to zero, even though there is a Cantor set of exceptional points for which the orbit remains bounded away from zero. (Problem 3-6.) This aroused great interest. After a hiatus dug the first world war, the subject was taken up in depth by Fatou, and also by Gaston Julia and others such as S. Lattés and J. F. Ritt. The most fundamental and incisive contributions were those of Fatou himself, although Julia developed much closely related material at more or less the same time. Julia, who had been badly wounded during the war, was awarded the "Grand Prix des Sciences mathématiques" by the Paris Academy of Sciences in 1918 for his work.

Definition. Let S be a Riemann surface, let $f : S \to S$ be a non-constant holomorphic mapping, and let $f^{\circ n} : S \to S$ be its n-fold iterate. Fixing some point $z_0 \in S$, we have the following basic dichotomy: *If there exists some neighborhood U of z_0 so that the sequence of iterates $\{f^{\circ n}\}$ restricted to U forms a normal family, then we say that z_0 is a* regular *or* normal *point, or that z_0 belongs to the* Fatou set *of f . Otherwise, if no such neighborhood exists, we say that z_0 belongs to the* Julia set *$J = J(f)$.* (For sharper formulations of this dichotomy in the rational case, see §11.8 and Problem 13-1.)

Thus, by its very definition, the Julia set J is a closed subset of S , while the Fatou set $S \smallsetminus J$ is the complementary open subset. (The choice as to which of these two sets should be named after Julia and which after Fatou is rather arbitrary. The term "Julia set" is firmly established, but the Fatou set is often called by other names, such as "stable set" or "normal set".) *Roughly speaking, z_0 belongs to the Fatou set if dynamics in some neighborhood of z_0 is relatively tame, and belongs to the Julia set, if dynamics in every neighborhood of z_0 is more wild.*

The classical example, and the one which we will emphasize, is the case where S is the Riemann sphere $\hat{\mathbf{C}} = \mathbf{C} \cup \infty$. Any holomorphic map $f : \hat{\mathbf{C}} \to \hat{\mathbf{C}}$ on the Riemann sphere can be expressed as a rational function, that is as the quotient $f(z) = p(z)/q(z)$ of two polynomials. Here we may assume that $p(z)$ and $q(z)$ have no common roots. The *degree* d of $f = p/q$ is then equal to the maximum of the degrees of p and q . In particular, for almost every choice of constant $c \in \hat{\mathbf{C}}$ the equation $f(z) = c$ has exactly d distinct solutions in $\hat{\mathbf{C}}$. (For every choice of c it has at least one solution, since we assume that $d > 0$.)

As a simple example, consider the squaring map $s : z \mapsto z^2$ on $\hat{\mathbf{C}}$. The entire open disk D is contained in the Fatou set of s , since successive iterates on any compact subset converge uniformly to zero. Similarly the exterior $\hat{\mathbf{C}} \smallsetminus \bar{D}$ is contained in the Fatou set, since the iterates of s converge to the constant function $z \mapsto \infty$ outside of $\bar{D}$. On the other hand, if z_0 belongs to the unit circle, then in any neighborhood of z_0 any limit of iterates $s^{\circ n}$ would necessarily have a jump discontinuity as we cross the unit circle. This shows that the Julia set $J(s)$ is precisely equal to the unit circle.

Such smooth Julia sets are rather exceptional. (Compare §5.) See Figure 1 for some much more typical examples of Julia sets for polynomial mappings. Figure 1a shows a rather wild Jordan curve, Figure 1b a rather thick Cantor set, Figure 1c a "dendrite", and Figure 1d a more complicated example, the "airplane", with a superattracting period 3 orbit. (Further examples of Julia sets are shown in Figures 2-5, 8-10, 12, and 17.)

We will also need the following concept.

Definition. By the *grand orbit* of a point z under $f : S \to S$ we mean the set $GO(z, f)$ consisting of all points $z' \in S$ whose orbits eventually intersect the orbit of z. Thus z and z' have the same grand orbit if and only if $f^{\circ m}(z) = f^{\circ n}(z')$ for some choice of $m \geq 0$ and $n \geq 0$.

Here are some basic properties of the Julia set.

3.1. Lemma. *The Julia set $J(f)$ of a holomorphic map $f : S \to S$ is fully invariant under f. That is, if z belongs to $J(f)$, then the entire grand orbit $GO(z, f)$ is contained in $J(f)$.*

Evidently it suffices to prove that $z \in J(f)$ if and only if $f(z) \in J(f)$. A completely equivalent statement is that the Fatou set is fully invariant. The proof, making use of the fact that a non-constant holomorphic map takes open sets to open sets, is completely straightforward and will be left to the reader. □

It follows that the Julia set possesses a great deal of *self-similarity*: *Whenever $f(z_1) = z_2$ in $J(f)$, with derivative $f'(z_1) \neq 0$, there is an induced conformal isomorphism from a neighborhood N_1 of z_1 to a neighborhood N_2 of z_2 which takes $N_1 \cap J(f)$ precisely onto $N_2 \cap J(f)$.* (Compare Problem 3-7.)

3.2. Lemma. *For any $n > 0$, the Julia set $J(f^{\circ n})$ of the n-fold iterate coincides with the Julia set $J(f)$.*

Again, the proof will be left to the reader. □

Now consider a *periodic orbit* or "*cycle*"

$$f \;:\; z_0 \mapsto z_1 \mapsto \;\cdots\; \mapsto z_{n-1} \mapsto z_n = z_0\,.$$

If the points $z_1, \ldots, z_n$ are all distinct, then the integer $n \geq 1$ is called the *period*. If the Riemann surface S is $\mathbf{C}$ (or an open subset of $\mathbf{C}$), then the derivative

$$\lambda \;=\; (f^{\circ n})'(z_i) \;=\; f'(z_1) \cdot f'(z_2) \cdots f'(z_n)$$

is a well defined complex number called the *multiplier* or the *eigenvalue* of this periodic orbit. More generally, for self-maps of an arbitrary Riemann surface the multiplier of a periodic orbit can be defined using a local coordinate chart around any point of the orbit. By definition, a periodic orbit is either *attracting* or *repelling* or *indifferent* (= *neutral*) according as its multiplier satisfies $|\lambda| < 1$ or $|\lambda| > 1$ or $|\lambda| = 1$. The orbit is called *superattracting* if $\lambda = 0$.

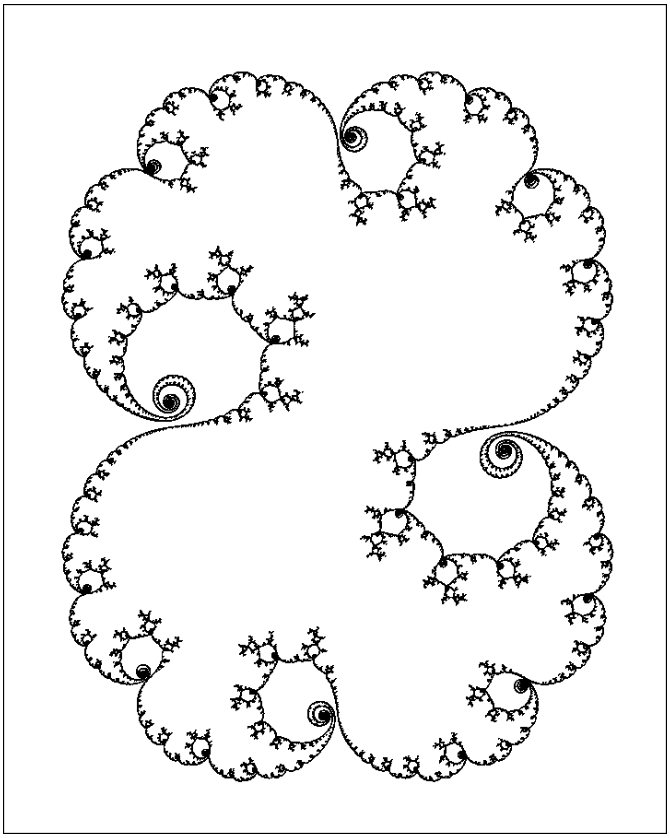


Figure 1a. A simple closed curve,
Julia set for $z \mapsto z^2 + (.99 + .14i)z$.

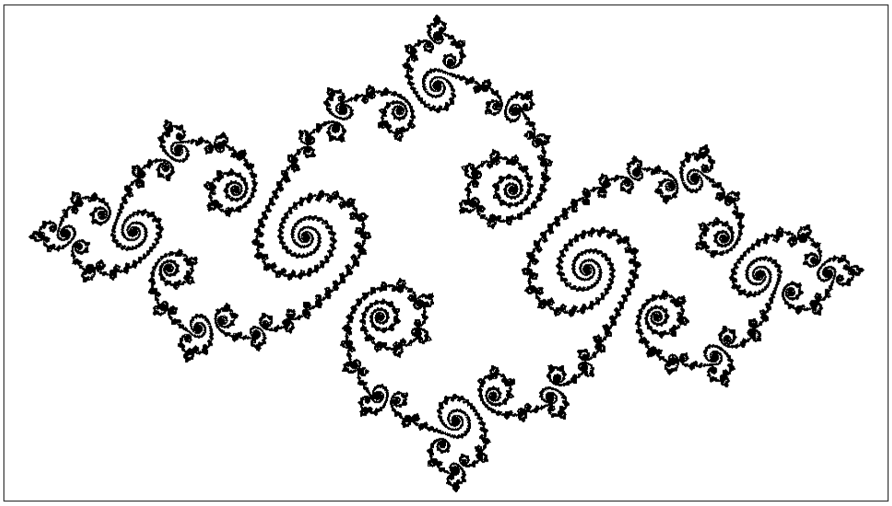


Figure 1b. A totally disconnected Julia set,
$z \mapsto z^2 + (-.765 + .12i)$.

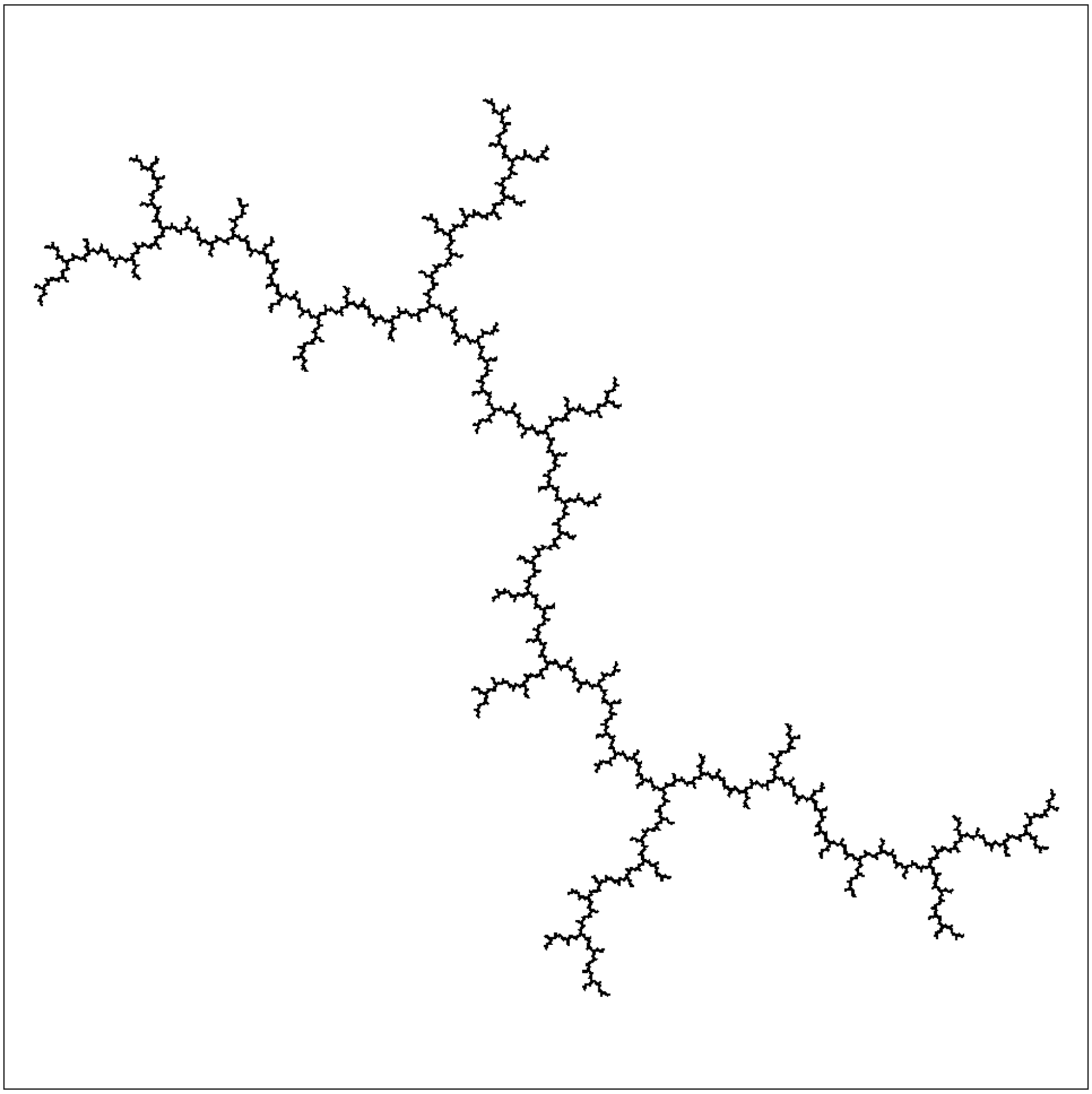


Figure 1c. A “dendrite”, Julia set for $z \mapsto z^2 + i$.

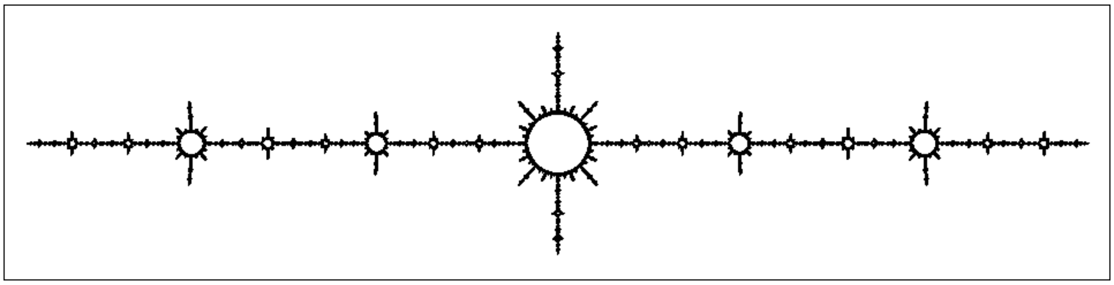


Figure 1d. Julia set for $z \mapsto z^2 - 1.75488\ldots$, the “airplane”.

Caution: In the special case where the point at infinity is periodic under a rational map, $f^{\circ n}(\infty) = \infty$, this definition may be confusing. The multiplier λ is *not* equal to the limit as $z \to \infty$ of the derivative of $f^{\circ n}(z)$, but is rather equal to the reciprocal of this number. In fact if we introduce the local uniformizing parameter $w = 1/z$ for z near ∞ , then it is easy to check that the derivative of $w \mapsto 1/f(1/w)$ at $w = 0$ is equal to $\lim_{z\to\infty} 1/(f^{\circ n})'(z)$. As an example, if $f(z) = 2z$ then ∞ is an attracting fixed point with multiplier $\lambda = 1/2$.

In the case of an attracting periodic orbit, we can define the *basin of attraction* to be the open set $\Omega \subset S$ consisting of all points $z \in S$ for which the successive iterates $f^{\circ n}(z)$, $f^{\circ 2n}(z)$, ... converge towards some point of the periodic orbit. In particular, this basin of attraction is defined in the superattracting case.

3.3. Theorem. *Every attracting periodic orbit is contained in the Fatou set. In fact the entire basin of attraction Ω for an attracting periodic orbit is contained in the Fatou set. However the boundary $\partial\Omega$ is contained in the Julia set, and every repelling periodic orbit is contained in the Julia set.*

Proof. In view of 3.2, we need only consider the case of a fixed point $f(z_0) = z_0$. If z_0 is attracting, then it follows from Taylor's Theorem that the successive iterates of f , restricted to a small neighborhood of z_0 , converge uniformly to the constant function $z \mapsto z_0$. The corresponding statement for any compact subset of the basin Ω then follows easily. On the other hand, around a boundary point of this basin, that is a point which belongs to the closure $\bar{\Omega}$ but not to Ω itself, it is clear that no sequence of iterates can converge to a continuous limit. (See Problem 3-2 for a sharper statement.) If z_0 is repelling, then no sequence of iterates can converge uniformly near z_0 , since the derivative $df^{\circ n}(z)/dz$ at z_0 takes the value λ^n , which diverges to infinity as $n \to \infty$. (Compare §1.4.) □

The case of an indifferent periodic point is much more difficult. (Compare §§7, 8.) One particularly important case is the following.

Definition. A periodic point $f^{\circ n}(z_0) = z_0$ is called *parabolic* if the multiplier λ at z_0 is equal to $+1$, yet $f^{\circ n}$ is not the identity map, or more generally if λ is a root of unity, yet no iterate of f is the identity map.

As an example, the two fixed points of the rational map $f(z) = z/(z-1)$ both have multiplier equal to -1 . These do not count as parabolic points since $f \circ f(z)$ is identically equal to z . This exclusion is necessary so that the following assertion will be true.

3.4. Lemma. *Every parabolic periodic point belongs to the Julia set.*

Proof. Let w be a local uniformizing parameter, with $w = 0$ corresponding to the periodic point. Then some iterate $f^{\circ m}$ corresponds to a local mapping of the w-plane with power series expansion of the form $w \mapsto w + a_k w^k + a_{k+1} w^{k+1} + \cdots$, where $k \geq 2$, $a_k \neq 0$. It follows that $f^{\circ mp}$ corresponds to a power series $w \mapsto w + p a_k w^k + \cdots$. Thus the k-th derivatives of $f^{\circ mp}$ diverge as $p \to \infty$. It follows from 1.4 that no subsequence can converge locally uniformly. □

Now and for the rest of §3, let us specialize to the case of a rational map $f : \hat{\mathbf{C}} \to \hat{\mathbf{C}}$ of degree $d \geq 2$.

3.5. Lemma. *If f is rational of degree two or more, then the Julia set $J(f)$ is non-vacuous.*

Proof. If $J(f)$ were vacuous, then some sequence of iterates $f^{\circ n(j)}$ would converge, uniformly over the entire sphere $\bar{\mathbf{C}}$, to a holomorphic limit $g : \hat{\mathbf{C}} \to \hat{\mathbf{C}}$. (Here we are using the fact that normality is a local property: Problem 2-5.) A standard topological argument would then show that the degree of $f^{\circ n(j)}$ is equal to the degree of g for large j . But the degree of $f^{\circ n}$ is equal to d^n , which diverges to infinity as $n \to \infty$. □

A different, more constructive proof of this Lemma will be given in §9.5.

Definition. A point $z \in \hat{\mathbf{C}}$ is called *grand orbit finite* or (to use the classical terminology) *exceptional* under f if its grand orbit $GO(z, f) \subset \hat{\mathbf{C}}$ is a finite set. Using Montel's Theorem, we prove the following.

3.6. Lemma. *If f is rational of degree two or more, then the set $\mathcal{E}(f)$ of grand orbit finite points can have at most two elements. These grand orbit finite points, if they exist, must always be critical points of f , and must belong to the Fatou set.*

Proof. (Compare Problem 3-3.) Note that f maps any grand orbit $GO(z, f)$ onto itself. Hence, any finite grand orbit must constitute a single periodic orbit under f . Each point z in this finite orbit must be critical (and in fact $(d-1)$-fold critical, where d is the degree), since otherwise $f(z)$ would have two or more pre-images. Therefore, such an orbit must be attracting, and hence contained in the Fatou set.

If there were three distinct grand orbit finite points, then the union of the grand orbits of these points would form a finite set whose complement U in $\hat{\mathbf{C}}$ would be Hyperbolic, with $f(U) = U$. Therefore, the set of iterates of f restricted to U would be normal by Montel's Theorem. Thus both U and its complement would be contained in the Fatou set, contradicting 3.5. □

3.7. Lemma. *Let z_1 be any point of the Julia set $J(f) \subset \hat{\mathbf{C}}$. If N is a sufficiently small neighborhood of z_1 , then the union U of the forward images $f^{\circ n}(N)$ is precisely equal to the complement $\hat{\mathbf{C}} \smallsetminus \mathcal{E}(f)$ of the set of grand orbit finite points.*

In particular, this union U contains the Julia set $J(f)$. (In §11.2 we will see that just one forward image $f^{\circ n}(N)$ actually contains the entire Julia set, provided that n is sufficiently large.)

Proof of 3.7. Let E be the complementary set $\hat{\mathbf{C}} \smallsetminus U$. We have $f(U) \subset U$, or equivalently $f^{-1}(E) \subset E$, by the construction. Since U intersects the Julia set, it follows from Montel's Theorem that its complement E has at most two points. Now since E is finite and f is onto, a counting argument shows that $f^{-1}(E) = E$, hence E is contained in the set $\mathcal{E}(f)$ of grand orbit finite points. If the initial neighborhood N is small enough to be disjoint from $\mathcal{E}(f)$, then it follows that $E = \mathcal{E}(f)$. □

3.8. Corollary. *If the Julia set contains an interior point* z_1 *, then it must be equal to the entire Riemann sphere.*

For if we choose a neighborhood $N \subset J$, then 3.7 shows that the union $U \subset J$ of forward images of N is everywhere dense on $\hat{\mathbf{C}}$. □

3.9. Theorem. *If* $z_0 \in J(f)$ *, then the set of all iterated pre-images*

$$\{z \; : \; f^{\circ n}(z) = z_0 \quad \text{for some} \quad n \geq 0\}$$

is everywhere dense in $J(f)$ *. In particular, it follows that the grand orbit* $GO(z_0, f)$ *is everywhere dense in* $J(f)$ *.*

For z_0 is not a grand orbit finite point, so 3.7 shows that every point $z_1 \in J(f)$ can be approximated arbitrarily closely by points z whose forward orbits contain z_0 . □

This Theorem suggests an algorithm for computing pictures of the Julia set: Starting with any $z_0 \in J(f)$, first compute all pre-images $f(z_1) = z_0$, then compute all pre-images $f(z_2) = z_1$, and so on, thus eventually coming arbitrarily close to every point of $J(f)$. This method is most often used in the quadratic case, since quadratic equations are very easy to solve. The method is very insensitive to round-off errors; since f tends to be expanding on its Julia set, so that f^{-1} tends to be contracting. (Compare Problem 3-4. ??)

3.10. Corollary. *If* f *has degree two or more, then* $J(f)$ *has no isolated points.*

Proof. Since $J(f)$ is fully invariant, it follows from 3.5 and 3.6 that $J(f)$ must be an infinite set. Hence it contains at least one limit point z_0 . Now the iterated pre-images of z_0 form a dense set of non-isolated points in $J(f)$. □

A property of a point of J is said to be true for *generic* $z \in J$ if it is true for all points in some countable intersection of dense open subsets $U_i \cap J \subset J$. (Compare §8. Here the notation is supposed to indicate that the U_i are open subsets of $\hat{\mathbf{C}}$ and that the closure of $U_i \cap J$ is equal to J .) By Baire's Theorem, any such countable intersection is itself a dense subset of J .

3.11. Corollary. *For generic* $z \in J(f)$ *, the forward orbit* $\{z \, , \, f(z) \, , \, f^{\circ 2}(z) \, , \, \dots \}$ *is everywhere dense in* $J(f)$ *.*

Proof. Let $\{B_j\}$ be a countable collection of open sets forming a basis for the topology of $\hat{\mathbf{C}}$. For each B_j which intersects $J = J(f)$, let U_j be the union of the iterated pre-images $f^{-n}(B_j)$ for $n \geq 0$. Then it follows from 3.9 that the closure of $U_j \cap J$ is equal to the entire Julia set J , and the conclusion follows. □

We will continue the study of rational Julia sets in §11.

Concluding problems.

Problem 3-1. If $f : \hat{\mathbf{C}} \to \hat{\mathbf{C}}$ is rational of degree $d = 1$, show that the Julia set $J(f)$ is either vacuous, or consists of a a single repelling or parabolic fixed point.

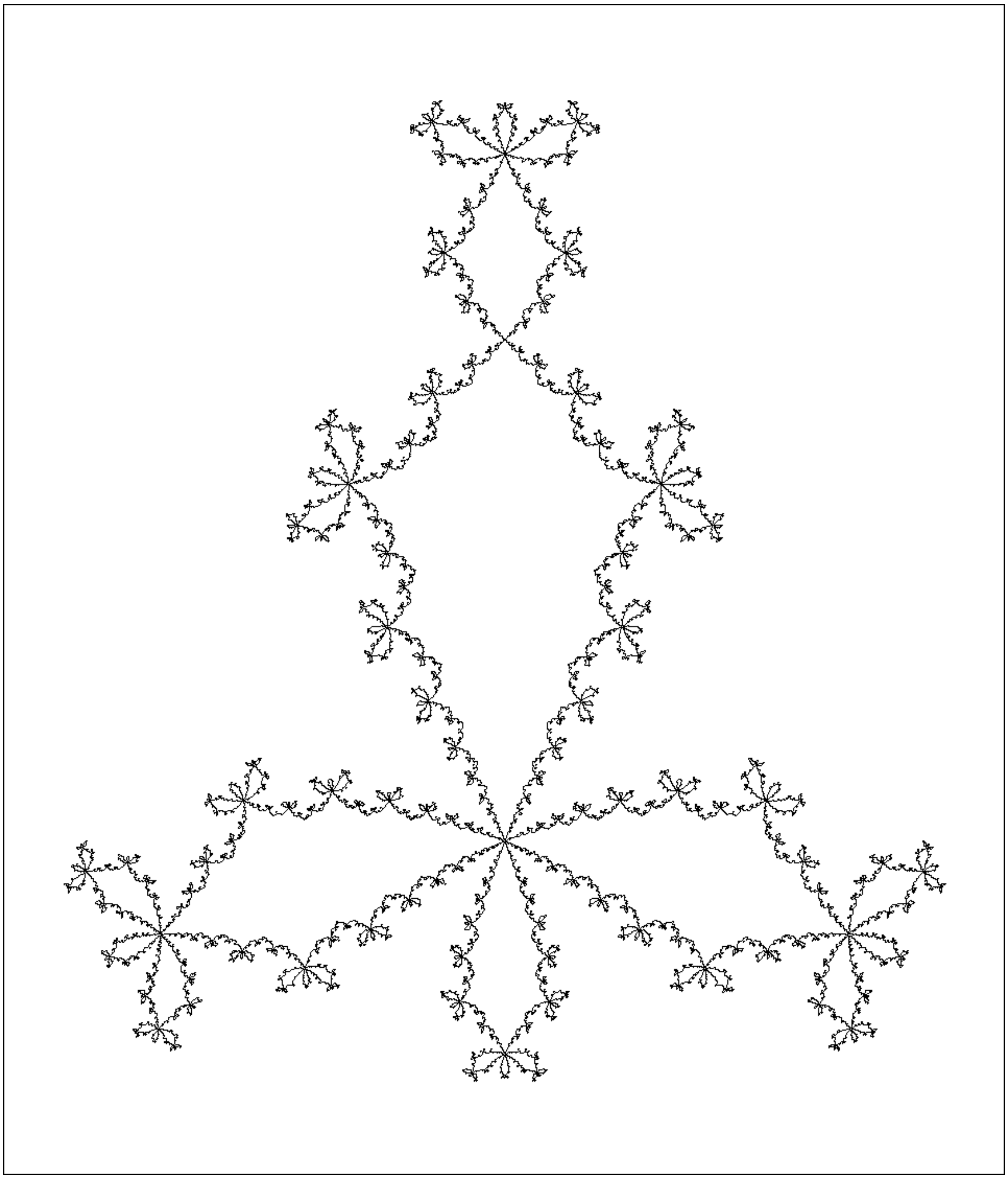


Figure 2. Julia set for $f(z) = z^3 + \frac{12}{25}z + \frac{116}{125}i$.

Problem 3-2. If $\Omega \subset \hat{\mathbf{C}}$ is the basin of attraction for an attracting periodic orbit, show that the boundary $\partial\Omega = \bar{\Omega} \smallsetminus \Omega$ is equal to the entire Julia set. (Compare Theorem 3.3.) Here it is essential that we include all connected components of this basin.

Problem 3-3. Show that a rational map f is actually a polynomial if and only if the point at infinity is a grand orbit finite fixed point for f. Show that f has both zero and infinity as grand orbit finite points if and only if $f(z) = \alpha z^n$, where n can be any non-zero integer, negative or positive, and where $\alpha \neq 0$. Conclude that a rational map has grand orbit finite points if and only if it is conjugate, under some fractional linear change of coordinates, either to a polynomial or to the map $z \mapsto 1/z^d$ for some $d \geq 2$.

Problem 3-4. Using a hand calculator if necessary, decide what maps to what in Figure 1d.

Problem 3-5. If $f(z) = z^2 - 6$, show that $J(f)$ is a Cantor set contained in the intervals $[-3, -\sqrt{3}] \cup [\sqrt{3}, 3]$. More precisely, show that a point in $J(f)$ with orbit $z_0 \mapsto z_1 \mapsto \cdots$ is uniquely determined by the sequence of signs $z_j/|z_j| = \pm 1$.

Problem 3-6. For Fatou's function $f(z) = z^2/(2+z^2)$, show that the entire completed real axis $\mathbf{R} \cup \infty$ is contained in the basin of attraction of the origin. Show that $J(f)$ is a Cantor set. More precisely, given any infinite sequence of signs $\epsilon_0, \epsilon_1, \ldots$ show that there is one and only one point $z = z(\epsilon_0, \epsilon_1, \ldots)$ which satisfies the condition that $f^{\circ n}(z)$ is uniformly bounded away from zero, and belongs to the half-plane $\epsilon_n H$ for each $n \geq 0$. To this end, consider the branch $g(z) = \sqrt{2z/(1-z)}$ of f^{-1}, mapping $\hat{\mathbf{C}} \smallsetminus [0,1]$ onto the upper half-plane H. Starting with any $z_0 \notin [0,1]$, show that the successive images $\epsilon_0 g(\epsilon_1 g(\cdots \epsilon_n g(z_0) \cdots))$ converge to the required point $z \in J(f)$. Since a Cantor set cannot separate the plane, show that the basin of attraction of the origin is equal to $\hat{\mathbf{C}} \smallsetminus J(f)$.

Problem 3-7. Self-similarity. With rare exceptions, any shape which is observed about one point of the Julia set will be observed infinitely often, throughout the Julia set. More precisely, for two points z and z' of $J = J(f)$, let us say that (J, z) is *locally conformally isomorphic* to (J, z') if there exists a conformal isomorphism from a neighborhood N of z onto a neighborhood N' of z' which carries z to z' and $J \cap N$ onto $J \cap N'$. For all but finitely many $z_0 \in J$, show that the set of z for which (J, z) is locally conformally isomorphic to (J, z_0) is everywhere dense in J.

As an example, consider the polynomial map $f(z) = z^3 + .48\,z + .928\,i$ of Figure 2, and explain why the fixed point $.8\,i$ looks different from all other points of the Julia set. (See also Example 2 of §5.) How many pre-images does this point have?

§4. Dynamics on Other Riemann Surfaces.

This section will try to say something about the theory of iterated holomorphic mappings on an arbitrary Riemann surface. However we will concentrate on the easy cases, and simply refer to the literature for the hard cases. It turns out that there are only three Riemann surfaces S for which the study of iterated mappings is really difficult, namely the sphere $\hat{\mathbf{C}}$, the plane $\mathbf{C}$, and the punctured plane $\mathbf{C} \setminus \{0\}$.

The theory of iterated holomorphic mappings from the *Riemann sphere* $\hat{\mathbf{C}}$ to itself has been outlined in §3, and will form the main goal of all of the subsequent sections.

We can distinguish two different classes of holomorphic maps from the *complex plane* $\mathbf{C}$ to itself. A *polynomial map* of $\mathbf{C}$ extends uniquely over the Riemann sphere $\hat{\mathbf{C}}$. Hence the theory of polynomial mappings can be subsumed as a special case of the theory of rational maps of $\hat{\mathbf{C}}$. (See especially §§17-18.) On the other hand, *transcendental mappings* from $\mathbf{C}$ to itself form an essential distinct and more difficult subject of study. Such mappings have been studied for more than sixty years by many authors, starting with Fatou himself. Even iteration of the exponential map $\exp : \mathbf{C} \to \mathbf{C}$ provides a number of quite challenging problems. (See for example Lyubich, and Rees.) A proof that the Julia set $J(\exp)$ is the entire plane $\mathbf{C}$ is included in Devaney [Dv1]. Further information about iterated transcendental functions may be found for example in Baker, Devaney [Dv2], Goldberg & Keen, and in Eremenko & Lyubich.

The study of iterated maps from the *punctured plane* $\mathbf{C} \setminus \{0\}$ to itself is also a difficult and interesting subject. (See for example Keen.)

For all of the uncountably many other Riemann surfaces, it turns out that the possible dynamical behavior is very restricted, and fairly easy to describe. We must distinguish two very different cases according as the surface is a *torus* or is *Hyperbolic*. First consider the case of a torus $\mathbf{T} = \mathbf{C}/\Lambda$, where Λ is a lattice in $\mathbf{C}$.

4.1. Theorem. *Every holomorphic map* $f : \mathbf{T} \to \mathbf{T}$ *is an affine map,* $f(z) \equiv \alpha z + c \pmod{\Lambda}$. *The corresponding Julia set* $J(f)$ *is either the empty set or the entire torus according as* $|\alpha| \le 1$ *or* $|\alpha| > 1$.

Proofs will be given at the end of this section. Here the possible values for the derivative α are sharply restricted. (Problem 4-1.) The dynamics of such iterated affine maps on the torus are of some interest. (See Problem 4-3, as well as Example 3 of §5.) However, the possibilities are so limited that this study cannot be considered very difficult.

Finally, suppose that S is Hyperbolic. Then we will see that any holomorphic self-map behaves in a rather dull manner under iteration. In particular, the Julia set is alway vacuous. Consider first the special case of the unit disk. The following was proved by Denjoy in 1926, sharpening an earlier result by Wolff.

4.2. Theorem. *Let* $f : D \to D$ *be any holomorphic map. Then either*

(1) f *is a "rotation" about some fixed point* $z_0 \in D$, *or else*

(2) *the successive iterates* $f^{\circ n}$ *converge, uniformly on compact subsets of* D, *to a constant function* $z \mapsto c_0 \in \bar{D}$.

Here the notation $f^{\circ n}$ stands for the n-fold composition $f \circ \cdots \circ f$ mapping D into itself. Note that the limiting value $\lim_{n\to\infty} f^{\circ n}(z)$ may belong to the boundary ∂D. Similar statements hold for maps of the upper half-plane H to itself. As examples, if $f(w) = 2w$ or if $f(w) = w + i$ for $w \in H$, then evidently $\lim_{n\to\infty} f^{\circ n}(w)$ is equal to the boundary point $+\infty$ for every $w \in H$. (Here we must measure convergence with respect to the spherical metric on the compact set $\bar{H} \subset \hat{\mathbf{C}}$.)

The corresponding assertion for an arbitrary Hyperbolic surface is only slightly more complicated to state. First some definitions. If $f : S \to S$, then the sequence of points $z \mapsto f(z) \mapsto f^{\circ 2}(z) \mapsto \cdots$ is called the *orbit* of the point $z \in S$. A fixed point $f(z_0) = z_0$ is said to be *attracting* if the derivative $\lambda = f'(z_0)$ satisfies $|\lambda| < 1$. (To be more precise, we must first choose some local coordinate in a neighborhood of the fixed point, and compute this derivative in terms of this local coordinate. We will be careless about this, since in practice our Riemann surfaces will usually be open subsets of $\mathbf{C}$.) For such an attracting fixed point, the *basin of attraction* $\Omega = \Omega(z_0)$ is defined to be the set of all points $z \in S$ for which the orbit $z \mapsto f(z) \mapsto f^{\circ 2}(z) \cdots$ converges towards z_0. It is not difficult to check that Ω is an open subset of S and that $z_0 \in \Omega$.

4.3. Theorem. *If S is a Hyperbolic Riemann surface, then for every holomorphic map $f : S \to S$ the Julia set $J(f)$ is vacuous. Furthermore either:*

(*a*) *every orbit converges towards a unique attracting fixed point $f(z_0) = z_0$,*

(*b*) *every orbit diverges to infinity with respect to the Poincaré metric on S,*

(*c*) *f is an automorphism of finite order, or*

(*d*) *S is isomorphic either to a disk D, a punctured disk $D \smallsetminus \{0\}$, or an annulus*

$$\mathcal{A}_r = \{z : 1 < |z| < r\},$$

and f corresponds to an irrational rotation $z \mapsto e^{2\pi i t} z$ with $t \notin \mathbf{Q}$.

Evidently these four possibilities are mutually exclusive. Much later, in §13, we will want to apply this Theorem to the case where S in an open subset of the sphere $\hat{\mathbf{C}}$ and f is a rational map carrying S into itself. In this case, as in 4.2, it is convenient to study limiting values which belong to the closure of S. We can then sharpen the statement in Case (b) as follows.

4.4. Addendum. *Suppose that U is a Hyperbolic open subset of $\hat{\mathbf{C}}$ and that $f : U \to U$ extends holomorphically throughout a neighborhood of the closure $\bar{U}$. Then in Case (b) above, all orbits in U must converge within $\hat{\mathbf{C}}$ to a single boundary fixed point*

$$f(\hat{z}) = \hat{z} \in \partial U.$$

This convergence is uniform on compact subsets of U.

(In the special case where U is the standard unit disk, of course we have this same result without the extra hypothesis that f extends over the boundary.)

Now let us begin the proofs.

Proof of Lemma 4.1. To fix our ideas, suppose that $\mathbf{T} = \mathbf{C}/\Lambda$ where the lattice $\Lambda \subset \mathbf{C}$ is spanned by the two numbers 1 and τ, and where $\tau \notin \mathbf{R}$. Any holomorphic map $f : \mathbf{T} \to \mathbf{T}$ lifts to a holomorphic map $F : \mathbf{C} \to \mathbf{C}$ on the universal covering surface. Note first that there exist two lattice elements λ_1, $\lambda_2 \in \Lambda$ so that

$$F(z+1) = F(z) + \lambda_1 , \quad F(z+\tau) = F(z) + \lambda_2$$

for every $z \in \mathbf{C}$. For example we certainly have $F(z+1) \equiv F(z) \pmod{\Lambda}$, and the difference function $F(z+1) - F(z) \in \Lambda$ must be constant since $\mathbf{C}$ is connected and the target space Λ is discrete. Now let $g(z) = F(z) - \lambda_1 z$, so that $g(z+1) = g(z)$. Then

$$g(z+\tau) = g(z) + (\lambda_2 - \lambda_1 \tau) .$$

Thus g gives rise to a map from the torus $\mathbf{T}$ to the infinite cylinder

$$\mathbf{C}/(\lambda_2 - \lambda_1\tau)\mathbf{Z} \cong \mathbf{C} \smallsetminus \{0\} ,$$

or from the torus to $\mathbf{C}$ itself if $\lambda_2 - \lambda_1 \tau = 0$. Using Liouville's Theorem or the Maximum Modulus Principle, we see easily that g must be constant Thus $g(z) \equiv c$, and $F(z) = \lambda_1 z + c$ as required. The computation of $J(f)$ will be left as an exercise. (See Problems 4-2 and 4-3 below.) $\square$

We will skip over 4.2 for the moment. The proof of Theorem 4.3 begins as follows. If we are not in Case (b), then some orbit $z_0 \mapsto z_1 \mapsto \cdots$ must possess an accumulation point $\hat{z} \in S$. That is, we can find integers $n(1) < n(2) < \cdots$ so that the sequence $\{z_{n(i)}\}$ converges to $\hat{z}$. Consider the sequence of maps $f^{\circ(n(i+1)-n(i))}$, carrying $z_{n(i)}$ to $z_{n(i+1)}$. By Montel's Theorem, in the version 2.6, there exists some subsequence which converges, uniformly on compact subsets, to a holomorphic map $h : S \to S$. Evidently $h(\hat{z}) = \hat{z}$.

First suppose that f strictly contracts the Poincaré metric. Then h must also contract the Poincaré metric, hence h cannot have two distinct fixed points. But f and h commute, since h is a limit of iterates of f; hence f must map fixed points of h to fixed points of h. Therefore the unique fixed point $\hat{z}$ of h must also be a fixed point of f. This fixed point is attracting, since f contracts the Poincaré metric. *It follows that every orbit under f converges to $\hat{z}$, so that we are in Case (a) of 4.3.* For otherwise, if z' were the closest point which did not belong to the attractive basin of $\hat{z}$, then the Poincaré distance $\rho(\hat{z}, z')$ could not strictly decrease under the map f.

Suppose, on the other hand, that f preserves the Poincaré metric. Still assuming that f has an orbit with a finite limit point, we will prove the following.

4.5. Lemma. *Under these hypotheses, some sequence of iterates $f^{\circ m(i)}$ converges, uniformly on compact subsets, to the identity map of S. It follows that f is necessarily an automorphism of S.*

Proof. Since f preserves the Poincaré metric, it lifts to an isometry from the universal covering $\tilde{S}$ to itself. It follows easily that f is either an automorphism or a covering map from S onto S. On the other hand, we know that some sequence of iterates of f converges to a map h which has a fixed point. We can lift h to a map $H : \tilde{S} \to \tilde{S}$

which has a corresponding fixed point. Since H must also preserve the Poincaré metric, it follows that H can only be a rotation about this fixed point. Therefore, some sequence of iterates of h converges to the identity map ι of S. It then follows easily that some sequence of iterates of f converges to ι. Therefore f is one-to-one. For if $f(z) = f(z')$, then $f^{\circ m}(z) = f^{\circ m}(z')$, and passing to the limit we have $\iota(z) = \iota(z')$, or in other words $z = z'$. Since a one-to-one covering map is clearly a conformal isomorphism, this proves the Lemma. □

To complete the proof of 4.3, we must prove the following.

4.6. Lemma. *If an automorphism f of a Hyperbolic surface S has iterates $f^{\circ m}$ which are arbitrarily close to the identity map (uniformly on compact subsets), then either f has finite order, or else S is isomorphic to D or $D\smallsetminus\{0\}$ or to an annulus, and f corresponds to an irrational rotation.*

(For this Lemma, we don't really need the hypothesis that S is Hyperbolic. In the non-Hyperbolic case, the corresponding statement would be that S is isomorphic to either $\mathbf{C}$ or $\mathbf{C}\smallsetminus\{0\}$ or $\hat{\mathbf{C}}$, and that f corresponds to an irrational rotation.)

Before proving 4.6, it will be convenient to briefly consider the more general situation of a map $f : S \to S'$ between different Riemann surfaces, where $S \cong \tilde{S}/\Gamma$ and $S' \cong \tilde{S}'/\Gamma'$. Such a map f lifts to a map $F : \tilde{S} \to \tilde{S}'$ which is unique up to composition with elements of the group Γ' of deck transformations of the target space. As in the proof of 4.1, to each deck transformation $\gamma \in \Gamma$ there corresponds one and only one deck transformation $\gamma' \in \Gamma'$ so that the identity

$$F(\gamma(z)) \;=\; \gamma'(F(z))$$

holds for all $z \in S$. In fact the correspondence $\gamma \mapsto \gamma'$ is a group homomorphism, which can be identified with the "induced homomorphism" $f_* : \pi_1(S) \to \pi_1(S')$ between fundamental groups.

We are interested in the special case $S = S'$, with universal covering $\tilde{S}$ isomorphic to the unit disk D. It will be convenient to choose some compact disk $K \subset S$ and a corresponding compact disk K^* in the universal covering surface $\tilde{S}$. If $f^{\circ m(j)}$ is uniformly close to the identity map on K, note that the lifted map $F^{\circ m(j)}$ may be far from the identity map on K^*. However, there must exist some deck transformation γ_j so that the composition $F_j = \gamma_j \circ F^{\circ m(j)}$ is uniformly close to the identity on K^*.

We will also need the following.

4.7. Lemma. *Let $\Gamma \subset G$ be any discrete subgroup of a topological group G. Then for each $\gamma \in \Gamma$ there exist a neighborhood N of the identity in G so that elements $g \in N$ satisfy $g \circ \gamma \circ g^{-1} \in \Gamma$ only if they commute with γ.*

The proof is straightforward. □

Proof of 4.6. By the discussion above, we have a sequence of automorphisms $F_j = \gamma_j \circ F^{\circ m(j)} \in G(\tilde{S})$ which converge locally uniformly to the identity automorphism. Furthermore, for each fixed $\gamma_0 \in \Gamma$ and each F_j there is a corresponding γ'_j so that

$$F_j \circ \gamma_0 \;= \gamma_j' \circ F_j \,.$$

By 4.7, it follows that F_j commutes with γ_0 for large j .

