Dynamics in One Complex Variable

Third Edition

John Milnor

ANNALS OF MATHEMATICS STUDIES

Annals of Mathematics Studies

Number 160

This page intentionally left blank

Dynamics in One Complex Variable

THIRD EDITION

by

John Milnor

PRINCETON UNIVERSITY PRESS

PRINCETON AND OXFORD

2006

Published by Princeton University Press, 41 William Street,
Princeton, New Jersey 08540
In the United Kingdom: Princeton University Press,
3 Market Place, Woodstock, Oxfordshire OX20 1SY

Library of Congress Cataloging-in-Publication Data

Milnor, John Willard, 1931–
Dynamics in one complex variable / John Milnor— 3rd ed.
p. cm. — (Annals of mathematics studies ; no. 160)
Includes bibliographical references and index
ISBN-13: 978-0-691-12487-2 (acid-free paper)
ISBN-10: 0-691-12487-6 (acid-free paper)
ISBN-13: 978-0-691-12488-9 (pbk. : acid-free paper)
ISBN-10: 0-691-12488-4 (pbk. : acid-free paper)
1. Functions of complex variables. 2. Holomorphic mappings.
3. Riemann surfaces. I. Title. II. Series.
QA331.7.M55 2006
515'.93—dc22
2005051060

British Library Cataloging-in-Publication Data is available.

The publisher would like to acknowledge the author of this
volume for providing the camera-ready copy from which this book was printed.

This book has been composed in Computer Modern Roman.

Printed on acid-free paper.∞
www.pup.princeton.edu

Printed in the United States of America

1 3 5 7 9 10 8 6 4 2

TABLE OF CONTENTS

LIST OF FIGURES

PREFACE TO THE THIRD EDITION

This book studies the dynamics of iterated holomorphic mappings from a Riemann surface to itself, concentrating on the classical case of rational maps of the Riemann sphere. It is based on introductory lectures given at Stony Brook during the fall term of 1989–90 and in later years. I am grateful to the audiences for a great deal of constructive criticism and to Bodil Branner, Adrien Douady, John Hubbard, and Mitsuhiro Shishikura, who taught me most of what I know in this field. Also, I want to thank a number of individuals for their extremely helpful criticisms and suggestions, especially Adam Epstein, Rodrigo Perez, Alfredo Poirier, Lasse Rempe, and Saeed Zakeri. Araceli Bonifant has been particularly helpful in the preparation of this third edition.

There have been a number of extremely useful surveys of holomorphic dynamics over the years. See the textbooks by Devaney [1989], Beardon [1991], Carleson and Gamelin [1993], Steinmetz [1993], and Berteloot and Mayer [2001], as well as expository articles by Brolin [1965], Douady [1982–83, 1986, 1987], Blanchard [1984], Lyubich [1986], Branner [1989], Keen [1989], Blanchard and Chiu [1991], and Eremenko and Lyubich [1990]. (See the list of references at the end of the book.)

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 2-dimensional differential geometry, as well as some basic topics from topology. The necessary material can be found for example in Ahlfors [1966], Hocking and Young [1961], Munkres [1975], Thurston [1997], and Willmore [1959]. However, two big theorems will be used here without proof, namely the Uniformization Theorem in §1 and the existence of solutions for the measurable Beltrami equation in Appendix F. (See the references in those sections.)

The basic outline of this third edition has not changed from previous editions, but there have been many improvements and additions. A brief historical survey has been added in §4.1, the definition of Lattès map has been made more inclusive in §7.4, the Écalle-Voronin theory of parabolic points is described in §10.12, the résidu itératif is studied in §12.9, the material on two complex variables in Appendix D has been expanded, and recent results on effective computability have been added in Appendix H. The list of references has also been updated and expanded.

John Milnor
Stony Brook, August 2005

CHRONOLOGICAL TABLE

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

Ernst Schröder	1841–1902
Hermann Amandus Schwarz	1843–1921
Henri Poincaré	1854–1912
Gabriel Kœnigs	1858–1931
Léopold Leau	1868–1940(?)
Lucjan Emil Böttcher	1872– ?
Samuel Lattès	1873–1918
Constantin Carathéodory	1873–1950
Paul Montel	1876–1975
Pierre Fatou	1878–1929
Paul Koebe	1882–1945
Arnaud Denjoy	1884–1974
Gaston Julia	1893–1978
Carl Ludwig Siegel	1896–1981
Hubert Cremer	1897–1983
Herbert Grötzsch	1902–1993
Charles B. Morrey	1907–1984
Lars Ahlfors	1907–1996
Lipman Bers	1914–1993
Irvine Noel Baker	1932–2001
Michael (Michel) R. Herman	1942–2000

Among the many present-day workers in the field, let me mention a few whose work is emphasized in these notes: Adrien Douady (b. 1935), Dennis P. Sullivan (b. 1941), Bodil Branner (b. 1943), John Hamal Hubbard (b. 1945), William P. Thurston (b. 1946), Mary Rees (b. 1953), Jean-Christophe Yoccoz (b. 1955), Curtis McMullen (b. 1958), Mikhail Y. Lyubich (b. 1959), and Mitsuhiro Shishikura (b. 1960).