Now let us identify $G(\tilde S)$ with the automorphism group $G(D)$, which operates not only on the open disk D but also on the closed disk $\bar D$. By 1.15 we know that two non-identity elements of $G(D)$ commute if and only if they have exactly the same fixed points in $\bar D$. Thus, if we fix some non-identity $\gamma \in \Gamma$, and if we exclude the case where some F_j is actually equal to the identity automorphism, then we can say that γ has exactly the same fixed points as F_j for j sufficiently large. *This implies that all non-identity elements of* Γ *have the same fixed points, so that* Γ *is a commutative group.* Evidently, either Γ is trivial and $S \cong D$, or Γ is free cyclic and S is an annulus or punctured disk. (Compare Problem 2-3.) Further details are straightforward, and will be left to the reader. □

Proof of Addendum 4.4. Starting at some arbitrary point z_0 in the connected open set $U \subset \hat{\mathbf{C}}$, choose a path $p:[0,1] \to U$ from the point $z_0 = p(0)$ to $f(z_0) = p(1)$, and continue this path inductively for all $t \ge 0$ by setting $p(t+1) = f(p(t))$. By hypothesis, the orbit $p(0) \mapsto p(1) \mapsto p(2) \mapsto \cdots$ converges to infinity with respect to the Poincaré metric on U . Hence it must tend to the boundary of the open set U , with respect to the spherical metric. Let δ be the diameter of the image $p[0,1]$ in the Poincaré metric for U . Then each successive image $p[n,\, n+1]$ must also have diameter less than δ . Since these sets converge to the boundary of U , it follows that the diameter of $p[n,\, n+1]$ in the spherical metric must tend to zero as $n \to \infty$. (§2.4.) Since $f(p(t)) = p(t+1)$, this implies that every accumulation point of the path $p(t)$ as $t \to \infty$ must be a fixed point of f . It is not difficult to check that the set of all such accumulation points must be a connected subset of the boundary ∂U . But our hypothesis that f continues holomorphically throughout a neighborhood of the closure $\bar U$ guarantees that f can have only finitely many fixed points in $\bar U$. This proves that the path $p(t)$ converges to just one fixed point $f(\hat z) = \hat z \in \partial U$. Thus the orbit of z_0 converges to $\hat z$ in the spherical metric. Using 2.2 and 2.4, it follows easily that every orbit in U converges to $\hat z$, and that this convergence is uniform on compact subsets of U . □

Proof of the Denjoy-Wolff Theorem 4.2. We now assume that S is the unit disk $D \subset \mathbf{C}$, but do not assume that f can be continued outside of the open disk. The following argument is taken from a lecture of Beardon, as communicated to me by Shishikura. For any $\epsilon > 0$, let us approximate f by the map $z \mapsto (1-\epsilon)f(z)$ from D into a proper subset of itself. Then it is not difficult to check that there is one and only one fixed point $z_\epsilon = (1-\epsilon)f(z_\epsilon)$. If f itself has no fixed point, then these z_ϵ must tend to the boundary of the disk as $\epsilon \to 0$. Let $r(\epsilon)$ be the Poincaré distance of z_ϵ from the origin, and consider the closed neighborhood $N_{r(\epsilon)}(z_\epsilon\,,\, \rho_D)$, which contains the origin as a boundary point. By Pick's Theorem 1.10, this neighborhood is necessarily carried into itself by the map $z \mapsto (1-\epsilon)f(z)$. These neighborhoods are actually round disks (although with a different center point) with respect to the Euclidean metric. (Problem 1-8.) After passing to a subsequence as $\epsilon \to 0$, we may assume that these disks $N_{r(\epsilon)}(z_\epsilon\,,\, \rho_D)$ tend to a limit disk N_0 , bounded by a circle (known as a "horocircle") which is tangent to the boundary of D at a single point $\hat{z}$. Now f must map N_0 into itself, hence the entire orbit of 0 under f must be contained in N_0 . Therefore, if this orbit has no limit point in the open disk D , then it must converge towards the point of tangency $\hat{z}$. The argument now proceeds as in 4.4. □

We conclude with problems for the reader.

Problem 4-1. Given some torus $\mathbf{T} = \mathbf{C}/\Lambda$ and some $\alpha \in \mathbf{C}$, show that there exists a holomorphic map $f(z) \equiv \alpha z + c$ from $\mathbf{T}$ to itself with derivative α if and only if $\alpha\Lambda \subset \Lambda$. Evidently an arbitrary rational integer $\alpha \in \mathbf{Z}$ will satisfy this condition. Show that there exists such a map with derivative $\alpha \notin \mathbf{Z}$ if and only the ratio of two generators of Λ satisfies a quadratic equation with integer coefficients. Such a torus is said to admit "*complex multiplications*".

Problem 4-2. If $|\alpha| = 1$ but $\alpha \neq 1$, show that f is an automorphism of finite order, and in fact of order either 2, 3, 4, or 6. Conclude that $J(f)$ is vacuous. Show that all four cases can occur, for suitably chosen Λ .

Problem 4-3. If $\alpha \neq 0$, show that any equation of the form $f(z) = z_0$ has exactly $|\alpha|^2$ solutions $z \in \mathbf{T}$. If $\alpha \neq 1$, show that f has exactly $|\alpha - 1|^2$ fixed points. (In particular, both $|\alpha|^2$ and $|\alpha - 1|^2$ are necessarily integers.) More generally, if $|\alpha| > 1$ show that the equation $f^{\circ n}(z) = z$ has exactly $|\alpha^n - 1|^2$ solutions in $\mathbf{T}$. Show that the periodic points of f are everywhere dense in $\mathbf{T}$ whenever $\alpha \neq 0$. Conclude that the Julia set $J(f)$ is the entire torus whenever $|\alpha| > 1$.

Problem 4-4. On the other hand, show that a map from a Hyperbolic surface to itself can have at most one periodic point (necessarily a fixed point), unless every point is periodic.

§5. Smooth Julia Sets.

Most Julia sets tend to be complicated fractal subsets of $\hat{\mathbf{C}}$. (Compare Brohlin [Br]. We will give a very partial explanation for this fact in §6.3.) This section, however, will be devoted to three exceptional Julia sets which are actually smooth manifolds.

Example 1: the Circle. As discussed already in §3, the unit circle appears as Julia set for the mapping $z \mapsto z^{\pm n}$ for any $n \geq 2$. Other rational maps with this same Julia set are described in Problem 5-1. Similarly, the real axis $\mathbf{R} \cup \infty$, as the image of the unit circle under a conformal automorphism, can appear as a Julia set. (Problem 5-2.)

Example 2: the Interval. Following Ulam and Von Neumann, consider the map $f(z) = z^2 - 2$, which carries the closed interval $I = [-2, 2]$ onto itself. We will show that the Julia set $J(f)$ is equal to this interval I , and that every point outside of I belongs to the attractive basin $\Omega(\infty)$ of the point at infinity. (For higher degree maps with this same property, see Problem 5-3.) For $z_0 \in I$, it is easy to check that both solutions of the equation $f(z) = z_0$ belong to this interval I . Since I contains a repelling fixed point $z = 2$, it follows from Theorem 3.9 that I contains the entire Julia set $J(f)$. On the other hand, the basin $\Omega(\infty)$ is a neighborhood of infinity whose boundary $\partial\Omega(\infty)$ is contained in $J(f) \subset I$ by Theorem 3.3. Hence everything outside of $J(f)$ must belong to this basin. Since $f(I) \subset I$, it follows that $J(f) = I$.

Here is a more constructive proof that $J(f) = I$. We make use of the substitution $g(w) = w + w^{-1}$, which carries the unit circle in a two-to-one manner onto $I = [-2, 2]$. For $z_0 \neq I$, the equation $g(w) = z_0$ has two solutions, one of which lies inside the unit circle and one of which lies outside. Hence g maps the exterior of the closed unit disk isomorphically onto the complement $\mathbf{C} \smallsetminus I$. Since the squaring map in the w-plane is related to f by the identity

$$g(w^2) = g(w)^2 - 2 = f(g(z)) ,$$

it follows easily that the orbit of z under f either remains bounded or diverges to infinity according as z does or does not belong to this interval. Again using 3.3, it follows that that $J(f) = I$. □

Example 3: all of $\hat{\mathbf{C}}$. The rest of this section will describe an example constructed by S. Lattès, shortly before his death in 1918. Given any lattice $\Lambda \subset \mathbf{C}$ we can form the quotient torus $\mathbf{T} = \mathbf{C}/\Lambda$, as in §4. Thus $\mathbf{T}$ is a compact Riemann surface, and is also an additive Lie group. Note that the automorphism $z \mapsto -z$ of this surface has just four fixed points. For example, if $\Lambda = \mathbf{Z} + \tau\mathbf{Z}$ is the lattice with basis 1 and τ , where $\tau \notin \mathbf{R}$, then the four fixed points are $0, 1/2, \tau/2,$ and $(1+\tau)/2$ modulo Λ .

Now form a new Riemann surface S as a quotient of $\mathbf{T}$ by identifying each $z \in T$ with $-z$. Evidently S inherits the structure of a Riemann surface (although it loses the group structure). In fact we can use $(z - z_j)^2$ as a local uniformizing parameter for S near each of the four fixed points z_j . Thus the natural map $\mathbf{T} \to S$ is two-to-one, except at the four ramification points. To compute the genus of S , we use the following.

5.1. Riemann-Hurwitz Formula. *Let* $\mathbf{T} \to S$ *be a branched covering map from one compact Riemann surface onto another. Then the number of branch points, counted with multiplicity, is equal to* $\chi(S)d - \chi(T)$ *, where* χ *is the Euler characteristic and* d *is the degree.*

Sketch of Proof. Choose some triangulation of S which includes all critical values (that is all images of ramification points) as vertices; and let $a_n(S)$ be the number of n-simplexes, so that $\chi(S) = a_2(S) - a_1(S) + a_0(S)$. In general, each simplex of S lifts to d distinct simplices in $\mathbf{T}$. However, if v is a critical value, then there are too few pre-images of v . The number of missing pre-images is precisely the number of ramification points over v , each counted with an appropriate multiplicity; and the conclusion follows. □

Remark. This proof works also for Riemann surfaces with smooth boundary. The Formula remains true for proper maps between non-compact Riemann surfaces, as can be verified, for example, by means of a direct limit argument.

In our example, since $\mathbf{T}$ is a torus with Euler characteristic $\chi(T) = 0$, and since there are exactly four simple branch points, we conclude that $2\chi(S) - \chi(T) = 4$, or $\chi(S) = 2$. *Using the standard formula* $\chi = 2 - 2g$ *, we conclude that* S *is a surface of genus zero, isomorphic to the Riemann sphere.* (Remark: The projection map from $\mathbf{T}$ to the sphere S is, up to a choice of normalization, known as the *Weierstrass* $\wp$*-function*.)

Now consider the doubling map $z \mapsto 2z$ on $\mathbf{T}$. This commutes with multiplication by -1 , and hence induces a map $f : S \to S$. Since the doubling map has degree four, it follows that f is a rational map of degree four. (More generally, in place of the doubling map, we could use any linear map which carries the lattice Λ into itself, as in Problem 4-1.)

5.2. Lattès Theorem. *The Julia set for this rational map* f *is the entire sphere* S *.*

Proof. Evidently the doubling map on $\mathbf{T}$ has the property that periodic points are everywhere dense. For example, if r and s are any rational numbers with odd denominator, then $r + s\tau$ is periodic. These periodic orbits are all repelling, since the multiplier is a power of two. Evidently f inherits the same property, and the conclusion follows by §3.3. (Alternatively, given a small open set $U \subset S$, it is not difficult to show that $f^{\circ n}(U)$ is equal to the entire sphere S whenever n is sufficiently large. Hence no sequence of iterates of f can converge to a limit on any open set.) □

In order to pin down just which rational map f has these properties, we must first label the points of S . The four branch points on $\mathbf{T}$ map to four "ramification points" on S , which will play a special role. Let us choose a conformal isomorphism from S onto $\hat{\mathbf{C}}$ which maps the first three of these points to $\infty, 0, 1$ respectively. The fourth ramification point must then map to some $a \in \mathbf{C} \smallsetminus \{0, 1\}$. In this way we construct a projection map $\pi : T \to \hat{\mathbf{C}}$ of degree two, which satisfies $\pi(-z) = \pi(z)$, and which has critical values

$$\pi(0) = \infty\,, \quad \pi(1/2) = 0\,, \quad \pi(\tau/2) = 1\,, \quad \pi((1+\tau)/2) = a\,.$$

Here a can be any number distinct from $0, 1, \infty$. In fact, given $a \in \mathbf{C} \smallsetminus \{0, 1\}$, it is not difficult to show that there is one and only one branched covering $\mathbf{T}' \to \hat{\mathbf{C}}$ of degree two with precisely $\{\infty, 0, 1, a\}$ as ramification points. (Compare Appendix E.) The Riemann-Hurwitz formula shows that this branched covering space $\mathbf{T}'$ is a torus, necessarily isomorphic to $\mathbf{C}/(\mathbf{Z} + \tau\mathbf{Z})$ for some $\tau \notin \mathbf{R}$. The unique deck transformation which interchanges the two pre-images of any point must preserve the linear structure, and hence be multiplication by -1.

Now the doubling map on $\mathbf{T}$ corresponds under π to a specific rational map $f_a : \hat{\mathbf{C}} \to \hat{\mathbf{C}}$, where

$$f_a(\pi(z)) = \pi(2z) ,$$

and where $J(f_a) = \hat{\mathbf{C}}$ by 5.2. A precise computation of this map f_a is described in Problem 5-5 below.

Remark. Mary Rees has proved the existence of many more rational maps with $J(f) = \hat{\mathbf{C}}$. (See also Herman.) For any degree $d \geq 2$, let $\mathrm{Rat}(d)$ be the complex manifold consisting of all rational maps of degree d. Rees shows that there is a subset of $\mathrm{Rat}(d)$ of positive measure consisting of maps f which are "ergodic". By definition, this means that any measurable subset of $\hat{\mathbf{C}}$ which is fully invariant under f must either have full measure or measure zero. It can be shown that any ergodic map must necessarily have $J(f) = \hat{\mathbf{C}}$.

We will study these smooth Julia sets further in §14.7.

Concluding problems.

Problem 5-1. For any $a \in D$ the map $\phi_a(z) = (z - a)/(1 - \bar{a}z)$ carries the unit disk isomorphically onto itself. (Problem 1-2.) A finite product of the form

$$f(z) \;=\; e^{i\theta}\phi_{a_1}(z)\phi_{a_2}(z)\cdots\phi_{a_n}(z)$$

with $a_j \in D$ is called a *Blaschke product* of degree n. Evidently any such f is a rational map which carries D onto D. If $n \geq 2$, and if one of the factors is $\phi_0(z) = z$ so that f has a fixed point at the origin, show that the Julia set $J(f) = J(1/f)$ is the unit circle.

Problem 5-2. Newton's method applied to the polynomial equation $f(z) = z^2 + 1 = 0$ yields the rational map

$$\nu(z) \;=\; z - f(z)/f'(z) \;=\; \tfrac{1}{2}(z - 1/z)$$

from $\hat{\mathbf{C}} = \mathbf{C} \cup \infty$ to itself. Show that $J(\nu) = \mathbf{R} \cup \infty$.

Problem 5-3. The *monic Tchebycheff polynomials*

$$P_1(z) = z\,, \quad P_2(z) = z^2 - 2\,, \quad P_3(z) = z^3 - 3z\,, \;\ldots$$

can be defined inductively by the formula $P_{n+1}(z) + P_{n-1}(z) = zP_n(z)$. Show that $P_n(w + w^{-1}) = w^n + w^{-n}$, or equivalently that $P_n(2\cos\theta) = 2\cos(n\theta)$, and show that $P_m \circ P_n = P_{mn}$. For $n \geq 2$ show that the Julia set of $\pm P_n$ is the interval $[-2, 2]$.

For $n \geq 3$ show that P_n has $n-1$ distinct critical points but only two critical values, namely ± 2 .

Problem 5-4. Show that the Julia sets studied in this section have the following extraordinary property. (Compare Problems 14-2, 14-1.) For all but one or two of the periodic orbits $z_0 \mapsto z_1 \mapsto \cdots \mapsto z_n = z_0$, the multiplier $\lambda = f'(z_1) \cdot \cdots \cdot f'(z_n)$ satisfies $|\lambda| = d^n$ when J is 1-dimensional, or $|\lambda| = d^{n/2}$ when $J = \hat{\mathbf{C}}$, where d is the degree.

Problem 5-5. For the torus $\mathbf{T} = \mathbf{C}/(\mathbf{Z}+\tau\mathbf{Z})$ of Example 3, show that the involution $z \mapsto z + 1/2$ of $\mathbf{T}$ corresponds under π to the involution $w \mapsto a/w$ of $\hat{\mathbf{C}}$, with fixed points $w = \pm\sqrt{a}$. Show that the rational map $f = f_a$ has poles at ∞ , 0 , 1 , a and double zeros at $\pm\sqrt{a}$. Show that f has a fixed point of multiplier $\lambda = 4$ at infinity, and conclude that

$$f(w) \;=\; \frac{(w^2 - a)^2}{4w(w-1)(w-a)} \;.$$

As an example, if $a = -1$ then

$$f(w) \;=\; \frac{(w^2 + 1)^2}{4w(w^2-1)} \;.$$

Show that the correspondence $\tau \mapsto a = a(\tau) \in \mathbf{C} \smallsetminus \{0, 1\}$ satisfies the equations

$$a(\tau + 1) = 1/a(\tau) \;, \quad a(-1/\tau) = 1 - a(\tau) \;,$$

and also $a(-\bar{\tau}) = \bar{a}(\tau)$. Conclude, for example, that $a(i) = 1/2$, and that $a((1+i)/2) = -1$. (This correspondence $\tau \mapsto a(\tau)$ is an example of an "elliptic modular function", and provides an explicit representation of the upper half-plane H as a universal covering of the thrice punctured sphere $\mathbf{C} \smallsetminus \{0, 1\}$. Compare Ahlfors [A1, pp.269-274].)

Problem 5-6. For each of the six critical points ω of f , show that $f(f(\omega))$ is the repelling fixed point at infinity. (According to §13.5, the fact that each critical orbit terminates on a repelling cycle is enough to imply that $J(f) = \hat{\mathbf{C}}$.)

LOCAL FIXED POINT THEORY

§6. Attracting and Repelling Fixed Points.

Consider a function

$$f(z) \;=\; \lambda z + a_2 z^2 + a_3 z^3 + \cdots \tag{1}$$

which is defined and holomorphic in some neighborhood of the origin, with a fixed point of multiplier λ at $z = 0$. If $|\lambda| \neq 1$, we will show that f can be reduced to a simple normal form by a suitable change of coordinates. First consider the case $\lambda \neq 0$, so that the origin is not a critical point. The following was proved by G. Kœnigs in 1884.

6.1. Kœnigs Linearization Theorem. *If the multiplier λ satisfies $|\lambda| \neq 0, 1$, then there exists a local holomorphic change of coordinate $w = \phi(z)$, with $\phi(0) = 0$, so that $\phi \circ f \circ \phi^{-1}$ is the linear map $w \mapsto \lambda w$ for all w in some neighborhood of the origin. Furthermore, ϕ is unique up to multiplication by a non-zero constant.*

(This functional equation $\phi \circ f \circ \phi^{-1}(w) \;=\; \lambda w$ had been studied some years earlier by Schröder, who believed however that it did not have many solutions.)

Proof of uniqueness. If there were two such maps ϕ and ψ, then the composition

$$\psi \circ \phi^{-1}(w) \;=\; b_1 w + b_2 w^2 + b_3 w^3 + \cdots$$

would commute with the map $w \mapsto \lambda w$. Comparing coefficients of the two resulting power series, we see that $\lambda b_n = b_n \lambda^n$ for all n. Since λ is neither zero nor a root of unity, this implies that $b_2 = b_3 = \cdots = 0$. Thus $\psi \circ \phi^{-1}(w) \;=\; b_1 w$, or in other words $\psi(z) = b_1 \phi(z)$.

Proof of existence when $0 < |\lambda| < 1$. Choose a constant $c < 1$ so that $c^2 < |\lambda| < c$, and choose a neighborhood D_r of the origin so that $|f(z)| \leq c|z|$ for $z \in D_r$. Thus for any starting point $z_0 \in D_r$, the orbit $z_0 \mapsto z_1 \mapsto \cdots$ under f converges geometrically towards the origin, with $|z_n| \leq rc^n$. By Taylor's Theorem,

$$|f(z) - \lambda z| \;\leq\; k|z^2|$$

for some constant k and for all $z \in D_r$, hence

$$|z_{n+1} - \lambda z_n| \;\leq\; kr^2 c^{2n}\,.$$

It follows that the numbers $w_n = z_n/\lambda^n$ satisfy

$$|w_{n+1} - w_n| \;\leq\; k'(c^2/|\lambda|)^n\,,$$

where $k' \;=\; kr^2/|\lambda|$. These differences converge uniformly and geometrically to zero.

Thus the holomorphic functions $z_0 \mapsto w_n(z_0)$ converge, uniformly throughout D_r, to a holomorphic limit $\phi(z_0) = \lim_{n\to\infty} z_n/\lambda^n$. The required identity $\phi(f(z)) = \lambda\phi(z)$ follows immediately. A similar argument shows that $|\phi(z) - z|$ is less than or equal to some constant times $|z^2|$. Therefore ϕ has derivative $\phi'(0) = 1$, and hence is a local conformal diffeomorphism.

Proof when $|\lambda| > 1$. Since $\lambda \neq 0$, the inverse map f^{-1} is locally well defined and holomorphic, having the origin as an attractive fixed point with multiplier λ^{-1}. Applying the above argument to f^{-1}, the conclusion follows. □

6.2. Remark. More generally, suppose that we consider a family of maps f_α of the form (1) which depend holomorphically on one (or more) complex parameters α and have multiplier $\lambda = \lambda(\alpha)$ satisfying $|\lambda(\alpha)| \neq 0, 1$. Then a similar argument shows that *the Kœnigs function $\phi(z) = \phi_\alpha(z)$ depends holomorphically on α.* This fact will be important in §8.5. To prove this statement, let us fix $0 < c < 1$ and suppose that $|\lambda(\alpha)|$ varies through some compact subset of the interval (c^2, c). Then it is easy to check that the convergence in the proof of 6.1 is uniform in α. The conclusion now follows easily. □

6.3. Remark. The Kœnigs Theorem helps us to understand why the Julia set $J(f)$ is so often a complicated "fractal" set. *Suppose that there exists a single repelling periodic point $\hat{z}$ for which the multiplier λ is not a real number. Then $J(f)$ cannot be a smooth manifold, unless it is all of $\hat{\mathbf{C}}$.* To see this, choose any point $z_0 \in J(f)$ which is close to $\hat{z}$, and let $w_0 = \phi(z_0)$. Then $J(f)$ must also contain an infinite sequence of points $z_1, z_2, \ldots$ with Kœnigs coordinates $\phi(z_n) = w_n/\lambda^n$ which lie along a logarithmic spiral and converge to zero. Evidently such a set can not lie in any smooth submanifold of $\mathbf{C}$. Furthermore, if we recall that the iterated pre-images of our fixed point are everywhere dense in $J(f)$, then we see that such sequences lying on logarithmic spirals are extremely pervasive. Compare Figures 3 and 4 which show typical examples of such spiral structures, associated with repelling points of periods 2 and 1 respectively.

We can restate 6.1 in a more global form as follows.

6.4. Corollary. *Suppose that $f : S \to S$ is a holomorphic map from a Riemann surface to itself with an attractive fixed point of multiplier $\lambda \neq 0$ at $\hat{z}$. Let $\Omega \subset S$ be the basin of attraction, consisting of all $z \in S$ for which $\lim_{n\to\infty} f^{\circ n}(z)$ exists and is equal to $\hat{z}$. Then there is a holomorphic map ϕ from Ω onto $\mathbf{C}$ so that the diagram*

$$\begin{array}{ccc} \Omega & \xrightarrow{f} & \Omega \\ \downarrow\phi & & \downarrow\phi \\ \mathbf{C} & \xrightarrow{\lambda\cdot} & \mathbf{C} \end{array} \qquad (2)$$

is commutative, and so that ϕ takes a neighborhood of $\hat{z}$ diffeomorphically onto a neighborhood of zero. Furthermore, ϕ is unique up to multiplication by a constant.

In fact, to compute $\phi(z_0)$ at an arbitrary point of Ω we must simply follow the orbit of z_0 until we reach some point z_n which is very close to $\hat{z}$, then evaluate the Kœnigs coordinate $\phi(z_n)$ and multiply by λ^{-n}. □

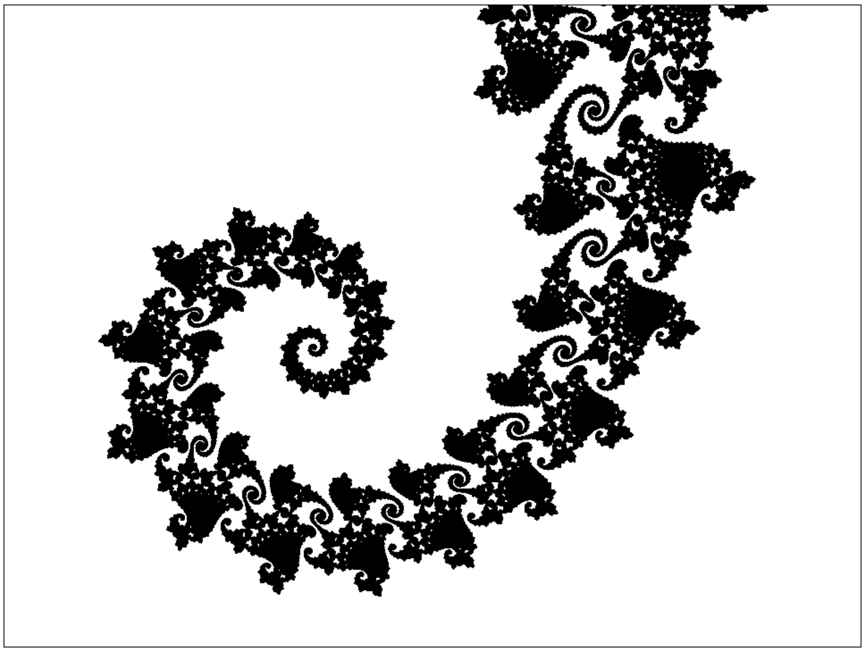


Figure 3. Detail of Julia set for $z \mapsto z^2 - .744336 + .121198i$.

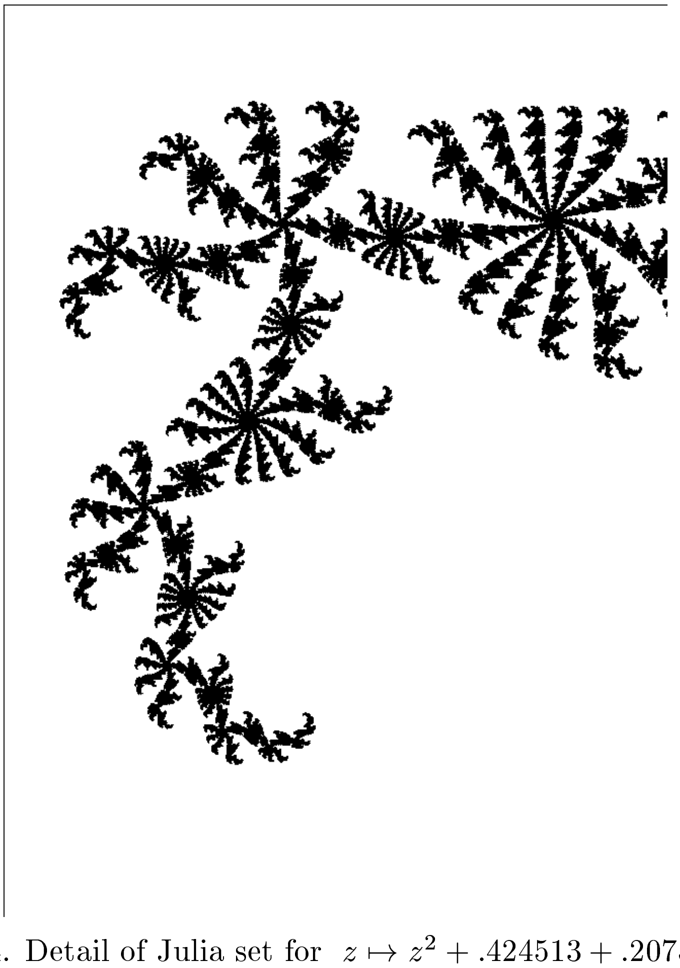


Figure 4. Detail of Julia set for $z \mapsto z^2 + .424513 + .207530i$.

As an example, Figure 5 illustrates the map $f(z) = z^2 + 0.7z$. Here the Julia set J is the outer Jordan curve, bounding the basin Ω of the attracting fixed point. The critical point $\omega = -0.35$ is at the center of the picture, and the attracting fixed point $\hat{z} = 0$ is directly above it. The curves $|\phi(z)| = \text{constant} = |\phi(\omega)/\lambda^n|$ have been drawn in. Note that ϕ has zeros at all iterated preimages of $\hat{z}$, and critical points at all iterated preimages of the critical point ω . The function $z \mapsto \phi(z)$ is unbounded, and oscillates wildly as z tends to $J = \partial\Omega$.

The statement is the repelling case is somewhat different:

6.5. Corollary. *If $\hat{z}$ is a repelling fixed point, then there is a holomorphic map $\psi : \mathbf{C} \to S$ in the opposite direction, so that the diagram*

$$\begin{array}{ccc} S & \xrightarrow{f} & S \\ \uparrow \psi & & \uparrow \psi \\ \mathbf{C} & \xrightarrow{\lambda\cdot} & \mathbf{C} \end{array}$$

is commutative, and so that ψ maps a neighborhood of zero diffeomorphically onto a neighborhood of $\hat{z}$. Here ψ is unique except that it may be replaced by $w \mapsto \psi(cw)$ for any constant $c \neq 0$.

To compute $\psi(w)$ we simply choose some $\lambda^{-n}w$ which is so small that $\phi^{-1}(\lambda^{-n}w)$ is defined, and then apply $f^{\circ n}$ to the result. $\square$

Now suppose that $f : \hat{\mathbf{C}} \to \hat{\mathbf{C}}$ is a rational function with an attracting fixed point $\hat{z}$. By the *immediate basin* $\Omega_0(\hat{z})$ we mean the connected component of $\hat{z}$ in the basin of attraction $\Omega = \Omega(\hat{z})$, or equivalently the connected component of $\hat{z}$ in the Fatou set $\hat{\mathbf{C}} \smallsetminus J$. (Compare 4.3.) The following is due to Fatou and Julia.

6.6. Theorem. *If f has degree two or more, then the immediate basin of any attracting fixed point $\hat{z}$ of f contains at least one critical point. Furthermore, if the multiplier λ is not zero, then there exists a unique compact neighborhood $\bar{U}$ of $\hat{z}$ in Ω_0 which:*

(a) maps bijectively onto some round disk $\bar{D}_r$ under the Kœnigs map ϕ , and

(b) has at least one critical point on its boundary ∂U .

Evidently $\bar{U}$ can be described as the largest neighborhood which maps bijectively to a round disk centered at the origin. As an example, in Figure 5 the region U is bounded by the top half of the central figure 8 shaped curve.

Proof of 6.6. If $\lambda = 0$, then $\hat{z}$ itself is critical. Thus we may assume that $\lambda \neq 0$ and apply 6.1. Evidently some branch ϕ_0^{-1} of the inverse map can be defined as a single valued holomorphic function over some small disk D_ϵ , with $\phi_0^{-1}(0) = \hat{z}$. Let us try to extend ϕ_0^{-1} by analytic continuation along radial lines through the origin in D_ϵ . It cannot be possible to extend indefinitely far in every direction; for then ϕ_0^{-1} would be a non-constant holomorphic map from the entire plane $\mathbf{C}$ into the basin $\Omega_0(\hat{z})$. This is impossible, since this basin is Hyperbolic. Thus there must exist some largest radius r so that ϕ_0^{-1} extends analytically throughout the open disk D_r . Let $U = \phi_0^{-1}(D_r)$. We

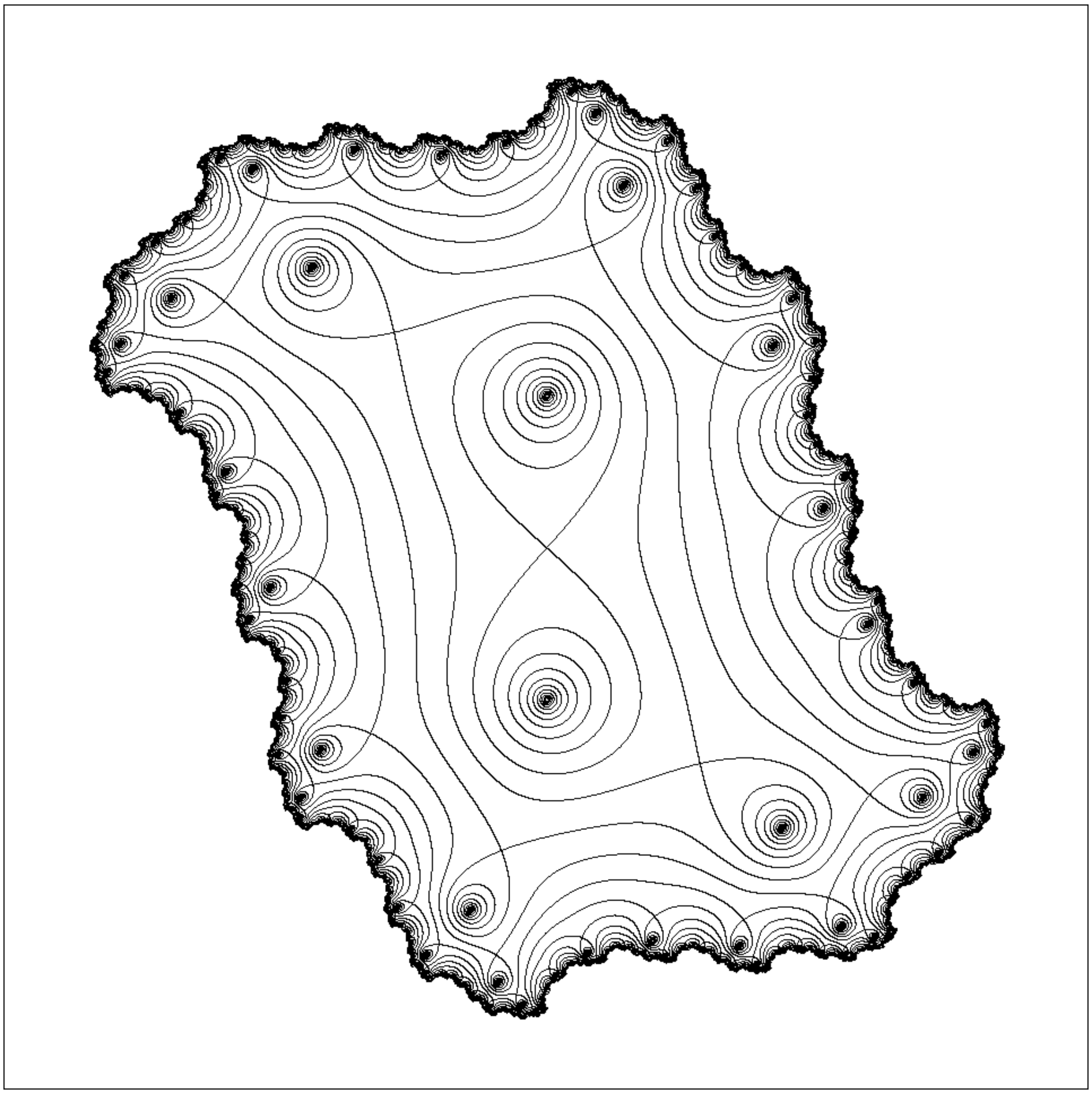


Figure 5. Julia set for $z \mapsto z^2 + .7z$, with curves $|\phi| = \text{constant}$.

must prove that the closure $\bar U$ is a compact subset of the basin $\Omega(\hat z)$, and also that there is at least one critical point of f on the boundary ∂U . If $z_1 \in \partial U$ is any boundary point, then using the identity $\phi(f(z)) = \lambda\phi(z) \in D_{|\lambda|r}$ for $z \in U$ arbitrarily close to z_1 , we see that $f(z_1)$ belongs to the open set $U \subset \Omega$. Therefore z_1 also belongs to the basin Ω , with $|\phi(z_1)| = r$ by continuity. Now at least one such z_1 must be a critical point of f . For whenever z_1 is non-critical we can continue ϕ_0^{-1} analytically throughout a neighborhood of the image point $\phi(z_1) \in \partial D_r$ simply by chasing around Diagram (2), composing the map $z \mapsto \phi_0^{-1}(\lambda z)$ with the branch of f^{-1} which carries $f(z_1)$ to z_1 .
□

For further information about the attractive basin Ω_0 see §13.4, and also §17.1.

Next let us consider the *superattracting* case $\lambda = 0$. The following was proved by Böttcher in 1904.

Historical Note: L. E. Böttcher was born in Warsaw in 1872. He took his doctorate in Leipzig in 1898, working in Iteration Theory, and then moved to Lvov. He published in Polish and Russian. (The Russian form of his name is Бётхеръ .)

6.7. Theorem of Böttcher. *Suppose that*

$$f(z) \;=\; a_n z^n \;+\; a_{n+1} z^{n+1} \;+\; \cdots \;,$$

where $n \ge 2$, $a_n \ne 0$. Then there exists a local holomorphic change of coordinate $w = \phi(z)$ which conjugates f to the n-th power map $w \mapsto w^n$ throughout some neighborhood of $\phi(0) = 0$. Furthermore, ϕ is unique up to multiplication by an $(n-1)$-st root of unity.

Thus near any critical fixed point, f is conjugate to a map of the form

$$\phi \circ f \circ \phi^{-1} \,:\, w \;\mapsto\; w^n \,,$$

with $n \ne 1$. This Theorem is most often applied in the case of a fixed point at infinity. For example, any polynomial map $p(z) = a_n z^n + a_{n-1} z^{n-1} + \cdots + a_0$ of degree $n \ge 2$ has a superattractive fixed point at infinity. It follows easily from 6.7 that there is a local holomorphic map ψ taking infinity to infinity which conjugates p to the map $w \mapsto w^n$ around $w = \infty$.

The proof is quite similar to the Kœnigs proof. It is only necessary to first make a logarithmic change of coordinates, and to be careful since the logarithm is not a single valued function. Suppose, to fix our ideas, that we consider a map having a fixed point at infinity, with a Laurent series expansion of the form

$$f(z) \;=\; a_n z^n + a_{n-1} z^{n-1} + \cdots + a_0 + a_{-1} z^{-1} + \cdots$$

with $n \ge 2$, convergent for $|z| > r$. Note first that the linearly conjugate map $z \mapsto \alpha f(z/\alpha)$, where $\alpha^{n-1} = a_n$, has leading coefficient equal to $+1$. Thus, without loss of generality, we may assume that f itself has leading coefficient $a_n = 1$. Then $f(z) = z^n(1 + \mathcal{O}|1/z|)$ for $|z|$ large, where $\mathcal{O}|1/z|$ stands for some expression which is bounded by a constant times $|1/z|$. Let us make the substitution $z = e^Z$, where Z

ranges over the half-plane $\mathcal{R}(Z) > \log(r)$. Then f lifts to a continuous map

$$F(Z) = \log f(e^Z)\ ,$$

which is uniquely defined up to addition of some multiple of $2\pi i$. With correct choice of this lifting F , it is not hard to check that $F(Z) = nZ + \mathcal{O}(e^{-\mathcal{R}(Z)})$ for $\mathcal{R}(Z)$ large. In fact, we will only need the weaker statement that

$$|F(Z) - nZ| < 1 \tag{3}$$

for $\mathcal{R}(Z)$ large. Let us choose $\sigma > 1$ to be large enough so that the inequality (3) is satisfied for all Z in the half-plane $\mathcal{R}(Z) > \sigma$; note that F necessarily maps this half-plane into itself. Note the identity $F(Z + 2\pi i) = F(Z) + 2\pi i n$, which follows since $F(Z+2\pi i) - F(Z)$ is a multiple of $2\pi i$ which differs from $n(Z+2\pi i) - nZ$ by at most 2. If $Z_0 \mapsto Z_1 \mapsto \cdots$ is any orbit under F in this half-plane, then we have $|Z_{k+1} - nZ_k| < 1$. Setting $W_k = Z_k/n^k$, it follows that

$$|W_{k+1} - W_k| < 1/n^{k+1}\ .$$

Thus the sequence of holomorphic functions $W_k = W_k(Z_0)$ converges uniformly and geometrically as $k \to \infty$ to a holomorphic limit $\Phi(Z_0) = \lim_{k\to\infty} W_k(Z_0)$. Evidently this mapping Φ satisfies the identity

$$\Phi(F(Z)) \;=\; n\Phi(Z)\ .$$

Note also that $\Phi(Z + 2\pi i) = \Phi(Z) + 2\pi i$. Therefore the mapping $\phi(z) = \exp(\Phi(\log z))$ is well defined near infinity, and satisfies the required identity $\phi(f(z)) = \phi(z)^n$.

To prove uniqueness, it suffices to study mappings $w \mapsto \eta(w)$ near infinity which satisfy $\eta(w^n) = \eta(w)^n$. Setting $\eta(w) = c_1 w + c_0 + c_{-1}w^{-1} + \cdots$, this becomes

$$c_1 w^n + c_0 + c_{-1}w^{-n} + \cdots \;=\; (c_1 w + c_0 + \cdots)^n \;=\; c_1^n w^n + n c_1^{n-1} c_0 w^{n-1} + \cdots\ .$$

Since $c_1 \neq 0$, c_1 must be an $(n-1)$-st root of unity, and an easy induction shows that the remaining coefficients are zero. □

Caution. In analogy with 6.4, one might hope that the change of coordinates $z \mapsto \phi(z)$ extends throughout the entire basin of attraction of the superattractive point as a holomorphic mapping. (Compare §17.3.) However, this is not always possible. Such an extension involves computing expressions of the form

$$z \mapsto \sqrt[n]{\phi(f(z))}\ ,$$

and this does not work in general since the n-th root cannot be defined as a single valued function. For example, there is trouble whenever some other point in the basin maps exactly onto the superattractive point, or whenever the basin is not simply-connected.

We conclude with a problem.

Problem 6-1. What maps to what in Figure 5?

§7. Parabolic Fixed Points: the Leau-Fatou Flower.

Again we consider functions $f(z) = \lambda z + a_2 z^2 + a_3 z^3 + \cdots$ which are defined and holomorphic in some neighborhood of the origin, but in this section we suppose that the multiplier λ at the fixed point is a root of unity, $\lambda^q = 1$. Such a fixed point is said to be *parabolic*, provided that $f^{\circ q}$ is not the identity map. (More generally, any periodic orbit with λ a root of unity is called parabolic, provided that no iterate of f is the identity map.) First consider the special case $\lambda = 1$. It will be convenient to write our map as

$$f(z) = z + az^{n+1} + \text{(higher terms)}, \tag{7.1}$$

with $a \neq 0$. The integer $n+1 \geq 2$ is called the *multiplicity* of the fixed point. (By definition, the " simple" fixed points with $\lambda \neq 1$ have multiplicity equal to 1.) Choose a neighborhood N of the origin which is small enough so that f maps N diffeomorphically onto some neighborhood N' of the origin.

Definition. A connected open set U, with compact closure $\bar{U} \subset N \cap N'$, will be called an *attracting petal* for f at the origin if

$$f(\bar{U}) \subset U \cup \{0\} \qquad \text{and} \qquad \bigcap_{k \geq 0} f^{\circ k}(\bar{U}) = \{0\}.$$

Similarly, $U' \subset N \cap N'$ is a *repelling petal* for f if U' is an attracting petal for f^{-1}.

7.2. Leau-Fatou Flower Theorem. *If the origin is a fixed point of multiplicity $n+1 \geq 2$, then there exist n disjoint attracting petals U_i and n disjoint repelling petals U'_i so that the union of these $2n$ petals, together with the origin itself, forms a neighborhood N_0 of the origin. These petals alternate with each other, as illustrated in Figure 6, so that each U_i intersects only U'_i and U'_{i-1} (where U'_0 is to be identified with U'_n).*

If U_i is an attracting petal, then evidently the sequence of maps $f^{\circ k}$ restricted to $\bar{U}_i$ converges uniformly to zero. On the other hand, if U'_i is a repelling petal, then every orbit $z_0 \mapsto z_1 \mapsto \cdots$ which starts out in U'_i must eventually leave U'_i, and in fact must leave the union $U'_1 \cup \cdots \cup U'_n$. (However it may later return, perhaps even infinitely often.) Here are three immediate consequences of 7.2.

7.3. Corollary. *There is no periodic orbit, other than the fixed point at the origin, which is completely contained within the neighborhood N_0.*

Now suppose that f is a globally defined rational function. We continue to assume that the origin is a fixed point with $\lambda = 1$. Each attracting petal U_i determines a parabolic *basin of attraction* Ω_i, consists of all z_0 for which the orbit $z_0 \mapsto z_1 \mapsto \cdots$ eventually lands in the attracting petal U_i, and hence converges to the fixed point through U_i. Evidently these basins $\Omega_1, \ldots, \Omega_n$ are disjoint open sets.

7.4 Corollary. *If we exclude the case of an orbit which exactly hits the fixed point, then an orbit $z_0 \mapsto z_1 \mapsto \cdots$ under f converges to the fixed point if and only if it eventually lands in one of the attracting petals U_i, and hence belongs to the associated basin Ω_i.*

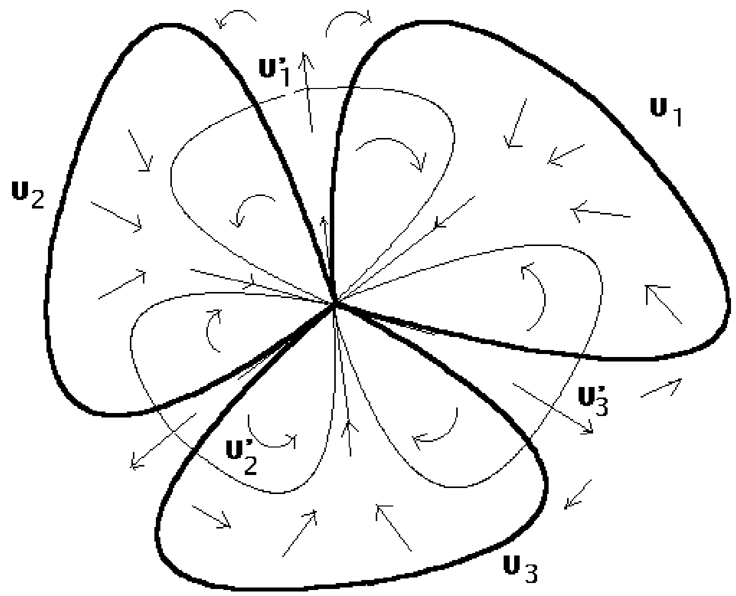


Figure 6. Leau-Fatou Flower with three attracting petals U_i and three repelling petals U_i' .

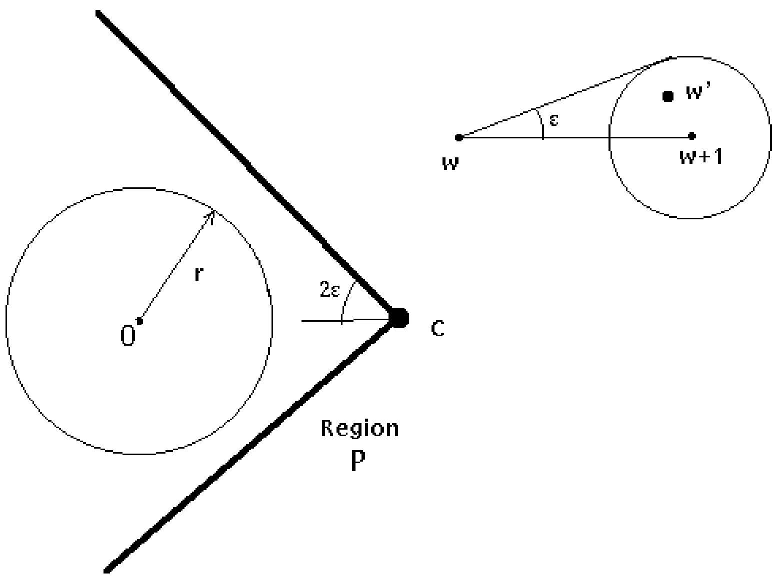


Figure 7.