RIEMANN SURFACES

§1. Simply Connected Surfaces

The first three sections will present an overview of some background material.

If $V \subset \mathbb{C}$ is an open set of complex numbers, a function $f : V \to \mathbb{C}$ is called *holomorphic* (or "complex analytic") if the first derivative

$$z \mapsto f'(z) = \lim_{h \to 0} (f(z+h) - f(z))/h$$

is defined and continuous as a function from V to $\mathbb{C}$, or equivalently if f has a power series expansion about any point $z_0 \in V$ which converges to f in some neighborhood of z_0. (See, for example, Ahlfors [1966].) Such a function is *conformal* if the derivative $f'(z)$ never vanishes. Thus our conformal maps must always preserve orientation. It is *univalent* (or *schlicht*) if it is conformal and one-to-one.

By a *Riemann surface* S we mean a connected complex analytic manifold of complex dimension 1. Thus S is a connected Hausdorff space. Furthermore, in some neighborhood U of an arbitrary point of S we can choose a *local uniformizing parameter* (or "coordinate chart") which maps U homeomorphically onto an open subset of the complex plane $\mathbb{C}$, with the following property: In the overlap $U \cap U'$ between two such neighborhoods, each of these local uniformizing parameters can be expressed as a holomorphic function of the other.

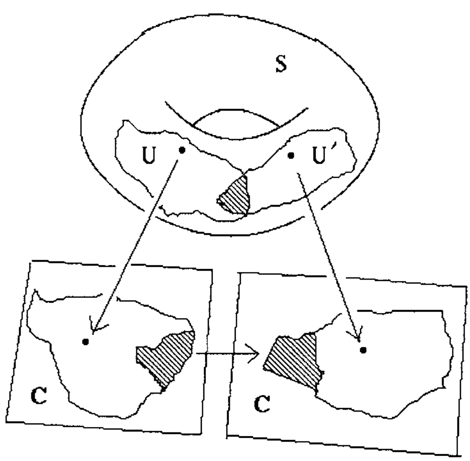


Figure 1. Overlapping coordinate neighborhoods.

By definition, two Riemann surfaces S and S' are *conformally isomorphic* (or *biholomorphic*) if and only if there is a homeomorphism from S onto S' which is holomorphic in terms of the respective local uniformizing parameters. (It is an easy exercise to show that the inverse map $S' \to S$ must then also be holomorphic.) In the special case $S = S'$, such a conformal isomorphism $S \to S$ is called a *conformal automorphism* of S.

Although there are uncountably many conformally distinct Riemann surfaces, there are only three distinct surfaces in the simply connected case. (By definition, the surface S is *simply connected* if every map from a circle into S can be continuously deformed to a constant map. Compare §2.) The following result is due to Poincaré and to Koebe.

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

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

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

(*c*) *to the Riemann sphere* $\widehat{\mathbb{C}}$ *consisting of* $\mathbb{C}$ *together with a point at infinity, using* $\zeta = 1/z$ *as local uniformizing parameter in a neighborhood of the point at infinity.*

This is a generalization of the classical Riemann Mapping Theorem. We will refer to these three cases as the *Euclidean, hyperbolic,* and *spherical* cases, respectively. (Compare §2.) The proof of Theorem 1.1 is nontrivial and will not be given here. However, proofs may be found in Koebe [1907], Ahlfors [1973], Beardon [1984], Farkas and Kra [1980], Nevanlinna [1967], and in Springer [1957]. (See also Fisher, Hubbard, and Wittner [1988].) By assuming this result, we will be able to pass more quickly to interesting ideas in holomorphic dynamics.

The Open Disk $\mathbb{D}$. For the rest of this section, we will discuss these three surfaces in more detail. We begin with a study of the unit disk $\mathbb{D}$.

Lemma 1.2 (Schwarz Lemma). *If* $f : \mathbb{D} \to \mathbb{D}$ *is a holomorphic map with* $f(0) = 0$, *then the derivative at the origin satisfies* $|f'(0)| \le 1$. *If equality holds,* $|f'(0)| = 1$, *then* f *is a rotation about the origin. That is,* $f(z) = cz$ *for some constant* $c = f'(0)$ *on the unit circle. On the other hand, if* $|f'(0)| < 1$, *then* $|f(z)| < |z|$ *for all* $z \neq 0$.

(The Schwarz Lemma was first proved, in this generality, by Carathéodory.)

Remarks. If $|f'(0)| = 1$, it follows that f is a conformal automorphism of the unit disk. But if $|f'(0)| < 1$ then f cannot be a conformal automorphism of $\mathbb{D}$, since the composition with any $g : (\mathbb{D}, 0) \to (\mathbb{D}, 0)$ would have derivative $g'(0)f'(0) \neq 1$. The example $f(z) = z^2$ shows that f may map $\mathbb{D}$ *onto* itself even when $|f(z)| < |z|$ for all $z \neq 0$ in $\mathbb{D}$.

Proof of Lemma 1.2. We use the *Maximum Modulus Principle*, which asserts that a nonconstant holomorphic function cannot attain its maximum absolute value at any interior point of its region of definition. First note that the quotient function $q(z) = f(z)/z$ is well defined and holomorphic throughout the disk $\mathbb{D}$, as one sees by dividing the local power series for f by z. Since $|q(z)| < 1/r$ when $|z| = r < 1$, it follows by the Maximum Modulus Principle that $|q(z)| < 1/r$ for all z in the disk $|z| \leq r$. Since this is true for all $r \to 1$, it follows that $|q(z)| \leq 1$ for all $z \in \mathbb{D}$. Again by the Maximum Modulus Principle, we see that the case $|q(z)| = 1$, for some z in the open disk, can occur only if the function $q(z)$ is constant. If we exclude this case $f(z)/z \equiv c$, then it follows that $|q(z)| = |f(z)/z| < 1$ for all $z \neq 0$, and similarly that $|q(0)| = |f'(0)| < 1$. $\square$

Here is a useful variant statement.

Lemma 1.2′ (Cauchy Derivative Estimate). *If f maps the disk of radius r about z_0 into some disk of radius s, then*

$$|f'(z_0)| \;\leq\; s/r\,.$$

Proof. This follows easily from the Cauchy integral formula (see, for example, Ahlfors [1966]): Set $g(z) = f(z + z_0) - f(z_0)$, so that g maps the disk $\mathbb{D}_r$ centered at the origin to the disk $\mathbb{D}_s$ centered at the origin. Then

$$f'(z_0) \;=\; g'(0) \;=\; \frac{1}{2\pi i}\oint_{|z|=r_1} \frac{g(z)\,dz}{z^2}$$

for all $r_1 < r$, and the conclusion follows easily. $\square$

(An alternative proof, based on the Schwarz Lemma, is described in Problem 1-a at the end of this section. With an extra factor of 2 on the right, this inequality would follow immediately from Lemma 1.2 simply by linear changes of variable, since the target disk of radius s must be contained in the disk of radius $2s$ centered at the image $f(z_0)$.)

As an easy corollary, we obtain the following.

Theorem 1.3 (Liouville Theorem). *A bounded function f which is defined and holomorphic everywhere on $\mathbb{C}$ must be constant.*

For in this case we have s fixed but r arbitrarily large, hence f' must be identically zero. □

As another corollary, we see that our three model surfaces really are distinct. There are natural inclusion maps $\mathbb{D} \to \mathbb{C} \to \widehat{\mathbb{C}}$. Yet it follows from the Maximum Modulus Principle that every holomorphic map $\widehat{\mathbb{C}} \to \mathbb{C}$ must be constant, and from Liouville's Theorem that every holomorphic map $\mathbb{C} \to \mathbb{D}$ must be constant.

Another closely related statement is the following. Let U be an open subset of $\mathbb{C}$.

Theorem 1.4 (Weierstrass Uniform Convergence Theorem). *If a sequence of holomorphic functions $f_n : U \to \mathbb{C}$ converges uniformly to the limit function f, then f itself is holomorphic. Furthermore, the sequence of derivatives f'_n converges, uniformly on any compact subset of U, to the derivative f'.*

It follows inductively that the sequence of second derivatives f''_n converges, uniformly on compact subsets, to f'', and so on.

Proof of Theorem 1.4. Note first that the sequence of first derivatives f'_n, restricted to any compact subset $K \subset U$, converges uniformly. For example, if $|f_n(z) - f_m(z)| < \epsilon$ for $m, n > N$, and if the r-neighborhood of any point of K is contained in U, then it follows from Lemma 1.2′ that $|f'_n(z) - f'_m(z)| < \epsilon/r$ for $m, n > N$ and for all $z \in K$. This proves uniform convergence of $\{f'_n\}$ restricted to K to some limit function g, which is necessarily continuous since any uniform limit of continuous functions is continuous. It follows that the integral of f'_n along any path in U converges to the integral of g along this path. Thus $f = \lim f_n$ is an indefinite integral of g, and hence g can be identified with the derivative of f. Thus f has a continuous complex first derivative and therefore is a holomorphic function. □

Conformal Automorphism Groups. For any Riemann surface S, the notation $\mathcal{G}(S)$ will be used for the group consisting of all conformal automorphisms of S. The identity map will be denoted by $I = I_S \in \mathcal{G}(S)$.