7.5. Corollary. *Each parabolic basin Ω_i is contained in the Fatou set $\hat{\mathbf{C}} \smallsetminus J(f)$, but each basin boundary $\partial\Omega_i$ is contained in the Julia set $J(f)$. It follows that each repelling petal U_i' must intersect $J(f)$.*

In particular, it is claimed that the parabolic fixed point $z = 0$ must belong to $J(f)$.

Proof of 7.5. We first show that $0 \in J(f)$. It follows from 7.1 that

$$f^{\circ k}(z) \;=\; z + kaz^{n+1} + (\text{higher terms}) .$$

Evidently no sequence of iterates $f^{\circ k}$ can converge uniformly in a neighborhood of the origin, since the corresponding $(n+1)$-st derivatives do not converge. (Compare 1.3.) Thus $0 \in J(f)$, and it follows that every point in the grand orbit of zero belongs to $J(f)$. If $z_1 \in \partial\Omega_i$ is not in the grand orbit of zero, then by 7.4 we can extract a subsequence from the orbit of z_1 which remains bounded away from zero. Since the sequence of iterates $f^{\circ k}$ converges to zero throughout the open set Ω_i , it follows that $\{f^{\circ k}\}$ can not be normal in any neighborhood of the boundary point z_1 . The proof is now straightforward. □

Proof of Theorem 7.2. We will say that a vector $v \in \mathbf{C}$ points in an *attracting direction* at the fixed point of 7.1 if the product av^n is real and negative. If we ignore higher order terms, then these are just the directions for which the vector from v to $f(v) \approx v(1 + av^n)$ points straight in towards the origin. Similarly, v points in a *repelling direction* if av^n is real and positive. Evidently there are n equally spaced attracting directions which are separated by the n equally spaced repelling directions.

We will make use of the substitution $w = b/z^n$ with inverse $z = \sqrt[n]{b/w}$, where $b = -1/(na)$. Evidently the sector between two repelling directions in the z-plane will correspond under this transformation to the entire w-plane, slit along the negative real axis. In particular, a neighborhood of zero in such a sector will correspond to a neighborhood of infinity in such a slit w-plane. Let us write the transformation 7.1 as

$$z \mapsto f(z) = z(1 + az^n + o|z^n|)$$

as $|z| \to 0$. Here $o|z^n|$ stands for a remainder term whose ratio to $|z^n|$ tends to zero. Substituting $z = (b/w)^{1/n}$, the corresponding self-transformation in the w-plane is

$$w \mapsto w' \;=\; b/f(z)^n \;=\; (b/z^n)(1 + az^n + o|z^n|)^{-n} \;=\; w(1 - naz^n + o|z^n|) .$$

But $z^n = b/w$ and $nab = -1$, so this can be written simply as

$$w' \;=\; w(1 + w^{-1} + o|w^{-1}|) \;=\; w + 1 + o(1)$$

as $|w| \to \infty$. In other words, given any small number, which it will be convenient to write as $\sin\epsilon > 0$, we can choose a radius r so that

$$|w' - w - 1| < \sin\epsilon \quad \text{for} \quad |w| > r .$$

It follows that the slope of the vector from w to w' satisfies $|\text{slope}| < \tan\epsilon$, as long as $|w| > r$. Now we can construct an "attracting petal for the point at infinity" in the w-plane as follows. Let P consist of all $w = u + iv$ with $|w| > r$, and with $u > c - |v|/\tan 2\epsilon$, where the constant c is large enough so that all points $w \in P$ satisfy $|w| > r$. (Figure 7.) Then an easy geometric argument shows that the closure $\bar{P}$ is mapped into P, and that every backward orbit starting in $\bar{P}$ must eventually leave $\bar{P}$. Translating these statements back to the z-plane, the proof can easily be completed. □

Now suppose that the multiplier λ is a q-th root of unity, say $\lambda = \exp(2\pi i p/q)$ where p/q is a fraction in lowest terms. Then we can apply the discussion above to the q-fold iterate $f^{\circ q}$.

7.6. Lemma. *If the multiplier λ at a fixed point $f(z_0) = z_0$ is a primitive q-th root of unity, then the number n of attractive petals around z_0 must be a multiple of q. In other words, the multiplicity $n+1$ of z_0 as a fixed point of $f^{\circ q}$ must be congruent to 1 modulo q.*

Intuitively, if we perturb f so as to change λ slightly, then the multiple fixed point of $f^{\circ q}$ will split up into one point which is still fixed by f together with some finite collection of orbits which have period q under f. This Lemma can be proved geometrically by showing that multiplication by $\lambda = f'(z_0)$ must permute the n attractive directions at z_0. It can be proved by a formal power series computation based on the observation that $f \circ f^{\circ q} = f^{\circ q} \circ f$. Details will be left to the reader. □

As an example, Figure 8 shows part of the Julia set for the polynomial $z \mapsto z^2 + \lambda z$ where λ is a seventh root of unity, $\lambda = e^{2\pi i t}$ with $t = 3/7$. There are seven attractive petals about the origin.

We can further describe the geometry around a parabolic fixed point as follows. As in 7.1, we consider a local analytic map with a fixed point of multiplier $\lambda = 1$. Let U be either one of the n attracting petals or one of the n repelling petals, as described in the Flower Theorem, §7.2. Form an identification space U/f from U by identifying z with $f(z)$ whenever both z and $f(z)$ belong to U. (This means that z is identified with $f(z)$ for every $z \in U$ in the case of an attracting petal, and for every $z \in U \cap f^{-1}(U)$ in the case of a repelling petal.) By definition, a holomorphic map $\alpha : U \to \mathbf{C}$ is *univalent* if distinct points of U correspond to distinct points of $\mathbf{C}$. The following was proved by Leau and Fatou.

7.7. Theorem. *The quotient manifold U/f is conformally isomorphic to the infinite cylinder $\mathbf{C}/\mathbf{Z}$. Hence there is one, and up to composition with a translation only one, univalent embedding α from U into the universal covering space $\mathbf{C}$ which satisfies the Abel functional equation*

$$\alpha(f(z)) = 1 + \alpha(z)$$

for all $z \in U \cap f^{-1}(U)$. With suitable choice of U, the image $\alpha(U) \subset \mathbf{C}$ will contain some right half-plane $\{w : \mathcal{R}(w) > c\}$ in the case of an attracting petal, or some left half-plane in the case of a repelling petal.

By definition, the quotient U/f is called an *Écalle cylinder* for U. (This term is due to Douady, suggested by the work of Écalle on holomorphic maps tangent to the identity.)

The proof of 7.7 begins as follows. To fix our ideas, we consider only the case of an attracting petal. As in the proof of 7.2, a substitution of the form $w = b/z^n$ will conjugate

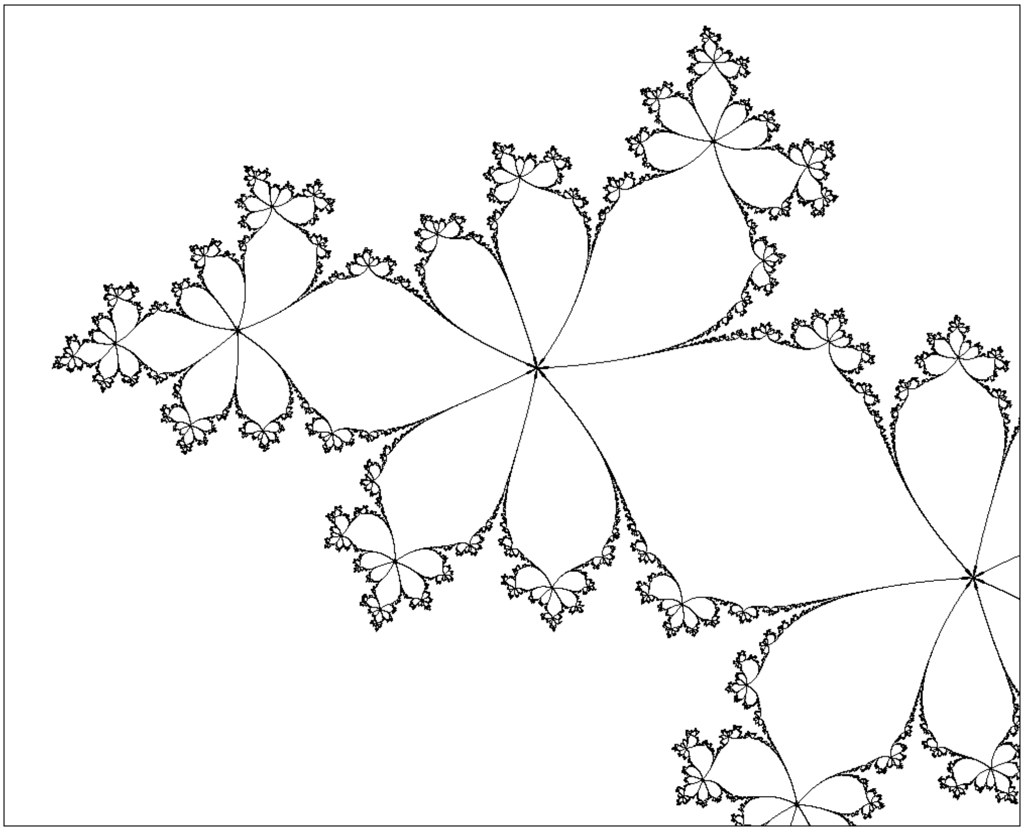


Figure 8. Julia set for $z \mapsto z^2 + e^{2\pi i t} z$ with $t = 3/7$.

the map f of §7.1 to a map which has the form

$$g(w) \;=\; w + 1 + a_1 w^{-1/n} + a_2 w^{-2/n} + \cdots \,.$$

Here w ranges over a neighborhood of infinity, with the negative real axis removed.

Definition. Let $\mathcal{G}$ be the group consisting of all holomorphic maps which are defined and univalent in some region of the form

$$\{\, u + iv \in \mathbf{C} \;:\; u > c_1 - c_2|v| \,\}\,, \tag{1}$$

and which are asymptotic to the identity map as $|u + iv| \to \infty$. Evidently our map $w \mapsto g(w)$ belongs to this group $\mathcal{G}$. Our object is to show that g is conjugate to the translation $w \mapsto w + 1$ within $\mathcal{G}$. More generally, we will prove the following.

7.8. Lemma. *If a transformation $g_0 \in \mathcal{G}$ has the form $g_0(w) = w + 1 + o(1)$ as $|w| \to \infty$, then g_0 is conjugate within $\mathcal{G}$ to the translation $w \mapsto w + 1$.*

Proof. We assume that g_0 can be written as $g_0(w) = w + 1 + \eta_0(w)$, where $\eta_0(w) \to 0$ as $|w| \to \infty$ within some region of the form (1). We will first make two preliminary transformations to improve the error bound. Let

$$F_0(w) = \int (1 + \eta_0(w))^{-1} dw$$

be any indefinite integral of $1/(1+\eta_0)$ within this region. Then it is not difficult to check that $F_0 \in \mathcal{G}$. Using the Schwarz Lemma (§1.3), note that $|\eta_0'(w)| = o(1/|w|)$ and hence

$$F_0''(w) = o(1/|w|),$$

within a smaller region of the same form. By Taylor's Theorem we have

$$\begin{aligned} F_0 \circ g_0(w) &= F_0(w + (1 + \eta_0(w))) = F_0(w) + F_0'(w)(1 + \eta_0(w)) + o(1/|w|) \\ &= F_0(w) + 1 + o(1/|w|) . \end{aligned}$$

In other words, setting $g_1 = F_0 \circ g_0 \circ F_0^{-1}$ we have $g_1(w) = w + 1 + o(1/|w|)$. Now repeating exactly this same construction, we see that g_1 is conjugate to a map $g_2 = F_1 \circ g_1 \circ F_1^{-1}$ within $\mathcal{G}$, where $g_2(w) = w + 1 + o(1/|w|^2)$ within some smaller region of the same form. In particular,

$$|g_2(w) - w - 1| \leq 1/|w|^2 \tag{2}$$

provided that $|w|$ is sufficiently large.

Starting with any w_0 in this region, consider the orbit $w_n = g_2^{\circ n}(w_0)$. We will prove that the differences $\{w_n - n\}$ form a Cauchy sequence which converges locally uniformly, so that the limit

$$w_0 \mapsto \phi(w_0) = \lim_{n \to \infty} (w_n - n)$$

defines a transformation ϕ which belongs to the group $\mathcal{G}$. Since $\phi \circ g_2(w)$ is evidently equal to $\phi(w) + 1$, this will prove Lemma 7.8, and hence prove 7.7.

As a preliminary remark, using the weaker inequality $g_2(w) = w + 1 + o(1)$, we see that for any $\epsilon > 0$ we have $|w_{n+1} - w_n - 1| < \epsilon$ for $|w_0|$ sufficiently large, and hence $|w_n - w_0 - n| < n\epsilon$. In particular, it is not difficult to check that $|w_n| \geq |w_0 + n|/2$, and hence

$$|w_{n+1} - w_n - 1| \leq 1/|w_n|^2 \leq 4/|w_0 + n|^2 ,$$

provided that $|w|$ is large. For $m > n \geq 0$, this implies that

$$|(w_m - m) - (w_n - n)| < \sum_{n \leq j < \infty} 4/|w_0 + j|^2 \approx 4 \int_n^\infty dj/|w_0 + j|^2 .$$

Setting $w_0 + n = re^{i\theta}$ with $|\theta| < \pi$, this integral can be evaluated as $\theta/(r \sin\theta) \leq c/r$, for some constant c depending on the region. This tends to zero as $r = |w_0 + n| \to \infty$, hence the $w_n - n$ form a Cauchy sequence. Further details will be left to the reader. □

Remark. Note that this preferred Fatou coordinate system is defined only within one of the $2n$ attracting or repelling petals. In order to described a full neighborhood of the parabolic fixed point, we would have to describe how these $2n$ Fatou coordinate systems are to be pasted together in pairs by means of univalent mappings. In fact each of the $2n$ required pasting maps has the form $w \mapsto w + \Upsilon(e^{\pm 2\pi i w})$, where Υ is defined and holomorphic in some neighborhood of the origin, and where the signs $\pm$ alternate. By studying this construction, one sees that there can be no normal form depending on only finitely many parameters for a general holomorphic map f in the neighborhood of a parabolic fixed point.

Now suppose that $f : \hat{\mathbf{C}} \to \hat{\mathbf{C}}$ is a globally defined rational map. Although attracting petals behave much like repelling petals in the local theory, they behave quite differently in the large.

7.9. Corollary. *If U is an attracting petal, then the Fatou map*

$$\alpha : U \to \mathbf{C}$$

extends uniquely to a map which is defined and holomorphic throughout the attractive basin Ω of U, still satisfying the Abel equation $\alpha(f(z)) = 1 + \alpha(z)$.

This extended map $\Omega \to \mathbf{C}$ is surjective. However, it is no longer univalent, but rather has critical points whenever some iterate $f \circ \cdots \circ f$ has a critical point. In fact, we have the following.

7.10. Corollary. *For each attracting petal U_i, the corresponding immediate basin $\Omega_i \supset U_i$ contains at least one critical point of f. Furthermore, there exists a unique preferred petal U_i^* for this basin which maps precisely onto a right half-plane under α, and which has at least one critical point on its boundary.*

The proofs are completely analogous to the corresponding proofs in §6.4 and §6.6. Thus α^{-1} can be defined on some right half-plane, and if we try to extend leftwards by analytic continuation then we must run into some obstruction, which can only be a critical point of f. (For an alternative proof that every attracting basin contains a critical point, see Milnor & Thurston, pp. 512-515.) □

As an example, Figure 9 illustrates the map $f(z) = z^2 + z$, with a parabolic fixed point of multiplier $\lambda = 1$ at $z = 0$, which is the cusp point at the right center of the picture. Here the Julia set J is the outer Jordan curve (the "cauliflower") bounding the basin of attraction Ω. The critical point $\omega = -1/2$ lies exactly at the center of the basin, and all orbits in this basin converge towards $z = 0$ to the right. The curves $\mathcal{R}(\alpha(z)) =$ constant $\in \mathbf{Z}$ have been drawn in. Thus the preferred petal U^*, with the critical point on ∂U^*, is bounded by the right half of the central figure ∞ shaped curve.

For a repelling petal, the corresponding statement is the following.

7.11. Corollary. *If U' is a repelling petal, then the inverse map*

$$\alpha^{-1} : \alpha(U') \to U'$$

extends uniquely to a globally defined holomorphic map $\beta : \mathbf{C} \to \hat{\mathbf{C}}$ which satisfies the corresponding equation $f(\beta(w)) = \beta(1 + w)$. The image $\beta(\mathbf{C})$ is equal to the finite plane $\mathbf{C}$ if f is a polynomial map, and is the entire sphere $\hat{\mathbf{C}}$ if f is not conjugate to any polynomial.

Again the proof is easily supplied. (Compare 6.5, together with 3.7 and Problem 3-3.) □

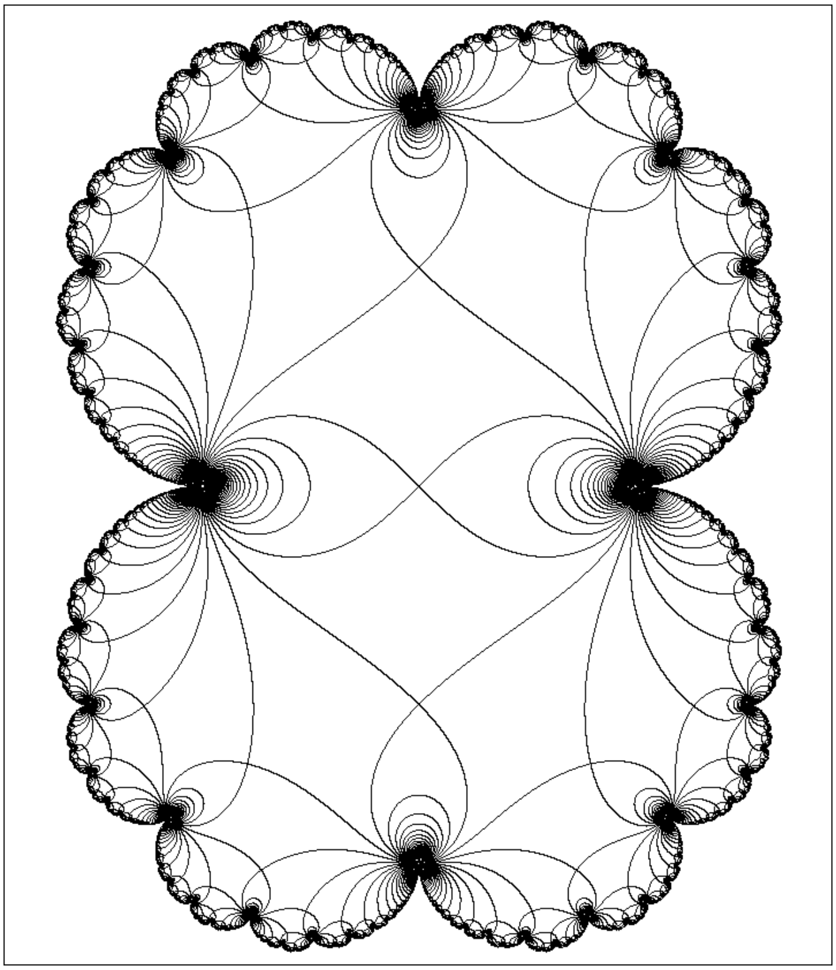


Figure 9. Julia set for $z \mapsto z^2 + z$, with the curves $\alpha(z) \in \mathbf{Z} + i\mathbf{R}$ drawn in.

Problem 7-1. If z_0 belongs to one of the basins of attraction Ω_i of Corollary 7.4, with orbit $z_0 \mapsto z_1 \mapsto z_2 \mapsto \cdots$, show that $\lim_{k\to\infty} z_k/|z_k|$ exists and is a unit vector which points in one of the n attracting directions.

Problem 7-2. Define two attracting petals U and V for f to be *equivalent* if every orbit for one intersects the other. Show that the petals which occur in 7.2 are unique up to equivalence. Show however that a petal as defined at the very beginning of §7 may be too small, so that it cannot occur in 7.2, and so that the quotient U/f is not a full cylinder.

§8. Cremer Points and Siegel Disks.

Once more we consider holomorphic maps of the form

$$f(z) = \lambda z + a_2 z^2 + a_3 z^3 + \cdots ,$$

defined throughout some neighborhood of the origin. In §6 we supposed that $|\lambda| \ne 1$, while in §7 we took λ to be a root of unity. This section considers the remaining cases where $|\lambda| = 1$ but λ is not a root of unity. Thus we assume that the multiplier λ can be written as

$$\lambda = e^{2\pi i \xi} \qquad \text{with} \quad \xi \quad \text{real and irrational}.$$

Briefly, we will say that the origin is an *irrationally indifferent* fixed point. The number $\xi \in \mathbf{R}/\mathbf{Z}$ may be descibed as the *angle of rotation* in the tangent space at the fixed point.

The central question here is whether or not there exists a local change of coordinate $z = h(w)$ which conjugates f to the irrational rotation $w \mapsto \lambda w$, so that

$$f(h(w)) = h(\lambda w)$$

near the origin. (Compare §6.) This is the so called "center problem". If such a linearization is possible, then a small disk $|w| < \epsilon$ in the w-plane corresponds to an open set U in the z-plane which is mapped bijectively onto itself by f. Evidently such a neighborhood U contains no periodic points of f other than the fixed point at zero. If f is a rational function, note that U is contained in its Fatou set $\hat{\mathbf{C}} \smallsetminus J$.

This section will first survey what is known about this problem, and then prove some of the easier results.

At the International Congress in 1912, E. Kasner conjectured that such a linearization is *always* possible. Five years later, G. A. Pfeiffer disproved this conjecture by giving a rather complicated description of certain holomorphic functions for which no linearization is possible. In 1919 Julia claimed to settle the question completely for rational functions of degree two or more by showing that such a linearization is *never* possible. His proof was incorrect. H. Cremer put the situation in much clearer perspective in 1927 with a beautiful note which proved the following.

Definition. It will be convenient to say that a property of a unit complex number is true for *generic* $\lambda \in S^1$ if the set of λ for which it is true contains a countable intersection of dense open subsets of the circle. According to Baire, such a countable intersection of dense open sets is necessarily dense and uncountably infinite.

Cremer Non-linearization Theorem. *For a generic choice of λ on the unit circle, the following is true. If z_0 is a fixed point of multiplier λ for a completely arbitrary rational function of degree two or more, then z_0 is the limit of an infinite sequence of periodic points. Hence there is no linearizing coordinate in a neighborhood of z_0.*

(See 8.5 below.) The question as to whether this statement is actually true for *all* numbers λ on the unit circle remained open until 1942, when Siegel proved the following. (Compare 8.4 and 8.6.)

Siegel Linearization Theorem. *For almost every λ on the unit circle (that is for every λ outside of a set with one-dimensional Lebesgue measure equal to zero) any germ of a holomorphic function with a fixed point of multiplier λ can be linearized by a local holomorphic change of coordinate.*

Remark. Thus there is a total contrast between behavior for generic λ and behavior for almost every λ. This contrast is quite startling, but is not uncommon in dynamics. Here is quite different and equally remarkable example. Consider the exponential function $\exp : \mathbf{C} \to \mathbf{C}$ as a dynamical system. *For a generic choice of $z \in \mathbf{C}$, the orbit of z is everywhere dense in $\mathbf{C}$. On the other hand, for almost every $z \in \mathbf{C}$ the set of all accumulation points for the orbit of z, consists only of those points*

$$0\,,\ 1\,,\ e\,,\ e^e\,,\ \ldots$$

which belong to the orbit of zero. (See Rees [R1], Lyubich [L2]. It is amusing to test this statement on a computer: From a random start, the iterated exponential usually either gets to zero to within computer accuracy, or else gets too big to compute, within five to ten iterations.) In applied dynamics, it usually understood that behavior which occurs for a set of parameter values of measure zero has no importance, and can be ignored. However, even in applied dynamics the study of generic behavior remains an extremely valuable tool.

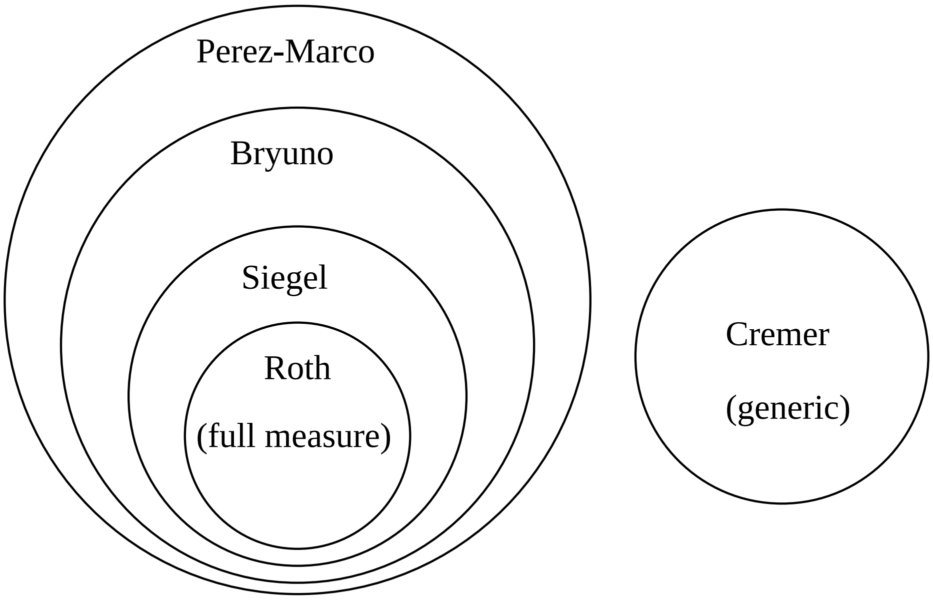
Definition. We will say that an irrationally indifferent fixed point is a *Siegel point* or a *Cremer point* according as a local linearization is possible or not. (In the classical literature, Siegel points are called "centers".)

8.1. Lemma. *An irrationally indifferent fixed point of a rational function is either a Cremer point or a Siegel point according as it belongs to the Julia set or not. In the case of a Siegel point z_0, the entire connected component U of the Fatou set $\hat{\mathbf{C}} \smallsetminus J$ which contains z_0 is conformally isomorphic to the open unit disk in such a way that the map f from U onto itself corresponds to the irrational rotation $w \mapsto \lambda w$ of the unit disk.*

By definition, such a component U is called a *Siegel disk*, or a *rotation disk*.

Proof of 8.1. If z_0 is a Siegel point, then the iterates of f in a neighborhood correspond to iterated rotations of a small disk, and hence form a normal family. Thus z_0 belongs to the Fatou set. Conversely, whenever z_0 belongs to the Fatou set, we see easily from Theorem 4.3 that z_0 must be a Siegel point. □

Both Cremer and Siegel proved theorems which are much sharper than the rough versions stated above. In order to state these precise results, and their more recent generalizations, it is convenient to introduce a number of different classes of irrational numbers, which are related to each other as indicated in the following schematic diagram.



Roughly speaking, the Cremer numbers are those which can be approximated extremely closely by rational numbers, while the Roth numbers are those which can only be approximated badly by rationals.

To give more precise definitions, given some fixed real number $\kappa \ge 2$ let us say that an irrational angle ξ satisfies a *Diophantine condition* of order κ if there exists some $\epsilon = \epsilon(\xi) > 0$ so that

$$\left|\xi - \frac{p}{q}\right| > \frac{\epsilon}{q^\kappa}$$

for every rational number p/q . Setting $\lambda = e^{2\pi i \xi}$ as above, since

$$|\lambda^q - 1| = |e^{2\pi i(q\xi - p)} - 1| \sim 2\pi q|\xi - p/q|$$

as $(q\xi - p) \to 0$, this is equivalent to the requirement that

$$|\lambda^q - 1| > \epsilon'/q^{\kappa-1}$$

for some $\epsilon' > 0$ which depends on λ , and for all positive integers q . Let $D_\kappa \subset \mathbf{R} \smallsetminus \mathbf{Q}$ be the set of all numbers ξ which satisfy such a condition. Note that $D_\kappa \subset D_\eta$ whenever $\kappa < \eta$. We define the set Si of *Siegel numbers* (also called "*Diophantine numbers*") to be the union of the D_κ . We can now make the following more precise statement.

Theorem of Siegel. *If the angle ξ belongs to this union $\mathrm{Si} = \bigcup D_\kappa$, then any holomorphic germ with multiplier $\lambda = e^{2\pi i \xi}$ is locally linearizable.*

Proofs may be found in Siegel, or Siegel and Moser, or Carleson.

A classical theorem of Liouville asserts that every algebraic number of degree d belongs to the class D_d . (Compare Problem 8-1.) Hence every irrational number outside of the class Si must be transcendental. Such numbers in the complement of Si are often called *Liouville numbers*.

Define the set of *Roth numbers* to be the intersection

$$\mathrm{Ro} = \bigcap_{\kappa > 2} D_\kappa \ .$$

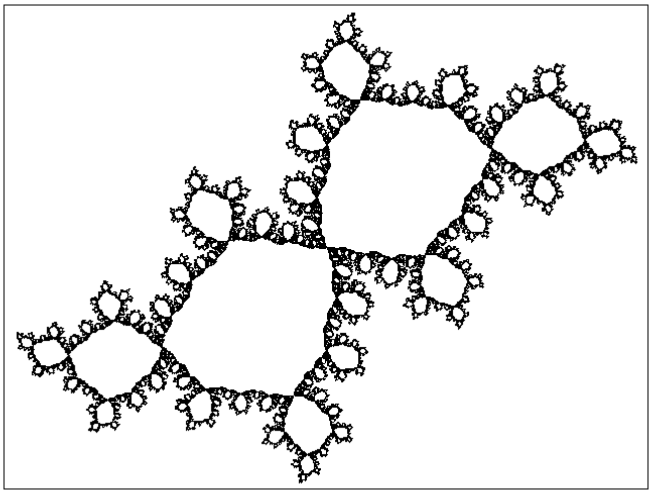


Figure 10a. Julia set for $z^2 + e^{2\pi i\xi} z$ with $\xi = \sqrt[3]{1/4} = .62996\cdots$.

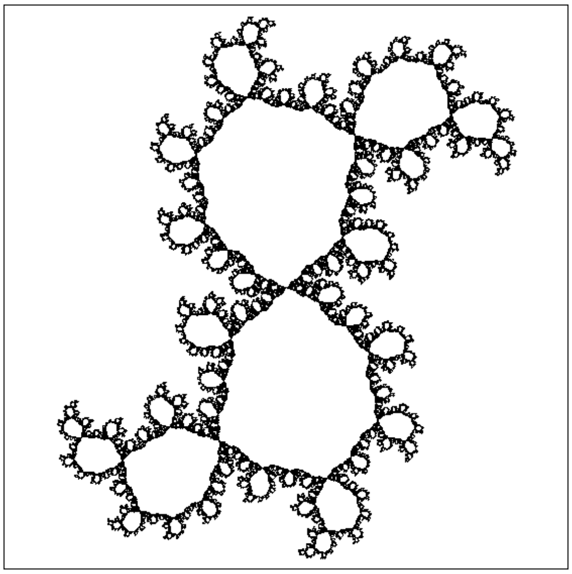


Figure 10b. Corresponding Julia set with a randomly chosen angle $\xi = .78705954039469$.

Roth, in 1955, proved the much sharper result that every algebraic number belongs to this intersection Ro . It is quite easy to check that: *Almost every real number belongs to* Ro . (See Problem 8-2.)

Thus if ξ *is a completely arbitrary irrational algebraic number, then any rational map, such as* $f(z) = z^2 + e^{2\pi i\xi} z$ *, which has a fixed point of multiplier* $e^{2\pi i\xi}$ *must have a Siegel disk.* Similarly, if ξ is a randomly chosen real number, then the same will be true with probability one. Examples illustrating both cases are shown in Figure 10.

For a more precise analysis of the approximation of an irrational number $\xi \in (0\,,\,1)$ by rationals, it is useful to consider the continued fraction expansion

$$\xi \;=\; \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cfrac{1}{a_3 + \cdots}}}$$

where the a_i are uniquely defined strictly positive integers. The rational numbers

$$\frac{p_n}{q_n} \;=\; \cfrac{1}{a_1 + \cfrac{1}{a_2 + \ddots + \cfrac{1}{a_{n-1}}}}$$

are called the *convergents* to ξ. The denominators q_n will play a particularly important role. These denominators always grow at least exponentially with n. In fact

$$q_{n+1} \;>\; q_n \;>\; ((\sqrt{5}+1)/2)^{n-2} \;>\; 1$$

for $n > 2$. We will need two basic facts, which are proved in Appendix C. Each p_n/q_n is the best approximation to ξ by rational numbers with denominator at most q_n. In fact, setting $\lambda = e^{2\pi i\xi}$ as usual, we have the following.

8.2. Assertion. $|\lambda^k - 1| \;>\; |\lambda^{q_n} - 1|$ *for* $k = 1\,,\,2\,,\,\ldots\,,\;q_n - 1$.

Furthermore, the error $|\lambda^{q_n} - 1|$ has the order of magnitude of $1/q_{n+1}$. That is:

8.3. Assertion. *There are constants* $0 < c_1 < c_2 < \infty$ *so that*

$$\frac{c_1}{q_{n+1}} \;\le\; |\lambda^{q_n} - 1| \;\le\; \frac{c_2}{q_{n+1}} \qquad \textit{in all cases.}$$

For example we can take $c_1 = 2$ and $c_2 = 2\pi$. Using these two facts, we can write the Roth condition as

Ro : $$\lim_{n\to\infty} \frac{\log q_{n+1}}{\log q_n} \;=\; 1\,,$$

and the Siegel condition as

Si : $$\sup \frac{\log q_{n+1}}{\log q_n} \;<\; \infty\,.$$

With these same notations, we now introduce the weaker *Bryuno condition*

Br : $$\sum_n \frac{\log(q_{n+1})}{q_n} \;<\; \infty\,,$$

and also the much weaker *Perez-Marco condition*

PM : $$\sum_n \frac{\log\log(q_{n+1})}{q_n} \;<\; \infty\,.$$

It is not difficult to check that **Ro** $\Rightarrow$ **Si** $\Rightarrow$ **Br** $\Rightarrow$ **PM** . Bryuno, in 1972, proved an extremely sharp version of Siegel's Theorem.

8.4. Theorem of Bryuno. *If the angle* ξ *satisfies the condition that* $\sum \log(q_{n+1})/q_n < \infty$, *then any holomorphic germ of the form*

$$f(z) \;=\; e^{2\pi i\xi} z + a_2 z^2 + \cdots$$

can be linearized by a local holomorphic change of variable.

A proof will be outlined at the end of this section. Yoccoz, in 1987, showed that this result is best possible.

Theorem of Yoccoz. *Conversely, if* $\sum \log(q_{n+1})/q_n = \infty$, *then the quadratic map*

$$f(z) \;=\; z^2 + e^{2\pi i\xi} z$$

has the property that every neighborhood of the origin contains infinitely many periodic orbits. Hence the origin is a Cremer point.

Briefly, we will sat that the fixed point can be approximated by "small cycles". Perez-Marco, in 1990, completely characterized the multipliers for which such small cycles must appear.

Theorem of Perez-Marco. *If* ξ *satisfies the condition that*

$$\sum \log\log(q_{n+1})/q_n < \infty \;,$$

then any non-linearizable germ with multiplier $e^{2\pi i\xi}$ *contains infinitely many periodic orbits in every neighborhood of the fixed point. However, whenever* $\sum \log\log(q_{n+1})/q_n \;=\; \infty$ *there exists a non-linearizable germ which has no periodic orbit other than the fixed point itself within some neighborhood of the fixed point.*

(In the special case of a rational function, it is not known whether every Cremer point can necessarily be approximated by small cycles. Compare 8.5. below.)

The rest of this section will provide a few proofs. We first prove a slightly sharper form of Cremer's Theorem, and then a rather weak form of Siegel's Theorem. Finally, we give a very rough outline proof for Bryuno's Theorem.

We begin with Cremer's Theorem. Let us say that an irrational angle ξ satisfies a "*Cremer condition*" of degree d if the associated $\lambda = e^{2\pi i\xi}$ satisfies

$$\mathbf{Cr}_d : \qquad \limsup_{q\to\infty} \frac{\log\log(1/|\lambda^q - 1|)}{q} \;>\; \log d \,.$$

Thus the error $|\lambda^q - 1|$ must tend to zero extremely rapidly for suitable large q . This is equivalent to the hypothesis that $\limsup_{q\to\infty} q^{-1}\log\log(1/|\xi - p/q|) \;>\; \log d$, or to the hypothesis that $\limsup\, (\log\log q_{n+1})/q_n \;>\; \log d$. *It is not difficult to show that a generic real number satisfies this condition* $\mathbf{Cr}_d$ *for every degree* d . (See Problem 8-3.)

8.5. Theorem. *If ξ satisfies $\mathbf{Cr}_d$ with $d \geq 2$, then for a completely arbitrary rational function of degree d , any neighborhood of a fixed point of multiplier $\lambda = e^{2\pi i\xi}$ must contain infinitely many periodic orbits. Hence no local linearization is possible.*

In particular, for a generic choice of ξ this statement will be true for non-linear rational functions of arbitrary degree.

The proof which follows is nearly all due to Cremer. However, Cremer used the slightly weaker hypothesis that $\liminf |\lambda^q|^{1/d^q} = 0$ and concluded only that the fixed point is a limit of periodic points, rather than full periodic orbits.

Proof of 8.5. First consider a monic polynomial $f(z) = z^d + \cdots + \lambda z$ of degree $d \geq 2$ with a fixed point of multiplier λ at the origin. Then $f^{\circ q}(z) = z^{d^q} + \cdots + \lambda^q z$, so the fixed points of $f^{\circ q}$ are the roots of the equation

$$z^{d^q} + \cdots + (\lambda^q - 1)z = 0 .$$

Therefore, the product of the $d^q - 1$ non-zero fixed points of $f^{\circ q}$ is equal to $\pm(\lambda^q - 1)$. If $|\lambda^q - 1| < 1$, then it follows that there exists at least one such fixed point z_q with

$$0 < |z_q| < |\lambda^q - 1|^{1/(d^q-1)} < |\lambda^q - 1|^{1/d^q} .$$

By hypothesis, for some $\epsilon > 0$, we can choose q arbitrarily large with

$$q^{-1} \log\log(1/|\lambda^q - 1|) > \log(d) + \epsilon ,$$

or in other words

$$|\lambda^q - 1|^{1/d^q} < \exp(-e^{\epsilon q}) .$$

This tends to zero as $q \to \infty$, so we certainly have periodic points $z_q \neq 0$ in every neighborhood of zero. By Taylor's Theorem, if $\delta > 0$ is sufficiently small, then $|f(z)| < e^{\epsilon}|z|$ whenever $|z| < \delta$. It follows that

$$|f^{\circ k}(z)| < \delta \quad \text{for} \quad 1 \leq k \leq q \quad \text{whenever} \quad |z| < e^{-\epsilon q}\delta .$$

Now note that there exist periodic points z_q which satisfy the inequality

$$|z_q| < \exp(-e^{\epsilon q}) < e^{-\epsilon q}\delta$$

for arbitrarily large values of q . It follows that the entire periodic orbit of such a point, with period at most q , is contained in the δ neighborhood of zero. Since δ can be arbitrarily small, this completes the proof of 8.5 in the polynomial case.

In order to extend this argument to the case of a rational function f , Cremer first notes that f must map at least one point $z_1 \neq 0$ to the fixed point $z = 0$. After conjugating by a Möbius transformation which carries z_1 to infinity, we may assume that $f(\infty) = f(0) = 0$. If we set $f(z) = P(z)/Q(z)$, this means that P is a polynomial of degree strictly less than d , Furthermore, after a scale change we may assume that $P(z) = (\text{higher terms}) + \lambda z$, and that $Q(z) = z^d + \cdots + 1$ is monic. A brief computation then shows that $f^{\circ q}(z) = P_q(z)/Q_q(z)$ where $P_q(z) = (\text{higher terms}) + \lambda^q z$ and where $Q_q(z)$ has the form $z^{d^q} + \cdots + 1$. Thus the equation for fixed points of $f^{\circ q}$ has the form

$$0 \;=\; zQ(z) - P(z) \;=\; z(z^{d^q} + \cdots + (1 - \lambda^q))\ .$$

The proof now proceeds just as in the polynomial case. □

For further information about Cremer points, see §11.5 and §18.6.

Let us next prove that Siegel disks really exist. We will describe a proof, due to Yoccoz, of the following special case of Siegel's Theorem. (Compare Herman [He2] or Douady [D2].)

8.6. Theorem. *For Lebesgue almost every angle $\xi \in \mathbf{R}/\mathbf{Z}$, taking $\lambda = e^{2\pi i \xi}$ as usual, the quadratic map $f_\lambda(z) = z^2 + \lambda z$ possesses a Siegel disk about the origin.*

Remark. Somewhat more precisely, we can define the *size* of a Siegel disk to be the largest number σ such that there exists a holomorphic embedding ψ of the disk of radius σ into the Fatou set $\hat{\mathbf{C}} \smallsetminus J(f_\lambda)$ so that $\psi'(0) = 1$, and so that $f_\lambda(\psi(w)) = \psi(\lambda w)$. If there is no Siegel disk, then we set $\sigma = 0$. Using a normal family argument, it is not difficult to show that this size σ is upper semicontinuous as a function of λ. (Compare the proof of 8.8 below.) *In other words, for any fixed $\epsilon > 0$ the set of $\lambda = e^{2\pi i \xi}$ with $\sigma(\lambda) \geq \epsilon$ is compact.* This set is totally disconnected, since it contains no roots of unity. As $\epsilon \to 0$, it grows larger, and the proof will show that its measure tends to the measure of the full unit circle.

The proof of 8.6 has three steps. The first two steps will be carried out here, while the third will be put off to Appendix A. Here is the first step. Consider the dynamics of f_λ for λ inside the open disk D. According to Kœnigs, for $\lambda \in D \smallsetminus \{0\}$, there exists a neighborhood U of zero and a holomorphic map $\phi_\lambda(z) = \lim_{k\to\infty} f_\lambda^{\circ k}(z)/\lambda^k$ which carries U diffeomorphically onto some disk D_ρ, so that f_λ on U corresponds to multiplication by λ on D_ρ, and so that ϕ_λ has derivative $+1$ at the origin.

8.7. Lemma. *We can choose the open set U so as to map diffeomorphically onto the disk D_ρ of radius $\rho = |\phi_\lambda(-\lambda/2)|$ under ϕ_λ. However no larger radius is possible. Furthermore, the correspondence*

$$\lambda \;\mapsto\; \eta(\lambda) = \phi_\lambda(-\lambda/2)$$

is bounded, holomorphic, and non-zero throughout the punctured disk $D \smallsetminus \{0\}$.

Proof. Note that $-\lambda/2$ is the unique critical point, that is the unique point at which the derivative f'_λ vanishes. Thus the first assertion of 8.7 follows immediately from §6.6. Furthermore, it follows from §6.2 that this correspondence $\lambda \mapsto \eta(\lambda)$ is holomorphic. To show that $\rho = |\eta|$ is bounded, note first that U must be contained in the disk D_2 of radius 2. For if $|z| > 2$ then an easy estimate shows that $|f_\lambda(z)| > |z|$, so the orbit of z cannot converge to zero. Thus ϕ^{-1} maps the disk D_ρ holomorphically onto $U \subset D_2$ with derivative 1 at the origin; so it follows from the Schwarz Lemma, §1.3, that $\rho \leq 2$.
□

Since this function η is bounded, it follows that η has a removable singularity at the origin; that is, it can be extended as a holomorphic function throughout the disk D. (See for example Ahlfors, 1966 p. 114, or 1973 p. 20.)

Now consider the *radial limit* of $\eta(r\exp(2\pi it))$ for fixed t, as $r \to 1$.

8.8. Lemma. *Suppose, for some fixed* $\lambda = e^{2\pi i\xi}$ *, that the quadratic map* f_λ *does not possess any Siegel disk. Then the radial limit*

$$\lim_{r\to 1} \eta(re^{2\pi i\xi})$$

must exist and be equal to zero.

Conversely, if the quantity $\rho = \limsup_{r\to 1} |\eta(re^{2\pi i\xi})|$ is strictly positive, the proof will show that $f_{\exp(2\pi i\xi)}$ admits a Siegel disk of "size" $\geq \rho$.

Remark. Yoccoz has shown that this estimate is best possible. That is, there exists a Siegel disk if and only if $\rho > 0$; and ρ is precisely the size of the maximal Siegel disk, as defined in the Remark following 8.6.

Proof of 8.8. If the $\limsup$ of $|\eta(r\exp(2\pi i\xi))|$ as $r \to 1$ is equal to $\rho_0 > 0$, then for any $\rho < \rho_0$ and for some sequence $\lambda_j \in D$ tending to $\lambda = \exp(2\pi i\xi) \in \bar{D}$, the inverse diffeomorphism $\phi_{\lambda_j}^{-1}$ mapping D_ρ into D_2 is well defined. By a normal family argument, we can choose a subsequence which converges, uniformly on compact sets, to a holomorphic limit $\psi : D_\rho \to \mathbf{C}$. It is easy to check that this limit ψ satisfies the required equation $\psi(\lambda w) = f_\lambda(\psi(w))$, and hence describes a Siegel disk. □

Finally, the third step in the proof of 8.6 is a classical theorem by F. and M. Riesz which asserts that such a radial limit cannot exist and be equal to zero for a set of ξ of positive Lebesgue measure. In other words the quantity

$$\limsup_{r\to 1} |\eta(re^{2\pi i\xi})|$$

must be strictly positive for almost every ξ. This theorem will be proved in Appendix A.3. Combining these three steps, we obtain a proof of the special case 8.6 of Siegel's Theorem. □

To conclude this section, here is a very rough outline of a proof of the Bryuno Theorem, due to Yoccoz. The proof is based on a "renormalization construction" due to Douady and Ghys. Consider first a map $f_1 : D \to \mathbf{C}$ which is univalent (that is, holomorphic and one-to-one) on the open unit disk D, with a fixed point of multiplier $\lambda_1 = e^{2\pi i\xi_1}$ at the origin. We introduce a new coordinate by setting $z = e^{2\pi iZ}$ where $Z = X + iY$ ranges over the half-plane $Y > 0$. Then f_1 corresponds to a map of the form

$$F_1(Z) \ = \ Z + \xi_1 + \sum_1^\infty a_n e^{2\pi inZ}$$

which is defined and univalent on this upper half-plane. This map F_1 commutes with the translation $T_1(Z) = Z+1$, and is approximately equal to the translation $T_{\xi_1}(Z) = Z+\xi_1$.