We first consider the case of the Riemann sphere $\widehat{\mathbb{C}}$ and show that $\mathcal{G}(\widehat{\mathbb{C}})$ can be identified with a well-known complex Lie group. Thus $\mathcal{G}(\widehat{\mathbb{C}})$ is not only a group, but also a complex manifold, and the product and inverse operations for this group are both holomorphic maps.

Lemma 1.5 (Möbius Transformations). *The group $\mathcal{G}(\hat{\mathbb{C}})$ of all conformal automorphisms of the Riemann sphere is equal to the group of all fractional linear transformations (also called Möbius transformations)*

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

where the coefficients are complex numbers with $ad-bc\neq 0$.

Here, if we multiply numerator and denominator by a common factor, then it is always possible to normalize so that the determinant $ad-bc$ is equal to $+1$. The resulting coefficients are well defined up to a simultaneous change of sign. *Thus it follows that the group $\mathcal{G}(\hat{\mathbb{C}})$ of conformal automorphisms can be identified with the complex 3-dimensional Lie group* $\mathbf{PSL}(2,\mathbb{C})$, *consisting of all 2×2 complex matrices with determinant $+1$ modulo the subgroup $\{\pm I\}$.* Since the complex dimension is 3, it follows that the real dimension of $\mathbf{PSL}(2,\mathbb{C})$ is 6.

Proof of Lemma 1.5. It is easy to check that $\mathcal{G}(\hat{\mathbb{C}})$ contains this group of fractional linear transformations as a subgroup. After composing the given $g\in\mathcal{G}(\hat{\mathbb{C}})$ 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 from $\mathbb{C}\smallsetminus\{0\}$ to itself. (In fact, $g(z)/z$ tends to the nonzero finite value $g'(0)$ as $z\to 0$. Setting $\zeta=1/z$ and $G(\zeta)=1/g(1/\zeta)$, evidently $g(z)/z=\zeta/G(\zeta)$ tends to the nonzero finite value $1/G'(0)$ as $z\to\infty$.) Setting $z=e^w$, it follows that the composition $w\mapsto g(e^w)/e^w$ is a bounded holomorphic function on $\mathbb{C}$. Hence it takes a constant value c by Liouville's Theorem. Therefore $g(z)=cz$ is linear, and hence g itself is an element of $\mathbf{PSL}(2,\mathbb{C})$. $\square$

Next we will show that both $\mathcal{G}(\mathbb{C})$ and $\mathcal{G}(\mathbb{D})$ can be considered as Lie subgroups of $\mathcal{G}(\hat{\mathbb{C}})$.

Corollary 1.6 (The Affine Group). *The group $\mathcal{G}(\mathbb{C})$ of all conformal automorphisms of the complex plane consists of all affine transformations*

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

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

Proof. First note that every conformal automorphism f of $\mathbb{C}$ extends uniquely to a conformal automorphism of $\hat{\mathbb{C}}$. In fact $\lim_{z\to\infty}f(z)=\infty$, so the singularity of $1/f(1/\zeta)$ at $\zeta=0$ is removable. (Compare Ahlfors [1966, p. 124].) It follows that $\mathcal{G}(\mathbb{C})$ can be identified with the subgroup of $\mathcal{G}(\hat{\mathbb{C}})$ consisting of Möbius transformations which fix the point ∞. Evidently this is just the complex 2-dimensional subgroup consisting of all complex affine transformations of $\mathbb{C}$. □

Theorem 1.7 (Automorphisms of $\mathbb{D}$). *The group $\mathcal{G}(\mathbb{D})$ of all conformal automorphisms of the unit disk can be identified with the subgroup of $\mathcal{G}(\widehat{\mathbb{C}})$ consisting of all maps*

$$f(z) = e^{i\theta}\frac{z-a}{1-\bar{a}z} \tag{1:1}$$

where a ranges over the open disk $\mathbb{D}$ and where $e^{i\theta}$ ranges over the unit circle $\partial\mathbb{D}$.

This is no longer a *complex* Lie group. However, $\mathcal{G}(\mathbb{D})$ is a *real* 3-dimensional Lie group, having the topology of a "solid torus" $\mathbb{D}\times\partial\mathbb{D}$.