More precisely, we have

$$F_1(Z) \;=\; T_{\xi_1}(Z) + o(1)\,, \qquad \text{uniformly in} \quad X \quad \text{as} \quad Y \to \infty\,,\,.$$

In fact, the $e^{2\pi i n Z}$ terms decrease exponentially fast as $Y \to \infty$, so that if Y is bounded well away from zero then F_1 is extremely close to the translation $Z \mapsto Z + \xi_1$. In particular, we can choose some height h_1 so that F_1 moves points $Z = X + iY$ definitely to the right, and has derivative close to 1, throughout the half-plane $Y > h_1$.

Construct a new Riemann surface S_1' as follows. Take a vertical strip S_1 in the Z-plane which is bounded on the left by the vertical line $L = \{iY : h_1 \le Y < \infty\}$, on the right by its image $F_1(L)$, and from below by the straight line from ih_1 to $F_1(ih_1)$. Now glue the left edge to the right edge by F_1. The resulting Riemann surface S_1' is conformally isomorphic to the punctured unit disk. Hence it can be parametrized by a variable $w \in D \smallsetminus \{0\}$. It is convenient to fill in the puncture point, $w = 0$, which corresponds to the improper points $Z = X + i\infty$ in the Z-plane. We now introduce a holomorphic map f_2 from a neighborhood of zero in S_1' into S_1' as follows. Starting from any point Z in the strip S_1 which is not too close to the bottom, let us iterate the map F_1 until we reach some point $1 + Z'$ of the translated strip $1 + S_1$. The corrrespondence $Z \mapsto Z'$ on S_1 now yields the required holomorphic map f_2 from a neighborhood of zero in the quotient surface S_1' to S_1'. Note that w is asymptotic to some constant times $e^{2\pi i Z/\xi_1}$ as $Y \to \infty$. If $1 + Z' = F_1^{\circ a}(Z) \approx Z + a\xi_1$ as $Y \to \infty$, then $2\pi i Z'/\xi_1 \approx 2\pi i Z/\xi_1 + a - 1/\xi_1$. Setting $1/\xi_1 \equiv \xi_2 \pmod 1$ with $0 < \xi_2 < 1$, it follows that the corresponding map $f_2(w) = w'$ in the disk S_1' is asymptotic to

$$w \;\mapsto\; w\, e^{-2\pi i \xi_2}\,.$$

Thus this Douady-Ghys construction relates a map f_1 with rotation angle ξ_1 to a map f_2 with rotation angle $-\xi_2 \equiv -1/\xi_1 \pmod 1$.

*** to be continued ***

See §11.4 and §12 for closely related results. We conclude this section with some problems.

Problem 8-1 (Liouville). Let f be a polynomial of degree d with integer coeficients, and suppose that $f(\xi) = 0$ where ξ is irrational. If every other root of this equation has distance at least ϵ from ξ, and if $|f'(x)| < K$ throughout the ϵ neighborhood of ξ, show that

$$K|\xi - p/q| \;\ge\; |f(p/q)| \;\ge\; 1/q^d$$

for every rational number p/q in the ϵ neighborhood of ξ. Conclude that $\xi \in D_d$, and hence that all irrational numbers in the complement of $\mathrm{Si} = \bigcup D_d$ must be transcendental.

Problem 8-2. If $\kappa > 2$ and $\epsilon > 0$, show that the set $S(\kappa, \epsilon)$ of numbers $\xi \in [0,1]$ which satisfy $|\xi - p/q| \le \epsilon/q^\kappa$ for some rational number p/q has measure less than or equal to $\epsilon \sum q/q^\kappa < \infty$. Since this tends to zero as $\epsilon \to 0$, conclude that almost every real number belongs to D_κ, and hence that almost every real number belongs to the Roth set $\mathrm{Ro} = \bigcap_{\kappa>2} D_\kappa$. (On the other hand, the subset D_2 has measure zero, and D_κ is vacuous for $\kappa < 2$. Compare Hardy and Wright. Numbers in D_2 are said to be "of constant type".)

Problem 8-3 (Cremer). Given a completely arbitrary function $q \mapsto \eta(q) > 0$, show that the set S_η, consisting of all irrational numbers ξ such that

$$|\xi - \frac{p}{q}| < \eta(q) \qquad \text{for infinitely many rational numbers } p/q,$$

is a countable intersection of dense open subsets of $\mathbf{R}$. As an example, taking $\phi(q) = 2^{-q!}$ conclude that a generic real number belongs to the set S_ϕ, which is contained in the Cremer class Cr_∞.

Problem 8-4 (Cremer 1938). If $f(z) = \lambda z + a_2 z^2 + a_3 z^3 + \cdots$, where λ is not zero and not a root of unity, show that there is one and only one formal power series of the form $h(z) = z + h_2 z^2 + h_3 z^3 + \cdots$ which formally satisfies the condition that $h(\lambda z) = f(h(z))$. In fact

$$h_n = \frac{a_n + X_n}{\lambda^n - \lambda}$$

for $n \ge 2$, where $X_n = X(a_2, \ldots, a_{n-1}, h_2, \ldots, h_{n-1})$ is a certain polynomial expression whose value can be computed inductively. Now suppose that we choose the a_n inductively, always equal to zero or one, so that $|a_n + X_n| \ge 1/2$. If

$$\liminf_{q\to\infty} |\lambda^q - 1|^{1/q} = 0,$$

show that the uniquely defined power series $h(z)$ has radius of convergence zero. Conclude that $f(z)$ is a holomorphic germ which is not locally linearizable. Choosing the a_n more carefully, show that we can even choose $f(z)$ to be an entire function.

GLOBAL FIXED POINT THEORY

§9. The Holomorphic Fixed Point Formula

First note the following.

9.1. Lemma. *Every rational map $f(z) \not\equiv z$ of degree d has exactly $d+1$ fixed points, counted with multiplicity.*

Here, by definition, the *multiplicity* of a fixed point is equal to one whenever the multplier satisfies $\lambda \neq 1$, and is strictly greater than one otherwise. (Compare §7.) As an example, in the special case of a polynomial map of degree $d \geq 2$, there is exactly one fixed point at infinity, and hence d finite fixed points counted with multiplicity.

Proof. Conjugating f by a fractional linear automorphism if necessary, we may assume that the point at infinity is not fixed by f . Hence we can write f as a quotient $f(z) = p(z)/q(z)$ where the polynomial $q(z)$ has degree equal to d and $p(z)$ has degree at most d . Now the equation $f(z) = z$ is equivalent to the polynomial equation $p(z) = zq(z)$ of degree $d+1$, and the assertion follows. □

Both Fatou and Julia made use of a "well known" relation between the multipliers at the fixed points of a rational map. Consider first an isolated fixed point $f(z_0) = z_0$ of a holomorphic map in one complex variable. If $z_0 \neq \infty$, we define the *holomorphic index* of f at z_0 to be the residue

$$\iota(f\,,\,z_0) \;=\; \frac{1}{2\pi i}\oint \frac{dz}{z - f(z)}$$

where we integrate in a small loop in the positive direction around z_0 . As usual, if $z_0 \in \hat{\mathbf{C}}$ happens to be the point at infinity, then we must first introduce the local uniformizing parameter $\zeta = \phi(z) = 1/z$. In this case, we define the residue of f at ∞ to be equal to the residue of $\phi \circ f \circ \phi^{-1}$ at the origin.

9.2. Theorem. *For any rational map $f : \hat{\mathbf{C}} \to \hat{\mathbf{C}}$ with $f(z)$ not identically equal to z , we have the relation*

$$\sum_{f(z)=z} \iota(f, z) \;=\; 1\,,$$

to be summed over all fixed points. In the case of a simple fixed point, where the multiplier $\lambda = f'(z_0)$ satisfies $\lambda \neq 1$, the index is given by

$$\iota(f, z_0) \;=\; \frac{1}{1-\lambda}\,.$$

In any case, the index at z_0 is a local analytic invariant. That is, if $g = \phi \circ f \circ \phi^{-1}$ where ϕ is a local holomorphic change of coordinate, then $\iota(f, z_0) = \iota(g, \phi(z_0))$.

As an application, if f has one fixed point with multiplier very close to 1, and hence with $|\iota|$ large, then it must have at least one other fixed point with $|\iota|$ large and hence with λ close to (or equal to) 1.

Proof of 9.2. Conjugating f by a linear fractional automorphism if necessary, we may assume that the point at infinity is not fixed by f. Then $f(z)$ remains bounded as $|z| \to \infty$, and there is a Laurent series expansion of the form

$$\big(z - f(z)\big)^{-1} \;=\; z^{-1} + c_2 z^{-2} + c_3 z^{-3} + \cdots$$

for large $|z|$. It follows that the integral of $\frac{1}{2\pi i} \cdot \frac{dz}{z - f(z)}$ in a large loop around the origin is equal to $+1$. Evidently this integral is equal to the sum of the residues $\iota(f, z_j)$ at the fixed points of f; hence the summation formula. The computation of $\iota(f, z_j)$ at a fixed point with multiplier $\lambda_j \ne 1$ is similar, using the Laurent series expansion of $\big(z - f(z)\big)^{-1}$ in a neighborhood of z_j.

The proof that f is a local analytic invariant, even at a multiple fixed point where $\lambda = 1$, can be sketched as follows. Choose a 1-parameter family of perturbations f_t of the given map f_0 so that f_t has distinct fixed points, all with multiplier different from 1, for all small $t \ne 0$. For example we can set $f_t(z) = f(z) + t$. Thus a fixed point z_0 of $f = f_0$ will split up into a cluster of nearby simple fixed points for $t \ne 0$. Since the integral in a fixed loop around z_0 varies continuously with t, and since this integral is a sum of residues which are evidently local analytic invariants when $t \ne 0$, it follows that $\iota(f_0\,, z_0)$ is also a local analytic invariant. □

For generalizations of this formula, the reader is referred to Atiyah and Bott.

Examples. A rational map $f(z) = c$ of degree zero has just one fixed point, with multiplier zero and hence with index $\iota(f, c) = 1$. A rational map of degree one usually has two distinct fixed points, and the relation

$$\frac{1}{1-\lambda_1} + \frac{1}{1-\lambda_2} \;=\; 1$$

simplifies to $\lambda_1 \lambda_2 = 1$. Any polynomial map $p(z)$ of degree two or more has a "super-attractive" fixed point at infinity, with multiplier zero and hence with index $\iota(p, \infty) = 1$. Thus the sum of the indices of the *finite* fixed points of a polynomial map of degree ≥ 2 is always zero. For a polynomial of degree exactly two, the relation

$$\frac{1}{1-\lambda_1} + \frac{1}{1-\lambda_2} \;=\; 0$$

for the finite fixed points simplifies to $\frac{1}{2}(\lambda_1 + \lambda_2) = 1$.

9.3. Lemma. *A fixed point with multiplier $\lambda \ne 1$ is attracting if and only if its index ι has real part $\mathcal{R}(\iota) > \frac{1}{2}$.*

Geometrically, this is proved by noting that inversion carries the disk $1 + D$ having the origin as boundary point to the appropriate half-plane. Computationally, it can be proved by noting that $\frac{1}{2} < \mathcal{R}(\frac{1}{1-\lambda})$ if and only if

$$1 \;<\; \frac{1}{1-\lambda} + \frac{1}{1-\bar{\lambda}}\,.$$

Clearing denominators, this reduces easily to the required inequality $\lambda\bar{\lambda} < 1$. □

One important consequence is the following.

9.4. Corollary. *Every rational map of degree two or more must have either a repelling fixed point, or a fixed point with* $\lambda = 1$ *, or both.*

Proof. If the $d+1$ fixed points were distinct and were all attracting or neutral, then each index would have real part $\mathcal{R}(\iota) \ge \frac{1}{2}$, hence the sum would have real part greater than or equal to $\frac{d+1}{2} > 1$; but this would contradict the Fixed Point Formula. □

Since repelling points and parabolic points both belong to the Julia set, this yields another proof of the following. (Compare §4.4, as well as §4.3 and §7.5.)

9.5. Corollary. *The Julia set, for any rational map of degree two or more, is always non-vacuous.*

Problem 9-1. If $f(z) = z + \alpha z^2 + \beta z^3 + (\text{higher terms})$, with $\alpha \ne 0$, show that the holomorphic index is given by $\iota(f, 0) = \beta/\alpha^2$. As an example, consider the one-parameter family of cubic maps

$$f_\alpha(z) \;=\; z + \alpha z^2 + z^3$$

with a double fixed point at the origin. Show that the remaining finite fixed point of f_α is attracting if and only if α^2 lies within a unit disk centered at -1 , or if and only if α lies within a figure 8 shaped region bounded by a lemniscate. Show that f_α can be perturbed so that the double fixed point at the origin splits up into two fixed points which are *both* attractive if and only if α^2 lies inside the disk of radius $1/2$ centered at $1/2$, or if and only if α lies within a region bounded by a lemniscate shaped like the symbol ∞ .

Problem 9-2. Any fixed point z_0 for f is evidently also a fixed point for $f^{\circ n}$. If z_0 is attracting [or repelling], show that $\iota(f^{\circ n}, z_0)$ tends to the limit 1 [or 0] as $n \to \infty$. If $f(z) = z + \alpha z^k + (\text{higher terms})$, show that $\iota(f^{\circ n}, 0)$ tends to the limit $k/2$.

§10. Most Periodic Orbits Repel.

This section will prove the following theorem of Fatou. By a *cycle* we will mean simply a periodic orbit of f. Recall that a cycle is called *attracting*, *neutral*, or *repelling* according as its multiplier λ satisfies $|\lambda| < 1$, $|\lambda| = 1$, or $|\lambda| > 1$.

10.1. Theorem. *Let $f : \mathbf{C} \to \mathbf{C}$ be a rational map of degree two or more. Then f has at most a finite number of cycles which are attracting or neutral.*

We will see in §11 that there always exist infinitely many repelling cycles. Shishikura has given the sharp upper bound of $2d-2$ for the number of attracting or neutral cycles, using methods of quasi-conformal surgery. However the classical proof, which is given here, shows only that this number is less than or equal to $6d-6$.

First consider the case of an attracting cycle $\hat{z}_0 \mapsto \hat{z}_1 \mapsto \cdots \mapsto \hat{z}_m = \hat{z}_0$. Each $\hat{z}_j$ is an attracting fixed point for the m-fold composition $f^{\circ m}$. If Ω_j is the immediate basin for $\hat{z}_j$ under $f^{\circ m}$, then the union $\Omega_0 \cup \cdots \cup \Omega_{m-1}$ will be called the *immediate basin* for our m-cycle. It can be described as the union of those components of the Fatou set which intersect the given attracting cycle. We continue to assume that f has degree two or more.

10.2. Lemma. *The immediate basin $\Omega_0 \cup \cdots \cup \Omega_{m-1}$ for any attracting cycle of f must contain at least one critical point of f.*

Proof. In the special case of an attracting fixed point, this was proved in §6.6. Applying this result to the m-fold iterate of f, we see that the immediate basin Ω_0 for the fixed point $\hat{z}_0$ of $f^{\circ m}$ must contain at least one critical point z_0 of $f^{\circ m}$. Let $z_0 \mapsto z_1 \mapsto \cdots$ be the orbit of this critical point under f. By the chain rule, at least one the the m points $z_0, z_1, \ldots, z_{m-1}$ must be a critical point for f. Since $z_j \in \Omega_j$, this proves the Lemma. □

Next consider a parabolic cycle $\hat{z}_0 \mapsto \hat{z}_1 \mapsto \cdots \mapsto \hat{z}_m = \hat{z}_0$, with multiplier λ equal to a q-th root of unity. Then the qm-fold iterate $f^{\circ qm}$ maps each $\hat{z}_j$ to itself with multiplier λ^q equal to 1. Recall that the number $n \geq 1$ of attracting petals around $\hat{z}_j$ must be some multiple of q. (Compare 7.2 and 7.6.)

Definition. The union of those components of the Fatou set which contain one of these n attracting petals around one of the m points of the cycle will be called the *immediate basin* of this parabolic cycle. Thus the number nm of connected components of this immediate basin is some multiple of qm.

10.3. Lemma. *If f has degree $d \geq 2$, then the immediate basin for any parabolic cycle must also contain at least one critical point.*

Proof. In the case of a fixed point with multiplier $\lambda = 1$, this was proved in §7.10. The general case follows easily, just as in the argument above. □

In fact, it is not difficult to check that there must be at least n/q distinct critical points in such an immediate basin. Combining 10.2 and 10.3 we obtain the following.

10.4. Lemma. *A rational map of degree* $d \geq 2$ *can have at most* $2d-2$ *cycles which are attracting or parabolic. Similarly, a polynomial map of degree* $d \geq 2$ *can have at most* $d-1$ *cycles in the finite plane which are attracting or parabolic.*

Proof. Since the immediate basins for distinct cycles are distinct by definition, it follows from 10.2 and 10.3 that the number of attracting or parabolic cycles is less than or equal to the number of critical points. In the case of a polynomial of degree d, the number of finite critical points is clearly at most $d-1$. In the case of a rational function of degree d, by the Riemann-Hurwitz formula §5.1 the number of critical ponts, counted with multiplicity, is equal to $2d-2$. Hence the number of distinct critical points is at most $2d-2$. In either case, the conclusion follows. □

10.5. Remark. This Lemma gives a practical computational procedure for finding all attracting or parabolic cycles. We must simply follow the orbits of all critical points, and test for convergence.

10.6. Lemma. *For a rational map of degree* $d \geq 2$, *the number of neutral cycles which have multiplier* $\lambda \neq 1$ *is at most* $4d-4$.

Evidently 10.4 and 10.6 together imply Theorem 10.1. The proof of 10.6 begins as follows. Following Fatou, we perturb the given map f and then applying 10.4. It will be convenient to compare f with the map $z \mapsto z^d$ of the same degree. Note that this model map has no neutral cycles. It has superattractive fixed points at zero and infinity, but otherwise all of its periodic points are strictly repelling, with $|\lambda| = d^m > 1$ where m is the period. Let $f(z) = p(z)/q(z)$ where $p(z)$ and $q(z)$ are polynomials, at least one of which has degree d. Consider the one-parameter family of rational maps

$$f_t(z) \;=\; \frac{(1-t)p(z) \,+\, tz^d}{(1-t)q(z) \,+\, t}$$

with $f_0(z) = f(z)$ and $f_1(z) = z^d$. There will be some finite set of exceptional values of t for which the numerator and denominator of this fraction have a common divisor. Algebraically this condition is expressed by setting the "resultant" of the numerator and denominator equal to zero. Geometrically, it means that a zero and pole of f_t crash together. If we exclude this finite set of bad values of t, then clearly $f_t(z)$ is holomorphic as a function of two variables, and each f_t has degree d.

Suppose that $f = f_0$ had $4d-3$ distinct neutral cycles with multipliers $\lambda_j \neq 1$. By the Implicit Function Theorem, we could follow each of these cycles under a small deformation of f_0. Thus, for small values of $|t|$, the map f_t would have corresponding cycles with multipliers $\lambda_j(t)$ which depend holomorphically on t, with $|\lambda_j(0)| = 1$.

10.7. Sub-Lemma. *None of these functions* $t \mapsto \lambda_j(t)$ *can be constant throughout a neighborhood of* $t = 0$.

Proof. Suppose that for some j the function $t \mapsto \lambda_j(t)$ were constant throughout a neighborhood of $t = 0$. Then we will show that it is possible to continue analytically along any path from 0 to 1 in the t-plane which misses the finitely many exceptional values of t. To prove this, we must check that the set of t for which we can continue is both open and closed. But it is closed since any limit point of periodic points with fixed multiplier $\lambda_j \neq 1$ is itself a periodic point with this same multiplier; and it is open since any such cycle varies smoothly with t, throughout some open neighborhood in the t-plane, by the Implicit Function Theorem. Now continuing analytically to $t = 1$, we see that the map $z \mapsto z^d$ must also have a cycle with mutiplier equal to λ_j, with $|\lambda_j| = 1$. But this is known to be false, which proves 10.7. $\square$

The proof of 10.6 continues as follows. We can express each of our $4d-3$ multipliers as a locally convergent power series

$$\lambda_j(t)/\lambda_j(0) \;=\; 1 + a_j t^{n(j)} + (\text{higher terms})\,,$$

where $a_j \neq 0$ and $n(j) \geq 1$. Hence $\log|\lambda_j(t)|$ is equal to the real part of $a_j t^{n(j)} + (\text{higher terms})$. If we ignore the higher order terms, this means that we can divide the t-plane into $n(j)$ sectors for which $|\lambda_j(t)| > 1$ and $n(j)$ congruent sectors for which $|\lambda_j(t)| < 1$. Taking account of the higher order terms, we have the following statement. Note that $\mathrm{sgn}(\log|\lambda|)$ is equal to $+1$ or -1 according as $|\lambda| > 1$ or $|\lambda| < 1$.

Assertion. *The step function*

$$\theta \;\mapsto\; \sigma_j(\theta) \;=\; \lim_{r\to 0}\ \mathrm{sgn}\ \log|\lambda_j(re^{i\theta})|$$

is well defined with value ± 1 except at finitely many jump discontinuities; and has average equal to zero.

Therefore the sum $\sigma_1(\theta)+\cdots+\sigma_{4d-3}(\theta)$ is also a well defined step function with average zero. Since this sum takes odd values almost everywhere, we can choose some θ for which $\sigma_1(\theta)+\cdots+\sigma_{4d-3}(\theta) \leq -1$. If we choose r sufficiently small and set $t = re^{i\theta}$, this means that f_t has at least $2d-1$ distinct cycles with multiplier satisfying $|\lambda_j| < 1$. But this is impossible by 10.4, which proves 10.6 and 10.1. $\square$

§11. Repelling Cycles are Dense in J .

We saw in §4.3 that every repelling cycle is contained in the Julia set. The following much sharper statement was proved by Fatou and by Julia. (Compare 11.9.)

11.1. Theorem. *The Julia set for any rational map of degree ≥ 2 is equal to the closure of its set of repelling periodic points.*

Since the proofs by Julia and by Fatou are interesting and different, we will give both.

Proof following Julia. We will use the Holomorphic Fixed Point Formula of §9. Recall from 9.4 that every rational map f of degree two or more has either a repelling fixed point, or a fixed point with $\lambda = 1$. In either case, this fixed point belongs to the Julia set $J(f)$. (Compare §4.3 and §7.5.)

Thus we can start with a fixed point z_0 in the Julia set. Let $U \subset \hat{\mathbf{C}}$ be any open set, disjoint from z_0 , which intersects $J(f)$. The next step is to construct a special orbit $\cdots \mapsto z_2 \mapsto z_1 \mapsto z_0$ which passes through U and terminates at this fixed point z_0 . By definition, such an orbit is called *homoclinic* if the backwards limit $\lim_{j\to\infty} z_j$ exists and is equal to the terminal point z_0 . To construct a homoclinic orbit, we will appeal to Theorem 4.8 which says that there exists an integer $r > 0$ and a point $z_r \in J(f) \cap U$ so that the r-th forward image $f^{\circ r}(z_r)$ is equal to z_0 . Given any neighborhood N_0 of z_0 , we can repeat this argument and conclude that there exists an integer $q > r$ and a point $z_q \in N_0$ so that $f^{\circ(q-r)}(z_q) = z_r$. (Figure 11.)

To be more explicit, in the case where z_0 is a repelling fixed point we choose N_0 to be a linearizing neighborhood, as in the Koenigs Theorem 6.1. In the parabolic case, we choose N_0 to be a flower neighborhood, as in 7.2. In either case, we choose N_0 small enough to be disjoint from z_r . It then follows that we can inductively choose preimages $\cdots \mapsto z_j \mapsto z_{j-1} \mapsto \cdots \mapsto z_q$, all inside of the neighborhood N_0 . *These preimages z_j will automatically converge to z_0 as $j \to \infty$.* If z_0 is repelling, this is clear. In the parabolic case, z_q cannot belong to an attracting petal; hence it must belong to a repelling petal, and again this statement is clear.

First suppose that none of the points $\cdots \mapsto z_j \mapsto \cdots \mapsto z_0$ in this homoclinic orbit are critical points of f . Then a sufficiently small disk neighborhood V_q of $z_q \in N_0$ will map diffeomorphically under $f^{\circ q}$ onto a neighborhood V_0 of z_0 . Pulling this neighborhood V_q back under iterates of f^{-1} , we obtain neighborhoods $z_j \in V_j$ for all j , shrinking down towards the limit point z_0 as $j \to \infty$. In particular, if we choose p sufficiently large, then $\bar{V}_p \subset V_0$. Now f^{-p} maps the simply connected open set V_0 holomorphically into a compact subset of itself. Hence it contracts the Poincaré metric on V_0 by a factor $c < 1$, and therefore must have a attractive fixed point z' within V_p . Evidently this point $z' \in V_p$ is a repelling periodic point of period p under the map f . Since the orbit of z' under f intersects the required open set U , the conclusion follows.

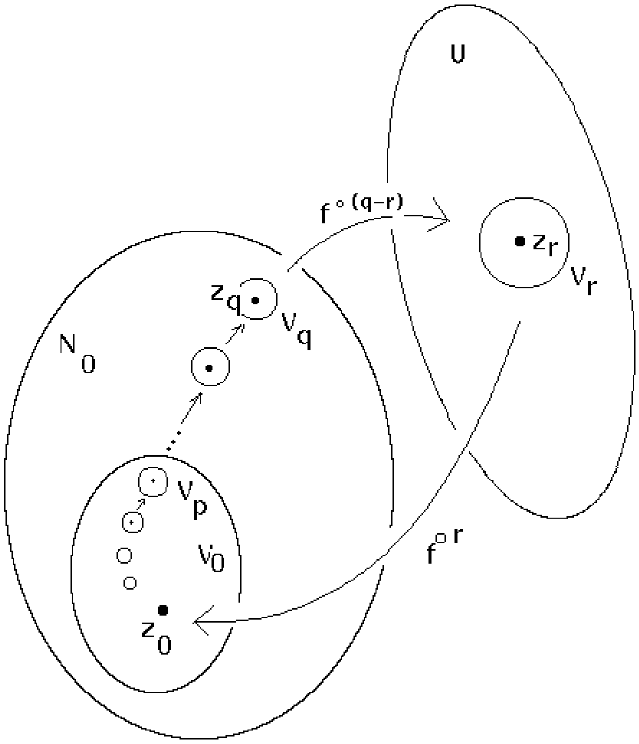


Figure 11. A homoclinic orbit.

If our homoclinic orbit contains critical points, then this argument must be modified very slightly as follows. We can still choose simply connected neighborhoods V_j of the z_j so that $\bar{V}_p \subset V_0$ for some large p, and so that f maps each V_j onto V_{j-1}. However, some finite number of these mappings will be branched. Choose a slit S in V_0 from the boundary to the midpoint z_0 so as to be disjoint from $\bar{V}_p$, and choose some sector in V_p which maps isomorphically onto $V_0 - S$ under $f^{\circ p}$. The proof now proceeds just as before. □

Proof of 11.1 following Fatou. In this case, the main idea is an easy application of Montel's Theorem (§2.5). However, we must use Theorem 10.1 to finish the argument.

To begin the proof, recall from 4.9 that the Julia set $J(f)$ has no isolated points. Hence we can exclude finitely many points of $J(f)$ without affecting the argument. Let z_0 be any point of $J(f)$ which is not a fixed point, and not a critical value. In other words, we assume that there are d preimages $z_1, \ldots, z_d$, which are distinct from each other and from z_0, where $d \geq 2$ is the degree. By the Inverse Function Theorem, we can find d holomorphic functions $z \mapsto \varphi_j(z)$ which are defined throughout some neighborhood N of z_0, and which satisfy $f(\varphi_j(z)) = z$, with $\varphi_j(z_0) = z_j$. We claim that for some $n > 0$ and for some $z \in N$ the function $f^{\circ n}(z)$ must take one of the three values z, $\varphi_1(z)$ or

$\varphi_2(z)$. For otherwise the family of holomorphic functions

$$g_n(z) \;=\; \frac{\big(f^{\circ n}(z)-\varphi_1(z)\big)\,\big(z-\varphi_2(z)\big)}{\big(f^{\circ n}(z)-\varphi_2(z)\big)\,\big(z-\varphi_1(z)\big)}$$

on N would avoid the three values 0, 1 and ∞, and hence be a normal family. It would then follow easily that $\{f^{\circ n}|N\}$ was also a normal family, contradicting the hypothesis that N intersects the Julia set. Thus we can find $z \in N$ so as to satisfy either $f^{\circ n}(z) = z$ or $f^{\circ n}(z) = \varphi_j(z)$. Clearly it follows that z is a periodic point of period n or $n+1$ respectively.

This shows that every point in $J(f)$ can be approximated arbitrarily closely by periodic points. Since all but finitely many of these periodic points must repel, this completes the proof. □

There are a number of interesting corollaries.

11.2. Corollary. *If U is an open set which intersects the Julia set J of f , then for n sufficiently large the image $f^{\circ n}(U \cap J)$ is equal to the entire Julia set J .*

Proof. We know that U contains a repelling periodic point z_0 of period say p . Thus z_0 is fixed by the iterate $g = f^{\circ p}$. Choose a small neighborhood $V \subset U$ of z_0 with the property that $V \subset g(V)$. Then clearly $V \subset g(V) \subset g^{\circ 2}(V) \subset \cdots$. But it follows from 4.6 or 4.8 that the union of the open sets $g^{\circ n}(V)$ contains the entire Julia set $J = J(f) = J(g)$. Since J is compact, this implies that $J \subset g^{\circ n}(V) \subset g^{\circ n}(U)$ for n sufficiently large, and the corresponding statement for f follows. □

11.3. Corollary. *If a Julia set J is not connected, then it has uncountably many distinct connected components.*

Proof. Suppose that J is the union $J_0 \cup J_1$ of two disjoint non-vacuous compact subsets. After replacing f by some iterate $g = f^{\circ n}$, we may assume by 11.2 that $g(J_0) = J$ and $g(J_1) = J$. Now to each point $z \in J$ we can assign an infinite sequence of symbols

$$\epsilon_0(z)\,,\ \epsilon_1(z)\,,\ \epsilon_2(z)\,,\ \ldots\ \in\ \{0,1\}$$

by setting $\epsilon_k(z)$ equal to zero or one according as $g^{\circ k}(z)$ belongs to J_0 or J_1 . It is not difficult to check that points with different symbol sequences must belong to different connected components of J , and that all possible symbol sequences actually occur. □

Remark. Evidently only countably many of the connected components of J can have positive Lebesgue measure. I don't know whether all but countably many of the components of J must be single points.

Siegel disks and Cremer points do not seem directly related to critical points, yet there is a connection as follows. By a *critical orbit* we will mean the forward orbit of some critical point.

11.4. Corollary. *Every boundary point of a maximal Siegel disk U belongs to the closure of some critical orbit.*

Proof. Otherwise, we could construct a small disk V around the given point $z_0 \in \partial U$ so that the forward orbits of all critical points avoid V. This would mean that every branch of the n-fold iterated inverse function f^{-n} could be defined as a single valued holomorphic function $f^{-n} : V \to \hat{\mathbf{C}}$. Let us choose that particular branch which carries the intersection $U \cap V$ into U by a "rotation" of U. Since the rotation number is irrational, we can choose some subsequence of these inverse maps which converges to the identity map on $U \cap V$. This is evidently a normal family, since it avoids the central part of U. Hence there is a sub-sub-sequence $\{f^{-n(i)}\}$ which converges on all of V, necessarily to the identity map of V. It follows easily that the corresponding sequence of forward iterates $f^{\circ n(i)}$ also converges to the identity on V. But this contradicts 11.2. □

11.5. Corollary. *Every Cremer point is a non-isolated point in the closure of some critical orbit.*

Proof. (This proof applies also to parabolic cycles. Compare §10.3.) Otherwise we could choose a small disk V around the given point z_0 so that no critical orbit intersects the punctured disk $V \smallsetminus \{z_0\}$. Replacing f by some iterate if necessary, we may assume that z_0 is fixed by f. Arguing as above, there exists a unique holomorphic branch $f^{-n} : V \to \hat{\mathbf{C}}$ of the n-fold iterated inverse function which fixes the point z_0. These inverse maps form a normal family since, for example, they avoid any periodic orbit which is disjoint from V. Thus we can choose a subsequence $f^{-n(i)}$ converging locally uniformly to some holomorphic map $h : V \to \hat{\mathbf{C}}$. By the Inverse function Theorem, since $|h'(z_0)| = 1$, this h must map some small neighborhood of z_0 isomorphically onto a neighborhood V' of z_0. It follows that the corresponding forward maps $f^{\circ n(i)}$ converge on V' to the inverse map $h^{-1} : V' \to V$. Again, this contradicts 11.2. □

By definition, the rational map f is called *post critically finite* (or is called a *Thurston map*) if it has the property that every critical orbit is finite, or in other words is either periodic or eventually periodic. According to Thurston, such a map can be uniquely specified by a finite topological description. (Compare Douady & Hubbard [DH1].)

11.6. Corollary. *If f is post critically finite, then every periodic orbit of f is either repelling or superattracting.*

11.7. Corollary. *More generally, suppose that f has the property that every critical orbit either is finite, or converges to an attracting periodic orbit. Then every periodic orbit of f is either attracting or repelling; there are no parabolic cycles, Cremer cycles, or Siegel cycles.*

(Compare 13.5, 14.4) The proofs are immediate, and will be left to the reader. □

By definition, z_0 belongs to the Julia set if *some* sequence of iterates of f has no subsequence which converges throughout a neighborhood of z_0. However, a priori, there could be other sequences which do converge.

11.8. Corollary. *If z_0 belongs to the Julia set, then no sequence of iterates of f can converge uniformly throughout a neighborhood of z_0.*

For if $\{f^{\circ n(i)}\}$ converges uniformly, with $n(i) \to \infty$, then according to Weierstrass (§1.4) the sequence of derivatives $df^{\circ n(i)}(z)/dz$ must also converge. But if z_1 is a repelling periodic point, then evidently the sequence of derivatives at z_1 diverges to infinity. □

11.9. Concluding Remark. The statement that the Julia set is equal to the closure of the set of repelling periodic points is actually true for an arbitrary holomorphic map of an arbitrary Riemann surface, providing that we exclude just one exceptional case. For transcendental functions this was proved by Baker [Ba1], and for maps of a torus or a Hyperbolic surface it follows easily from §4. The unique exceptional case occurs for degree one maps of $\hat{\mathbf{C}}$ which have just one parabolic fixed point — for example the map $f(z) = z + 1$ with $J(f) = \{\infty\}$.

On the other hand, for a holomorphic map of a complex 1-dimensional manifold with two or more connected components (for example where both components map into one), the Julia set can clearly be much bigger than the repelling orbit closure.

STRUCTURE OF THE FATOU SET

§12. Herman Rings.

The next two sections will be surveys only, with no proofs for several major statements. This section will describe a close relative of the Siegel disk.

Definition. A component U of the Fatou set $\hat{\mathbf{C}} \smallsetminus J(f)$ is called a *Herman ring* if U is conformally isomorphic to some annulus $\mathcal{A}_r = \{z : 1 < |z| < r\}$, and if f (or some itterate of f) corresponds to an irrational rotation of this annulus.

Remark. These objects are sometimes called "Arnold-Herman rings", since the existence of some examples would follow easily from Arnold's work in 1965. Siegel disks and Herman rings are often collectively called "*rotation domains*".

There are two known methods for constructing Herman rings. The original method, due to Herman, is based on a careful analysis of real analytic diffeomorphisms of the circle, as first studied by Arnold. An alternative method, due to Shishikura, uses quasiconformal surgery to cut and paste together two Siegel disks in order to fabricate such a ring.

The original method can be outlined as follows. (Compare [S1], [He1].) First a number of definitions. If $f : \mathbf{R}/\mathbf{Z} \to \mathbf{R}/\mathbf{Z}$ is an orientation preserving homeomorphism, then we can lift to a homeomorphism $F : \mathbf{R} \to \mathbf{R}$ which satisfies the identity $F(t+1) = F(t)+1$, and is uniquely defined up to addition of an integer constant.

Definition. Following Poincaré, the *rotation number* of the lifted map F is defined to be the real number

$$\mathrm{Rot}(F) \;=\; \lim_{n\to\infty} \frac{F^{\circ n}(t_0)}{n}$$

for any constant t_0, while the rotation number $\mathrm{rot}(f) \in \mathbf{R}/\mathbf{Z}$ of the circle map f is the residue class of $\mathrm{Rot}(F)$ modulo 1.

It is well known that this construction is well defined, and invariant under orientation preserving topological conjugacy, and that it has the following properties. (Compare Coddington and Levinson, or de Melo.)

12.1. Lemma. *The homeomorphism f has a periodic point with period q if and only if its rotation number is rational with denominator q.*

12.2. Denjoy's Theorem. *If f is smooth of class C^2, and if the rotation number $\rho = \mathrm{rot}(f)$ is irrational, then f is topologically conjugate to the rotation $t \mapsto t+\rho \pmod 1$.*

12.3. Lemma. *Consider a one-parameter family of lifted maps of the form*

$$F_\alpha(t) \;=\; F_0(t) + \alpha\,.$$

Then the rotation number $\mathrm{Rot}(F_\alpha)$ increases continuously and monotonically with α, increasing by $+1$ as α increases by $+1$. (However, this dependence is not strictly monotone. Rather, there is an interval of constancy corresponding to each rational value of $\mathrm{Rot}(F_\alpha)$.)

In the real analytic case, Denjoy's Theorem has an analog which can be stated as follows. Recall from §8 that a real number ξ is said to be *Diophantine* if there exist a (large) number n and a (small) number ϵ so that the distance of ξ from every rational number p/q satisfies $|\xi - p/q| > \epsilon/q^n$. The following was proved in a local version by Arnold, and sharpened first by Herman and then by Yoccoz.

12.4. Herman-Yoccoz Theorem. *If f is a real analytic diffeomorphism of $\mathbf{R}/\mathbf{Z}$ and if the rotation number ρ is Diophantine, then f is real analytically conjugate to the rotation $t \mapsto t + \rho \pmod 1$.*

I will not attempt to give a proof. (In the C^∞ case, Herman and Yoccoz prove a corresponding if and only if statement: *Every C^∞ diffeomorphism with rotation number ρ is C^∞-conjugate to a rotation if and only if ρ is Diophantine.*)

Next we will need the concept of a *Blaschke product.* (Compare Problems 1-2 and 5-1.) Given any constant $a \in \hat{\mathbf{C}}$ with $|a| \neq 1$, it is not difficult to show that there is one and only one fractional linear transformation $z \mapsto \beta_a(z)$ which maps the unit circle ∂D onto itself fixing the base point $z = 1$, and which maps a to $\beta_a(a) = 0$. For example $\beta_0(z) = z$, $\beta_\infty(z) = 1/z$, and in general

$$\beta_a(z) \;=\; \frac{1-\bar{a}}{1-a} \cdot \frac{z-a}{1-\bar{a}z}$$

whenever $a \neq \infty$. If $|a| < 1$, then β_a preserves orientation on the circle, and maps the unit disk into itself. On the other hand, if $|a| > 1$, then β_a reverses orientation on ∂D and maps D to its complement.

12.5. Lemma. *A rational map of degree d carries the unit circle into itself if and only if it can be written as a "Blaschke product"*

$$f(z) \;=\; e^{2\pi i t}\beta_{a_1}(z) \cdots \beta_{a_d}(z) \tag{$*$}$$

for some constants $e^{2\pi i t} \in \partial D$ and $a_1, \ldots, a_d \in \hat{\mathbf{C}} \smallsetminus \partial D$.

Here the a_i must satisfy the conditions that $a_j \bar{a}_k \neq 1$ for all j and k . For if $a\bar{b} = 1$, then a brief computation shows that $\beta_a(z)\beta_b(z) \equiv 1$. Evidently the expression in 12.5 is unique, since the constants $e^{2\pi i t} = f(1)$ and $\{a_1, \ldots, a_d\} = f^{-1}(0)$ are uniquely determined by f . The proof of 12.5 is not difficult: Given f , one simply chooses any solution to the equation $f(a) = 0$, then divides $f(z)$ by $\beta_a(z)$ to obtain a rational map of lower degree, and continues inductively. □

Such a Blaschke product carries the unit disk into itself if and only if all of the a_j satisfy $|a_j| < 1$. (Compare Problems 5-1, 12-3.) However, we will rather be interested in the mixed case, where some of the a_j are inside the unit disk and some are outside.

12.6. Theorem. *For any odd degree $d \geq 3$ we can choose a Blaschke product f of degree d which carries the unit circle ∂D into itself by a diffeomorphism with any desired rotation number ρ . If this rotation number ρ is Diophantine, then f possesses a Herman ring.*

Proof Outline. Let $d = 2n+1$, and choose the a_j so that $n+1$ of them are close to zero while the remaining n are close to ∞. Then it is easy to check that the Blaschke product $z \mapsto \beta_{a_1}(z) \cdots \beta_{a_d}(z)$ is C^1-close to the identity map on the unit circle ∂D. In particular, it induces an orientation preserving diffeomorphism of ∂D. Now multiplying by $e^{2\pi it}$ and using 12.3, we can adjust the rotation number to be any desired constant. If this rotation number ρ is Diophantine, then there is a real analytic diffeomorphism h of ∂D which conjugates f to the rotation $z \mapsto e^{2\pi i\rho} z$. Since h is real analytic, it extends to a complex analytic diffeomorphism on some small neighborhood of ∂D, and the conclusion follows. $\square$

As an example, Figure 12 shows the Julia set for the cubic rational map $f(z) = e^{2\pi it} z^2(z-4)/(1-4z)$ with zeros at $0, 0, 4$, where the constant $t = .6151732\cdots$ is adjusted so that the rotation number will be equal to $(\sqrt{5}-1)/2$. There is a critical point near the center of this picture, with a Herman ring to its left, surrounding the super-attractive basin about the origin in the left center. This is the simplest kind of example one can find, since Shishikura has shown that such a ring can exist only if the degree d is at least three, and since it is easy to check that a polynomial map cannot have any Herman ring. (Problem 12-1 or §17.1.)

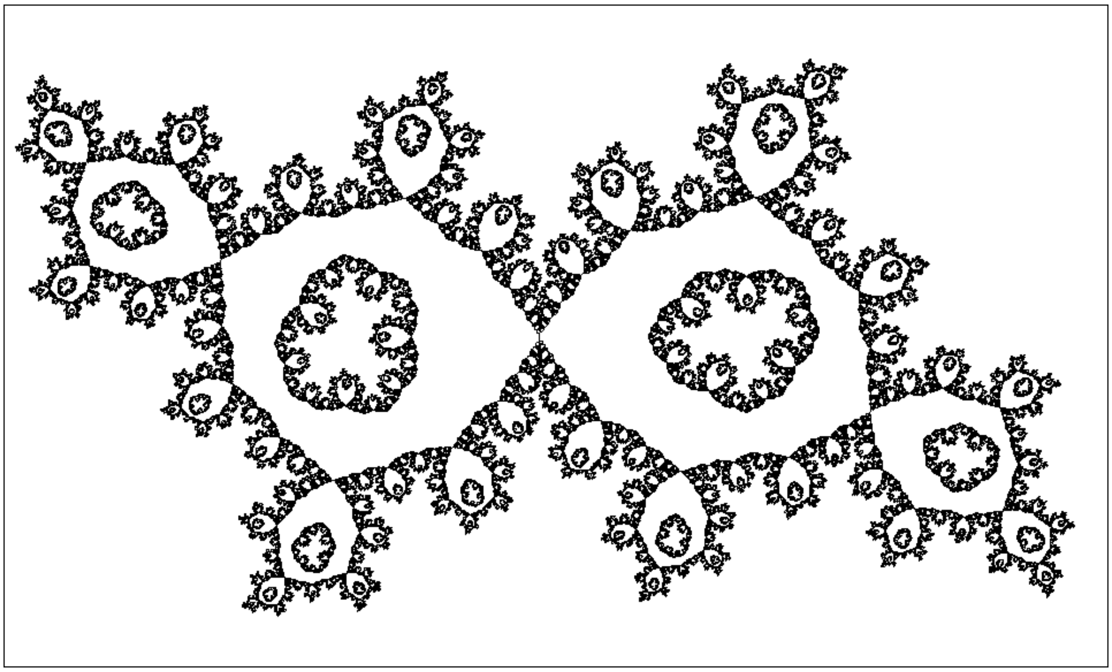


Figure 12. Julia set for a cubic rational map possessing a Herman ring.

The rings constructed in this way are very special in that they are symmetric about the unit circle, with $f(1/\bar{z}) = 1/\bar{f}(z)$. However Herman's original construction, based on work of Helson and Sarason, was more flexible. Shishikura's more general construction also avoids the need for symmetry. Furthermore Shishikura's construction makes it clear that the possible rotation numbers for Herman rings are exactly the same as the possible rotation numbers for Siegel disks. In particular, any number satisfying the Bryuno condition of §8.4 can occur. The idea is roughly that one starts with two rational maps having Siegel disks with rotation numbers $+\rho$ and $-\rho$ respectively. One cuts out a small concentric disk from each, and then glues the resulting boundaries together. After making corresponding modifications at each of the infinitely many iterated pre-images of each of the Siegel disks, Shishikura applies the Morrey-Ahlfors-Bers Measurable Riemann Mapping Theorem in order to conjugate the resulting topological picture to an actual rational map.

Although Herman rings do not contain any critical points, none-the-less they are closely associated with critical points.

12.7. Lemma. *If U is a Herman ring, then every boundary point of U belongs to the closure of the orbit of some critical point. The boundary ∂U has two connected components, each of which is an infinite set.*

The proof is almost identical to the proof of 11.4. □

Problem 12-1. Using the maximum modulus principle, show that no polynomial map can have a Herman ring.

Problem 12-2. For any Blaschke product $f : \hat{\mathbf{C}} \to \hat{\mathbf{C}}$ show that z is a critical point of f if and only if $1/\bar{z}$ is a critical point, and show that z is a zero of f if and only if $1/\bar{z}$ is a pole.

Problem 12-3. A holomorphic map $f : D \to D$ is said to be *proper* if the inverse image of any compact subset of D is compact. Show that any proper holomorphic map from D onto itself can be expressed uniquely as a Blaschke product $(*)$, with $a_j \in D$.