Proof of Theorem 1.7. Evidently the map f defined by (1:1) carries the entire Riemann sphere $\widehat{\mathbb{C}}$ conformally onto itself. A brief computation shows that

$$\begin{aligned} |f(z)|<1 \quad &\Longleftrightarrow \quad (z-a)(\bar{z}-\bar{a}) < (1-\bar{a}z)(1-a\bar{z}) \\ &\Longleftrightarrow \quad (1-z\bar{z})\,(1-a\bar{a}) > 0\,. \end{aligned}$$

For any $a\in\mathbb{D}$, it follows that $|f(z)|<1 \Longleftrightarrow |z|<1$. Hence f maps $\mathbb{D}$ onto itself. Now if $g:\mathbb{D}\xrightarrow{\cong}\mathbb{D}$ is an arbitrary conformal automorphism and $a\in\mathbb{D}$ is the unique solution to the equation $g(a)=0$, then we can consider $f(z)=(z-a)/(1-\bar{a}z)$, which also maps a to zero. The composition $g\circ f^{-1}$ is an automorphism fixing the origin, hence it has the form $g\circ f^{-1}(z)=e^{i\theta}z$ by the Schwarz Lemma, and $g(z)=e^{i\theta}f(z)$, as required. □

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

Lemma 1.8 ($\mathbb{D}\cong\mathbb{H}$). *The half-plane $\mathbb{H}$ is conformally isomorphic to the disk $\mathbb{D}$ under the holomorphic mapping*

$$w \mapsto (i-w)/(i+w),$$

with inverse

$$z \mapsto i(1-z)/(1+z),$$

where $z\in\mathbb{D}$ and $w\in\mathbb{H}$.

Proof. If z and $w=u+iv$ are complex numbers related by these formulas, then $|z|^2<1$ 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 $v>0$. □

Corollary 1.9 (Automorphisms of $\mathbb{H}$). *The group $\mathcal{G}(\mathbb{H})$ consisting of all conformal automorphisms of the upper half-plane can be identified with the group of all fractional linear transformations $w \mapsto (aw+b)/(cw+d)$, where the coefficients a, b, c, d are real 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 $\mathcal{G}(\mathbb{H})$ is isomorphic to the group $\mathbf{PSL}(2,\mathbb{R})$, consisting of all 2×2 real matrices with determinant $+1$ modulo the subgroup $\{\pm I\}$.*

Proof of Corollary 1.9. If $f(w) = (aw+b)/(cw+d)$ with real coefficients and nonzero determinant, then it is easy to check that f maps $\mathbb{R}\cup\infty$ homeomorphically onto itself. Note that the image

$$f(i) \;=\; (ai+b)(-ci+d)/(c^2+d^2)$$

lies in the upper half-plane $\mathbb{H}$ if and only if $ad-bc>0$. It follows easily that this group $\mathbf{PSL}(2,\mathbb{R})$ of positive real fractional linear transformations acts as a group of conformal automorphisms of $\mathbb{H}$. This group acts transitively. In fact the subgroup consisting of all $w\mapsto aw+b$ with $a>0$ already acts transitively, since such a map carries the point i to a completely arbitrary point $ai+b\in\mathbb{H}$. Furthermore, $\mathbf{PSL}(2,\mathbb{R})$ contains the group of "rotations"

$$g(w) = (w\cos\theta+\sin\theta)/(-w\sin\theta+\cos\theta), \qquad (1:2)$$

which fix the point $g(i)=i$ with derivative $g'(i)=e^{2i\theta}$. By Lemmas 1.2 and 1.8, there can be no further automorphisms fixing i, and it follows easily that $\mathcal{G}(\mathbb{H})\cong\mathbf{PSL}(2,\mathbb{R})$. □

To conclude this section, we will try to say something more about the structure of these three groups. For any map $f:X\to X$, it will be convenient to use the notation $\mathrm{Fix}(f)\subset X$ for the set of all fixed points $x=f(x)$. If f and g are commuting maps from X to itself, $f\circ g = g\circ f$, note that

$$f\big(\mathrm{Fix}(g)\big) \;\subset\; \mathrm{Fix}(g). \qquad (1:3)$$

For if $x\in\mathrm{Fix}(g)$, then $f(x)=f\circ g(x)=g\circ f(x)$, hence $f(x)\in\mathrm{Fix}(g)$. We first apply these ideas to the group $\mathcal{G}(\mathbb{C})$ of all affine transformations of $\mathbb{C}$.

Lemma 1.10 (Commuting Elements of $\mathcal{G}(\mathbb{C})$). *Two non-identity affine transformations of $\mathbb{C}$ commute if and only if they have the same fixed point set.*

It follows easily that any $g \neq I$ in the group $\mathcal{G}(\mathbb{C})$ is contained in a unique maximal abelian subgroup consisting of all f with $\mathrm{Fix}(f) = \mathrm{Fix}(g)$, together with the identity element.