Problem 12-4. Show that the rotation number $\mathrm{rot}(f)$, as well as its continued fraction expansion, can be deduced directly from the cyclic order relations on a single orbit, without passing to the universal covering, as follows. Let $0 = t_0 \mapsto t_1 \mapsto t_2 \mapsto \cdots$ be the orbit of zero, where we can choose representatives modulo 1 so that $t_1 \le t_j < t_1+1$ for all j . Let us define t_j to be a "closest return on the left" if $t_j < 0$, and if none of the numbers t_k with $k < j$ belong to the open interval $(t_j\,,\,0)$. Similarly, t_j is a "closest return on the right" if the interval $(0, t_j)$ does not contain any t_k with $k < j$. If the sequence $t_1\,, t_2\,, \ldots$ contains first n_1 closest returns on the left then n_2 closest returns on the right, and so on, show that $\mathrm{rot}(f) = 1/(n_1 + (1/n_2 + 1/(n_3 + \cdots)))$. (Compare Appendix C.)

§13. The Sullivan Classification of Fatou Components.

The results in this section are due in part to Fatou and Julia, but with very major contributions by Sullivan.

By a *Fatou component* we will mean any connected component of the Fatou set $\hat{\mathbf{C}} \smallsetminus J(f)$. Evidently f carries each Fatou component U onto some Fatou component U' by a proper holomorphic map. First consider the special case $U = U'$.

13.1. Theorem. *If f maps the Fatou component U onto itself, then there are just four possibilities, as follows. Either U is the immediate attractive basin of an attracting fixed point, or of one petal of a parabolic fixed point, or else U is a Siegel disk or Herman ring.*

Here we are lumping together the case of a superattracting fixed point, with multiplier $\lambda = 0$, and the case of an ordinary attracting fixed point, with $\lambda \neq 0$. Note that immediate attractive basins always contain critical points by 10.2 and 10.3, while rotation domains (that is Siegel disks and Herman rings) evidently cannot contain critical points.

Much of the proof of 13.1 has already been carried out in §4. In fact, according to 4.3 and 4.4, a priori there are just four possibilities. Either:

(a) U contains an attractive fixed point;

(b) all orbits in U converge to a boundary fixed point;

(c) f is an automorphism of finite order; or

(d) f is conjugate to an irrational rotation of a disk, punctured disk, or annulus.

In Case (a) we are done. Case (c) cannot occur, since our standing hypothesis that the degree is two or more guarantees that there are only countably many periodic points. In Case (d) we cannot have a punctured disk, since the puncture point would have to be a fixed point belonging to the Fatou set, so that we would just have a Siegel disk with its center point incorrectly removed. Thus, in order to prove 13.1, we need only show that the boundary fixed point in Case (b) must be parabolic. But this boundary fixed point certainly cannot be an attracting point or a Siegel point, since it belongs to the Julia set. Furthermore, it cannot be repelling, since it attracts all orbits in U. The only other possibility, which we must exclude, is that it might be a Cremer point.

The proof will be based on the following statement, which is due to Douady and Sullivan. (Compare Sullivan [S1], or Douady-Hubbard [DH2, p. 70]. For a more classical alternative, see Lyubich [L1, p. 72].) Let

$$f(z) = \lambda z + a_2 z^2 + a_3 z^3 + \cdots$$

be a map which is defined and holomorphic in some neighborhood U of the origin, and which has a fixed point with multiplier λ at $z = 0$. By a *path converging to the origin* in U we will mean a continuous map $p : (0, \infty) \to U \smallsetminus \{0\}$, where $(0, \infty)$ is the open interval consisting of all positive real numbers, satisfying the condition that $p(t)$ tends to zero as $t \to \infty$.

13.2. Snail Lemma. *Let* $p : (0,\infty) \to U \smallsetminus \{0\}$ *be a path which converges to the origin, and suppose that* f *maps* p *into itself so that* $f(p(t)) = p(t+1)$ *. Then either the origin is either an attracting or a parabolic fixed point. More precisely, the multiplier* λ *satisfies either* $|\lambda| < 1$ *or* $\lambda = 1$ *.*

Replacing f by f^{-1} , we obtain the following completely equivalent formulation, which will be useful in §18.

13.3. Corollary. *If* $p : (0,\infty) \to U \smallsetminus \{0\}$ *is a path which converges to the origin, with* $f(p(t)) = p(t-1)$ *for* $t > 1$ *, then either* $|\lambda| > 1$ *or* $\lambda = 1$ *.*

Proof of 13.2. By hypothesis, the orbit $p(0) \mapsto p(1) \mapsto p(2) \mapsto \cdots$ in $U \smallsetminus \{0\}$ converges towards the origin. Thus the origin certainly cannot be a repelling fixed point: we must have $|\lambda| \le 1$. Let us assume that $|\lambda| = 1$ with $\lambda \ne 1$, and show that this hypothesis leads to a contradiction.

The intuitive idea can be described as follows. As the path $t \mapsto p(t)$ winds closer and closer to the origin, the behavior of the map f on $p(t)$ is more and more dominated by the linear term $z \mapsto \lambda z$. Thus $p(t+1) \approx \lambda p(t)$ for large t , and the image must resemble a very tight spiral as shown in Figure 13. Draw a radial segment E joining two turns of this spiral, as shown. Then the region V bounded by E together with a segment of the spiral will be mapped strictly into itself by f . Therefore, by the Schwarz Lemma, the fixed point of f at the point $0 \in V$ must be strictly attracting; which contradicts the hypothesis that $|\lambda| = 1$.

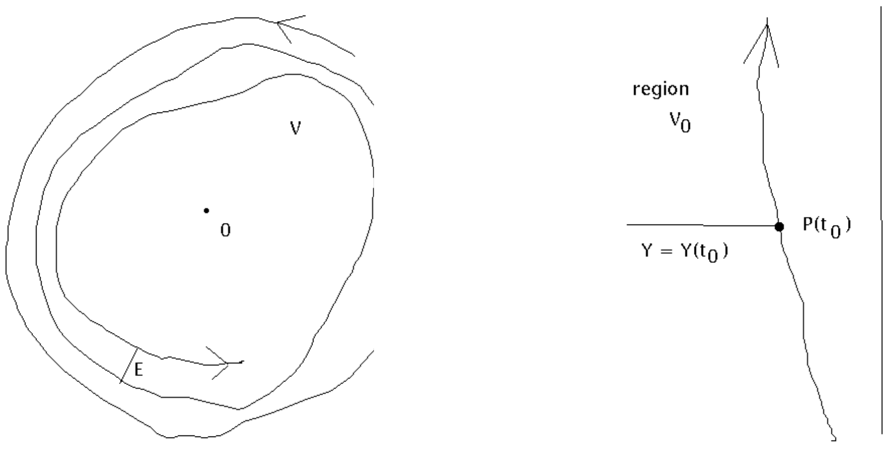


Figure 13. Diagram in the z-plane (left) and in the $Z = \log(z)$ plane (right).

In order to fill in the details of this argument, it is convenient to set $Z = \log z$, and to lift p to a continuous path $Z = P(t) = X(t) + iY(t)$, with $e^{P(t)} = p(t)$. Then evidently as $t \to \infty$ we have $X(t) \to -\infty$ and

$$P(t+1) \;=\; P(t) + ic + o(1) \,.$$

Here ic is a pure imaginary constant with $e^{ic} = \lambda$, and $o(1)$ stands for a remainder term which tends to zero. Similarly, we can lift f to a map $Z \mapsto F(Z)$ which is defined on some left half-plane $X <$ constant , and which satisfies

$$F(Z) \ = \ Z + ic + o(1)$$

as the real part $X = \mathcal{R}(Z)$ tends to $-\infty$, with the same non-zero constant ic . Suppose, to fix our ideas, that $c > 0$. Then, for $Z = X + iY$ within some half-plane $X < X_0$, we may assume that F is univalent, and that the imaginary part Y increases by at least $c/2$ under each iteration of F . Choose t_0 large enough so that $X(t) < X_0$ for $t \geq t_0$. Then the path $t \mapsto P(t)$ cuts the half-plane $Y > Y(t_0)$ into two or more connected components. Let V_0 be that component whose intersection with each horizontal line is unbounded to the left. Thus V_0 is contained in the quarter-plane $X < X_0\,,\ Y > Y(t_0)$. Evidently F maps V_0 diffeomorphically into itself. Furthermore, the imaginary part of $X + iY$ increases by at least $c/2$ under each iteration, hence the real part must decrease towards $-\infty$ under iteration.

Let $V = \exp(V_0) \cup \{0\}$ be the corresponding open set in the z-plane. Then it follows that f maps V into itself, and that every orbit of f in V converges towards the origin. Since the open set V is Hyperbolic, it follows by the Schwarz Lemma that the origin is an attractive fixed point; which contradicts our hypothesis that $|\lambda| = 1$. □

To prove 13.3, we simply note that the orbit

$$\cdots \mapsto p(2) \mapsto p(1) \mapsto p(0)$$

is repelled by the origin, so the multiplier λ cannot be zero. Hence f^{-1} is defined and holomorphic near the origin. Applying 13.2 to f^{-1} , the conclusion follows. □

Proof of 13.1. Let U be a Fatou component which is mapped into itself by f in such a way that all orbits converge to a boundary fixed point w_0 . Choose any base point z_0 in U , and choose any path $p : [0, 1] \to U$ from $p(0) = z_0$ to $p(1) = f(z_0)$. Extending for all $t \geq 0$ by setting $p(t + 1) = f(p(t)$, we obtain a path in U which converges to the boundary point w_0 as $t \to \infty$. Therefore, according to 13.2, the fixed point w_0 must be either parabolic or attracting. But w_0 belongs to the Julia set, and hence cannot be attracting. □

Thus we have classified the Fatou components which are mapped onto themselves by f . There is a completely analogous description of Fatou components which cycle periodically under f . These are just the Fatou components which are fixed by some iterate of f . Hence each such component is either

(1) the immediate attractive basin for some attracting periodic point,

(2) the immediate attractive basin for some petal of a parabolic periodic point,

(3) one member of a cycle of Siegel disks, or

(4) one member of a cycle of Herman rings.

In Cases (3) and (4), the topological type of the domain U is specified by this description. In Cases (1) and (2) it can be described as follows.

13.4. Lemma. *Every immediate attractive basin is either simply connected or infinitely connected.*

Proof. If U were a non-simply-connected region of finite connectivity, then it would have a finite Euler characteristic $\chi(U) \le 0$. By 10.2 or 10.3, f maps U onto itself by a branched covering map with at least one branch point, and hence with degree $n \ge 2$. But the precise number of branch points, counted with multiplicity, is equal to $(n-1)\chi(U) \le 0$ by the Riemann-Hurwitz formula of §5.1. This contradiction completes the proof. □

Sullivan showed that there can be at most a finite number of such periodic Fatou components. *In fact, according to Shishikura, there can be at most 2d-2 distinct cycles of Fatou components.*

In order to complete the picture, we need the fundamental theorem that there are no "wandering" Fatou components. (Compare [S2], [C].)

13.5. Theorem of Sullivan. *Every Fatou component U is eventually periodic. That is, there necessarily exist integers $n \ge 0$ and $p \ge 1$ so that the n-th forward image $f^{\circ n}(U)$ is mapped onto itself by $f^{\circ p}$.*

Thus every Fatou component is a preimage, under some iterate of f , of one of the four types described above. The proof, by quasiconformal deformation, is beyond the scope of these notes. Roughly speaking, if a wandering Fatou component were to exist, then one could construct an infinite dimensional space of deformations, all of which would have to be rational maps of the same degree. But the space of rational maps of fixed degree is finite dimensional.

Recall from §11.6 that f is *post critically finite* (or a *Thurston map*) if every critical orbit is finite. Combining 13.1 and 13.4 with 11.6 and 12.7, we easily obtain the following. (Compare Problem 5-7.)

13.6. Corollary. *If a post critically finite rational map has no superattractive periodic orbit, then its Julia set is the entire sphere $\hat{\mathbf{C}}$.*

We will give a different proof for this statement in 14.6.

We can sharpen the defining property of the Fatou set as follows.

Problem 13-1. If V is a connected open subset of the Fatou set $\hat{\mathbf{C}} \smallsetminus J$, show that the set of all limits of successive iterates $f^{\circ n}|V$ as $n \to \infty$ is either (1) a finite set of constant maps from V into an attracting or parabolic periodic orbit, or (2) a one-parameter family of maps, consisting of all compositions $R_\theta \circ f^{\circ k}|V$ where $f^{\circ k}$ is some fixed iterate with values in a rotation domain and R_θ is the rotation of this domain through angle θ .

§14. Sub-hyperbolic and hyperbolic Maps.

This section will discuss two important classes of rational maps. The exposition will be based on Douady-Hubbard [DH2].

First some standard definitions from the theory of smooth dynamical systems. Let $f : M \to M$ be a C^1-smooth map from a smooth Riemannian manifold to itself, and let $X \subset M$ be a compact f-invariant compact subset, $f(X) \subset X$. Let $Df_x : T_xM \to T_{f(x)}M$, or briefly Df, be the *first derivative map* at x, that is the induced linear map on the tangent space at x, and let $\|v\|$ be the Riemannian norm of a vector $v \in T_xM$. By definition, the map f is *expanding* on X if the length of any tangent vector at a point of X expands exponentially under iteration of Df, that is, if there exist constants $c > 0$ and $k > 1$ so that

$$\|Df^{\circ n}(v)\| \geq ck^n\|v\|$$

for every $x \in X$ and $v \in T_xM$, and every $n \geq 0$. Since X is compact, a completely equivalent requirement is that there exists some fixed $n \geq 1$ so that

$$\|Df^{\circ n}(v)\| > \|v\|$$

for all non-zero tangent vectors v over X. Similarly, f is *contracting* on X if there are constants $c > 0$ and $k < 1$ so that $\|Df^{\circ n}(v)\| \leq ck^n\|v\|$. It is not difficult to check that these conditions do not depend on the particular choice of Riemannian metric.

In the higher dimensional case, a map is called *hyperbolic* on X if the tangent space of M restricted to X splits as the Whitney sum of two sub-bundles, each invariant under Df, so that Df is expanding on one sub-bundle and contracting on the other. In the one-dimensional case, this concept simplifies as follows.

Definition: Let f be a holomorphic map from a Riemann surface to itself, and let X be a compact f-invariant subset. The map f is *hyperbolic* on X if f is expanding on X or contracting on X, or if X is the disjoint union of a compact f-invariant subset X^+ on which f is expanding and a compact f-invariant subset X^- on which f is contracting.

14.1. Theorem. *For a rational map $f : \hat{\mathbf{C}} \to \hat{\mathbf{C}}$ of degree $d \geq 2$, the following two conditions are equivalent:*

(1) *f is expanding and hence hyperbolic on its Julia set J.*

(2) *The forward orbit of each critical point of f converges towards some attracting periodic orbit.*

Map satisfying these conditions are briefly called *hyperbolic* rational maps. As examples, Figures 1a, 1b, 1d and 5 show the Julia sets of hyperbolic maps, but the other figures do not. **Caution:** This use of the word "hyperbolic" has nothing to do with the concept of "Hyperbolic Riemann surface", which we write with a capital *H.* (Compare §2.)

The proof that $(1) \Rightarrow (2)$ proceeds as follows. We suppose given some smooth Riemannian metric on $\hat{\mathbf{C}}$ so that the lengths of tangent vectors to $\hat{\mathbf{C}}$ at points of J increase exponentially under iteration of Df. After replacing f by some fixed iterate $f^{\circ p}$ (or alternatively after carefully modifying the metric), we may assume that $\|Df(v)\| \geq k\|v\|$ for all tangent vectors at points within some neighborhood V of J, where $k > 1$. In particular, there can be no critical points within this neighborhood V.

14.2. Lemma. *If there exists such an expanding metric in a neighborhood of the Julia set, then every orbit outside of the Julia set converges towards an attracting periodic orbit.*

Proof. We first show that there exists a constant $\epsilon > 0$ so that the neighborhood N_ϵ, consisting of all points at Riemannian distance less than ϵ from J, has the following property:

If we iterate the mapping f starting at any point which belongs to the neighborhood N_ϵ but not to J, then the distance from J will be multiplied by k or more at each iteration until we leave the set N_ϵ. Furthermore, no point outside of N_ϵ maps into N_ϵ.

In fact we need only choose ϵ small enough so that the neighborhood N_ϵ is contained in V, but disjoint from the compact set $f(\hat{\mathbf{C}} \smallsetminus V)$. It then follows that $f^{-1}(N_\epsilon) \subset V$. In fact we will prove the stronger statement that $f^{-1}(N_\epsilon) \subset N_{\epsilon/k} \subset N_\epsilon$. For if $f(z) \in N_\epsilon$, then $f(z)$ can be joined to J within N_ϵ by a geodesic of length equal to the Riemannian distance $\rho(f(z), J) < \epsilon$. We can then lift this back to a smooth curve which joins z to J within V. The length of this pulled back curve is evidently $\leq \rho(f(z), J)/k < \epsilon/k$.

Since $f^{-1}(N_\epsilon) \subset N_\epsilon$, it follows that the complementary compact set $L = \hat{\mathbf{C}} \smallsetminus N_\epsilon$ satisfies $f(L) \subset L$. We have shown that every orbit which starts outside of J must eventually be absorbed by this invariant set L. Since L is a compact subset of the Fatou set $\hat{\mathbf{C}} \smallsetminus J$, it must be covered by finitely many of the connected components of $\hat{\mathbf{C}} \smallsetminus J$. Let $U_1, \cdots, U_p$ be those components of $\hat{\mathbf{C}} \smallsetminus J$ which intersect L. Since $f(L) \subset L$, it follows that f carries each one of these U_i onto some U_j. As we follows such an orbit, we must eventually reach a component U_j which is mapped onto itself by some iterate $f^{\circ q}$.

Now let us examine the four possibilities for such a map $f^{\circ q} : U_j \to U_j$, as listed in Theorem 3.1. This map cannot be a rotation of U_j, or an automorphism of finite order, since all orbits in U_j are eventually absorbed by the compact subset $L \cap U_j$. Similarly, orbits cannot diverge to infinity with respect to the Poincaré metric on U_j. Thus the only possibility is that all orbits of $f^{\circ q}$ within U_j converge to an attracting fixed point of $f^{\circ q}$. This proves the Lemma, and hence completes the proof that $(1) \Rightarrow (2)$.

Proof that $(2) \Rightarrow (1)$**.** Let P be the *post-critical set*, that is the union of the forward orbits of the critical points. Then Condition (2) clearly implies that the closure $\bar{P}$ is disjoint from the Julia set. Conversely, if $\bar{P} \cap J = \emptyset$, then we will construct an expanding metric near the Julia set. Let U be the open set $\hat{\mathbf{C}} \smallsetminus \bar{P}$. Since $f(\bar{P}) \subset \bar{P}$, we have $f^{-1}(U) \subset U$. There are no critical points in U, so it follows that f^{-1} lifts to a single valued analytic function from the universal covering surface $\tilde{U}$ into itself. But

U contains the Julia set, which contains at least one repelling fixed point. Therefore this inverse map on $\tilde{U}$ can be chosen so as to contain an attracting fixed point. If $\tilde{U}$ is isomorphic to the unit disk, then this implies that f^{-1} strictly contracts the Poincaré metric on $\tilde{U}$, so that f must strictly increase the Poincaré metric on U. This proves that f is expanding on the compact subset $J \subset U$. On the other hand, if $\tilde{U}$ is not isomorphic to D, then the complement $\bar{P} = \hat{\mathbf{C}} \smallsetminus U$ can contain at most two points. It is then easy to check that f must be conjugate to a map of the form $z \mapsto z^{\pm d}$. In this case, clearly f is expanding with respect to the Euclidian metric on U. Thus $(2) \Rightarrow (1)$, which completes the proof. □

One invariant set of particular interest is the "non-wandering set" of f. By definition, a point is *wandering* if it has a neighborhood U so that the forward images $f(U)$, $f^{\circ 2}(U)$, $\ldots$ are all disjoint from U. Otherwise, it is *non-wandering.* The closed set consisting of all non-wandering points is called the *non-wandering set* $\Omega = \Omega(f)$. The map f is said to satisfy Smale's *Axiom A* if f is hyperbolic on Ω, and if Ω is precisely equal to the closure of the set of periodic points.

14.3. Corollary. *A rational map satisfies Axiom A if and only if it is hyperbolic.*

The proof is straightforward, since we know by §11 that the repelling periodic points are dense in the Julia set. Details will be left to the reader. (Compare Problem 14-1.) □

Remark. These hyperbolic maps have other extremely important properties. If f is hyperbolic, then every nearby map is also hyperbolic. Furthermore, according to Mañé, Sad, Sullivan, and also Lyubich, the Julia set $J(f)$ varies continuously under a deformation of f through hyperbolic maps. In the non-hyperbolic case, a small change in f may well lead to a drastic alteration of $J(f)$. It is generally conjectured that every rational map can be approximated arbitrarily closely by a hyperbolic map.

Douady and Hubbard, using ideas of Thurston, also consider a wider class of mapping which they call *sub-hyperbolic.* The only change in the definition is that the Riemannian metric is now allowed to have a finite number of relatively mild singularities, like that of the metric $|d\sqrt[n]{z}|$ at the origin. Such singularities have the property that the Riemannian distance between points still tends to zero as the points approach each other.

Definition. A conformal metric on a Riemann surface, with the expression $\gamma(z)|dz|$ in terms of a local uniformizing parameter z, will be called an *orbifold metric* if the function $\gamma(z)$ is smooth and non-zero except at a locally finite collection of points $a_1, a_2, \ldots$ where it blows up in the following special way. There should be integers $\nu_j \geq 2$ called the *ramification indices* at the points a_j with the following property. If we take a local branched covering by setting $z(w) = a_j + w^{\nu_j}$, then the induced metric $\gamma(z(w))|(dz/dw)dw|$ on the w-plane should be smooth and non-singular throughout some neighborhood of the origin.

The rational map f is *sub-hyperbolic* if it is expanding with respect to some orbifold metric on a neighborhood of its Julia set.

14.4. Theorem. *A rational map is sub-hyperbolic if and only if:*

(α) *every critical orbit in the Julia set is eventually periodic, and*

(β) *every critical orbit outside of the Julia set converges to an attracting periodic orbit.*

(Compare 11.7.) As examples, Figures 1-5 show sub-hyperbolic maps, but 8-10, 12, 17 do not.

The proof in one direction is completely analogous to the proof above. If f is expanding with respect to some orbifold metric, defined near the Julia set, then, just as in 14.2 above, every orbit outside the Julia set J will be pushed away from J, and can only converge to an attractive periodic orbit. On the other hand, if ω is a critical point inside the Julia set, then every post-critical point $f^{\circ n}(\omega)$, $n > 0$, must be one of the singular points a_j for our orbifold metric, since the map $f^{\circ n}$ is critical, and yet is required to be expanding, at the point ω. Thus the forward orbit of ω cannot have any limit points, and hence must be finite.

For the proof in the other direction, we must introduce the concept of the "universal branched covering" of an orbifold. Compare Appendix E. For our purposes, an *orbifold* (S, ν) will just mean a Riemann surface S, together with a locally finite collection of marked points a_j, each of which is assigned a *ramification index* $\nu_j = \nu(a_j) \ge 2$ as above. For any point z which is not one of the a_j we set $\nu(z) = 1$.

Definition. To every rational map f which satisfies Conditions (α) and (β) of Theorem 14.4, we assign the *canonical orbifold* (S, ν) as follows. As underlying Riemann surface S we take the Riemann sphere $\hat{\mathbf{C}}$ with all attracting periodic orbits removed. As ramified points a_j we take all strictly *post-critical* points, that is, all points which have the form $a_j = f^{\circ n}(\omega)$ for some $n > 0$, where ω is a critical point of f. Using (α) and (β), we see easily that this collection of points a_j is locally finite in S (although perhaps not in $\hat{\mathbf{C}}$). In order to specify the corresponding ramification indices $\nu_j = \nu(a_j)$, we will need another definition. If $f(z_1) = z_2$, with local power series development

$$f(z) = z_2 + c(z - z_1)^n + (\text{higher terms}),$$

where $c \ne 0$ and $n \ge 1$, then the integer $n = n(f, z_1)$ is called the *local degree* or the *branch index* of f at z_1. *Now choose the $\nu(a_j) \ge 2$ to be the smallest integers which satisfy the following:*

Condition ($*$). For any $z \in S$, the ramification index $\nu(f(z))$ at the image point must be a multiple of the product $n(f, z)\nu(z)$.

In fact, to construct these integers $\nu(a_j)$ we simply consider all critical pre-images

$$f^{\circ m}(\omega) = a_j, \quad f'(\omega) = 0,$$

and let $\nu(a_j)$ be the least common multiple of the corresponding branch indices $n(f^{\circ m}, \omega)$. Since we have removed all attracting periodic orbits, it is not difficult to check that this least common multiple is finite, and that it provides a minimal solution to the required condition ($*$).

As in Appendix E, we consider the *universal covering surface*

$$\tilde{S}_\nu \to (S\,,\,\nu)$$

for this canonical orbifold, that is the unique regular branched covering of S which has the given ν as ramification function, and which is simply connected. In order to determine the geometry of $\tilde{S}_\nu$, we introduce the *orbifold Euler characteristic*

$$\chi(S\,,\,\nu) \;=\; \chi(S) + \sum\left(\frac{1}{\nu(a_j)} - 1\right) \;=\; 2 - k + \sum\left(\frac{1}{\nu(a_j)} - 1\right),$$

to be summed over all ramified points, where k is the number of attracting periodic points. According to Lemma E.1 in the Appendix, such a universal covering nearly always exists; and the few pathological cases where it does not exist necessarily have Euler characteristic $\chi(S\,,\,\nu) > 0$. Furthermore, by E.4, the surface $\tilde{S}_\nu$ is either Spherical, Hyperbolic, or Euclidean according as $\chi(S\,,\,\nu)$ is positive, negative, or zero.

Still assuming Conditions (α) and (β) of 14.4, we will prove the following.

14.5. Lemma. *This canonical orbifold has Euler characteristic $\chi(S\,,\,\nu) \le 0$. Hence its universal covering $\tilde{S}_\nu$ has either a unique Poincaré metric or a Euclidean metric which is unique up to a scale change. In either case, there is an induced orbifold metric on S, and f is expanding with respect to this orbifold metric throughout any compact subset of S.*

Evidently this will complete the proof of Theorem 14.4. In the Euclidean case, the proof yields much more. Let z be a uniformizing parameter on the covering surface $\tilde{S}_\nu \cong \mathbf{C}$.

14.6. Theorem. *If $\chi(S\,,\,\nu) = 0$, then f induces a linear isomorphism $z \mapsto az + b$ from the Euclidean covering space $\tilde{S}_\nu \cong \mathbf{C}$ onto itself. In this case, the Julia set is either a circle or line segment, or the entire Riemann sphere. The coefficient of expansion $|a|$ satisfies $|a| = d$ in the first two cases, and $|a| = \sqrt{d}$ in the last case, where d is the degree of f.*

(Caution: The coefficient a itself is not uniquely determined; for the lifting of f to the covering surface is determined only up to composition with a deck transformation. The deck transformations may well have fixed points, since we are dealing with a branched covering, but necessarily have the form $z \mapsto \omega z + c$ where ω is some root of unity.)

Proof of 14.5. First note that Condition $(*)$ above is exactly what is needed in order to assert that f^{-1} lifts, locally at least, to a holomorphic map from the universal covering $\tilde{S}_\nu$ to itself. But since $\tilde{S}_\nu$ is simply connected, there is then no obstruction to constructing a globally lifted inverse

$$\tilde{f}^{-1} : \tilde{S}_\nu \to \tilde{S}_\nu\,.$$

The proof now divides into three cases.

Spherical Case. Note that the composition

$$\tilde{S}_\nu \xrightarrow{\tilde{f}^{-1}} \tilde{S}_\nu \xrightarrow{p} S \xrightarrow{f} S$$

always coincides with the projection map $p : \tilde{S}_\nu \to S$. If $\tilde{S}_\nu$ were Spherical, then this composition would have a well defined degree, which would have to be at least d times the degree of p . This is impossible, hence the Spherical case cannot occur.

Hyperbolic Case. If $\tilde{S}_\nu$ is Hyperbolic, then it has a uniquely defined Poincaré metric, and the proof proceeds just as in the proof that $(2) \Rightarrow (1)$ in 14.1 above.

Euclidean Case. Suppose that $\tilde{S}_\nu$ is Euclidean, or equivalently that $\chi(S\,,\,\nu) = 0$. Construct a new orbifold $(S'\,,\,\mu)$ as follows. Let $S' \subset S$ be the open set S with all immediate pre-images of attracting periodic points removed, and define $\mu = f^*(\nu)$ by the formula

$$\mu(z) \;=\; \nu(f(z))/n(f\,,\,z)$$

where $n(f\,,\,z)$ is the branch index. Note that $\mu(z) \geq \nu(z)$ for all $z \in S'$. Evidently it follows that

$$\chi(S'\,,\,\mu) \;\leq\; \chi(S\,,\,\nu)$$

with equality only if $S' = S$ and $\mu = \nu$. But by Lemma E.2, since the map $f : (S'\,,\,\mu) \to (S\,,\,\nu)$ is a "d-fold covering" of orbifolds, we conclude that f induces an isomorphism $\tilde{S}'_\mu \to \tilde{S}_\nu$ of universal covering surfaces, and also that the Riemann-Hurwitz formula takes the form

$$\chi(S'\,,\,\mu) \;=\; \chi(S\,,\,\nu)d\,.$$

Thus $\chi(S'\,,\,\mu)$ is also zero. We conclude that $S' = S$ and $\mu = \nu$, so that f must lift to a (necessarily linear) isomorphism from the Euclidean covering space $\tilde{S}_\nu$ to itself. Further details will be left to the reader. This completes the proof of 14.4 through 14.6. □

As a corollary, we obtain another proof of 13.6. We continue to assume that every critical orbit either converges to an attracting orbit or is eventually periodic, according as it belongs to the Fatou set or the Julia set.

14.7. Corollary. *If f is sub-hyperbolic with no attracting periodic orbits, so that S is the entire Riemann sphere, then f is expanding on the entire sphere, and it follows that $J(f)$ is the entire sphere.*

Problem 14-1. Using the results of §13, show that the non-wandering set for a rational map f is the (disjoint) union of its Julia set, its rotation domains (if any), and its set of attracting periodic points.

Problem 14-2. Show that the Julia set for the rational map $z \mapsto (1-2/z)^{2n}$ is the entire Riemann sphere. Show that the orbifold metric for this example is Euclidean when $n = 1$, but is Hyperbolic for $n > 1$.

Problem 14-3. For any sub-hyperbolic map whose canonical orbifold metric is Euclidean, show that every periodic orbit outside of the finite post-critical set has multiplier λ satisfying $|\lambda| = d^{p/\delta}$ where d is the degree, p is the period, and δ is the dimension (1 or 2) of the Julia set.

Problem 14-4. A rational map f is said to be *expansive* in a neighborhood of its Julia set if there exists $\epsilon > 0$ so that, for any two points $x \neq y$ whose orbits remain in the neighborhood forever, there exists some $n \geq 0$ so that $f^{\circ n}(x)$ and $f^{\circ n}(y)$ have distance greater than ϵ. Using Sullivan's results from §13, show that this condition is satisfied if and only if f is hyperbolic. (However, a map with a parabolic fixed point may be expansive on the Julia set itself.)

CARATHÉODORY THEORY

§15. Prime Ends.

Let U be a simply connected subset of $\hat{\mathbf{C}}$ such that the complement $\hat{\mathbf{C}} - U$ is infinite. Then the Riemann Mapping Theorem asserts that there is a conformal isomorphism

$$\phi : U \xrightarrow{\approx} D \, .$$

In some cases, ϕ will extend to a homeomorphism from the closure $\bar{U}$ onto the closed disk $\bar{D}$. (Compare Figures 1a and 14a, together with §16.7.) However, this is certainly not true in general, since the boundary ∂U may be an extremely complicated object. As an example, Figure 14b shows a region U such that one point of ∂U corresponds to a Cantor set of distinct points of the circle ∂D, and Figures 14c, 14d show examples for which an entire interval of points of ∂U corresponds to a single point of the circle. An effective analysis of the relationship between the compact set ∂U and the boundary circle ∂D was carried out by Carathéodory in 1913, and will be described here.

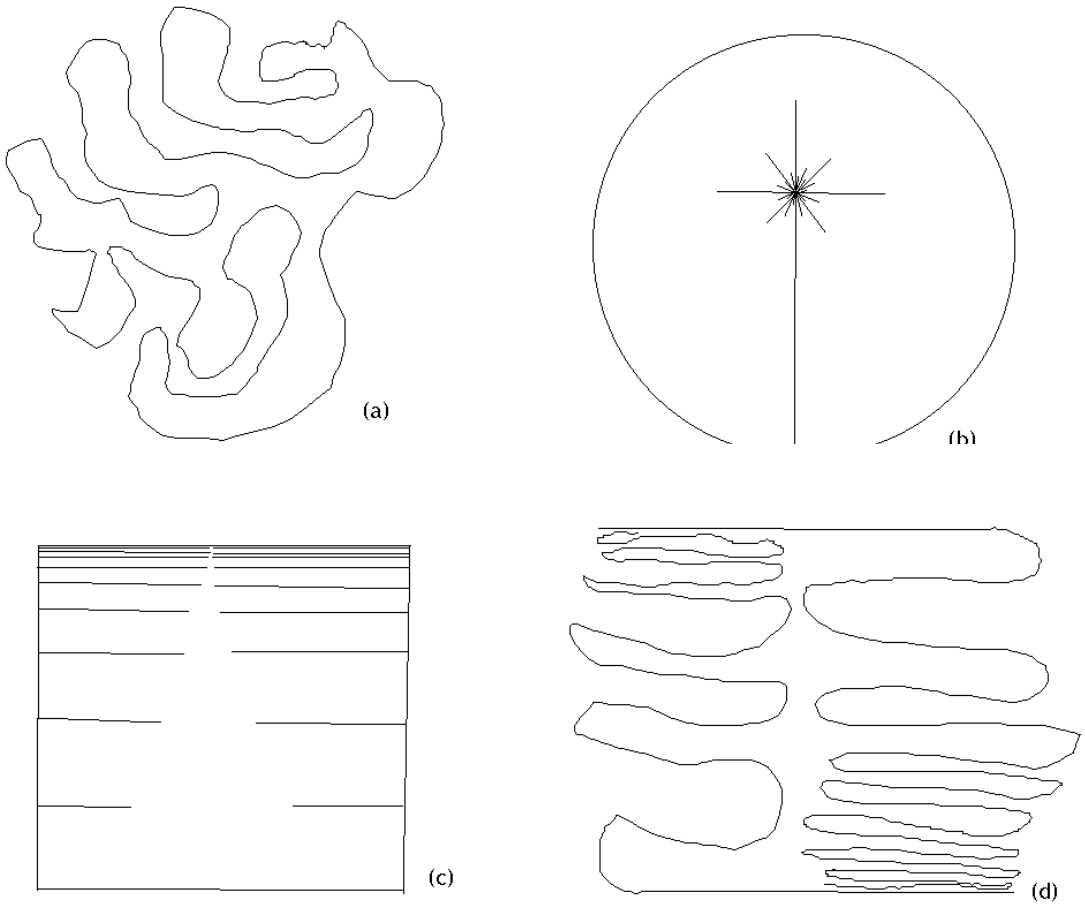


Figure 14. The boundaries of four simply connected regions in $\mathbf{C}$.

The final construction will be purely topological, but we must first use analytic methods, as in Appendix B, to prove several lemmas about the existence of short arcs. Let $I = (0, \delta)$ be an open interval of real numbers, and let $I^2 \subset \mathbf{C}$ be the open square, consisting of all $z = x + iy$ with $x,\, y \in I$. We will always use the spherical metric on $\hat{\mathbf{C}}$ when measuring either arclengths or areas. Recall from §1(6) that this metric has the form $ds = \eta(z)|dz|$ with $\eta > 0$. (In fact $\eta(z) = 2/(1 + |z|^2)$.) Setting $z = x + iy$, the corresponding spherical area element on $\hat{\mathbf{C}}$ can be written as $dA = \eta^2(z)dxdy$, with total area $\iint dA = 4\pi < \infty$.

15.1. Lemma. *If $f : I^2 \to V$ is a conformal isomorphism from the open square onto an open subset of $\hat{\mathbf{C}}$, then for Lebesgue almost every $x \in I$ the arc $f(x \times I)$ has finite spherical arclength.*

Proof. (Compare B.1 in the Appendix.) The length $L(x)$ of this image $f(x \times I)$ can be expressed as $\int_I |\eta f'|dy$ where the function η is to be evaluated at $f(x + iy)$, while the area of V can be expressed as $A = \int_I a(x)dx$ with $a(x) = \int_I |\eta f'|^2 dy$. By the Schwarz inequality we have

$$\Big(\int_I 1 \cdot |\eta f'|\, dy\Big)^2 \;\le\; \Big(\int_I 1^2 dy\Big) \cdot \Big(\int_I |\eta f'|^2 dy\Big),$$

or in other words $L(x)^2 \le \delta a(x)$. Since the integral of $a(x)$ is finite, it follows that $a(x)$ is finite almost everywhere, hence $L(x)$ is finite almost everywhere. □

Here is a more quantitative version of this statement. Again let A be the spherical area of $f(I^2) = V$.

15.2. Lemma. *With $f : I^2 \overset{\approx}{\to} V$ as above, the length $L(x)$ of the curve $f(x \times I)$ satisfies $L(x) < 2\sqrt{A}$ for more than half of the points $x \in I$. Similarly, the length of $f(I \times y)$ is less than $2\sqrt{A}$ for more than half of the points $y \in I$.*

Proof. We must show that the Lebesgue measure of $S = \{x \in I : L(x) \ge 2\sqrt{A}\}$ satisfies $\ell(S) < \frac{1}{2}\ell(I) = \frac{1}{2}\delta$. It is clear that $4A\ell(S) = (2\sqrt{A})^2\ell(S) \le \int_I L(x)^2 dx$. On the other hand, the discussion above shows that $\int_I L(x)^2 dx \le \delta A$. Combining these two inequalities and dividing by $4A$, we obtain $\ell(S) \le \delta/4$, as required. □

Now consider a simply connected open set $U \subset \hat{\mathbf{C}}$ and some choice of Riemann map $\phi : U \to D$, with inverse $\psi : D \to U$.

15.3. Theorem. *For almost every point $e^{i\theta}$ of the circle ∂D the radial line $r \mapsto re^{i\theta}$ maps under ψ to a curve of finite spherical length in U. In particular, the radial limit*

$$\lim_{r \to 1} \psi(re^{i\theta}) \in \partial U$$

exists for Lebesgue almost every θ. Furthermore, if we fix θ, then for Lebesgue almost every θ' the radial limit of $\psi(re^{i\theta'})$ is different from the radial limit of $\psi(re^{i\theta})$.

We will say briefly that almost every image curve $r \mapsto \psi(re^{i\theta})$ in U *lands* at some single point of ∂U , and that different values of θ almost always correspond to distinct landing points.

Remark. The first part of this Lemma is the univalent version of a basic *Theorem of Fatou*, which Carathéodory used as the starting point of his theory. Fatou's Theorem says that any bounded holomorphic function on D has radial limits in almost all directions, whether or not it is univalent. (See for example Hoffman, p. 38.) However the univalent case is all that we will need, and is easier to prove than the general theorem.

Proof of 15.3. Map the upper half-plane onto $D - \{0\}$ by the exponential map $w \mapsto e^{iw}$. In terms of polar coordinates $r,\ \theta$ in D , this means that we set $w = \theta - i \log r$, so that $e^{iw} = re^{i\theta}$. Applying the argument of 15.1 to the composition $w \mapsto \psi(e^{iw})$, we see that ψ has radial limits in almost all directions $e^{i\theta}$.

If this map ψ is bounded, then a Theorem of F. and M. Riesz asserts that any given radial limit can occur only for a set of directions $e^{i\theta}$ of measure zero. A proof may be found in Appendix A, Theorem A.3. Now for any univalent ψ , we can reduce to the bounded case in two steps, as follows. First suppose that the image $\psi(D) = U$ omits an entire neighborhood of some point z_0 of $\hat{\mathbf{C}}$. Then by composing ψ with a fractional linear transformation which carries z_0 to ∞ , we reduce to the bounded case. In general, $\psi(D)$ must omit at least two values, which we may take to be 0 and ∞ . Then $\sqrt{\psi}$ can be defined as a single valued function which omits an entire open set of points; and we are reduced to the previous case. □

Similarly, using the change of coordinates $z = e^{iw}$ together with 15.2, we obtain the following. Let $\psi : D \overset{\approx}{\to} U$ as above.

15.4. Lemma. *Any point of* ∂D *has a nested sequence of neighborhoods* $\mathcal{N}_1 \supset \mathcal{N}_2 \supset \cdots$ *in* $\bar{D}$ *with the following properties:*

(1) *each boundary* $\overline{\mathcal{N}_k} \cap \overline{D - \mathcal{N}_k}$ *consists of an open arc* $\mathcal{A}_k$ *in the interior of* D *together with two end points in* ∂D *,*

(2) *these boundaries* $\bar{\mathcal{A}}_k$ *are pairwise disjoint,*

(3) *both* $\mathcal{A}_k$ *and its image* $\psi(\mathcal{A}_k) \subset U$ *have finite length which tends to zero as* $k \to \infty$ *, and*

(4) *each image arc* $\psi(\mathcal{A}_k)$ *has two distinct boundary points in* ∂U *, and the closures* $\overline{\psi(\mathcal{A}_k)}$ *are pairwise disjoint.*

Proof. We choose a neighborhood bounded by two short radial line segments going out to the boundary of D together with an arc of a circle $|z| = \text{constant}$. Using 15.2 we see that the image of such a curve $\mathcal{A}$ can have length less than any given epsilon, and using the last part of 15.3 we see that $\mathcal{A}$ can be chosen so that the two endpoints of $\psi(\mathcal{A})$ in ∂U are distinct, and so that successive arcs have distinct endpoints. □

Conversely, let us start with an embedded arc $\alpha : [0, 1) \to U$ which lands at a point z_0 of ∂U . That is, we suppose that $\lim_{t \to 1} \alpha(t) = z_0$.

15.5. Theorem. *Any arc in* U *which lands at one point* z_0 *of* ∂U *corresponds, under the Riemann map, to an arc in* D *which lands at one point of* ∂D *. Furthermore, arcs which land at distinct points of* ∂U *necessarily correspond to arcs which land at distinct points of* ∂D *.*

Proof. If the image arc $t \mapsto \phi(\alpha(t))$ did not land at a single point of D , then it would have to oscillate back and forth (or round and round), accumulating towards some connected set of limit points on ∂D . It would then follow, for some set of $e^{i\theta} \in \partial D$ of positive measure, that the radial limit of $\psi(re^{i\theta})$ could only be z_0 , which is impossible by the Riesz Theorem.

If arcs landing at two distinct points of ∂U corresponded to arcs landing at a single point of ∂D . Then, using 15.4, we could cut both of these arcs by the image of an arc which is arbitrarily short and arbitrarily close to the boundary both in D and in U . Evidently this is impossible. □

Using these simple facts, Carathéodory showed that we can reconstruct the topology of the closed disk $\bar{D}$ from the topology of the embedding of U into $\bar{U}$. There are a number of possible variations on his basic definitions. (Compare Ahlfors [1973], Epstein, Ohtsuka.) We will make use of a variation which is fairly close to the original construction.

A *transverse arc* (or "crosscut") in U is a set $A \subset \bar{U}$ which is homeomorphic to the closed interval $[0, 1]$, and which intersects the boundary ∂U only in the two end points of the interval.

Note that it is very easy to construct examples of transverse arcs. For example we can start with any short line segment inside U and extend in both directions until it first hits the boundary.

15.6. Lemma. *Any transverse arc* A *cuts* U *into two connected components.*

Proof. The quotient space $\bar{U}/\partial U$, in which the boundary is identified to a point, is evidently homeomorphic to the 2-sphere. Since A corresponds to a Jordan curve in this quotient 2-sphere, the conclusion follows from the Jordan Curve Theorem. (See for example Munkres.) □

It will be convenient to distinguish these two components of $U - A$ by choosing some base point $b_0 \in U$. Then for any transverse arc A which is disjoint from b_0 we define the neighborhood $N(A)$ which is "cut off" by A to be the component of $U - A$ which does not contain b_0 .

Definition. A *fundamental chain* is a infinite sequence A_1 , A_2 , ... of disjoint transverse arcs with the property that the corresponding neighborhoods are nested:

$$N(A_1) \supset N(A_2) \supset N(A_3) \supset \cdots ,$$

and with the property that the diameter of A_i tends to zero as $i \to \infty$.

Note however that the neighborhoods $N(A_i)$ are not required to become small as $i \to \infty$. (Compare Figure 14c, d.) By the *impression* (or *support*) of a fundamental chain $\{A_i\}$ we mean the intersection of the closures $\bar{N}(A_i)$. It is not difficult to check that this impression is always a compact connected subset of ∂U . Evidently the impression consists of a single point if and only if the diameter of $N(A_i)$ tends to zero as $i \to \infty$.

Two such fundamental chains $\{A_i\}$ and $\{A'_j\}$ are said to be *equivalent* if each $N(A_i)$ contains some $N(A'_j)$ and each $N(A'_j)$ contains some $N(A_i)$. An equivalence class of fundamental chains is also called a *prime end* $\mathcal{E}$ in U , or a point in the *Carathéodory boundary* of U .

Two fundamental chains $\{A_i\}$ and $\{A'_j\}$ are said to be *disjoint* if $N(A_i) \cap N(A'_j) = \emptyset$ for some i and j .

15.7. Lemma. *Any two fundamental chains are either equivalent or disjoint.*

Proof. Given A_i we can choose j large enough so that the diameter of A'_j is less than the distance between A_i and A_{i+1} . It then follows easily that $N(A'_j)$ must be either contained in $N(A_i)$ or disjoint from $N(A_{i+1})$. □

We can now define the *Carathéodory completion* $\hat{U}$ of U to be the disjoint union of the set U and the set of all prime ends of U , topologized as follows. For any transverse arc $A \subset \bar{U} - \{b_0\}$ we define the neighborhood $\mathcal{N}_A$ to be the neighborhood $N(A) \subset U$ together with the set consisting of all prime ends $\mathcal{E}$ for which some representative fundamental chain $\{A_i\}$ satisfies $A_1 = A$. These neighborhoods $\mathcal{N}_A$, together with the open subsets of U , form a basis for the required topology.

15.8. Lemma. *In the case of the unit disk D , the identity map of D extends uniquely to a homeomorphism between the closure $\bar{D} \subset \mathbf{C}$ and the Carathéodory completion $\hat{D}$.*

(Compare 16.7.) In particular, the prime ends of D are in one-to-one correspondence with the points of ∂D . More precisely, the impression of any prime end of D is a single point of ∂D , and each point of ∂D is the impression of one and only one prime end. The proof is not difficult. □

For an arbitrary simply connected and Hyperbolic $U \subset \hat{\mathbf{C}}$ we now have the following.