Proof of Lemma 1.10. Clearly an affine transformation with two fixed points must be the identity map. If g has just one fixed point z_0, then it follows from (1:3) that any f which commutes with g fixes this same point. The set of all such f forms a commutative group, consisting of all $f(z) = z_0 + \lambda(z - z_0)$ with $\lambda \neq 0$. Similarly, if $\mathrm{Fix}(g)$ is the empty set, then g is a translation $z \mapsto z + c$, and $f \circ g = g \circ f$ if and only if f is also a translation. □

Now consider the group $\mathcal{G}(\hat{\mathbb{C}})$ of automorphisms of the Riemann sphere. By definition, an automorphism g is called an *involution* if $g \circ g = I$, but $g \neq I$.

Theorem 1.11 (Commuting Elements of $\mathcal{G}(\hat{\mathbb{C}})$). *For every $f \neq I$ in $\mathcal{G}(\hat{\mathbb{C}})$, the set $\mathrm{Fix}(f) \subset \hat{\mathbb{C}}$ contains either one point or two points. In general, two nonidentity elements $f, g \in \mathcal{G}(\hat{\mathbb{C}})$ commute if and only if $\mathrm{Fix}(f) = \mathrm{Fix}(g)$. The only exceptions to this statement are provided by pairs of commuting involutions, each of which interchanges the two fixed points of the other.*

(Compare Problem 1-c. As an example, the involution $f(z) = -z$ with $\mathrm{Fix}(f) = \{0, \infty\}$ commutes with the involution $g(z) = 1/z$ with $\mathrm{Fix}(g) = \{\pm 1\}$.)

Proof of Theorem 1.11. 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 $\hat{\mathbb{C}}$. (If an automorphism of $\hat{\mathbb{C}}$ fixes three distinct points, then it must be the identity map.)

If f commutes with g, which has exactly two fixed points, then since $f(\mathrm{Fix}(g)) = \mathrm{Fix}(g)$ by (1:3), it follows that f either must have the same two fixed points or must interchange the two fixed points of g. In the first case, taking the fixed points to be 0 and ∞, it follows that both f and g belong to the commutative group consisting of all linear maps $z \mapsto \lambda z$ with $\lambda \in \mathbb{C} \smallsetminus \{0\}$. In the second case, if f interchanges 0 and ∞, then it is necessarily a transformation of the form $f(z) = \eta/z$, with $f \circ f(z) = z$. Setting $g(z) = \lambda z$, the equation $g \circ f = f \circ g$ reduces to $\lambda^2 = 1$, so that g must also be an involution.

Finally, suppose that g has just one fixed point, which we may take to be the point at infinity. Then by (1:3) any f which commutes with g

must also fix the point at infinity. Hence we are reduced to the situation of Lemma 1.10, and both f and g must be translations $z \mapsto z + c$. (Such automorphisms with just one fixed point, at which the first derivative is necessarily $+1$, are called *parabolic* automorphisms.) This completes the proof. $\square$

We want a corresponding statement for the open disk $\mathbb{D}$. However, it is better to work with the closed disk $\overline{\mathbb{D}}$, in order to obtain a richer set of fixed points. Using Theorem 1.7, we see easily that every automorphism of the open disk extends uniquely to an automorphism of the closed disk.

Theorem 1.12 (Commuting Elements of $\mathcal{G}(\mathbb{D})$). *For every $f \neq I$ in $\mathcal{G}(\mathbb{D}) \cong \mathcal{G}(\overline{\mathbb{D}})$, the set $\mathrm{Fix}(f) \subset \overline{\mathbb{D}}$ consists of either a single point of the open disk $\mathbb{D}$, a single point of the boundary circle $\partial\mathbb{D}$, or two points of $\partial\mathbb{D}$. Two nonidentity automorphisms $f, g \in \mathcal{G}(\mathbb{D})$ commute if and only if they have the same fixed point set in $\overline{\mathbb{D}}$.*

Remark 1.13. (Compare Problem 1-d.) An automorphism of $\overline{\mathbb{D}}$ is often described as "elliptic," "parabolic," or "hyperbolic" according to whether it has one interior fixed point, one boundary fixed point, or two boundary fixed points. We can describe these transformations geometrically as follows. In the elliptic case, after conjugating by a transformation which carries the fixed point to the origin, we may assume that $0 = g(0)$. It then follows from the Schwarz Lemma that g is just a rotation about the origin. In the parabolic case, it is convenient to replace $\mathbb{D}$ by the upper half-plane, choosing the isomorphism $\mathbb{D} \cong \mathbb{H}$ so that the boundary fixed point corresponds to the point at infinity. Using Corollary 1.9, we see that g must correspond to a linear transformation $w \mapsto aw+b$ with a, b real and $a > 0$. Since there are no fixed points in $\mathbb{R} \subset \partial\mathbb{H}$, it follows that $a = 1$, so that we have a horizontal translation. Similarly, in the hyperbolic case, taking the fixed points to be $0, \infty \in \partial\mathbb{H}$, we see that g must correspond to a linear map of the form $w \mapsto aw$ with $a > 0$. (It is rather inelegant that we must extend to the boundary in order to distinguish between the parabolic and hyperbolic cases. For a more intrinsic interpretation of this dichotomy see Problem 1-f, or Problem 2-e in §2.)

Proof of Theorem 1.12. In fact every automorphism of $\mathbb{D}$ or $\overline{\mathbb{D}}$ is a Möbius transformation and hence extends uniquely to an automorphism F of the entire Riemann sphere. This extension commutes with the inversion map $\alpha(z) = 1/\bar{z}$. In fact the composition $\alpha \circ F \circ \alpha$ is a holomorphic map which coincides with F on the unit circle and hence coincides with F everywhere. Thus F has a fixed point z in the open disk $\mathbb{D}$ if and only if it has a corresponding fixed point $\alpha(z)$ in the exterior $\widehat{\mathbb{C}} \smallsetminus \overline{\mathbb{D}}$. It now follows from Theorem 1.11 that two elements of $\mathcal{G}(\overline{\mathbb{D}})$ commute if and only if they have the same fixed point set in $\overline{\mathbb{D}}$, providing that we can exclude the possibility of two commuting involutions. However, if $F \in \mathcal{G}(\widehat{\mathbb{C}})$ is an involution, note that the derivative $F'(z)$ at each of the two fixed points must be -1. Thus, if F maps $\mathbb{D}$ onto itself, neither of these fixed points can be on the boundary circle, hence one fixed point must be in $\mathbb{D}$ and one in $\widehat{\mathbb{C}} \smallsetminus \overline{\mathbb{D}}$. Therefore, a second involution which commutes with F and interchanges these two fixed points cannot map $\mathbb{D}$ onto itself. This completes the proof. $\square$

Concluding Problems

Problem 1-a. Alternate proof of Lemma 1.2′. **(1)** Check that an arbitrary conformal automorphism

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

of the unit disk satisfies $|g'(0)| = |1 - a\bar{a}| \le 1$. **(2)** Since any holomorphic map $f : \mathbb{D} \to \mathbb{D}$ can be written as a composition $g \circ h$ where g is an automorphism mapping 0 to $f(0)$ and where h is a holomorphic map which fixes the origin, conclude using Lemma 1.2 that $|f'(0)| \le 1$ even when $f(0) \ne 0$. **(3)** More generally, if f maps the disk of radius r centered at z into some disk of radius s, show that $|f'(z)| \le s/r$.

Problem 1-b. Triple transitivity. **(1)** Show that the action of the group $\mathcal{G}(\widehat{\mathbb{C}})$ on $\widehat{\mathbb{C}}$ is *simply 3-transitive*. That is, there is one and only one automorphism which carries three distinct specified points of $\widehat{\mathbb{C}}$ into three other specified points. **(2)** Similarly, show that the action of $\mathcal{G}(\mathbb{C})$ on $\mathbb{C}$ is simply 2-transitive. (For corresponding statements for the disk $\mathbb{D}$, see Problem 2-d.)

Problem 1-c. Cross-ratios. **(1)** Show that the group $\mathcal{G}(\widehat{\mathbb{C}})$ is generated by the subgroup of affine transformations $z \mapsto az + b$ together with the inversion $z \mapsto 1/z$. **(2)** Given four distinct points z_j in $\widehat{\mathbb{C}}$, show that the *cross-ratio**

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

---

*Caution: Several variant notations for cross-ratios are in common use. This particular version, characterized by the property that $\chi(0,1,z,\infty) = z$, is particularly convenient for our purposes. Compare Problem 2-c and note that $\chi(a,b,c,d) > 1$ whenever a,b,c,d are real with $a < b < c < d$.

is invariant under fractional linear transformations. (If one of the z_j is the point at infinity, this definition extends by continuity.) **(3)** Show that χ is real if and only if the four points lie on a straight line or circle. **(4)** Given two points $z_1 \neq z_2$ show that there is one and only one involution f with $\mathrm{Fix}(f) = \{z_1, z_2\}$ and show that a second involution g with $\mathrm{Fix}(g) = \{z_1', z_2'\}$ commutes with f if and only if $\chi(z_1, z_2, z_1', z_2') = 1/2$.