15.9. Theorem. *The Riemann map $\phi : U \to D$ extends uniquely to a homeomorphism between the completion $\hat{U}$ and the completion $\hat{D} \cong \bar{D}$.*

Proof. It follows immediately from 15.5 that every transverse arc in U corresponds to a transverse arc in D , and that disjoint transverse arcs correspond to disjoint transverse arcs. Thus every end in U gives rise to an end in D . On the other hand, given any point of ∂D , we can find a nested sequence of neighborhoods in $\bar{D}$, each bounded by a figure consisting of two radial segments and an arc which is completely inside D . As in 15.4, we can choose these figures so as to correspond to transverse arcs in $\bar{U}$ which are pairwise disjoint, with lengths converging to zero. The resulting fundamental chain in U determines the required prime end. Further details will be left to the reader. □

§16. Local Connectivity.

This will be a continuation of the preceeding section, describing Carathéodory's theory in the locally connected case. First a number of definitions and well known lemmas. We will always assume that our topological spaces are Hausdorff. By definition, a topological space X is *locally connected* at the point $x \in X$ if there exist arbitrarily small connected neighborhoods of x in X.

Here our "neighborhoods" are not required to be open subsets of X, but only to be sets which contain a (relatively) open set containing x. Unfortunately, some authors make a slightly different definition by considering only open neighborhoods. Let us say that X is *openly locally connected* at a point x if there exist arbitrarily small connected open neighborhoods of x in X. This is a stronger condition (compare Problem 16-1); however we have the following.

16.1. Lemma. *The space X is locally connected at every point $x \in X$ if and only if every open subset of X is a union of connected open subsets of X.*

If this conditions is satisfied, then evidently X is openly locally connected at every point. Such spaces are simply said to be *locally connected.* The proof is straightforward, and will be left to the reader. □

16.2. Remark. If X is compact metric, with distance function $\rho(x, x')$, then an equivalent condition is that:

For any $\epsilon > 0$ there exists $\delta > 0$ so that any two points with distance $\rho(x, x') < \delta$ are contained in a connected set of diameter less than ϵ.

Again the proof is straightforward, and will be omitted.

The space X is *path connected* if there exists a continuous map from the unit interval $[0, 1]$ into X which joins any two given points. It is *arcwise connected* if there is a topological embedding of $[0, 1]$ into X which joins any two given distinct points.

16.3. Lemma. *A space is path connected if and only if it is arcwise connected.*

(Recall that all spaces must be Hausdorff.) Proofs will be deferred until the end of this section. There is a corresponding concept of *"locally path connected"* or equivalently *"locally arcwise connected"*. In general, a connected space need not be path connected (Figure 14d). However:

16.4. Lemma. *If a compact metric space X is locally connected, then it is locally path connected. Hence every connected component of X is actually path connected.*

We will also need the following statement.

16.5. Lemma. *Any continuous image of a compact locally connected space is compact and locally connected.*

The following is the principal result of this section. Let $U \subset \hat{\mathbf{C}}$ be an open set which is conformally isomorphic to the unit disk.

16.6. Theorem of Carathéodory. *The inverse Riemann map $\psi : D \stackrel{\approx}{\to} U$ extends continuously to a map from the closed disk $\bar{D}$ onto $\bar{U}$ if and only if the boundary ∂U is locally connected, or if and only if the complement $\hat{\mathbf{C}} - U$ is locally connected.*

Carathéodory also proved the following.

16.7. Theorem. *If the boundary of U is a Jordan curve, then the Riemann map extends to a homeomorphism from the closure $\bar{U}$ onto the closed disk $\bar{D}$.*

The proofs begin as follows.

Proof of 16.3. Let $f = f_0 : [0,1] \to X$ be any continuous path with $f(0) \neq f(1)$. We must construct an embedded arc $A \subset X$ from $f(0)$ to $f(1)$. Choose a closed subinterval $I_1 = [a_1 , b_1] \subset [0,1]$ whose length $0 \le \ell(I_1) = b_1 - a_1 < 1$ is as large as possible, subject to the condition that $f(a_1) = f(b_1)$. Now, among all subintervals of $[0,1]$ which are disjoint from I_1, choose an interval $I_2 = [a_2 , b_2]$ of maximal length subject to the condition $f(a_2) = f(b_2)$. Continue this process inductively, constructing disjoint subintervals of maximal lengths $\ell(I_1) \ge \ell(I_2) \ge \cdots \ge 0$ subject to the condition that f is constant on the boundary of each I_j.

Let $\alpha : [0,1] \to X$ be the unique map which is constant on each closed interval I_j, and which coincides with f outside of these subintervals. Thus $\alpha(t)$ must coincide with $f(t)$ for $t \in \partial I_j$. Then it is easy to check that α is continuous, and that for each point $x \in \alpha([0,1])$ the preimage $\alpha^{-1}(x) \subset [0,1]$ is a (possibly degenerate) closed interval of real numbers. *We will show that these conditions, with α non-constant, imply that the image $A = \alpha([0,1]) \subset X$ is an embedded arc joining $\alpha(0)$ to $\alpha(1)$.*

Choose a countable dense subset $\{t_1 , t_2 , \ldots\} \subset [0,1]$. Define a total ordering of the image A by specifying that $\alpha(s) < \alpha(t)$ if and only if $\alpha(s)$ and $\alpha(t)$ are distinct points with $s < t$. An order preserving homeomorphism $h : [0,1] \stackrel{\approx}{\to} A$ can now be constructed inductively as follows. We must set $h(0) = \alpha(0)$ and $h(1) = \alpha(1)$. For each dyadic fraction $0 < m/2^k < 1$ with m odd, let us assume inductively that $h((m-1)/2^k)$ and $h((m+1)/2^k)$ have already been defined. Then choose the smallest index j so that

$$h\Big(\frac{m-1}{2^k}\Big) \;<\; \alpha(t_j) \;<\; h\Big(\frac{m+1}{2^k}\Big)$$

and set $h(m/2^k) = \alpha(t_j)$. It is not difficult to check that the h constructed in this way on dyadic fractions extends uniquely to an order preserving map from $[0,1]$ to A, which is necessarily a homeomorphism. □

Proof of 16.4. Let X be compact metric and locally connected. Given $\epsilon > 0$, it follows from 16.2 that we can choose a sequence of numbers $\delta_n > 0$ so that any two points with distance $\rho(x, x') < \delta_n$ are contained in a connected set of diameter less than $\epsilon/2^n$. *We will prove that any two points $x(0)$ and $x(1)$ with distance $\rho(x(0), x(1)) < \delta_0$ can be joined by a path of diameter at most 4ϵ.* The plan of attack is as follows. We will choose a sequence of denominators $1 = k_0 < k_1 < k_2 < \cdots$, each of which divides the next. Also, for each fraction of the form i/k_n between 0 and 1 we will choose an intermediate point $x(i/k_n)$. These are to satisfy two conditions:

(1) Any two consecutive fractions i/k_n and $(i+1)/k_n$ must correspond to points $x(i/k_n)$ and $x((i+1)/k_n)$ which have distance less than δ_n from each other.

Hence any two such points are contained in a connected set $C(i,k_n)$ which has diameter less than $\epsilon/2^n$. The second condition is that:

(2) Each point of the form $x(j/k_{n+1})$, where j/k_{n+1} lies between i/k_n and $(i+1)/k_n$, must belong to this set $C(i,k_n)$, and hence have distance less than $\epsilon/2^n$ from $x(i/k_n)$ and from $x((i+1)/k_n)$.

The construction of such denominators k_n and intermediate points $x(i/k_n)$, by induction on n , is completely straightforward, and will be left to the reader. Thus we may assume that $x(r)$ has been defined for a dense set of rational numbers r in the unit interval.

Next we will prove that this densely defined correspondence $r \mapsto x(r)$ is uniformly continuous. Let r and r' be any two rational numbers for which $x(r)$ and $x(r')$ are defined. Suppose that $|r-r'|<1/k_n$. Then choosing i/k_n as close as possible to both r and r' , it is easy to show that $x(r)$ and $x(r')$ have distance less than $2\epsilon/2^n$ from $x(i/k_n)$. Hence they have distance less than $4\epsilon/2^n$ from each other. This proves uniform continuity; and it follows that there is a unique continuous extension $t \mapsto x(t)$ which is defined for all $t \in [0,1]$. In this way, we have constructed the required path of diameter $\le 4\epsilon$ from $x(0)$ to $x(1)$. Thus X is locally path connected; and the rest of the argument is straightforward. $\square$

Proof of 16.5. Let $f(X)=Y$ where X is compact and locally connected. Then Y is clearly compact, and we must show that it is locally.connected. Given any point $y \in Y$ and open neighborhood $V \subset Y$, choose a smaller neighborhood U with $\bar{U} \subset V$. Then $f^{-1}(V)$ is a union of disjoint connected open sets, and finitely many of these connected open sets, say $V_1, \ldots, V_q$, will suffice to cover the compact subset $K=f^{-1}(\bar{U})$. Let N be the union of those $f(V_i)$ which contain the given point y . Evidently N is connected and contained in V . In order to prove that N is a neighborhood of y , let L be the union of those compact images $f(K \cap V_i)$ which do not contain y . Then $U-L$ is an open neighborhood of y which is contained in N . Thus $N \subset V$ is a connected neighborhood of y , which proves that Y is locally connected. $\square$

Proof of Theorem 16.6. We will always use the spherical metric on $\hat{\mathbf{C}}$. If either ∂U or $\hat{\mathbf{C}}-U$ is locally connected, then given ϵ we can choose δ so that any two points of distance less than δ in ∂U are joined by an arc A of diameter less than ϵ in $\hat{\mathbf{C}}-U$. Now suppose that we start with some point $w_0 \in \partial D$. According to Lemma 15.4, we can find a neighborhood of diameter less than δ in $\bar{D}$ which is bounded by an arc which maps under $\psi : D \overset{\approx}{\to} U$ to an arc $A' \subset U$ which has length less than δ . Furthermore, we can choose this arc so that its two endpoints in ∂U are distinct. Let $A \subset \hat{\mathbf{C}}-U$ be an arc in the complement which has diameter less than ϵ , and has the same endpoints. Then the union $A \cup A'$ is a Jordan curve of diameter less than $\epsilon+\delta$ in $\hat{\mathbf{C}}$. Hence it bounds a region of diameter less than $\epsilon+\delta$. Now as $\epsilon \to 0$ this region must converge to a point in ∂U , which we define to be $\psi(w_0)$. Thus we have constructed a function from the closed disk $\bar{D}$ to $\bar{U}$ which extends $\psi : D \to U$. It is not difficult to check that this extension is continuous.

Conversely, if ψ extends continuously over $\bar{D}$, the the boundary ∂U is a continuous image of the circle ∂D . Hence it follows from Lemma 16.5 that ∂U is locally connected. To see that $\hat{\mathbf{C}}-U$ must also be locally connected, let N be an arbitrarily small connected neighborhood of a point z within ∂U . Choose $\delta > 0$ so that the δ-neighborhood of z within ∂U is contained in N . Then the δ-neighborhood of z within $\hat{\mathbf{C}} - U$, together with N , is clearly also connected. Therefore $\hat{\mathbf{C}} - U$ is locally connected. □

Proof of Theorem 16.7. If ∂U is a Jordan curve, that is a homeomorphic image of the circle, then we certainly have a continuous extension $\psi : \bar{D} \to \bar{U}$ by the preceeding theorem. Suppose that two distinct points of the circle ∂D mapped to the same point z_0 of ∂U . Then the straight line segment A joining these two points in $\bar{D}$ would map to a Jordan curve $\Gamma \subset \hat{\mathbf{C}}$ under ψ . Such a Jordan curve must separate $\hat{\mathbf{C}}$ into two components, of which Γ is the common boundary. Note that the two components of $D-A$ must map into the two distinct components of $\hat{\mathbf{C}} - \Gamma$ under ψ . Since Γ intersects the Jordan curve ∂U in a single point z_0 , it follows that the entire connected set $\partial U - \{z_0\}$ must be contained in just one of the two connected components of $\hat{\mathbf{C}} - \Gamma$. But this means that one of the two components of $D - A$ must map into the other component of $\hat{\mathbf{C}} - \Gamma$, which is disjoint from ∂U . Hence all of the corresponding boundary points in ∂D can only map to the single point $z_0 \in \Gamma \cap \partial U$ under ψ . This is impossible by the Riesz Theorem A.3 of Appendix A. □

Problem 16-1. Let $X \subset \mathbf{C}$ be the compact connected set which is obtained from the unit interval $[0, 1]$ by drawing line segments from 1 to the points $\frac{1}{2}(1 + i/n)$ for $n = 1, 2, 3, \ldots$ and then adjoining the successive images of this configuration under the map $z \mapsto z/2$. (Figure 15.) Show that X is locally connected at the origin, but not openly locally connected.

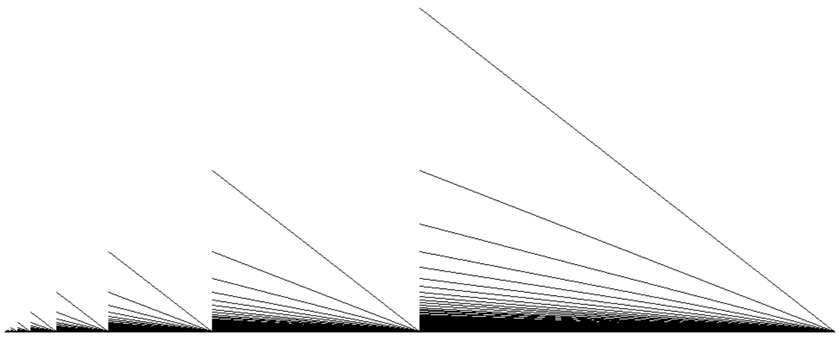


Figure 15. The witch's broom.

POLYNOMIAL MAPS

§17. The Filled Julia Set.

Let $f : \hat{\mathbf{C}} \to \hat{\mathbf{C}}$ be a polynomial map of degree $d \ge 2$, say

$$f(z) = a_d z^d + \cdots + a_1 z + a_0$$

with $a_d \ne 0$. Then f has a superattracting fixed point at infinity. In particular, there exists a constant $k = k_f$ so that any point z with $|z| > k_f$ belongs to the basin of attraction of infinity.

Definition. The complement of the attractive basin of infinity, that is the set of all points $z \in \mathbf{C}$ with bounded forward orbit under f , is called the *filled Julia set* $K = K(f)$.

17.1. Lemma. *This filled Julia set K is a compact set consisting of Julia set J together with all of the bounded components of the complement $\mathbf{C} \smallsetminus J$. These bounded components (if any) are all simply connected. The Julia set J is equal to the topological boundary ∂K .*

Proof. (Compare Problem 4-1.) If z is a boundary point of K , then z itself has bounded orbit, but points arbitrarily close have orbits which converge to the point at infinity. Thus the family of iterates of f cannot be normal in any neighborhood, hence $z \in J$. Conversely, if $z \in J$, then z has bounded orbit, but it follows from 4.6 that points arbitrarily close to f have unbounded orbit, which therefore converges to infinity. Thus $J = \partial K \subset K$, and it follows that the unbounded component of the complement $\mathbf{C} \smallsetminus J$ consists of points attracted to infinity. But if z belongs to a bounded component of $\mathbf{C} \smallsetminus J$, then it follows from the Maximum Modulus Principle that every point in the orbit of z belongs to the closed disk of radius k_f , hence $z \in K$. Finally, any bounded component U of $\mathbf{C} \smallsetminus J$ must be simply connected, since for any Jordan curve $\Gamma \subset U$ the entire bounded component of $\mathbf{C} \smallsetminus \Gamma$ must also be contained in U , again by the Maximum Modulus Principle. $\square$

It follows from the Böttcher Theorem §6.7 that exists a new coordinate $w = \phi(z)$ which is defined throughout some neighborhood $|z| > c_f$ of infinity, with $w \to \infty$ as $z \to \infty$, and which satisfies

$$\phi \circ f \circ \phi^{-1}(w) \;=\; w^d .$$

Further, ϕ is unique up to multiplication with a $(d-1)$-st root of unity.

The function $G(z) = \log|\phi(z)|$ is called the canonical *potential function* or *Green's function* for $K(f)$. Note the identity

$$G(z) \;=\; G(f(z))/d . \qquad (*)$$

Although we have so far defined G only in a neighborhood of infinity, there is one and only one extension to all of $\mathbf{C}$ which is continuous and satisfies (*). In fact we define $G(z) = 0$ for $z \in K(f)$, and $G(z) = G(f^{\circ n}(z))/d^n$ otherwise, where n is large enough so that $|f^{\circ n}(z)| > c_f$. Alternatively, we can simply set

$$G(z) \;=\; \lim_{n\to\infty} \log^+ |f^{\circ n}(z)|/d^n ,$$

where $\log^+(x) = \log(x)$ for $x \geq 1$ and $\log^+(x) = 0$ for $0 \leq x \leq 1$. Note that G is a smooth real analytic function outside of the Julia set J, but is only continuous at points of J. The verification of these facts will be left to the reader.

The lines $|\phi(z)| = \text{constant} > 0$ are called *equipotential curves.* Their orthogonal trajectories, that is the images under ϕ^{-1} of the radial lines extending out towards infinity from the unit disk are called *external rays* for f. Both of these families of curves are clearly smooth except at critical points of G. *Note that a point $z \in \mathbf{C} \smallsetminus K$ is a critical point of G if and only if it is a pre-critical point of f, that is a critical point of some iterate $f \circ \cdots \circ f$.* Again proofs are easily supplied.

For each $r > 1$ let $V(r)$ be the compact region with boundary consisting of all z with $G(z) \leq \log(r)$. Evidently the boundary $\partial V(r)$ is just the equipotential curve $\{z : G(z) = \log r\}$.

17.2. Lemma. *Suppose that there are no critical points of G (or pre-critical points of f) on the boundary $\partial V(r)$. If $m \geq 0$ is the algebraic number of critical points of G in the complement $\mathbf{C} \smallsetminus V(r)$, then $V(r)$ is a disjoint union of $m+1$ closed topological disks, each of which intersects the Julia set $J(f)$.*

Proof. If $\partial V(r)$ contains no critical points (or equivalently if $\log r$ is a regular value of the map G), then evidently $V(r)$ is a smooth compact manifold with boundary. Each component must be a closed topological disk. For otherwise, the complement $\mathbf{C} \smallsetminus V(r)$ would have a bounded component, which is impossible since the function G would have to take its maximum on the boundary of such a component. Similarly, the minimum value of G on each component of $V(r)$ must occur at a point of $K(f)$. Hence each such component must intersect the Julia set $\partial K(f)$.

Given $r > 1$ as above, let us choose n so large that the region $V(r^{d^n})$ is a single topological disk. It will be convenient to set $s = r^{d^n}$. Then the composition $f^{\circ n}$ maps $V(r)$ onto $V(s)$ by a d^n-fold branched covering map. According to the Riemann-Hurwitz formula, §5.1, the algebraic number of branch points of $f^{\circ n}$ in $V(r)$ is equal to $d^n \chi(V(s)) - \chi(V(r)) = d^n - \chi(V(r))$, where in our case χ is just the number of connected components. If m is the number of branch points of $f^{\circ n}$ outside of $V(r)$, then, since the total number of branch points is $d^n - 1$, we can write this formula as

$$d^n - 1 - m \;=\; d^n - \chi(V(r)),$$

or in other words $\chi(V(r)) = m + 1$. This completes the proof, since critical points of G are the same as critical points of $f^{\circ n}$ for large n. $\square$

Following Brown, a compact set in Euclidean space is said to be *cellular* if it is equal to the intersection of a nested sequence of closed embedded topological disks, each of which contains the next in its interior. The following result is essentially due to Fatou and Julia.

17.3. Theorem. *For a polynomial map f of degree $d \geq 2$, there are just two mutually exclusive possibilities: If the filled Julia set K contains all of the finite critical points of f, then K and $J = \partial K$ are connected, and in fact K is a cellular set. Furthermore, the Böttcher map near infinity extends to a conformal* *isomorphism*

$$\phi : \mathbf{C} \smallsetminus K \xrightarrow{\approx} \mathbf{C} \smallsetminus \bar{D} .$$

On the other hand, if f has at least one critical point in $\mathbf{C} \smallsetminus K$, then both J and K have uncountably many connected components.

Proof of 17.3. If K contains all of the finite critical points, then it follows from 17.2 that K is the intersection of a strictly nested family of embedded disks $V(r)$. Thus K is cellular, and hence connected. Furthermore, the proof shows that each $\mathbf{C} \smallsetminus V(r)$ is a d^n-fold unramified covering of some $\mathbf{C} \smallsetminus V(s)$, which is isomorphic to $\mathbf{C} \smallsetminus \bar{D}_s$ under the Böttcher map. We can then lift to a Böttcher isomorphism $\mathbf{C} \smallsetminus V(r) \xrightarrow{\approx} \mathbf{C} \smallsetminus \bar{D}_r$ which is compatible near infinity and satisfies

$$\phi(f(z)) = \phi(z)^d .$$

Passing to the limit as $r \to 1$, we see that ϕ can be defined throughout $\mathbf{C} \smallsetminus K$. Finally, note that the Julia set J can be expressed as the intersection of the closures of the bounded connected annuli $\phi^{-1}(\{w : 1 < |w| < r\})$; hence J is also connected.

Conversely, if f has at least one critical point in $\mathbf{C} \smallsetminus K$, then the argument above shows that some $V(r)$ has two or more connected components, and that each of these components intersects the subsets $J \subset K \subset V(r)$. Thus J has uncountably many components by 11.3. The analogous proof for K will be left to the reader. □

For information on the structure of K when it is not connected, see Branner & Hubbard, Blanchard. As an immediate consequence of Carathéodory's Theorem 16.6 we have the following. (Note that a compact set which has infinitely many components clearly cannot be locally connected.)

17.4. Corollary. *With f as above, the following three conditions are equivalent:*

(*a*) *The set $J(f)$ is locally connected.*

(*b*) *The set $K(f)$ is locally connected.*

(*c*) *K(f) is connected, and the inverse Riemann mapping $\psi : \hat{\mathbf{C}} \smallsetminus \bar{D} \xrightarrow{\approx} \hat{\mathbf{C}} \smallsetminus K$ extends continuously over the boundary, yielding a continuous map from the unit circle onto J satisfying the identity $f(\psi(w)) = \psi(w^d)$.*

Following Douady and Hubbard, we have the following.

17.5. Theorem. *If the polynomial map f is hyperbolic or subhyperbolic, with $K(f)$ connected, then $K(f)$ is locally connected.*

For non locally connected examples, see 18.6.

Proof of 17.5. We may assume that there is a smooth Riemannian metric or orbifold metric whose restriction to some small neighborhood U of J is strictly increasing under f , say by a factor of at least $k > 1$. Choose $r > 1$ small enough so that the neighborhood $V(r)$ of §17.2 is contained in $K \cup U$, and let $s = \sqrt[d]{r}$. Consider the (half-closed) annulus

$$A_0 = V(r) \smallsetminus V(s) ,$$

which is isomorphic under ϕ to the "round" annulus $\bar{D}_r \smallsetminus \bar{D}_s$. Then $V(r) \smallsetminus K$ can be expressed as the countable union $A_0 \cup A_1 \cup \cdots$, where A_n is the annulus $f^{-n}(A_0)$. Similarly, $\bar{D}_r \smallsetminus \bar{D}_1$ is the union of a corresponding family of round annuli.

Each radial line segment in $\bar{D}_r \smallsetminus \bar{D}_s$ corresponds to a smooth curve in A_0, which is a segment of some external ray. Let ℓ_0 be the maximum of the lengths of these curve segments in A_0, and let ℓ_n be the maximum length of the corresponding curve segments in A_n. Since $f^{\circ n}$ maps A_n onto A_0, preserving this foliation by external rays, and stretching all distances by at least k^n, it follows that $\ell_n \le \ell_0 / k^n$. Therefore, the total length of any external ray intersected with $V(r)$ is uniformly bounded by the quantity $\ell_0(1 + k^{-1} + k^{-2} + \cdots) = \ell_0/(k-1) < \infty$. It then follows easily that the maps

$$e^{i\theta} \mapsto \psi(re^{i\theta})$$

from the unit circle onto $\partial V(r)$ converge uniformly as $r \to 1$ to the required limit mapping, which carries the unit circle continuously onto the Julia set. $\square$

§18. External Rays and Periodic Points.

To fix our ideas and simplify the discussion, let us assume that the filled Julia set K is connected. (For the general case, see [DH2].) By 17.3 the complement $\mathbf{C}\smallsetminus K$ is isomorphic to the complement $\mathbf{C}\smallsetminus\bar{D}$ under a conformal isomorphism

$$\phi : \mathbf{C}\smallsetminus K \xrightarrow{\approx} \mathbf{C}\smallsetminus\bar{D}$$

which is compatible with the dynamics:

$$\phi(f(z)) = \phi(z)^d .$$

Here ϕ is unique up to multiplication by a $(d-1)$-st root of unity. In practice, it is customary to make a linear change of coordinates so that the polynomial f is *monic*, ie., has leading coefficient equal to $+1$. There is then a preferred choice of Böttcher coordinate $w=\phi(z)$, determined by the requirement that $w/z\to 1$ as $|z|\to\infty$.

Definition. Let $\psi : \mathbf{C}\smallsetminus\bar{D} \xrightarrow{\approx} \mathbf{C}\smallsetminus K$ be the inverse map. The image under ψ of a radial line

$$\{re^{2\pi it} : r>1\}$$

in $\mathbf{C}\smallsetminus\bar{D}$ is called the *external ray* R_t at *angle* t in $\mathbf{C}\smallsetminus K$. Note that our *angles* are elements of $\mathbf{R}/\mathbf{Z}$, measured in fractions of a full turn and not in radians. Evidently

$$f(R_t) = R_{dt} .$$

That is: *f maps the external ray at angle t to the external ray at angle dt.* In particular, if t is a fraction of the form $m/(d-1)$ then $f(R_t)=R_t$.

By definition, the external ray R_t *lands* at some point z_t, which necessarily belongs to the Julia set J, if the points $\psi(re^{2\pi it})\in\mathbf{C}$ tend to a well defined limit z_t as $r\to 1$. If J, or equivalently K, is locally connected, then according to 17.4 *every* external ray lands, at a point z_t which depends continuously on t. Douady and Hubbard call the resulting parametrization $t\mapsto z_t$ the *Carathéodory loop* on the Julia set. However, in this section we will usually not assume that J is locally connected.

Remark. A priori, it is possible that every external ray may land even if K is not locally connected. An example of a compact set (but not a filled Julia set) with this property is shown in Figure 16. Evidently, in such a case, the correspondence $t\mapsto z_t$ cannot be continuous.

Definition. An external ray R_t will be called *rational* if its angle $t\in\mathbf{R}/\mathbf{Z}$ is rational, and *periodic* if t is periodic under multiplication by the degree d, so that $d^n t\equiv t \pmod 1$ for some $n\ge 1$.

Note that R_t is eventually periodic under multiplication by d if and only if t is rational, and is periodic if and only if the number t is rational with denominator relatively prime to d. (If t is rational with denominator m, then the successive images of R_t under f have angles $dt, d^2t, d^3t, \ldots \pmod 1$ with denominators dividing m. Since there are only finitely many such fractions modulo 1, this sequence must eventually repeat. In the special case where m is relatively prime to d, the fractions with denominator m

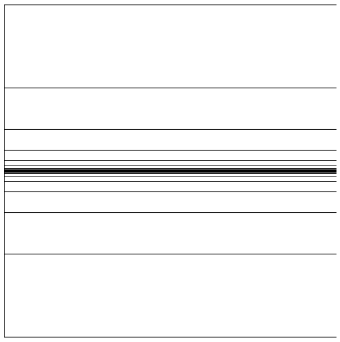


Figure 16. The double comb $[0,2] \times \{0\,,\, \pm 2^{-n}\} \,\cup\, 0 \times [-1,1]$

are permuted under multiplication by d modulo 1, so the landing point z_t is actually periodic.)

The following result is due to Sullivan, Douady and Hubbard. We do not assume local connectivity.

18.1. Theorem. *If $K(f)$ is connected, then every periodic external ray lands at a periodic point which is either repelling or parabolic.*

The converse result, due to Douady and Yoccoz, is more difficult:

18.2. Theorem. *Still assuming that $K(f)$ is connected, every repelling or parabolic periodic point is the landing point of at least one external ray, which is necessarily periodic.*

In fact we will only prove the parabolic case. For the repelling case, which is quite similar, the reader is referred to Petersen. Here is a complementary result.

18.3. Lemma. *If a periodic ray lands at z_t, then only finitely many external rays, all periodic of the same period, can land at z_t.*

The following is an immediate consequence of 18.1.

18.4. Corollary. *If t is rational but not periodic, then R_t lands at a point z_t which is eventually periodic but not periodic.*

As an example, Figure 17 shows the Julia set for the cubic map $f(z) = z^3 - iz^2 + z$. Here the rays R_0 and $R_{1/2}$ are mapped into themselves by f. Both must land at the parabolic fixed point $z = 0$, since the only other fixed point, at $z = i$, is superattracting and hence does not belong to the Julia set. On the other hand, the $1/6$, $1/3$, $2/3$ and $5/6$ rays have denominator divisible by 3. These four rays land at the two disjoint pre-images of zero.

The following result of Sullivan and Douady is closely related. (Sullivan 1983. For an independent proof, see Lyubich, p. 85.)

18.5. Theorem. *If the Julia set J of a polynomial map f is locally connected, then every periodic point in J is either repelling or parabolic.*

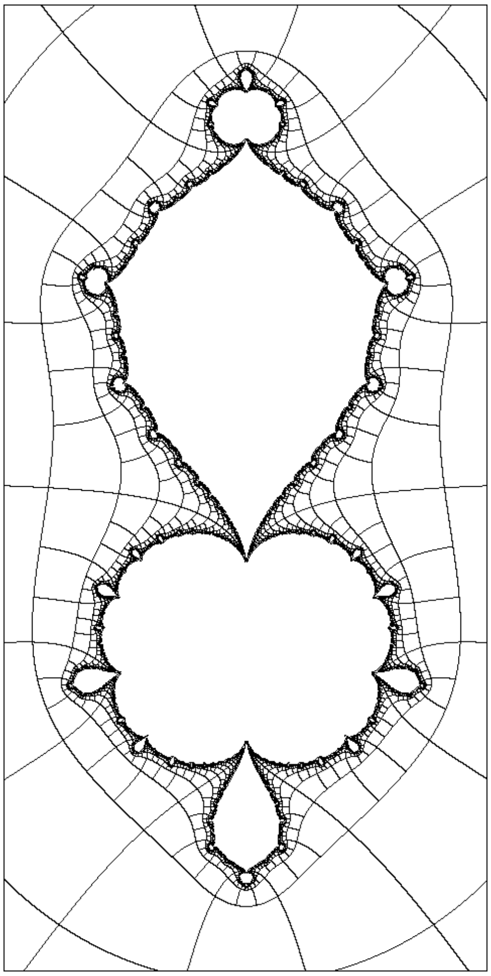


Figure 17. Julia set for $z \mapsto z^3 - iz^2 + z$
with some external rays and equipotentials indicated.

Recall from §8 that a *Cremer point* can be characterized as a periodic point which belongs to the Julia set but is neither repelling nor parabolic. Thus the following is a completely equivalent formulation.

18.6. Corollary. *If f is a polynomial map with a Cremer point, then the Julia set $J(f)$ is not locally connected.*

The proofs begin as follows.

18.7. Lemma. *The ray R_t lands at a single point z_t of the Julia set if and only if R_{dt} lands at a single point z_{dt}. Furthermore, the image $f(z_t)$ is necessarily equal to z_{dt}.*

Proof. The set $L(t)$ consisting of all limit points of $\psi(re^{2\pi it})$ as r tends to 1 from above is certainly compact and connected, with $f(L(t)) = L(dt)$. It follows easily that $L(t)$ is a single point if and only if $L(dt)$ is a single point. □

Thus if the ray R_t is periodic, and if R_t lands at a point z_t , then it follows that z_t is a periodic point of f .

Proof of 18.3. First consider the special case $t = 0$. Then f maps the ray R_0 to itself, hence $f(z_0) = z_0$. If another ray R_s also lands at z_0 , then we will prove that s is a rational number of the form $j/(d-1)$. This will show that there are at most $d-1$ external rays landing at z_0 , and that they are all periodic (and in fact fixed) under multiplication by d , since the numbers $s = j/(d-1)$ are exactly the solutions to the congruence $s \equiv ds \pmod 1$ where d is the degree.

If a ray R_s which is not of this form lands at z_0 , then we have $s \not\equiv ds \pmod 1$. Define a sequence of angles $0 \le s_n < 1$ by the condition that $s_0 \equiv s$ and $s_{n+1} \equiv ds_n$ modulo 1. By hypothesis, the numbers 0 , s_0 and s_1 are distinct. Suppose, to fix our ideas, that $0 < s_0 < s_1 < 1$. Since f is a local diffeomorphism near z_0 , it must preserve the cyclic order of the rays landing at z_0 . Thus the images of 0 , s_0 and s_1 must satisfy $0 < s_1 < s_2 < 1$. Continuing inductively, it follows that $0 < s_0 < s_1 < s_2 < \cdots < 1$. Thus the s_n must converge to some angle s_∞ , which is necessarily a fixed point of the map $s \mapsto ds \pmod 1$. But this is impossible, since this map has only strictly repelling fixed points.

More generally, if R_t is any ray which is mapped into itself by f , so that $t = j/(d-1)$, then we can translate all angles by $-t$ and apply the argument above. The proof in the general case now follows easily. We can replace f by some iterate $g = f^{\circ n}$, where $d^n s \equiv s \pmod 1$ so that $g(R_s) = R_s$. The argument then proceeds as above. This proves 18.3. □

The proof of 18.1 begins as follows. We will make use of the Poincaré metric on $\mathbf{C} \smallsetminus K$. *Note first that the map f is a local isometry for this metric.* In fact the universal covering of $\mathbf{C} \smallsetminus \bar{D}$ is isomorphic to the right half-plane $\{W = U + iV : U > 0\}$ under the exponential map, and the map f on $\mathbf{C} \smallsetminus K$ corresponds to the d-th power map on $\mathbf{C} \smallsetminus \bar{D}$, which corresponds to the automorphism $W \mapsto dW$ on this right half-plane.

We suppose that t has denominator prime to d , so that some iterate $g = f^{\circ n}$ maps R_t to itself. This ray R_t can be expressed as a union

$$R_t \;=\; \cdots \cup S(-1) \cup S(0) \cup S(1) \cup \cdots$$

of finite segments, where g maps each $S(j)$ isomorphically onto $S(j+1)$. In fact let $S(j)$ correspond under ϕ to the set of $re^{2\pi it}$ with $d^{jn} \le \log r \le d^{(j+1)n}$. Thus the $S(j)$ all have the same Poincaré length. But as $j \to -\infty$ the segments $S(j)$ converge towards the boundary J of $\mathbf{C} \smallsetminus K$. Hence, according to §2.4, the Euclidean length of $S(j)$ must converge to zero. This means that any limit point of the sequence of sets $S(j)$, as $j \to -\infty$, must be a fixed point of g . But, as noted in the proof of 18.7, the set of all limit points must be connected. Since g has only finitely many fixed points, this proves that the ray R_t lands at a single fixed point z_t of the map $g = f^{\circ n}$.

Now let us apply the Snail Lemma of §13. After translating coordinates so that our fixed landing point lies at the origin, the map g and the path $x \mapsto \psi(\exp(d^{nx} + 2\pi it))$ will satisfies the hypothesis of 13.3. Hence it follows that this fixed point z_t of g is either repelling, or parabolic with multiplier $g'(z_t) = 1$. In terms of the original map f, it follows that z_t is a periodic point, and that its periodic orbit is either repelling or has a root of unity as multiplier. This proves 18.1. □

Proof of 18.4. Consider a rational angle t with denominator which is not relatively prime to d. In this case the multiple dt will have smaller denominator, but the sequence of angles $d^n t$ modulo 1 must certainly be eventually periodic. Thus some forward image $f^{\circ n} R_t$ must have a well defined landing point. But this implies that R_t itself has a well defined landing point by 18.7. □

The proof of 18.5 and 18.6 will be based on 18.1, together with the following. Let $n \geq 2$ be an integer.

18.8. Lemma. *Let $L \subset \mathbf{R}/\mathbf{Z}$ be a compact set which is mapped homeomorphically onto itself by the map $t \mapsto nt (\mathrm{mod}\ 1)$. Then L is finite.*

Proof. In fact we will prove the following more general statement. Let X be a compact metric space with distance function $\rho(x, y)$, and let $h : X \to X$ be a homeomorphism which is expanding in the following sense: There should exist numbers $\epsilon > 0$ and $k > 1$ so that

$$\rho(h(x), h(y)) \ \geq \ k\rho(x, y)$$

whenever $\rho(x, y) < \epsilon$. Then we will show that X is finite. Evidently this hypothesis is satisfied with $X = L \subset \mathbf{R}/\mathbf{Z}$, so this argument will prove 18.8.

Since h^{-1} is uniformly continuous, we can choose $\delta > 0$ so that $\rho(x, y) < \epsilon$ whenever $\rho(h(x), h(y)) < \delta$. But this implies that $\rho(x, y) < \delta/k$. Since X is compact, we can choose some finite number, say n, of balls of radius δ which cover X. Applying h^{-p}, we obtain n balls of radius δ/k^p which cover X. Since p can be arbitrarily large, this proves that X can have at most n distinct points. □

Proof of 18.5. Let z_0 be any periodic point. After replacing the given polynomial map by some iterate, we may assume that z_0 is fixed by f. Furthermore, after a linear change of coordinates, we may assume that f is monic. Since $J(f)$ is locally connected, according to 17.4 the inverse Riemann map

$$\psi \,:\, \hat{\mathbf{C}} \smallsetminus \bar{D} \ \stackrel{\approx}{\longrightarrow} \ \hat{\mathbf{C}} \smallsetminus K(f)$$

extends continuously over the boundary, yielding a map from the unit circle onto $J(f)$ which semi-conjugates the d-th power map on the circle to the map f on $J(f)$. In other words, identifying the unit circle with $\mathbf{R}/\mathbf{Z}$, we have a map

$$\Psi : \mathbf{R}/\mathbf{Z} \ \to \ J(f)$$

which satisfies $\Psi(dt) = f(\Psi(t))$, where d is the degree. Let X be the compact set $\Psi^{-1}(z_0) \subset \mathbf{R}/\mathbf{Z}$. Since z_0 is fixed by f, it follows immediately that the map $t \mapsto td (\mathrm{mod}\ 1)$ carries X homeomorphically onto itself. Hence X is finite by 18.8, and the conclusion follows. □

We will prove the parabolic case of Theorem 18.2 in the following slightly sharper form. For the proof in the repelling case, the reader is referred to Petersen.

18.9. Theorem. *Let f be a polynomial of degree $d \geq 2$ with $K(f)$ connected, and let z_0 be a parabolic fixed point with multiplier $\lambda = 1$. Then for each repelling petal P at z_0 there exists at least one external ray R_t which lands at z_0 through P, where the angle t is a rational number of the form $m/(d-1)$, so that R_t is mapped onto itself by f.*

Very roughly, we will show that the basin $\mathbf{C} \smallsetminus K$ of the point at infinity corresponds to a union of one or more annuli in the Écalle cylinder P/f. Each of these annuli contains a unique simple closed geodesic Γ (think of placing a rubber band around a napkin ring); and Γ will lift to the required external ray in $\mathbf{C} \smallsetminus K$. Before giving details, let us look at two examples.

Example 1. Consider the map $f(z) = z^2 + \exp(2\pi i \cdot 3/7)z$ of Figure 8. Here the seventh iterate is a map of degree $2^7 = 128$ with seven repelling petals about the origin. Hence there must be at least seven external rays landing at the origin, and their angles must be rational numbers of the form $m/127$, so as to map into themselves under multiplication by 128 modulo 1. In fact a little experimentation shows that only the ray with angle $21/127$ and its successive iterates under doubling modulo 1 will fit in the right order around the origin. (Compare Goldberg.) Thus there are just seven rays which land at zero, one in each repelling petal. The numerators of the corresponding angles are $21, 37, 41, 42, 74, 82, 84$.

Example 2. Now consider the cubic map $f(z) = z^3 - iz^2 + z$ of Figure 17. Evidently there is only one repelling petal at the parabolic fixed point $z = 0$. Yet we saw above (following 18.4) that the two rays R_0 and $R_{1/2}$ must both land at this point.

The proof of 18.9 begins as follows. According to §7.7, there is an essentially unique map $\beta = \alpha^{-1}$ which is defined and univalent in some left half-plane $\mathcal{R}(w) < c$, taking values in the repelling petal P, and which satisfies the Abel equation $\beta(w+1) = f(\beta(w))$ or equivalently:

$$\beta(w) = f(\beta(w-1)).$$

As in 7.11, we can extend to an analytic map from $\mathbf{C}$ to $\mathbf{C}$ satisfying this same equation by setting $\beta(w) = f^{\circ n}(\beta(w-n))$ for large n. **Definition:** Let K_P be the inverse image of $K = K(f)$ under β. Since K is closed and f-invariant, it follows that K_P is closed and periodic: $K_P = K_P + 1$. Furthermore, it is not hard to show that K_P contains all points $w = u + iv$ with $|v|$ large. For points which are far from the real axis must correspond, under β, to points which belong to one of the two neighboring attracting petals of f.

18.10. Yoccoz Lemma. *Each component U of $C \smallsetminus K_P$ is a universal covering of $C \smallsetminus K$ with projection map β.*

Proof. Consider the topological space E consisting of all left infinite sequences $\mathbf{w} = (\ldots, w_{-2}, w_{-1}, w_0)$ of points in the Fatou set $\mathbf{C} \setminus J$ satisfying

$$\cdots \overset{f}{\mapsto} w_{-2} \overset{f}{\mapsto} w_{-1} \overset{f}{\mapsto} w_0$$

and $\lim_{n\to-\infty} w_n = z_0$. We can give each component of E the structure of a Riemann surface in such a way that the projection $\pi : (\ldots, w_{-1}, w_0) \mapsto w_0$ is a local conformal isomorphism. In fact, for any such $\mathbf{w}$, we can choose $N < 0$ so that w_n belongs to some fixed repelling petal P for $n \le N$. The coordinate $w_N \in P$ then serves as a local uniformizing parameter for E, and π corresponds to the map $f^{|N|} : w_N \mapsto w_0$. (It is important here that there are no critical points in the finite part of the Fatou set.)

To show that $\pi : E \to \mathbf{C} \setminus K$ is actually a covering map, we start with any simply connected neighborhood V of $\pi(\mathbf{w}) = w_0$ in $\mathbf{C} \setminus K$. Then we can find a unique single valued branch f_0^{-n} of f^{-n} on V which maps w_0 to w_{-n}. These maps f_0^{-n} must constitute a normal family on V, since they take values in $\mathbf{C} \setminus K$. Since the sequence f_0^{-n} converges uniformly to z_0 throughout some small neighborhood of w_0, it follows easily that this sequence converges to z_0 throughout the given neighborhood V. Thus for each $w \in V$ the sequence $\mathbf{w}(w) = \big(\cdots, f_0^{-2}(w), f_0^{-1}(w), w\big)$ belongs to E, hence V is evenly covered under the projection π.

It is not difficult to check that $\mathbf{C} \setminus K_P$ embeds naturally as an open and closed subset of our covering space E. But each component U of $\mathbf{C} \setminus K_P$ is simply connected. For otherwise, there would be a compact component of $\mathbf{C} \setminus U$ which, after left translation by a large integer, would embed into a compact of K in the petal P. This is impossible, since K is connected. Thus each such U is a universal covering surface of $\mathbf{C} \setminus K$. □

Remark. Although U is a universal covering of $\mathbf{C} \setminus K$, the proof does not tell us just how the free cyclic group of deck transformations acts on U. We will show in 18.12 that there is also a natural action of the free cyclic group $\mathbf{Z}$ by integer translations of U, but that action is quite different, and has a quotient surface $U/\mathbf{Z}$ which is an annulus, naturally embedded in the Écalle cylinder P/f.

Let $\hat{G} = G \circ \beta$ be the potential function on $\mathbf{C} \setminus K$ lifted to $\mathbf{C} \setminus K_P$, so that $\hat{G}(w+1) = \hat{G}(w)d$ where d is the degree. Let $\hat{G}_0$ be the maximum value of $\hat{G}$ on the imaginary axis, attained say at w_0. Then $\hat{G}(w) \le G_0$ for $\mathcal{R}(w) \le 0$ by the maximum modulus principle for harmonic maps, hence

$$G(w) \le d^n \hat{G}_0 \qquad \text{whenever} \qquad \mathcal{R}(w) \le n. \tag{$*$}$$

18.11. Lemma. *The real part of w must take values tending to ∞ within each component U.*

This follows, since $\hat{G}$ must be unbounded on each U by 18.10. □

18.12. Main Lemma. *Each such component U is periodic: $U = U + 1$.*

Proof. Otherwise the translates $U - n$ would have to be pairwise disjoint. First consider the component U on which the maximum $\hat{G}_0$ is attained. Let k_n be the vertical width of the intersection of $U - n$ with the imaginary axis. Then the sum of k_n is finite, so k_n tends to zero as $n \to +\infty$. But k_n is also the vertical width of the intersection of U with the line $\mathcal{R}(w) = n$. *Since these numbers tend to zero, the Poincaré distance of the imaginary axis from* $\mathcal{R}(w) = n$ *within* U *must grow more than linearly with* n *.*

On the other hand, an orthogonal trajectory of the curves $\hat{G} = \text{constant}$, of Poincaré arclength $n \log(d)$ within U , will get from w_0 on the imaginary axis to a point w_n with $\hat{G}(w_n) = (d^n)\hat{G}_0$. By $(*)$ above, the real part of w_n is $\geq n$. Thus this distance grows at most linearly with n . This contradiction proves the Main Lemma 18.12 when w_0 belongs to U . To get a proof for arbitrary U , we must simply use the union of translates $U - n$ in place of the entire open set $C \smallsetminus K_P$ in the argument above. This proves 18.12. $\square$

Proof of Theorem 18.9. We know that each component U of $\mathbf{C} \smallsetminus K_P$ is a universal covering of $\mathbf{C} \smallsetminus K$, and that the unit translation $w \mapsto w + 1$ of U corresponds to the map f on $\mathbf{C} \smallsetminus K$. Furthermore, it follows easily from Corollary 7.5 that the complement $\mathbf{C} \smallsetminus K_P$ is non-vacuous, so that there exists at least one such component U . Since the imaginary coordinate is bounded throughout U , the quotient $U/\mathbf{Z}$ is clearly an annulus. Hence it has a unique simple closed geodesic with respect to the Poincaré metric. (Compare Problem 2-3. We can model the annulus as the upper half-plane H modulo the identification $w \sim t_0 w$ for some fixed $t_0 > 1$. The imaginary axis in H then covers the unique closed geodesic.)

This closed geodesic $\Gamma \subset U/\mathbf{Z}$ lifts to an infinite $\mathbf{Z}$-invariant geodesic $\tilde{\Gamma} \subset U$, and hence to an infinite f-invariant geodesic $\Gamma' = \beta(\tilde{\Gamma})$ in $\mathbf{C} \smallsetminus K$. In the negative direction, this geodesic Γ' converges to the parabolic fixed point z_0 . For β was constructed in such a way that for any compact subset $L \subset U$ the images $\beta(L - n)$ converge to z_0 as $n \to \infty$. On the other hand, in the positive direction Γ' converges to the point at infinity, since every compact subset $L \subset U$ has the property that the images $\beta(L+n) = f^{\circ n} \circ \beta(L)$ converge to ∞ as $n \to \infty$. But a Poincaré geodesic in $\mathbf{C} \smallsetminus K$ which leads to the point at infinity is necessarily an external ray. $\square$

Appendix A. Theorems from Classical Analysis.

This appendix will describe some miscellaneous theorems from classical complex variable theory. We first complete the arguments from §8.6 and §15.3 by proving Jensen's inequality and the Riesz brothers' theorem. We then describe results from the theory of univalent functions, due to Gronwall and Bieberbach, in order to prove the Koebe Quarter Theorem for use in Appendix G. (By definition, a function of one complex variable is called *univalent* if it is holomorphic and injective.)

We begin with a discussion of Jensen's inequality. (J. L. W. V. Jensen was the president of the Danish Telephone Company, and a noted amateur mathematician.) Let $f : D \to \mathbf{C}$ be a holomorphic function on the open disk which is not identically zero. Given any radius $0 < r < 1$, we can form the average

$$A(f, r) \;=\; \frac{1}{2\pi}\int_0^{2\pi} \log|f(re^{i\theta})|d\theta$$

of the quantity $\log|f(z)|$ over the circle $|z| = r$.

A.1. Jensen's Inequality. *This average $A(f,r)$ is monotone increasing (and also convex upwards) as a function of r. Hence $A(f,r)$ either converges to a finite limit or diverges to $+\infty$ as $r \to 1$.*

In fact the proof will show something much more precise.

A.2. Lemma. *If we consider $A(f,r)$ as a function of $\log r$, then it is piecewise linear, with slope $dA(f,r)/d\log r$ equal to the number of roots of f inside the disk D_r of radius r, where each root is to be counted with its appropriate multiplicity.*

In particular, the function $A(f,r)$ is determined, up to an additive constant, by the location of the roots of f. To prove this Lemma, note first that we can write $d\theta = dz/iz$ around any loop $|z| = r$. Consider an annulus $\mathcal{A} = \{z : r_0 < |z| < r_1\}$ which contains no zeros of f. According to the "Argument Principle", the integral

$$n \;=\; \frac{1}{2\pi i}\oint_{|z|=r} d\log f(z) \;=\; \frac{1}{2\pi i}\oint \frac{f'(z)dz}{f(z)}$$

measures the number of zeros of f inside the disk D_r. It follows that the difference $\log f(z) - \log z^n$ can be defined as a single valued holomorphic function throughout this annulus $\mathcal{A}$. Therefore, the integral of $(\log f(z) - \log z^n)dz/iz$ around a loop $|z| = r$ must be independent of r, as long as $r_0 < r < r_1$. Taking the real part, it follows that the difference $A(f,r) - A(z^n, r)$ is a constant, independent of r. Since $A(z^n, r) = n\log r$, this proves that the function $\log r \mapsto A(f,r)$ is linear with slope n for $r_0 < r < r_1$.

Finally, note that the average $A(f,r)$ takes a well defined finite value even when f has one or more zeros on the circle $|z| = r$, since the singularity of $\log|f(z)|$ at a zero of f is relatively mild. Continuity of $A(f,r)$ as r varies through such a singularity is not difficult, and will be left to the reader. □

A.3. Theorem of F. and M. Riesz. *Suppose that* $f : D \to \mathbf{C}$ *is bounded and holomorphic on the open unit disk. If the radial limit*

$$\lim_{r\to 1} f(re^{i\theta})$$

exists and is equal to zero for θ *belonging to a set* $E \subset [0, 2\pi]$ *of positive Lebesgue measure, then* f *must be identically zero.*

(Compare the discussion of Fatou's Theorem in §15.3.)

Proof. Let $E(\epsilon, \delta)$ be the measurable set consisting of all $\theta \in E$ such that

$$|f(re^{i\theta})| < \epsilon \qquad \text{whenever} \qquad 1 - \delta < r < 1\,.$$

Evidently, for each fixed ϵ, the union of the nested family of sets $E(\epsilon, \delta)$ contains E. Therefore, the Lebesgue measure $\ell(E(\epsilon, \delta))$ must tend to a limit which is $\geq \ell(E)$ as $\delta \to 0$. In particular, given ϵ we can choose δ so that $\ell(E(\epsilon, \delta)) > \ell(E)/2$. Now consider the average $A(f, r)$ of Jensen's Inequality, where $r > 1 - \delta$. Multiplying f by a constant if necessary, we may assume that $|f(z)| < 1$ for all $z \in D$. Thus the expression $\log |f(re^{i\theta})|$ is less than or equal to zero everywhere, and less than or equal to $\log \epsilon$ throughout a set of measure at least $\ell(E)/2$. This proves that

$$2\pi A(f, r) \;<\; \log(\epsilon)\, \ell(E)/2$$

whenever r is sufficiently close to 1. Since ϵ can be arbitrarily small, this implies that $\lim_{r\to 1} A(f, r) = -\infty$, which contradicts A.1 unless f is identically zero. □

Now consider the following situation. Let K be a compact connected subset of $\mathbf{C}$, and suppose that the complement $\mathbf{C} \smallsetminus K$ is conformally diffeomorphic to the complement $\mathbf{C} \smallsetminus \bar{D}$.

A.4 Gronwall Area Inequality. *Let* $\phi : \mathbf{C} \smallsetminus \bar{D} \to \mathbf{C} \smallsetminus K$ *be a conformal isomorphism and let*

$$\phi(w) \;=\; b_1 w + b_0 + b_{-1}/w + b_{-2}/w^2 + \cdots$$

be its Laurent expansion. Then $|b_1| \geq |b_{-1}|$, *with equality if and only if* K *is a straight line segment.*

Proof. For any $r > 1$ consider the image under ϕ of the circle $|w| = r$. This will be some embedded circle in $\mathbf{C}$ which encloses a region of area say $A(r)$. We can compute this area by Green's Theorem, as follows. Let $\phi(re^{i\theta}) = z = x + iy$. Then

$$A(r) \;=\; \oint x dy \;=\; -\oint y dx \;=\; \tfrac{1}{2i} \oint \bar{z} dz\,,$$

to be integrated around the image of $|w| = r$. Substituting the Laurent series $z = \sum_{n\leq 1} b_n w^n$, with $w = r^n e^{ni\theta}$, this yields

$$A(r) \;=\; \tfrac{1}{2} \sum_{m,n\le 1} n\bar{b}_m b_n r^{m+n} \oint e^{(n-m)i\theta}\,d\theta\,.$$

Since the integral equals 2π if $m=n$, and is zero otherwise, we obtain

$$A(r) \;=\; \pi \sum_{n\le 1} n|b_n|^2 r^{2n}\,.$$

Therefore, taking the limit as $r\to 1$, we obtain the simple formula

$$A(1) \;=\; |b_1|^2 - |b_{-1}|^2 - 2|b_{-2}|^2 - 3|b_{-3}|^2 - \cdots \qquad (A.5)$$

for the area (that is the 2-dimensional Lebesgue measure) of the compact set K. Evidently it follows that $|b_1| \ge |b_{-1}|$. Furthermore, if equality holds then all of the remaining coefficients must be zero: $b_{-2}=b_{-3}=\cdots=0$. After a rotation of the w coordinate and a linear transformation of the z coordinate, the Laurent series will reduce to the simple formula $z = w + w^{-1}$. As noted in §5, Example 2, this transformation carries $\mathbf{C}\smallsetminus\bar{D}$ diffeomorphically onto the complement of the interval $[-2,2]$. $\square$

Now consider an open set $U\subset\mathbf{C}$ which contains the origin and is conformally isomorphic to the open disk.

A.6. Bieberbach Theorem. *If $\psi: D\to U$ is a conformal isomorphism with power series expansion $\psi(\eta) = \sum_{n\ge 1} a_n\eta^n$, then $|a_2|\le 2|a_1|$, with equality if and only if $\mathbf{C}\smallsetminus U$ is a closed half-line pointing towards the origin.*

Remark. The Bieberbach conjecture, recently proved by DeBrange, asserts that $|a_n|\le n|a_1|$ for all $n\ge 2$. Again, equality holds if $\mathbf{C}\smallsetminus U$ is a closed half-line pointing towards the origin.

Proof of A.6. After composing ψ with a linear transformation, we may assume that $a_1=1$. Let us set $\eta = 1/w^2$, so that each point $\eta\ne 0$ in D corresponds to two points $w\in\mathbf{C}\smallsetminus\bar{D}$. Similarly, set $\psi(\eta)=\zeta=1/z^2$, so that each $\zeta\ne 0$ in U corresponds to two points z in some centrally symmetric neighborhood N of infinity. A brief computation shows that ψ corresponds to a Laurent series

$$w\mapsto z = 1/\sqrt{\psi(1/w^2)} \;=\; w - \tfrac{1}{2}a_2/w + (\text{higher terms})$$

which maps $\mathbf{C}\smallsetminus\bar{D}$ diffeomorphically onto N. Thus $|a_2|\le 2$ by Gronwall's Inequality, with equality if and only if N is the complement of a line segment, necessarily centered at the origin. Expressing this condition on N in terms of the coordinate $\zeta = 1/z^2$, we see that equality holds if and only if U is the complement of a half-line pointing towards the origin. $\square$

A.7. Koebe-Bieberbach Quarter Theorem. *Again suppose that the map*

$$\eta \mapsto \psi(\eta) = a_1\eta + a_2\eta^2 + \cdots$$

carries the unit disk D diffeomorphically onto an open set $U \subset \mathbf{C}$. Then the distance r between the origin and the boundary of U satisfies

$$\tfrac{1}{4}|a_1| \le r \le |a_1| .$$

Here the first equality holds if and only if $\mathbf{C}\smallsetminus U$ is a half-line pointing towards the origin, and the second equality holds if and only if U is a disk centered at the origin.

In particular, in the special case $a_1 = 1$ the open set U necessarily contains the disk of radius $1/4$ centered at the origin. The left hand inequality was conjectured and partially proved by Koebe, and later completely proved by Bieberbach. The right hand inequality is an easy consequence of the Schwarz Lemma.

Here is an interesting restatement of the Quarter Theorem. Let $ds = \gamma(z)|dz|$ be the Poincaré metric on the open set U, and let $r = r(z)$ be the distance from z to the boundary of U.

A.8. Corollary. *If $U \subset \mathbf{C}$ is simply connected, then the Poincaré metric $ds = \gamma(z)|dz|$ on U agrees with the metric $|dz|/r(z)$ up to a factor of two in either direction. That is*

$$\frac{1}{2r(z)} \le \gamma(z) \le \frac{2}{r(z)}$$

for all $z \in U$. Again, the left equality holds if and only if $\mathbf{C}\smallsetminus U$ is a half-line pointing towards the point $z \in U$, and the right equality holds if and only if U is a round disk centered at z.

As an example, if U is a half-plane, then the Poincaré metric precisely agrees with the $1/r$ metric $|dz|/r$.

Proof of A.7. Without loss of generality, we may assume that $a_1 = 1$. If $z_0 \in \partial U$ be a boundary point with minimal distance r from the origin, then we must prove that $\frac{1}{4} \le r \le 1$. We will compose ψ with the linear fractional transformation $z \mapsto z/(1-z/z_0)$ which maps z_0 to infinity. Then the composition has the form

$$\eta \mapsto \psi(\eta)/(1-\psi(\eta)/z_0) = \eta + (a_2 + 1/z_0)\eta^2 + \cdots .$$

By Bieberbach's Theorem we have $|a_2| \le 2$ and $|a_2 + 1/z_0| \le 2$, hence $|1/z_0| = 1/r \le 4$, or $r \ge 1/4$. Here equality holds only if $|a_2| = 2$ and $|a_2 + 1/z_0| = 2$. The exact description of U then follows easily.

Now suppose that $r \ge 1$. Then the inverse mapping ψ^{-1} is defined and holomorphic throughout the unit disk D, and takes values in D. Since its derivative at zero is 1, it follows from the Schwarz Inequality that ψ is the identity map. □

Appendix B. Length-Area-Modulus Inequalities.

The most basic length-area inequality is the following. Let $I^2 \subset \mathbf{C}$ be the open unit square consisting of all $z = x + iy$ with $0 < x < 1$ and $0 < y < 1$. By a *conformal metric* on I^2 we mean a metric of the form

$$ds = \rho(z)|dz|$$

where $z \mapsto \rho(z) > 0$ is any strictly positive continuous real valued function on the open square. In terms of such a metric, the *length* of a smooth curve $\gamma : (a, b) \to I^2$ is defined to be the integral

$$\mathbf{L}_\rho(\gamma) = \int_a^b \rho(\gamma(t))|d\gamma(t)| ,$$

and the *area* of a region $U \subset I^2$ is defined to be

$$\mathrm{area}_\rho(U) = \int\int_U \rho(x + iy)^2 dx\, dy .$$

In the special case of the Euclidean metric $ds = |dz|$, with $\rho(z)$ identically equal to 1, the subscript ρ will be omitted.

Theorem B.1. *If the integral* $\mathrm{area}_\rho(I^2)$ *over the entire square is finite, then for Lebesgue almost every* $y \in (0, 1)$ *the length* $\mathbf{L}_\rho(\gamma_y)$ *of the horizontal line* $\gamma_y : t \mapsto (t, y)$ *at height* y *is finite. Furthermore, there exists* y *so that*

$$\mathbf{L}_\rho(\gamma_y)^2 \le \mathrm{area}_\rho(I^2) . \tag{1}$$

In fact, the set consisting of all $y \in (0, 1)$ *for which this inequality is satisfied has positive Lebesgue measure.*

Remark 1. Evidently this inequality is best posible. For in the case of the Euclidean metric $ds = |dz|$ we have

$$\mathbf{L}(\gamma_y)^2 = \mathrm{area}(I^2) = 1 .$$

Remark 2. It is essential here that we use a square, rather than a rectangle. If we consider instead a rectangle R with base Δx and height Δy, then the corresponding inequality would be

$$\mathbf{L}_\rho(\gamma_y)^2 \le \frac{\Delta x}{\Delta y}\mathrm{area}_\rho(R) \tag{2}$$

for a set of y with positive measure.

Proof of B.1. We use the Schwarz inequality

$$\left(\int_a^b f(x)g(x)\, dx\right)^2 \le \left(\int_a^b f(x)^2\, dx\right) \cdot \left(\int_a^b g(x)^2\, dx\right) ,$$

which says (after taking a square root) that the inner product of any two vectors in the Euclidean vector space of square integrable real functions on an interval is less than or equal to the product of their norms. We may as well consider the more general case of a rectangle $R = (0,\, \Delta x) \times (0,\, \Delta y)$. Taking $f(x) \equiv 1$ and $g(x) = \rho(x, y)$ for some fixed y , we obtain

$$\left(\int_0^{\Delta x} \rho(x, y)\, dx \right)^2 \;\le\; \Delta x \int_0^{\Delta x} \rho(x, y)^2\, dx \;,$$

or in other words

$$L_\rho(\gamma_y)^2 \;\le\; \Delta x \int_0^{\Delta x} \rho(x, y)^2 dx \;,$$

for each constant height y . Integrating this inequality over the interval $0 < y < \Delta y$ and then dividing by Δy , we get

$$\frac{1}{\Delta y} \int_0^{\Delta y} L_\rho(\gamma_y)^2 dy \;\le\; \frac{\Delta x}{\Delta y}\, \mathrm{area}_\rho(A) \;. \tag{3}$$

In other words, the *average* over all y in the interval $(0,\, \Delta y)$ of $L_\rho(\gamma_y)^2$ is less than or equal to $\frac{\Delta x}{\Delta y}\, \mathrm{area}_\rho(A)$. Further details of the proof are straightforward. $\square$

Now let us form a cylinder C of circumference Δx and height Δy by gluing the left and right edges of our rectangle together. More precisely, let C by the quotient space which is obtained from the infinitely wide strip $0 < y < \Delta y$ in the z-plane by identifying each point $z = x + iy$ with its translate $z + \Delta x$. Define the *modulus* $\mathrm{mod}(C)$ of such a cylinder to be the ratio $\Delta y / \Delta x$ of height to circumference. By the *winding number* of a closed curve γ in C we mean the integer

$$w \;=\; \frac{1}{\Delta x} \oint_\gamma dx \;.$$

Theorem B.2 (Length-Area Inequality for Cylinders). *For any conformal metric $\rho(z)|dz|$ on the cylinder C there exists some simple closed curve γ with winding number $+1$ whose length $\mathbf{L}_\rho(\gamma) = \oint_\gamma \rho(z)|dz|$ satisfies the inequality*

$$\mathbf{L}_\rho(\gamma)^2 \;\le\; \mathrm{area}_\rho(A) / \mathrm{mod}(A) \;. \tag{4}$$

Furthermore, this result is best possible: If we use the Euclidean metric $|dz|$ then

$$\mathbf{L}(\gamma)^2 \;\ge\; \mathrm{area}(A) / \mathrm{mod}(A) \tag{5}$$

for every such curve γ .

Proof. Just as in the proof of B.1, we find a horizontal curve γ_y with

$$\mathbf{L}_\rho(\gamma_y)^2 \;\le\; \frac{\Delta x}{\Delta y} \mathrm{area}_\rho(C) \;=\; \frac{\mathrm{area}_\rho(C)}{\mathrm{mod}(C)} \;.$$

On the other hand in the Euclidean case, for any closed curve γ of winding number one we have

$$\mathbf{L}(\gamma) \;=\; \oint_\gamma |dz| \;\ge\; \oint_\gamma dx \;=\; \Delta x \;,$$

hence $\mathbf{L}(\gamma)^2 \;\ge\; (\Delta x)^2 \;=\; \mathrm{area}(C) / \mathrm{mod}(C)$. $\square$

Definitions. A Riemann surface A is said to be an *annulus* if it is conformally isomorphic to some cylinder. An embedded annulus $A \subset C$ is said to be *essentially embedded* if it contains a curve which has winding number one around C.

Here is an important consequence of Theorem B.2.

Corollary B.3 (An Area-Modulus Inequality). *Let $A \subset C$ be an essentially embedded annulus in the cylinder C, and suppose that A is conformally isomorphic to a cylinder C'. Then*

$$\mathrm{mod}(C') \;\le\; \frac{\mathrm{area}(A)}{\mathrm{area}(C)}\,\mathrm{mod}(C)\;. \tag{6}$$

In particular:

$$\mathrm{mod}(C') \;\le\; \mathrm{mod}(C)\;. \tag{7}$$

Proof. Let $\zeta \mapsto z$ be the embedding of C' onto $A \subset C$. The Euclidean metric $|dz|$ on C, restricted to A, pulls back to some conformal metric $\rho(\zeta)|d\zeta|$ on C', where $\rho(\zeta) = |dz/d\zeta|$. According to B.2, there exists a curve γ' with winding number 1 about C' whose length satisfies

$$L_\rho(\gamma')^2 \le \mathrm{area}_\rho(C')/\mathrm{mod}(C')\;.$$

This length coincides with the Euclidean length $\mathbf{L}(\gamma)$ of the corresponding curve γ in $A \subset C$, and $\mathrm{area}_\rho(C')$ is equal to the Euclidean area $\mathrm{area}(A)$, so we can write this inequality as

$$L(\gamma)^2 \le \mathrm{area}(A)/\mathrm{mod}(C')\;.$$

But according to (5) we have

$$\mathrm{area}(C)/\mathrm{mod}(C) \;\le\; L(\gamma)^2\;.$$

Combining these two inequalities, we obtain

$$\mathrm{area}(C)/\mathrm{mod}(C) \;\le\; \mathrm{area}(A)/\mathrm{mod}(C')\;,$$

which is equivalent to the required inequality (6). $\square$

Corollary B.4. *The modulus of a cylinder is a well defined conformal invariant.*

Proof. If C' is conformally isomorphic to C then (7) asserts that $\mathrm{mod}(C') \le \mathrm{mod}(C)$, and similarly $\mathrm{mod}(C) \le \mathrm{mod}(C')$. $\square$

It follows that the *modulus* of an annulus A can be defined as the modulus of any conformally isomorphic cylinder. Furthermore, if A is essentially embedded in some other annulus A', then $\mathrm{mod}(A) \le \mathrm{mod}(A')$.

Corollary B.5 (Grötzsch Inequality). *Suppose that $A' \subset A$ and $A'' \subset A$ are two disjoint annuli, each essentailly embedded in A. Then*

$$\mathrm{mod}(A') + \mathrm{mod}(A'') \;\le\; \mathrm{mod}(A)\;.$$

Proof. We may assume that A is a cylinder C. According to (6) we have

$$\mathrm{mod}(A') \;\le\; \frac{\mathrm{area}(A')}{\mathrm{area}(C)}\,\mathrm{mod}(C)\;, \qquad \mathrm{mod}(A'') \;\le\; \frac{\mathrm{area}(A'')}{\mathrm{area}(C)}\,\mathrm{mod}(C)\;.$$

where all areas are Euclidean. Using the inequality

$$\mathrm{area}(A') + \mathrm{area}(A'') \le \mathrm{area}(C)\;,$$

the conclusion follows. □

Now consider a *flat torus* $\mathbf{T} = \mathbf{C}/\Lambda$. Here $\Lambda \subset \mathbf{C}$ is to be a 2-dimensional *lattice*, that is an additive subgroup of the complex numbers, spanned by two elements λ_1 and λ_2 where $\lambda_1/\lambda_2 \notin \mathbf{R}$. Let $A \subset \mathbf{T}$ be an embedded annulus.

By the "*winding number*" of A in $\mathbf{T}$ we will mean the lattice element $w \in \Lambda$ which is constructed as follows. Under the universal covering map $\mathbf{C} \to \mathbf{T}$, the central curve of A lifts to a curve segment which joins some point $z_0 \in \mathbf{C}$ to a translate $z_0 + w$ by the required lattice element. We say that $A \subset \mathbf{T}$ is an *essentially embedded annulus* if $w \ne 0$.

Corollary B.6 (Bers Inequality). *If the annulus A is embedded in the flat torus $\mathbf{T} = \mathbf{C}/\Lambda$ with winding number $w \in \Lambda$, then*

$$\mathrm{mod}(A) \;\le\; \frac{\mathrm{area}(T)}{|w|^2}\;. \tag{8}$$

Roughly speaking, if A winds many times around the torus, so that $|w|$ is large, then A must be very skinny. A slightly sharper version of this inequality is given in Problem B-2 below.

Proof. Choose a cylinder C' which is conformally isomorphic to A. The Euclidean metric $|dz|$ on $A \subset \mathbf{T}$ corresponds to some metric $\rho(\zeta)|d\zeta|$ on C', with

$$\mathrm{area}_\rho(C') \;=\; \mathrm{area}(A)\;.$$

By B.2 we can choose a curve γ' of winding number one on C', or a corresponding curve γ on $A \subset \mathbf{T}$, with

$$\mathbf{L}(\gamma)^2 \;=\; \mathbf{L}_\rho(\gamma')^2 \;\le\; \frac{\mathrm{area}_\rho(C')}{\mathrm{mod}(C')} \;=\; \frac{\mathrm{area}(A)}{\mathrm{mod}(A)} \;\le\; \frac{\mathrm{area}(T)}{\mathrm{mod}(A)}\;.$$

Now if we lift γ to the universal covering space $\mathbf{C}$ then it will join some point z_0 to $z_0 + w$. Hence its Euclidean length $\mathbf{L}(\gamma)$ must satisfy $\mathbf{L}(\gamma) \ge |w|$. Thus

$$|w|^2 \;\le\; \frac{\mathrm{area}(T)}{\mathrm{mod}(A)}\;,$$

which is equivalent to the required inequality (8). □

Now consider the following situation. Let $U \subset \mathbf{C}$ be a bounded simply connected open set, and let $K \subset U$ be a compact subset so that the difference $A = U \smallsetminus K$ is a topological annulus. As noted in §2, such an annulus must be conformally isomorphic to a finite or infinite cylinder. By definition an infinite cylinder, that is a cylinder of infinite height, has modulus zero. (Such an infinite cylinder may be either one-sided infinite, conformally isomophic to a punctured disk, or two-sided infinite, conformally isomorphic to the punctured plane.)

Corollary B.7. *Suppose that $K \subset U$ as described above. Then K reduces to a single point if and only if the annulus $A = U \smallsetminus K$ has infinite modulus. Furthermore, the diameter of K is bounded by the inequality*

$$4\,\mathrm{diam}(K)^2 \;\le\; \frac{\mathrm{area}(A)}{\mathrm{mod}(A)} \;\le\; \frac{\mathrm{area}(U)}{\mathrm{mod}(A)}\;. \tag{9}$$

Proof. According to B.2, there exists a curve with winding number one about A whose length satisfies $L^2 \le \mathrm{area}(A)/\mathrm{mod}(A)$. Since K is enclosed within this curve, it follows easily that $\mathrm{diam}(K) \le \mathbf{L}/2$, and the inequality (9) follows. Conversely, if K is a single point then using (7) we see easily that $\mathrm{mod}(A) = \infty$. □

The following ideas are due to McMullen. (Compare [BH, II].) The *isoperimetric inequality* asserts that the area enclosed by a plane curve of length $\mathbf{L}$ is at most $\mathbf{L}^2/(4\pi)$, with equality if and only if the curve is a round circle. (See for example [CR].) Combining this with the argument above, we see that

$$\mathrm{area}(K) \;\le\; \frac{\mathbf{L}^2}{4\pi} \;\le\; \frac{\mathrm{area}(A)}{4\pi\,\mathrm{mod}(A)}\;.$$

Writing this inequality as $4\pi\,\mathrm{mod}(A) \le \mathrm{area}(A)/\mathrm{area}(K)$ and adding $+1$ to both sides we obtain the completely equivalent inequality $1 + 4\pi\,\mathrm{mod}(A) \le \mathrm{area}(U)/\mathrm{area}(K)$, or in other words

$$\mathrm{area}(K) \;\le\; \frac{\mathrm{area}(U)}{1 + 4\pi\,\mathrm{mod}(A)}\;. \tag{10}$$

This can be sharpened as follows:

Corollary B.8 (McMullen Inequality). *If $A = U \smallsetminus K$ as above, then*

$$\mathrm{area}(K) \;\le\; \mathrm{area}(U)/e^{4\pi\,\mathrm{mod}(A)}\;.$$

Proof. Cut the annulus A up into n concentric annuli A_i , each of modulus equal to $\mathrm{mod}(A)/n$. Let K_i be the bounded component of the complement of A_i , and assume that these annuli are nested so that $A_i \cup K_i = K_{i+1}$ with $K_1 = K$, and let $K_{n+1} = A \cup K = U$. Then $\mathrm{area}(K_{i+1})/\mathrm{area}(K_i) \ge 1 + 4\pi\,\mathrm{mod}(A)/n$ by (10), hence

$$\mathrm{area}(U)/\mathrm{area}(K) \;\ge\; (\,1 \,+\, 4\pi\,\mathrm{mod}(A)/n\,)^n\;,$$

where the right hand side converges to $e^{4\pi\,\mathrm{mod}(A)}$ as $n \to \infty$. □

Concluding Problems:

Problem B-1. In the situation of Theorem B.1, show that more than half of the horizontal curves γ_y have length $L_\rho(\gamma_y) \le \sqrt{2\,\mathrm{area}_\rho(I^2)}$. (Here "more than half" is to be interpreted in the sense of Lebesgue measure.)

Problem B-2 (Sharper Bers Inequality). If the flat torus $\mathbf{T} = \mathbf{C}/\Lambda$ contains several disjoint annuli A_i , all with the same "winding number" $w \in \Lambda$, show that

$$\sum \mathrm{mod}(A_i) \;\le\; \mathrm{area}(\mathbf{T})/|w|^2 \,.$$

If two essentially embedded annuli are disjoint, show that they necessarily have the same winding number.

Problem B-3 (Branner-Hubbard). Let $K_1 \supset K_2 \supset K_3 \supset \cdots$ be compact subsets of $\mathbf{C}$ with each K_{n+1} contained in the interior of K_n . Suppose further that each interior K_n^o is simply connected, and that each difference $A_n = K_n^o \smallsetminus K_{n+1}$ is an annulus. If $\sum_1^\infty \mathrm{mod}(A_n)$ is infinite, show that the intersection $\bigcap K_n$ reduces to a single point. Show that the converse statement is false. (For example, do this by showing that a closed disk $\overline{D}'$ of radius $1/2$ can be embedded in the open unit disk D so that the complementary annulus $A = D \smallsetminus \overline{D}'$ has modulus arbitrarily close to zero.)

Appendix C. Continued Fractions.

Suppose that we start with a real number r_1 in the open interval $(0,1)$. Then $1/r_1$ can be written uniquely as the sum $a_1 + r_2$ of an integer $a_1 \geq 1$ and a remainder term $0 \leq r_2 < 1$. If $r_2 > 0$, then similarly we can set $1/r_2 = a_2 + r_3$, and so on, so that

$$1/r_n \;=\; a_n + r_{n+1}\,, \qquad 0 \leq r_{n+1} < 1\,, \tag{1}$$

where each a_n is a strictly positive integer. If r_1 is rational, then r_2 is rational with smaller denominator, so this construction must stop with $r_{n+1} = 0$ after finitely many steps. But if r_1 is irrational then the construction continues indefinitely:

$$r_1 \;=\; \frac{1}{a_1 + r_2} \;=\; \frac{1}{a_1 + \dfrac{1}{a_2 + r_3}} \;=\; \cdots \;=\; \cfrac{1}{a_1 + \cfrac{1}{a_2 + \ddots + \cfrac{1}{a_{n-1} + r_n}}} \;=\; \cdots .$$

Closely related is the *Euclidean algorithm*, a procedure for finding the greatest common divisor of two real numbers, when it exists. Suppose that we start with two numbers $\xi_0 > \xi_1 > 0$. Then we can express ξ_0 as an integral multiple of ξ_1 plus a strictly smaller remainder term,

$$\xi_0 = a_1 \xi_1 + \xi_2\,, \quad \text{with} \quad \xi_1 > \xi_2 \geq 0\,,$$

where a_1 is a strictly positive integer. If $\xi_2 > 0$, then similarly we can set $\xi_1 = a_2\xi_2 + \xi_3$ with $\xi_2 > \xi_3 \geq 0$, and so on. If the ratio $r_1 = \xi_1/\xi_0$ is rational, then this procedure terminates with

$$\xi_n = \text{greatest common divisor}\,, \qquad \xi_{n+1} = 0$$

after finitely many steps. However, we will rather assume that r_1 is irrational so that the procedure can be continued indefinitely, yielding an infinite sequence of numbers

$$\xi_0 > \xi_1 > \xi_2 > \cdots > 0 \qquad \text{with} \qquad \xi_{n-1} \;=\; a_n \xi_n + \xi_{n+1}\;, \tag{2}$$

where each a_n is a strictly positive integer. If we set $r_n = \xi_n/\xi_{n-1} \in (0,1)$, then dividing equation (2) by ξ_n we obtain $1/r_n = a_n + r_{n+1}$, as above.

In practice, it will be convenient to set $x_n = (-1)^{n+1}\xi_n$ so that the x_n alternate in sign with $|x_n| = \xi_n$. Then

$$x_{n+1} \;=\; x_{n-1} + a_n x_n\;. \tag{3}$$

In matrix notation we can write this as

$$\begin{bmatrix} x_n \\ x_{n+1} \end{bmatrix} \;=\; \begin{bmatrix} 0 & 1 \\ 1 & a_n \end{bmatrix} \begin{bmatrix} x_{n-1} \\ x_n \end{bmatrix}$$

and therefore

$$\begin{bmatrix} x_n \\ x_{n+1} \end{bmatrix} \;=\; \begin{bmatrix} 0 & 1 \\ 1 & a_n \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 1 & a_{n-1} \end{bmatrix} \cdots \begin{bmatrix} 0 & 1 \\ 1 & a_1 \end{bmatrix} \begin{bmatrix} x_0 \\ x_1 \end{bmatrix} \;=\; \begin{bmatrix} p_n & q_n \\ p_{n+1} & q_{n+1} \end{bmatrix} \begin{bmatrix} x_0 \\ x_1 \end{bmatrix} \tag{4}$$

where

$$\begin{bmatrix} p_n & q_n \\ p_{n+1} & q_{n+1} \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 1 & a_n \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 1 & a_{n-1} \end{bmatrix} \cdots \begin{bmatrix} 0 & 1 \\ 1 & a_1 \end{bmatrix} . \tag{5}$$

It follows inductively that

$$\begin{array}{lllll} p_0 = 1 \,, & p_1 = 0 \,, & p_2 = 1 \,, & \ldots \,, & p_{n+1} = p_{n-1} + a_n p_n \,, \\ q_0 = 0 \,, & q_1 = 1 \,, & q_2 = a_1 \,, & \ldots \,, & q_{n+1} = q_{n-1} + a_n q_n \,. \end{array} \tag{6}$$

Here the integers $0 = p_1 < p_2 < p_3 < \cdots$ and $0 = q_0 < q_1 < q_2 < \cdots$ grow at least exponentially with n, since

$$q_{n+2} \;\ge\; q_n + q_{n+1} \;\ge\; q_n + (q_{n-1} + q_n) \;\ge\; 2q_n \,,$$

and similarly $p_{n+2} \ge 2p_n$.

Suppose, to fix our ideas, that we start with some irrational number $\xi = r_1 \in (0,1)$ and set $\xi_0 = 1$, $\xi_1 = \xi$. Then we can write equation (4) as

$$x_n \;=\; p_n x_0 + q_n x_1 \;=\; -p_n + q_n \xi \,,$$

or in other words

$$\xi \;=\; \frac{p_n}{q_n} + \frac{x_n}{q_n} \,. \tag{7}$$

Thus the irrational number ξ *is equal to the rational number* p_n/q_n *plus a remainder term* x_n/q_n. In the literature, these numbers p_n/q_n are called the ***convergents*** to the continued fraction expansion of ξ. (Compare Problem C-5.) The numbers $|x_n|$ converge at least geometrically to zero as $n \to \infty$ (see Problem C-1 or inequality (8) below), so the remainder term x_n/q_n in equation (7) tends to zero quite rapidly as $n \to \infty$.

It follows from (7) that the successive convergents p_n/q_n to ξ are ordered as follows:

$$0 = \frac{p_1}{q_1} < \frac{p_3}{q_3} < \cdots < \xi < \cdots < \frac{p_4}{q_4} < \frac{p_2}{q_2} = \frac{1}{a_1} \le 1 \,.$$

In order to estimate the error term x_n/q_n in terms of the q_i, note that the product matrix (5) has determinant $(-1)^n$. Therefore

$$\frac{p_n}{q_n} - \frac{p_{n+1}}{q_{n+1}} \;=\; \frac{(-1)^n}{q_n q_{n+1}} \,.$$

Since ξ lies between p_n/q_n and p_{n+1}/q_{n+1} but is closer to p_{n+1}/q_{n+1}, it follows that the absolute value of the error term x_n/q_n lies between $1/(2q_nq_{n+1})$ and $1/(q_nq_{n+1})$. In other words:

$$\frac{1}{2q_{n+1}} \;<\; |x_n| \;<\; \frac{1}{q_{n+1}} \,. \tag{8}$$

Now let us set $\lambda = e^{2\pi i \xi}$ on the unit circle $S^1 \subset \mathbf{C}$. Using the discussion above, we will study the orbit $1 \mapsto \lambda \mapsto \lambda^2 \mapsto \cdots$ under the rotation $z \mapsto \lambda z$ of this circle.

Definition. We say that a point λ^q on this orbit is a *closest return* to 1 if

$$|\lambda^q - 1| \;<\; |\lambda^m - 1|$$

for every m with $0 < m < q$, so that λ^q is closer to 1 than any preceding point on the orbit. We will prove the following.

Lemma C.1. *The point* $\lambda^q = e^{2\pi i \xi q}$ *is a closest return to 1 along the orbit*

$$1 \;\mapsto\; \lambda \;\mapsto\; \lambda^2 \;\mapsto\; \cdots$$

if and only if q *is one of the denominators* $1 = q_1 \le q_2 < q_3 < q_4 < \cdots$ *in the continued fraction approximations to* ξ. *Furthermore, if* $q = q_n$ *with* $n \ge 2$ *then the order of magnitude of the distance* $|\lambda^q - 1|$ *is given by*

$$\frac{2}{q_{n+1}} \;<\; |\lambda^{q_n} - 1| \;<\; \frac{2\pi}{q_{n+1}} \;. \tag{9}$$

The proof will be based on the following. Instead of studying the multiplicative group $S^1 \subset \mathbf{C}$, we can equally well work with the additive group $\mathbf{R}/\mathbf{Z}$ and the orbit $0 \mapsto \xi \mapsto 2\xi \mapsto \cdots$ under the translation $t \mapsto t + \xi \pmod{\mathbf{Z}}$. It will be convenient to introduce the following special notation. For each real number x let

$$\ll x \gg \;=\; \mathrm{dist}(x\,,\,\mathbf{Z}) \;=\; \mathrm{Min}\,\{\, |x+n| \;:\; n \in \mathbf{Z} \,\}$$

be the distance to the nearest integer. Then an easy geometric argument shows that

$$|\lambda^m - 1| \;=\; |e^{2\pi i m\xi} - 1| \;=\; 2\,\sin\big(\pi \ll m\xi \gg\big)\,.$$

Since $4 < 2\sin(\pi t)/t < 2\pi$ for $t \in (0\,,\ 1/2)$, it follows that

$$4 \;<\; \frac{|\lambda^m - 1|}{\ll m\xi \gg} \;<\; 2\,\pi \;. \tag{10}$$

In the special case $m = q_n$ note that

$$q_n\,\xi \;\;\equiv\;\; x_n \pmod{\mathbf{Z}}$$

by equation (7). If $n \ge 2$, then $|x_n| < 1/2$. (Compare Problem C-1.) Hence it follows that $|x_n|$ *is equal to* $\ll q_n\xi \gg$. The inequality (9) now follows from (8) and (10).

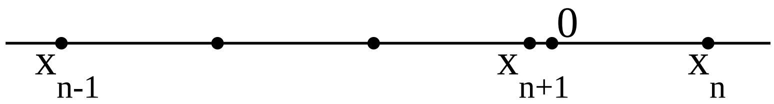
Suppose, to fix our ideas, that n is odd, so that

$$x_{n-1} \;<\; 0 \;<\; x_n \;<\; -x_{n-1} \;.$$

(The case n even is completely analogous, but with all inequalities reversed.) Then it is easy to check that

$$x_{n-1} \;<\; x_{n-1} + x_n \;<\; x_{n-1} + 2x_n \;<\; \cdots \;<\; x_{n-1} + a_n x_n \;<\; 0 \;<\; x_n \tag{11}$$

where $x_{n-1} + a_n x_n$ is equal to x_{n+1}. For example:



Note that each $x_{n-1}+jx_n$ is a representative mod $\mathbf{Z}$ for a point $m\xi$ on our orbit, where $m = q_{n-1}+jq_n$. For each $n \ge 1$ we claim the following.

Assertion $\mathbf{A}_n$. *No point $m\xi$ with $0<m<q_n$ has a representative* mod $\mathbf{Z}$ *which lies strictly between x_{n-1} and x_n .*

As part of the proof, we will also show the following sharper statement.

Assertion $\mathbf{B}_n$. *For m in the range $0<m\le q_{n+1}$, the only points $m\xi$ which have a representative η_m* mod $\mathbf{Z}$ *which lies strictly between x_{n-1} and x_n are the points $x_{n-1}+jx_n$ which are listed in formula* (11) .

(In each case, these numbers must be ordered appropriately, according as n is even or odd.) The proof will be by a double induction on n and m . The assertion $\mathbf{A}_1$ is trivially true, since $q_1=1$. We will show that $\mathbf{A}_n \Rightarrow \mathbf{B}_n$. Since it is easy to check that $\mathbf{B}_n \Rightarrow \mathbf{A}_{n+1}$, this will prove inductively that the assertions $\mathbf{A}_1 \Rightarrow \mathbf{B}_1 \Rightarrow \mathbf{A}_2 \Rightarrow \mathbf{B}_2 \Rightarrow \cdots$ are all true.

Suppose then that $\mathbf{A}_n$ is true, and suppose, for some m between 0 and q_{n+1} , that $m\xi$ has a representative η_m mod $\mathbf{Z}$ which lies strictly between x_{n-1} and x_n . We will show by induction on m that m must have the form $q_{n-1}+jq_n$. According to $\mathbf{A}_n$ we must have $m>q_n$. We divide the proof into two cases according as η_m lies between x_{n-1} and $x_{n-1}+x_n$ or between $x_{n-1}+x_n$ and x_n . In the former case, we see that $(m-q_{n-1})\,\xi$ has a representative $\eta_m - x_{n-1}$ mod $\mathbf{Z}$ which lies between 0 and x_n , thus contradicting our induction hypothesis on m . In the later case, we see that $(m-q_n)\,\xi$ has a representative $\eta_m - x_n$ mod $\mathbf{Z}$ which lies between x_{n-1} and 0 . Therefore, by induction, $m-q_n$ has the form $q_{n-1}+jq_n$, hence m itself also has this form. This completes the double induction, proving $\mathbf{A}_n$ and $\mathbf{B}_n$. Evidently Lemma C.1 follows easily from the Assertions $\mathbf{A}_n$. □

Concluding Problems.

Problem C-1. With r_n as in equation (1) or (3), show that the product

$$r_n r_{n+1} \;=\; \xi_{n+1}/\xi_{n-1}$$

is always smaller than $1/2$. (If both r_n and r_{n+1} are greater than $1/2$ then $r_n r_{n+1} = 1-r_{n+1} < 1/2$.) Conclude that the numbers

$$\xi_{2n} \;<\; \xi_0/2^n$$

converge rapidly to zero as $n\to\infty$.

Problem C-2. In the simplest possible case $a_1=a_2=\cdots=1$, show that

$$\{q_n\} = \{p_{n+1}\} = \{0\,,\ 1\,,\ 1\,,\ 2\,,\ 3\,,\ 5\,,\ 8\,,\ 13\,,\ 21\,,\ \dots\}\ ,$$

yielding the sequence of Fibonacci numbers. Prove the asymptotic formula $q_n \sim \gamma^n/\sqrt{5}$ as $n\to\infty$, and hence $p_n/q_n \to 1/\gamma$ as $n\to\infty$, where $\gamma=(\sqrt{5}+1)/2$. Show that this special case corresponds to the slowest possible growth for the coefficients p_n and q_n .

Problem C-3. Using Lemma C.1, show for $n \ge 2$ that the convergent p_n/q_n is the *best* rational approximation to ξ among all fractions with denominator $m \le q_n$. In particular,

$$\left|\xi - \frac{p_n}{q_n}\right| < \left|\xi - \frac{i}{m}\right|$$

for any integers i and m with $0 < m < q_n$.

Problem C-4. Define a polynomial $\mathcal{P}(x_1, \ldots, x_n)$ in n variables inductively by the formula

$$\mathcal{P}(x_1, \ldots, x_n) = \mathcal{P}(x_1, \ldots, x_{n-2}) + \mathcal{P}(x_1, \ldots, x_{n-1})\, x_n \ ,$$

starting with

$$\mathcal{P}(\,) = 1\,, \quad \mathcal{P}(x) = x\,, \quad \mathcal{P}(x,y) = 1 + xy\,, \quad \mathcal{P}(x,y,z) = x + z + xyz\,, \quad \ldots .$$

Show that the p_n and q_n can be expressed as polynomial functions of the a_i by the formulas

$$p_{n+1} = \mathcal{P}(a_2, \ldots, a_n)\,, \qquad q_{n+1} = \mathcal{P}(a_1, \ldots, a_n)\ .$$

Using the inverse of the matrix equation (5), show that

$$\mathcal{P}(a_1, \ldots, a_n) = \mathcal{P}(a_n, \ldots, a_1)\ .$$

Conclude that we can also write

$$\mathcal{P}(x_1, \ldots, x_n) = x_1\,\mathcal{P}(x_2, \ldots, x_n) + \mathcal{P}(x_3, \ldots, x_n)\ . \tag{12}$$

Show that $\mathcal{P}(x_1, \ldots, x_n)$ is equal to the sum of all distinct monomials which can be formed out of the product $x_1 \cdots x_n$ by striking out any number of consecutive pairs. Show that the number of such monomials is equal to the n-th Fibonacci number.

Problem C-5. Using (12), give an inductive proof of the finite continued fraction equation

$$\frac{p_{n+1}}{q_{n+1}} = \frac{\mathcal{P}(a_2, \ldots, a_n)}{\mathcal{P}(a_1, \ldots, a_n)} = \cfrac{1}{a_1 + \cfrac{1}{a_2 + \ddots + \cfrac{1}{a_{n-1} + \cfrac{1}{a_n}}}}\ .$$

Appendix D. Remarks Concerning Two Complex Variables.

Many of the arguments in these lectures are strictly one-dimensional. In fact several of our underlying principles break down completely in the two variable case.

In order to illustrate these differences, it is useful to consider the family of (generalized) *Hénon maps*, which can be described as follows. Choose a complex constant $\delta \neq 0$ and a polynomial map $f : \mathbf{C} \to \mathbf{C}$ of degree $d \geq 2$, and consider doubly infinite sequences of complex numbers $\ldots, z_{-1}, z_0, z_1, z_2, \ldots$ satisfying the recurrence relation

$$z_{n+1} - f(z_n) + \delta\, z_{n-1} \;=\; 0\,.$$

Evidently we can solve for $(z_n\,, z_{n+1})$ as a polynomial function $F(z_{n-1}\,, z_n)$, where the transformation $F(z_{n-1}\,, z_n) = (z_n\,, f(z_n) - \delta z_{n-1})$ has Jacobian matrix

$$\begin{bmatrix} 0 & 1 \\ -\delta & f'(z_n) \end{bmatrix}$$

with constant determinant δ, and with trace $f'(z_n)$. Similarly we can solve for $(z_{n-1}\,, z_n)$ as a polynomial function $F^{-1}(z_n\,, z_{n+1})$. As an example, consider the quadratic polynomial $f(z) = z^2 + \lambda z$, having a fixed point of multiplier λ at the origin. Then

$$F(x\,, y) \;=\; (y\,, y^2 + \lambda y - \delta x)\,, \qquad (*)$$

where x and y are complex variables. This map has a fixed point at $(0,0)$, where the eigenvalues λ_1 and λ_2 of the Jacobian matrix satisfy $\lambda_1 + \lambda_2 = \lambda$ and $\lambda_1\lambda_2 = \delta$. Evidently, by the appropriate choice of λ and δ, we can realize any desired non-zero λ_1 and λ_2. If $\lambda_1 \neq \lambda_2$, then we can diagonalize this Jacobian matrix by a linear change of coordinates.

Lemma D.1. *Consider any holomorphic transformation $F(x,y) = (x', y')$ in two complex variables, with*

$$x' = \lambda_1 x + \mathcal{O}(|x^2| + |y^2|)\,, \quad y' = \lambda_2 y + \mathcal{O}(|x^2| + |y^2|)$$

as $(x,y) \to (0,0)$. If the eigenvalues λ_1 and λ_2 of the derivative at the origin satisfy $1 > |\lambda_1| \geq |\lambda_2| > |\lambda_1^2|$, then F is conjugate, under a local holomorphic change of coordinates, to the linear map $L(u,v) = (\lambda_1 u\,, \lambda_2 v)$.

Proof. We must show that there exists a change of coordinates $(u,v) = \phi(x,y)$, defined and holomorphic throughout a neighborhood of the origin, so that $\phi\circ F\circ\phi^{-1} = L$. As in the proof of the Koenigs Theorem, §6.1, we first choose a constant c so that $1 > c > |\lambda_1| \geq |\lambda_2| > c^2$. To any orbit

$$(x_0\,, y_0) \overset{F}{\mapsto} (x_1\,, y_1) \overset{F}{\mapsto} \cdots$$

near the origin, we associate the sequence of points

$$(u_n\,, v_n) \;=\; L^{-n}(x_n\,, y_n) \;=\; (x_n/\lambda_1^n\,, y_n/\lambda_2^n)$$

and show, using Taylor's Theorem, that it converges geometrically to the required limit

$\phi(x_0\,,\,y_0)$, with successive differences bounded by a constant times $(c^2/\lambda_2)^n$. Details will be left to the reader. $\square$

Remarks. Some such restriction on the eigenvalues is essential. As an example, for the map

$$f(x,y) \;=\; (\lambda x,\, \lambda^2 y + x^2)\,,$$

with eigenvalues λ and λ^2 , there is no such local holomorphic change of coordinates. (Problem D-1.) For a much sharper statement as to when linearization is possible, even when one or both eigenvalues lie on the unit circle, compare Zehnder.

Now consider a Hénon map $F : \mathbf{C}^2 \stackrel{\approx}{\to} \mathbf{C}^2$ as in $(*)$ above, with eigenvalues satisfying the conditions of D.1. Let Ω be the attractive basin of the origin. We claim that ϕ extends to a global diffeomorphism $\Phi : \Omega \stackrel{\approx}{\to} \mathbf{C}^2$. In fact, for any $(x,y) \in \Omega$ we set $\Phi(x,y) = L^{-n} \circ \phi \circ F^{\circ n}(x,y)$. If n is sufficiently large, then $F^n(x,y)$ is close to the origin, so that this expression is defined. Similarly, $\Phi^{-1}(u,v) = F^{-n} \circ \phi^{-1} \circ L^{\circ n}$ is well defined for large n . This shows that Φ is a holomorphic diffeomorphism with holomorphic inverse.

Note that this basin Ω is not the entire space $\mathbf{C}^2$. For example, if $|z_1|$ is sufficiently large compared with $|z_0|$, and if

$$(z_0\,,\,z_1) \stackrel{F}{\mapsto} (z_1\,,\,z_2) \stackrel{F}{\mapsto} (z_2\,,\,z_3) \stackrel{F}{\mapsto} \cdots$$

then it is not difficult to check that $|z_1| < |z_2| < |z_3| < \cdots$, so that $(z_0\,,\,z_1)$ is not in Ω . *Thus we have constructed:*

(1) *a proper subset* $\Omega \subset \mathbf{C}^2$ *which is analytically diffeomorphic to all of* $\mathbf{C}^2$ *, and*

(2) *a non-linear map with an attractive basin which contains no critical points.*

Evidently neither phenomenon can occur in one complex variable. Open sets satisfying (1) are called *Fatou-Bieberbach domains* since they were first constructed by Fatou, by an easy argument similar to that given here, and then later independently by Bieberbach, who had a much more difficult construction.

The proof that there are only finitely many attracting cycles also breaks down in two variables. Compare Newhouse.

Problem D-1. For the map $F(x,y) = (\lambda x\,,\,\lambda^2 y + x^2)$, where $\lambda \neq 0\,,\,1$, show that there is only one smooth F-invariant curve through the origin, namely $x = 0$. By way of contrast, for the associated linear map $L(x,y) = (\lambda x\,,\,\lambda^2 y)$ note that there are infinitely many F-invariant curves $y = cx^2$. Conclude that F is not locally holomorphically conjugate to a linear map.

Appendix E. Branched Coverings and Orbifolds.

This will be an outline of definitions and results, without proofs. (See the list of references at the end.) We will use "branch point" as a synonym for "critical point" and "ramified point" as a synonym for "critical value". Thus if $f(z_0) = w_0$ with $f'(z_0) = 0$, then z_0 is called a *branch point* and the image $f(z_0) = w_0$ is called a *ramified point*. More precisely, if

$$f(z) \;=\; w_0 + c(z - z_0)^n + (\text{higher terms}) \,,$$

with $n \geq 1$ and $c \neq 0$, then the integer $n = n(z_0)$ is called the *branch index* or the *local degree* of f at the point z_0. Thus $n(z) \geq 2$ if z is a branch point, and $n(z) = 1$ otherwise.

A holomorphic map $p : S' \to S$ between Riemann surfaces is called a *covering map* if each point of S has a connected neighborhood U which is *evenly covered*, in that each connected component of $p^{-1}(U) \subset S'$ maps onto U by a conformal isomorphism. A map $p : S' \to S$ is *proper* if the inverse image $p^{-1}(K)$ of any compact subset of S is a compact subset of S'. Note that every proper map is finite-to-one, and has a well defined finite degree $d \geq 1$. Such a map may also be called a d*-fold branched covering*. On the other hand, a covering map may well be infinite-to-one. Combining these two concepts, we obtain the following more general concept.

Definition. A holomorphic map $p : S' \to S$ between Riemann surfaces will be called a *branched covering map* if every point of S has a connected neighborhood U so that each connected component of $p^{-1}(U)$ maps onto U by a proper map.

Such a branched covering is said to be *regular* or *normal* if there exists a group Γ of conformal automorphisms of S', so that two points z_1 and z_2 of S' have the same image in S if and only if there is a group element γ with $\gamma(z_1) = z_2$. In this case we can identify S with the quotient manifold S'/Γ. In fact it is not difficult to check that the conformal structure of such a quotient manifold is uniquely determined. This Γ is called the group of *deck transformations* of the covering.

Regular branched covering maps have several special properties. For example, each ramified point is isolated, so that the set of all ramified points is a discrete subset of S. Furthermore, the branch index $n(z)$ depends only on the target point $f(z)$, that is, $n(z_1) = n(z_2)$ whenever $f(z_1) = f(z_2)$. Thus we can define the *ramification function* $\nu : S \to \{1,\, 2,\, 3,\, \ldots\}$ by setting $\nu(w)$ equal to the common value of $n(z)$ for all points z in the pre-image $f^{-1}(w)$. By definition, $\nu(w) \geq 2$ if w is a ramified point, and $\nu(w) = 1$ otherwise.

Definition. A pair (S, ν) consisting of a Riemann surface S and a "ramification function" $\nu : S \to \{1,\, 2,\, 3,\, \ldots\}$ which takes the value $\nu(w) = 1$ except at isolated points will be called a Riemann surface *orbifold*.

(Remark: Thurston's general concept of orbifold involves a structure which is locally modeled on the quotient of a coordinate space by a finite group. However, in the Riemann surface case only cyclic groups can occur, so a simpler definition can be used.)

Definition. If S' is simply connected, then a regular branched covering $p : S' \to S$ with ramification function ν will be called the *universal covering* for the orbifold $(S,\, \nu)$.

We will use the notation $\tilde S_\nu \to S$ for this universal branched covering. The associated group Γ of deck transformations is called the *fundamental group* $\pi_1(S,\,\nu)$ of the orbifold.

Lemma E.1. *With the following exceptions, every Riemann surface orbifold $(S,\,\nu)$ has a universal covering surface $\tilde S_\nu$ which is unique up to conformal isomorphism over S. The only exceptions are given by:*

(1) a surface $S \approx \hat{\mathbf{C}}$ with just one ramified point, or

(2) a surface $S \approx \hat{\mathbf{C}}$ with two ramified points for which $\nu(w_1) \neq \nu(w_2)$.

In these exceptional cases, no such universal covering exists.

By definition, the *Euler characteristic* of an orbifold $(S,\,\nu)$ is the rational number

$$\chi(S,\,\nu) \;=\; \chi(S) + \sum \big(\frac{1}{\nu(w_j)} - 1\big)\,,$$

to be summed over all ramified points, where $\chi(S)$ is the usual Euler characteristic of S. Intuitively speaking, each ramified point w_j makes a contribution of $+1$ to the usual Euler characteristic $\chi(S)$, but a smaller contribution of $1/\nu(w_j)$ to the orbifold Euler characteristic. Thus $\chi(S,\,\nu) < \chi(S) \le 2$, or more precisely

$$\chi(S) - r \;<\; \chi(S,\,\nu) \;\le\; \chi(S) - r/2$$

where r is the number of ramified points. *As an example, if $\chi(S,\,\nu) \ge 0$, with at least one ramified point, then it follows that $\chi(S) > 0$, so the base surface S can only be $D,\,\mathbf{C}$ or $\hat{\mathbf{C}}$, up to isomorphism.* Compare E.5 below.

If there are infinitely many ramified points, note that we must set $\chi(S,\,\nu) = -\infty$. Similarly, if S is a surface which is not of finite type, then we must set $\chi(S,\,\nu) = \chi(S) = -\infty$.

If S' and S are provided with ramification functions μ and ν respectively, then a branched covering map $f : S' \to S$ is said to yield a *covering map* $(S'\,,\,\mu) \to (S\,,\,\nu)$ *between orbifolds* if the identity

$$n(z)\mu(z) \;=\; \nu(f(z))$$

is satisfied for all $z \in S'$, where $n(z)$ is the branch index. As an example, the universal covering map $\tilde S_\nu \to (S\,,\,\nu)$ is always a covering map of orbifolds, where $\tilde S_\nu$ is provided with the trivial ramification function $\mu \equiv 1$.

Lemma E.2. *$f : (S'\,,\,\mu) \to (S\,,\,\nu)$ is a covering map between orbifolds if and only if it lifts to a conformal isomorphism from the universal covering $\tilde S'_\mu$ onto $\tilde S_\nu$. If f is a covering in this sense, and has finite degree d, then the Riemann-Hurwitz formula of §5.1 takes the form*

$$\chi(S'\,,\,\mu) \;=\; \chi(S\,,\,\nu)d\,.$$

In particular, if the universal covering of $(S\,,\,\nu)$ is a covering of finite degree d, then $\chi(\tilde S_\nu) = \chi(S,\,\nu)d$.

The fundamental group and the Euler characteristic are related to each other as follows.

Lemma E.3. *Let* (S, ν) *be any Riemann surface orbifold which possesses a universal covering:*

If $\chi(S, \nu) > 0$ *, then the fundamental group* $\pi_1(S, \nu)$ *is finite.*

If $\chi(S, \nu) = 0$ *, then the fundamental group contains either* $\mathbf{Z}$ *or* $\mathbf{Z} \oplus \mathbf{Z}$ *as a subgroup of finite index.*

If $\chi(S, \nu) < 0$ *, then the fundamental group contains a non-abelian free product* $\mathbf{Z} * \mathbf{Z}$ *, and hence cannot contain any abelian subgroup of finite index.*

The Euler characteristic and the geometry of $\tilde{S}_\nu$ are related as follows.

Lemma E.4. *Suppose that* S *is obtained from a compact Riemann surface by removing at most a finite number of points. Then* $\tilde{S}_\nu$ *is either spherical, hyperbolic or Euclidean according as* $\chi(S\,,\,\nu)$ *is positive, negative or zero.*

Remark. In all other cases, that is whenever S is not isomorphic to a finitely punctured compact surface, S is hyperbolic, and it follows that the universal covering $\tilde{S}_\nu$ must also be hyperbolic.

Examples. If $S = \mathbf{C}$ with two ramified points $\nu(1) = \nu(-1) = 2$, then the map $z \mapsto \cos(2\pi z)$ provides a universal covering $\mathbf{C} \to (\mathbf{C}, \nu)$. The Euler characteristic $\chi(\mathbf{C}, \nu)$ is zero, and the fundamental group $\pi_1(S, \nu)$ consists of all transformations of the form $\gamma : z \mapsto n \pm z$.

For $S = \hat{\mathbf{C}}$ with three ramified points $\nu(0) = \nu(1) = \nu(\infty) = 2$, the rational map $\pi(z) = -4z^2/(z^2-1)^2$ provides a universal covering $\hat{\mathbf{C}} \to (\hat{\mathbf{C}}, \nu)$. In this case we have $\chi(\hat{\mathbf{C}}, \nu) = 1/2$, and the degree is equal to $\chi(\hat{\mathbf{C}})/\chi(\hat{\mathbf{C}}, \nu) = 4$. The fundamental group consists of all transformations $\gamma : z \mapsto \pm z^{\pm 1}$.

For $S = \hat{\mathbf{C}}$ with four ramified points of index $\nu(w_j) = 2$, then the torus T described in §5 provides a regular 2-fold branched covering. Its universal covering $\tilde{T}$ can be identified with the universal covering of $(\hat{\mathbf{C}}, \nu)$. In this case, the Euler characteristic $\chi(\hat{\mathbf{C}}, \nu)$ is zero.

Remark E.5. There are relatively few cases in which $\chi(S\,,\,\nu) \geq 0$. In fact all of these cases can be listed quite explicitly as follows. The unramified cases are very well known, namely the sphere, plane or disk with $\chi > 0$; and the punctured plane or disk, and the infinite families consisting of annuli and tori, all with $\chi = 0$.

By the “ramification indices” we will mean the list of values of the ramification function at the r ramified points, ordered for example so that $\nu(w_1) \leq \cdots \leq \nu(w_r)$.

If $\chi(\hat{\mathbf{C}}, \nu) > 0$ with $r > 0$, then the ramification indices must be either (n, n) or $(2, 2, n)$ for some $n \geq 2$, or $(2, 3, 3)$, $(2, 3, 4)$ or $(2, 3, 5)$. These five possibilities correspond to the five types of finite rotation groups of the 2-sphere; namely to the cyclic, dihedral, tetrahedral, octahedral and icosahedral groups respectively.

If $\chi(\hat{\mathbf{C}}, \nu) = 0$, then the ramification indices must be either $(2, 4, 4)$, $(2, 3, 6)$, $(3, 3, 3)$ or $(2, 2, 2, 2)$. These correspond to the automorphism groups of the tilings of $\mathbf{C}$ by squares, equilateral triangles, alternately colored equilateral triangles, and parallelograms respectively. In the parallelogram case, note that there is actually a one complex parameter family of distinct possible shapes, corresponding to the cross-ratio of the four ramified points.

Similarly, if $\chi(\mathbf{C},\,\nu)$ or $\chi(D,\,\nu)$ is strictly positive, then we must have $r \le 1$; while if $\chi(\mathbf{C},\,\nu)$ or $\chi(D,\,\nu)$ is zero, then we must have $r = 2$ with ramification indices $(2, 2)$. This is the complete list.

Appendix F. Parameter Space.

A very important part of complex dynamics, which has barely been mentioned in these notes, is the study of parametrized families of mappings. As an example, consider the family of all quadratic polynomial maps. A priori, a quadratic polynomial is specified by three complex parameters; however any such polynomial can be put into the unique normal form

$$f(z) \;=\; z^2 + c \tag{1}$$

by an affine change of coordinates. (Other normal forms which have been used are $\omega \mapsto \omega^2 + \lambda\omega$, with a preferred fixed point of multiplier λ at the origin, or

$$w \;\mapsto\; \lambda w(1-w) \tag{2}$$

which is more or less equivalent provided that $\lambda \neq 0$. Here $4c = \lambda(2-\lambda)$ and $\omega = z - \lambda/2 = -\lambda w$.) Using such a normal form, we can make a computer picture in the *parameter space* consisting of all complex constants c or λ. Each pixel in such a picture, corresponding to a small square in the parameter space, is to be assigned some color, perhaps only black or white, which depends on the dynamics of the corresponding quadratic map.

The first crude pictures of this type were made by Brooks and Matelski, as part of a study of Kleinian groups. They used the normal form (1), and introduced the open set consisting of all points of the c-plane for which the corresponding quadratic map has an attracting periodic orbit in the finite plane. I will use the notation $\mathcal{H}$ (standing for "hyperbolic") for this Brooks-Matelski set. At about the same time, Hubbard (unpublished) made much better pictures of a quite different parameter space arising from Newton's method for cubics. Mandelbrot, perhaps inspired by Hubbard, made corresponding pictures for quadratic polynomials, using the normal form (2) and also a variant of (1). In order to avoid confusion, let me translate all of Mandelbrot's definitions to the normal form (1). He introduced two different sets, which I will call $Q \subset M$. Mandelbrot did not give these sets different names, since he believed that they were identical. By definition, a parameter value c belongs to Q if the corresponding filled Julia set contains an interior point, and belongs to M if its filled Julia set contains the critical point $z = 0$ (or equivalently is connected). The Brooks-Matelski set $\mathcal{H}$ is a subset of $Q \subset M$. Mandelbrot made quite good computer pictures, which seemed to show a number of isolated "islands". Therefore, he conjectured that Q [or M] has many distinct connected components. (The editors of the journal thought that his islands were specks of dirt, and carefully removed them from the pictures.) Mandelbrot also described an important smaller set $\mathcal{L} \subset Q$ which he believed to be the principal connected component of Q. This set $\mathcal{L}$ consists of a central cardioid $\mathcal{L}_0$ with some (but not all) boundary points included, together with countably many smaller nearly-round disks which are pasted on inductively, in an explicitly described pattern.

Although Mandelbrot's statements in this first paper were not completely right, he deserves a great deal of credit for being the first to point out the extremely complicated geometry associated with the parameter space for quadratic maps. His major achievement has been to demonstrate to a very wide audience that such complicated "fractal" objects play an important role in a number of mathematical sciences.

The first real mathematical breakthrough came with Douady and Hubbard's work in 1982. They introduced the name *Mandelbrot set* for the compact set M described above, and provided a firm foundation for its mathematical study, proving for example that M is connected with connected complement. (Meanwhile, Mandelbrot had decided empirically that his isolated islands were actually connected to the mainland by very thin filaments.) Already in this first paper, they showed that each hyperbolic component of the interior of M can be canonically parametrized, and showed that the boundary ∂M can be profitably studied by following external rays.

It may be of interest to compare the three sets $\mathcal{H} \subset Q \subset M$ in parameter space. They are certainly different since $\mathcal{H}$ is open, M is compact, and Q is neither. Using Sullivan's work (§13), we can say that Q consists of $\mathcal{H}$ together with a very sparse set of boundary points, namely those for which the corresponding map has either a parabolic orbit or a Siegel disk. Quite likely, there is no difference between these three sets as far as computer graphics are concerned, since it is widely conjectured* that the Brooks-Matelski set $\mathcal{H}$ is equal to the interior of M, and that M is equal to the closure of $\mathcal{H}$. However, as far as practical computing is concerned, it should be noted that it is quite easy to test (at least roughly) whether a parameter value belongs to M, but somewhat harder to decide whether it belongs to $\mathcal{H}$ (compare §10.5), and very difficult to decide whether it belongs to Q. (Compare Appendix G. Here I am only speaking of approximate tests. To decide precisely whether a given point belongs to M or to whether a given point belongs to $\mathcal{H}$ may be very difficult. As a specific example, I have no idea whether or not the point $c = -1.5$ belongs to $\mathcal{H}$.)

Another important development came in 1983, with the work of Mañé, Sad and Sullivan on stability of the Julia set $J(f)$ under deformation of f. (Compare the discussion following §14.3.) These results were obtained independently by Lyubich. The study of parameter space for higher degree polynomials began some five years later with the work of Branner and Hubbard. Using the normal form

$$f(z) \;=\; z^3 - 3a^2 z + b\,,$$

with the two critical points at $z = \pm a$, they proved that the *cubic connectedness locus*, consisting of all parameter pairs (a, b) for which $J(f)$ is connected, is a *cellular set.* (Compare §17.3.) In particular, this set is compact and connected. A corresponding result for polynomials of higher degree has recently been obtained by Lavaurs. Parameter space studies for rational maps are more awkward, since there is no obvious normal form. However, an important beginning has been made by Rees. All of these studies are very new, and much remains to be done.

Here are two problems for the reader.

---

* Douady and Hubbard have shown that these conjectures are true if the set M is locally connected. Recent work of Yoccoz lends very strong support to the belief that M is indeed locally connected.

Problem F-1. Show that every polynomial map of degree $d \geq 2$ is conjugate, under an affine change of coordinates, to one in the "Fatou normal form"

$$f(z) \;=\; z^d + a_{d-2}z^{d-2} + \cdots + a_1 z + a_0\,.$$

Let $P(d) \cong \mathbf{C}^{d-1}$ be the space of all such maps. Show that the cyclic group $Z(d-1)$ of $(d-1)$-st roots of unity acts on $P(d)$ by linear conjugation, replacing $f(z)$ by $f(\omega z)/\omega$, and show that the quotient $P(d)/Z(d-1)$ can be identified with the "moduli space" of degree d polynomials up to affine conjugation. If $d \geq 4$, show that this moduli space is not a manifold.

Problem F-2. Show that every quadratic rational map is conjugate, under a fractional linear change of coordinates, for example to one in the form

$$f(z) \;=\; 1 + (z^2 - 1)/(az^2 - b)$$

where $a \neq b$, with critical points at zero and ∞ and a fixed point at $z = 1$. (In general this form is not unique, since such a map actually has three fixed points, and since the roles of zero and infinity can be interchanged.)

Appendix G. Remarks on Computer Graphics.

In order to make a computer picture of some complicated compact set $L \subset \mathbf{C}$, for example a Julia set, we must compute a matrix of small integers, where the (i,j)-th entry describes the color (perhaps only black or white) which is assigned to the (i,j)-th "pixel" of the computer screen. Each pixel represents a small square in the complex plane, and the color which is assigned must tell us something about intersection of L with this square.

In the case of a quadratic Julia set $J(f)$, one very fast method involves following iterates of the inverse map f^{-1} , taking all possible branches. (Compare §3.9.) As Mandelbrot points out, this method yields an excellent picture of the outer parts of the $J(f)$ but shows very little detail in the inner parts. If we think of the electrostatic field produced by an electric charge on $J(f)$, this method will emphasize only those parts of the Julia set at which "lines of force" (or "external rays" in the language of §18) tend to land.

A slower but much better procedure for plotting $J(f)$ involves iterating the map f for some large number (perhaps 50 to 50000) of times, starting at the midpoint of each pixel. If the orbit "escapes" from a large disk after n iterations, then the corresponding pixel is assigned a color which depends on n . In more refined versions of this method, one computes not only the value of the n-th iterate of f but also the absolute value of its derivative. Compare the discussion below. Similar remarks apply to the *Mandelbrot set* M , as defined by Douady and Hubbard. (Compare Appendix F). In this case one takes the quadratic map corresponding to the midpoint of the square and follows the orbit of its critical point.

Remark. In order to understand some of the limitations of this method, consider the situation near a fixed point $f(z_0) = z_0$ in the Julia set. First consider the repelling case, with multiplier say of absolute value two. If we start at at point z at distance $1/1000$ from z_0 , then the distance from z_0 will roughly double with each iteration. Hence, after only ten iterations the image of z will move substantially away from z_0 . The result will be a computer picture which is quite sharp and accurate near z_0 . (Figures 3, 4.)

Now suppose that we try to construct a picture for $z \mapsto z + z^4$ by the same method. Again start with a point $z \in J$ at a distance of $\epsilon \approx 1/1000$ from the fixed point at zero. Examining the proof of 7.2, we see that the associated coordinate $w_0 = -1/3z_0^3$, which increases by one under each iteration, is roughly equal to $-1/3\epsilon^3 \approx 300,000,000$. *Thus we would have to follow such an orbit for some* $300,000,000$ *iterations in order to escape from a neighborhood of* $z = 0$. The result in practice will be a false picture which shows everything near the origin to be in the Fatou set. (This difficulty was eliminated in Figure 8 by using a special computer program, which extrapolated iterates of f in order to distinguish different Fatou components. Similarly, Figures 10, 12 were plotted with special purpose programs.)

For the fixed points of Cremer type, the situation is much worse. As far as I know, no useful computer picture of such a point has ever been produced!

Many Julia sets are made up of very fine filaments. For such sets, it is essential to make some kind of distance estimate in order to obtain a sharp picture. For all of the center points of our pixels will quite likely lie outside of these filaments, and hence correspond to escaping orbits. But a good distance estimate can tell us that our pixel intersects the set $J(f)$, even though its center point is outside. In the case of the Mandelbrot set M , this procedure is even more important, since M contains both large regions and also very fine filaments. Indeed, it was precisely the difficulty of seeing such filaments which led to Mandelbrot's incorrect belief that M has many components.

Here is an example of how first derivatives can be used to make such distance estimates, following Fisher. Consider a rational map $f : \hat{\mathbf{C}} \to \hat{\mathbf{C}}$ with a superattractive fixed point at the origin. Let Ω be the basin of attraction of this fixed point. Let us assume, to simplify the discussion, that this basin is connected, simply connected, and contains no other critical point of f . Then it is not difficult to show that the Bötkher coordinate of §6.7 can be defined throughout Ω , and yields a conformal isomorphism $\phi : \Omega \to D$ with $\phi(f(z)) = \phi(z)^n$. Define the *canonical potential function* $G : \Omega - \{0\} \to \mathbf{R}$ by

$$G(z) = \log|\phi(z)| < 0 .$$

(Compare §17.) Denoting the gradient vector of G by G' , we will show that:

(1) *the function G and the norm $\|G'\| = |\phi'(z)/\phi(z)|$ are easy to compute, and*

(2) *the distance of z from the boundary of Ω can be computed, up to a factor of two, from a knowledge of G and $\|G'\|$.*

In fact, for any orbit $z_0 \mapsto z_1 \mapsto \cdots$ in Ω , it is easy to check that

$$G(z_0) = \lim_{k\to\infty} \log|z_k|/n^k .$$

Since the convergence is locally uniform, we can also write

$$\|G'(z_0)\| = \lim_{k\to\infty} \frac{|dz_k/dz_0|}{n^k|z_k|} .$$

In both cases, the successive terms can easily be computed inductively, and we obtain good approximations by iterating until $|z_k|$ is small. (If many iterations do not yield any small z_k , then we must assume that $z_0 \neq \Omega$, and set $G = 0$.)

Setting $\phi(z) = w$, a brief computation shows that the Poincaré metric on Ω can be written as

$$\frac{2|dw|}{1-|w|^2} = \frac{2|\phi'(z)dz|}{1-|\phi(z)|^2} = \frac{\|G'(z)dz\|}{|\sinh G(z)|} .$$

As an immediate consequence of the Quarter Theorem, §A.7, A.8, we obtain the following.

Corollary. *The distance between z and the boundary of Ω is equal to $|\sinh(G)|/\|G'\|$, up to a factor of two.*

If z is very close to $\partial\Omega$, then G is small and this distance estimate is very close to the ratio $|G|/\|G'\|$. It is interesting to note that this is just the step size which would be prescribed if we tried to solve the equation $G(z) = 0$ by Newton's method.

There are similar estimates in the case of a superattractive fixed point at infinity, or in other words for the basin $\mathbf{C}\smallsetminus K(f)$ of a polynomial map. For further information, the reader is referred to Fisher, to Milnor [M5], and to Peitgen.

REFERENCES

For general background, see for example:

[A1] L. Ahlfors, Complex Analysis, McGraw-Hill 1966.

[Mu] J. Munkres, Topology: A First Course, Prentice-Hall 1975.

[Wi] T. Willmore, An Introduction to Differential Geometry, Clarendon, 1959.

Surveys of conformal dynamics.

[Bl1] P. Blanchard, Complex analytic dynamics on the Riemann sphere, Bull. Amer. Math. Soc. **11** (1984), 85-141.

[BlC] P. Blanchard and A. Chiu, Conformal dynamics: an informal discussion, Lecture Notes, Boston University 1990.

[Br] H. Brolin, Invariant sets under iteration of rational functions, Arkiv för Mat. **6** (1965) 103-144.

[C] L. Carleson, Complex Dynamics, Lecture Notes UCLA 1990.

[Dv1] R. Devaney, An Introduction to Chaotic Dynamical Systems, 2^{nd} ed., Addison-Wesley, 1989.

[D1] A. Douady, Systèmes dynamiques holomorphes, Séminar Bourbaki, 35^e année 1982-83, n^o 599.

[Fa] K. Falconer, Fractal Geometry: Mathematical Foundations and Applications, Wiley 1990. (Ch. 14.)

[K1] L. Keen, Julia sets, pp. 57-74 of "Chaos and Fractals, the Mathematics behind the Computer Graphics", edit. Devaney and Keen, Proc. Symp. Appl. Math. **39**, Amer. Math. Soc. 1989.

[L1] M. Lyubich, The dynamics of rational transforms: the topological picture, Russian Math. Surveys **41:4** (1986), 43-117.

§1. Simply connected surfaces; uniformization.

[A2] L. Ahlfors, Conformal Invariants, McGraw-Hill 1973.

[AS] L. Ahlfors and L. Sario, Riemann Surfaces, Princeton U. Press 1960.

[Be] A. Beardon, A Primer on Riemann Surfaces, Cambridge U. Press 1984.

[FK] H. Farkas and I. Kra, Riemann Surfaces, Springer 1980.

[Sp] G. Springer, Introduction to Riemann Surfaces, Addison-Wesley 1957.

§2. Montel's theorem.

[Mo] P. Montel, Leçons sur les Familles Normales, Gauthier-Villars 1927.

§3. Fatou and Julia.

[Ca] A. Cayley, Application of the Newton-Fourier method to an immaginary root of an equation, Quart. J. Pure Appl. Math. **16** (1879) 179-185.

[F1] P. Fatou, Sur les solutions uniformes de certaines équations fonctionnelle, C. R. Acad. Sci. Paris **143** (1906) 546-548.

[F2] P. Fatou, Sur les équations fonctionnelles, Bull. Soc. math. France **47** (1919) 161-271, and **48** (1920) 33-94, 208-314.

[J] G. Julia, Memoire sur l'iteration des fonctions rationelles, J. Math. Pure Appl. **8** (1918) 47-245.

[Ri] J. F. Ritt, On the iteration of rational functions, Trans. Amer. Math. Soc. **21** (1920) 348-356.

§4. Iterated maps (Hyperbolic and Euclidean cases).

[Ba1] I. N. Baker, Repulsive fixedpoints of entire functions, Math. Zeit. **104** (1968) 252-256.

[Ba2], ———, An entire function which has wandering domains, J. Austral. Math. Soc. **22** (1976) 173-176.

[Dn1] A. Denjoy, Sur l'itération des fonctions analytiques, C.R.Acad.Sci. Paris **182** (1926) 255-257.

[Dv2] R. Devaney, Exploding Julia sets, pp. 141-154 of "Chaotic Dynamics and Fractals", edit. Barnsley and Demko, Academic Press 1986.

[EL] A. Eremenko and M. Lyubich, Dynamical properties of some classes of entire functions, Stony Brook Institute for Mathematical Sciences Preprint 1990#4. (See also Sov. Math. Dokl. **30** (1984) 592-594; Func. Anal. Appl. **19** (1985) 323-324; and J. Lond. Math. Soc. **36** (1987) 458-468.)

[F3] P. Fatou, Sur l'itération des fonctions transcendantes entières, Acta Math. **47** (1926) 337-370.

[GK] L. Goldberg and L. Keen, A finiteness theorem for a dynamical class of entire functions, Erg. Th. & Dy. Sy. **6** (1986) 183-192.

[K2] L. Keen, The dynamics of holomorphic self-maps of $\mathbf{C}^*$, in "Holomorphic Functions and Moduli", edit. Drasin et al., Springer 1988.

[L2] M. Lyubich, The measurable dynamics of the exponential map, Siber. J. Math. **28** (1987) 111-127. (See also Sov. Math. Dokl. **35** (1987) 223-226.)

[R1] M. Rees, The exponential map is not recurrent, Math. Zeit. **191** (1986) 593-598.

§5. Smooth Julia sets.

[La] S. Lattès, Sur l'iteration des substitutions rationelles et les fonctions de Poincaré , C. R. Acad. Sci. Paris **16** (1918) 26-28.

[He1] M. Herman, Exemples de fractions rationnelles ayant une orbite dense sur la sphère de Riemann, Bull. Soc. Math. France **112** (1984) 93-142.

[R2] M. Rees, Ergodic rational maps with dense critical point forward orbit, Erg. Th. & Dy. Sy. **4** (1984) 311-322.

[R3] M. Rees, Positive measure sets of ergodic rational maps, Ann. Sci. École Norm. Sup. (4) **19** (1986) 383-407.

[UvN] S. Ulam and J. von Neumann, On combinations of stochastic and deterministic processes, Bull. Amer. Math. Soc. **53** (1947) 1120.

§6. Attracting and repelling fixed points.

[Sch] E. Schröder, Ueber iterirte Functionen, Math. Ann. **3** (1871); see p. 303.

[Kœ] G. Kœnigs, Recherches sur les integrals de certains equations fonctionelles, Ann. Sci. Éc. Norm. Sup. (3^e ser.) **1** (1884) supplém. 1-41.

[Bö] L. E. Böttcher (Бётхеръ), The principal laws of convergence of iterates and their application to analysis (Russian), Izv. Kazan. Fiz.-Mat. Obshch. **14** (1904) 155-234.

§7. Parabolic fixed points.

[Le] L. Leau, Étude sur les equations fonctionelles à une ou plusièrs variables, Ann. Fac. Sci. Toulouse **11** (1897).

[Éc] J. Écalle, Théorie itérative: introduction a la théorie des invariants holomorphes, J. Math. Pure Appl. **54** (1975) 183-258.

[Ca] C. Camacho, On the local structure of conformal mappings and holomorphic vector fields, Astérisque **59-60** (1978) 83-94.

[MT] J. Milnor and W. Thurston, Iterated maps of the interval, pp. 465-563 of "Dynamical Systems (Maryland 1986-87)", edit. J.C.Alexander, Lect. Notes Math. **1342**, Springer 1988.

(*See also* [Bö, pp.201-234] *and* [F3].)

§8. Cremer points and Siegel disks.

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

[Cr1] H. Cremer, Zum Zentrumproblem, Math. Ann. **98** (1927) 151-163.

[Cr2] ————, Über die Häufigkeit der Nichtzentren, Math. Ann. **115** (1938) 573-580.

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

[HW] G. H. Hardy and E. M. Wright, An Introduction to the Theory of Numbers, Clarendon Press 1938, 1945, 1954.

[Br] A. D. Bryuno, Convergence of transformations of differential equations to normal forms, Dokl. Akad. Nauk USSR **165** (1965) 987-989.

[SiM] C. L. Siegel and J. Moser, Lectures on Celestial Mechanics, Springer 1971.

[He2] M. Herman, Recent results and some open questions on Siegel's linearization theorem of germs of complex analytic diffeomorphisms of C^n near a fixed point, pp. 138-198 of Proc 8^{th} Int. Cong. Math. Phys., World Sci. 1986.

[D2] A. Douady, Disques de Siegel et anneaux de Herman, Sém. Bourbaki 39^e année (1986-87) n^o 677.

[Y1] J.-C. Yoccoz, Linéarisation des germes de difféomorphismes holomorphes de $(\mathbf{C}, 0)$, C. R. Acad. Sci. Paris **306** (1988) 55-58.

§9. Holomorphic fixed point formula.

[AB] M. Atiyah and R. Bott, A Lefschetz fixed point formula for elliptic differential operators, Bull.A.M.S. **72** (1966) 245-250.

§10. Most periodic orbits repel.

[Shi] M. Shishikura, On the quasiconformal surgery of rational functions, Ann. Sci. Éc. Norm. Sup. **20** (1987) 1-29.

§12. Herman rings

[Ar] V. Arnold, Small denominators I, on the mappings of the circumference into itself, Amer. Math. Soc. Transl. (2) **46** (1965) 213-284.

[CL] E. Coddington and N. Levinson, Theory of Ordinary Differential Equations, McGraw-Hill 1955.

[Dn2] A. Denjoy, Sur les courbes définies par les équations différentielles à la surface du tore, Journ. de Math. **11** (1932) 333-375.

[He3] M. Herman, Sur la conjugation différentiables des difféomorphismes du cercle à les rotations, Pub. I.H.E.S. **49** (1979) 5-233.

[dM] W. de Melo, Lectures on One-Dimensional Dynamics, 17^o Col. Brasil. Mat., IMPA 1990.

[Y2] J.-C. Yoccoz, Conjugation différentiable des difféomorphismes du cercle dont le nombre de rotation vérifie une condition diophantienne, Ann. Sci. E.N.S. Paris (4) **17** (1984) 333-359.

(*See also* [He1], [D2], [Shi].)

§13. Classification of Fatou components.

[S1] D. Sullivan, Conformal dynamical systems, pp. 725-752 of "Geometric Dynamics", edit. Palis, Lecture Notes Math. **1007** Springer 1983.

[S2] D. Sullivan, Quasiconformal homeomorphisms and dynamics I, solution of the Fatou-Julia problem on wandering domains, Ann. Math. **122** (1985) 401-418.

(*See also* [DH1] ?? *and* [DH2] *below.*)

§14. Sub-hyperbolic and hyperbolic maps.

[Sm] S. Smale, Differentiable dynamical systems, Bull. Amer. Math. Soc. **73** (1967) 747-817.

[MSS] R. Mañé, P. Sad and D. Sullivan, On the dynamics of rational maps, Ann. Sci. Éc. Norm. Sup. Paris (4) **16** (1983) 193-217.

[L3] M. Lyubich, Some typical properties of the dynamics of rational maps, Russian Math. Surveys **38** (1983) 154-155. (See also Sov. Math. Dokl. **27** (1983) 22-25/

[L4] M. Lyubich, An analysis of the stability of the dynamics of rational functions, Selecta Math. Sovietica **9** (1990) 69-90. (Russian original published in 1984.)

[DH1] A. Douady and J. H. Hubbard, A proof of Thurston's topological characterization of rational functions, preprint, Mittag-Leffler 1984.

[Shu] M. Shub, Global Stability of Dynamical Systems, Springer 1987.

Basic reference for §§14 and 16-19:

[DH2] A. Douady and J. H. Hubbard, Étude dynamique des polynômes complexes I & II, Publ. Math. Orsay (1984-85).

§15. Prime ends.

[Ca] C. Carathéodory, Über die Begrenzung einfach zusammenhängender Gebiete, Math. Ann. **73** (1913) 323-370. (Gesam. Math. Schr., v. 4.)

[Ep] D.B.A.Epstein, Prime Ends, preprint, Univ. Warwick 1978.

[Ho] K. Hoffman, Banach Spaces of Analytic Functions, Prentice-Hall 1962.

[Ma] J. Mather, Topological proofs of some purely topological consequences of Carathéodory's theory of prime ends, pp. 225-255 of "Selected Studies", edit. T and G. Rassias, North-Holland 1982.

[Oh] M. Ohtsuka, Dirichlet Problem, Extremal Length and Prime Ends, van Nostrand 1970.

(*See also* [A2].)

§16. Local connectivity.

[Ku] K. Kuratowski, Topologie, Warsaw 1958-61.

[HY] Hocking and Young, Topology, Addison-Wesley 1961.

(*See also* [Mu].)

§17. The filled Julia set.

[Bro] M. Brown, A proof of the generalized Schoenflies theorem, Bulletin A.M.S. **66** (1960) 74-76. (See also "The monotone union of open n-cells is an open n-cell", Proc.A.M.S. **12** (1961) 812-814.)

[Bl2] P. Blanchard, Disconnected Julia sets, pp. 181-201 of "Chaotic Dynamics and Fractals", edit. Barnsley and Demko, Academic Press 1986.

[BH1] B. Branner and J. H. Hubbard, The iteration of cubic polynomials, Part I: the global topology of parameter space, Acta Math. **160** (1988) 143-206.

[BH2] B. Branner and J. H. Hubbard, The iteration of cubic polynomials, Part II: patterns and parapatterns, Acta Math., to appear.

§18. External rays and periodic points.

[Pe] C. Petersen, On the Pommerenke-Levin-Yoccoz inequality, preprint, IHES 1991.

[Po] C. Pommerenke, On conformal mapping and iteration of rational functions, Complex Var. Th. & Appl. **5**, 1986.

[G] L. Goldberg, Rotation cycles on the unit circle, Stony Brook IMS preprint 1990/14.

[GM] L. Goldberg and J. Milnor, Fixed point portraits of polynomial maps, Stony Brook IMS preprint 1990/14.

Appendix A. Theorems from classical analysis.

[Je] J. L. W. V. Jensen, Sur un nouvel et important théoreme de la théorie des fonctions, Acta Math. **22** (1899) 219-251.

[RR] F. and M. Riesz, Über Randwerte einer analytischen Funktionen, Quatr. Congr. Math. Scand. Stockholm 1916, 27-44.

[Ko] P. Koebe, Über die Uniformizierung beliebiger analytischer Kurven, Nachr. Akad. Wiss. Göttingen, Math.-Phys. Kl. (1907) 191-210.

[Gr] T. H. Gronwall, Some remarks on conformal representation, Ann. of Math. **16** (1914-15) 72-76.

[Bi] L. Bieberbach, Über die Koeffizienten derjenigen Potenzreihen welche eine schlichte Abbildung des Einheitskreises vermitteln, S.-B. Preuss. Akad. Wiss. (1916) 940-955.

[dB] L. de Brange, A proof of the Bieberbach conjecture, Acta Math. **154** (1985) 137-152.

Appendix B. Length-area-modulus inequalities.

[CR] R. Courant and H. Robbins, What is Mathematics?, Oxford U. Press 1941.

(*See also* [A2], [BH2].)

Appendix C. Continued fractions.

(*See* Hardy and Wright [HW].)

Appendix D. Two complex variables.

[Ze] E. Zehnder, A simple proof of a generalization of a theorem by C. L. Siegel, in "Geometry and Topology III", edit. do Carmo and Palis, Lecture Notes Math. **597**, Springer 1977.

[Ne] S. Newhouse, Diffeomorphisms with infinitely many sinks, Topology **16** (1974) 9-18.

[H] J. Hubbard, The Hénon mapping in the complex domain, pp. 101-111 of Chaotic Dynamics and Fractals, M. Barnsley and S. Demko, (ed.), Academic Press 1986.

[FM] S. Friedland and J. Milnor, Dynamical properties of plane polynomial automorphisms, Erg. Th. & Dy. Sy. **9** (1989) 67-99.

[Be] E. Bedford, Iteration of polynomial automorphisms of $\mathbf{C}^2$, Preprint, Purdue 1990.

Appendix E. Branched coverings and orbifolds

[Mas] B. Maskit, Kleinian Groups, Grundl. math. Wiss. **287** Springer 1987.

[M1] J. Milnor, On the 3-dimensional Brieskorn manifolds $M(p,q,r)$, pp. 175-225 of "Knots, Groups, and 3-Manifolds", edit. Neuwirth, Ann. Math. Studies **84**, Princeton U. Press 1975. (See §2.)

[T] W. Thurston, Three-dimensional Geometry and Topology, to appear.

Appendix F. Parameter space.

[BM] R. Brooks and P. Matelski, The dynamics of 2-generator subgroups of $PSL(2, \mathbf{C})$, pp. 65-71 of "Riemann Surfaces and Related Topics", Proceedings 1978 Stony Brook Conference, edit. Kra and Maskit, Ann. Math. Stud. **97** Princeton U. Press 1981.

[Man] B. Mandelbrot, Fractal aspects of the iteration of $z \mapsto \lambda z(1-z)$ for complex λ, z, Annals NY Acad. Sci. **357** (1980) 249-259.

[DH3] A. Douady and J. H. Hubbard, Itération des polynômes quadratiques complexes, CRAS Paris **294** (1982) 123-126.

[DH4] A. Douady and J. H. Hubbard, On the dynamics of polynomial-like mappings, Ann. Sci. Ec. Norm. Sup. (Paris) **18** (1985), 287-343.

[D3] A. Douady, Julia sets and the Mandelbrot set, pp. 161-173 of "The Beauty of Fractals", edit. Peitgen and Richter, Springer 1986.

[B1] B. Branner, The Mandelbrot set, pp. 75-105 of "Chaos and Fractals", edit. Devaney and Keen, Proc. Symp. Applied Math. **39**, Amer. Math. Soc. 1989.

[B2] B. Branner, The parameter space for complex cubic polynomials, pp. 169-179 of "Chaotic Dynamics and fractals", edit. Barnsley and Demko, Academic Press 1986.

[M2] J. Milnor, Remarks on iterated cubic maps, Stony Brook IMS preprint 1990/6.

[M3] J. Milnor, Hyperbolic components in spaces of polynomial maps, in preparation.

[M4] J. Milnor, On cubic polynomials with periodic critical point, in preparation.

[R4] M. Rees, Components of degree two hyperbolic rational maps, Invent. Math. **100** 1990, 357-382.

[R5] M. Rees, A partial description of parameter space of rational maps of degree two, Part I: preprint, Univ. Liverpool 1990; and Part II: Stony Brook IMS preprints 1991/4.

(*See also* [BH1].)

Appendix G. Computer graphics.

[M5] J. Milnor, Self-similarity and hairiness in the Mandelbrot set, pp. 211-257 of "Computers in Geometry and Topology", edit. Tangora, Lect. Notes Pure Appl. Math. **114**, Dekker 1989 (cf. p. 218 and §5).

[Fi] Y. Fisher, Exploring the Mandelbrot set, pp. 287-296 of The Science of Fractal Images, edit. Peitgen and Saupe, Springer 1989.

[Pei] H.-O. Peitgen, Fantastic deterministic fractals, ibid. pp. 169-218.

INDEX