Problem 1-d. Conjugacy classes in $\mathcal{G}(\mathbb{H})$. By definition, a conformal automorphism of $\mathbb{D}$ or $\mathbb{H}$ is *elliptic* if it has a fixed point, and otherwise is *parabolic* or *hyperbolic* according to whether its extension to the boundary circle has one or two fixed points. **(1)** Classify conjugacy classes in the group $\mathcal{G}(\mathbb{H}) \cong \mathbf{PSL}(2, \mathbb{R})$ as follows. Show that every automorphism of $\mathbb{H}$ 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 $w_0 \in \mathbb{H}$ is uniquely determined by the derivative $\lambda = g'(w_0)$, where $|\lambda| = 1$. **(2)** Show also that each nonidentity element of $\mathbf{PSL}(2, \mathbb{R})$ 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 to whether its elements are parabolic or hyperbolic or elliptic. Here t ranges over the additive group of real numbers.

Problem 1-e. The Euclidean case. Show that the conjugacy class of a nonidentity automorphism $g(z) = \lambda z + c$ in the group $\mathcal{G}(\mathbb{C})$ is uniquely determined by its image under the derivative homomorphism $g \mapsto g' \equiv \lambda \in \mathbb{C} \smallsetminus \{0\}$.

Problem 1-f. Antiholomorphic involutions. By an *antiholomorphic* mapping from one Riemann surface to another, we mean a transformation which, in terms of local coordinates z and w, takes the form $z \mapsto w = \eta(\bar{z})$ where η is holomorphic. By an *antiholomorphic involution* of S we mean an antiholomorphic map $\alpha : S \to S$ such that $\alpha \circ \alpha$ is the identity map. **(1)** If L is a straight line in $\mathbb{C}$, show that there is one and only one antiholomorphic involution of $\mathbb{C}$ having L as fixed point set and show that no other fixed point sets can occur. **(2)** Show that the automorphism group $\mathcal{G}(\mathbb{C})$ acts transitively on the set of straight lines in $\mathbb{C}$. **(3)** Similarly, if L is *either* a straight line or a circle in $\widehat{\mathbb{C}}$, show that there is one and only one antiholomorphic involution of $\widehat{\mathbb{C}}$ having L as fixed point set and show that no other nonvacuous fixed point sets can occur. **(4)** Show that the automorphism group $\mathcal{G}(\widehat{\mathbb{C}})$ acts transitively on the set of straight lines and circles in $\widehat{\mathbb{C}}$. **(5)** For an antiholomorphic involution of $\mathbb{D}$, show that the fixed point set is either a diameter of $\mathbb{D}$ or a circle arc meeting the boundary $\partial\mathbb{D}$ orthogonally, and show that $\mathcal{G}(\mathbb{D})$ acts transitively on the set of all such diameters and circle arcs. **(6)** Finally, show that an automorphism of $\mathbb{D}$ without interior fixed point is hyperbolic if and only if it commutes with some antiholomorphic involution, or if and only if it carries some such diameter or circle arc into itself.

Problem 1-g. Fixed points of Möbius transformations. **(1)** For a nonidentity automorphism $g \in \mathcal{G}(\widehat{\mathbb{C}})$, show that the derivatives $g'(z)$ at the two fixed points are reciprocals, say λ and λ^{-1}. **(2)** Show that the average $\alpha = (\lambda + \lambda^{-1})/2$ is a complete conjugacy class invariant which can take any value in $\mathbb{C}$. (In the special case of a fixed point at infinity, one must evaluate the derivative using the local uniformizing parameter $\zeta = 1/z$.) **(3)** Show that $\alpha = 1$ if and only if the two fixed points coincide and that $-1 \le \alpha = \cos\theta < 1$ if and only if g is conjugate to a rotation through angle θ.

Problem 1-h. Convergence to zero. **(1)** If a holomorphic map $f : \mathbb{D} \to \mathbb{D}$ 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 $\mathbb{D}$. **(2)** Prove that this convergence is uniform on compact subsets of $\mathbb{D}$. (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 $\mathbb{D}$.)

§2. Universal Coverings and the Poincaré Metric

First recall some standard topological constructions. (Compare Munkres [1975], as well as Appendix E.) A map $p : M \to N$ between connected manifolds is called a *covering map* if every point of N has a connected open neighborhood U within N which is *evenly covered*; that is, each component of $p^{-1}(U)$ must map onto U by a homeomorphism. The manifold N is *simply connected* if it has no nontrivial coverings, that is, if every such covering map $M \to N$ is a homeomorphism. (Equivalently, N is simply connected if and only if every map from a circle to N can be continuously deformed to a point.) For any connected manifold N, there exists a covering map $\widetilde{N} \to N$ such that $\widetilde{N}$ is simply connected. This is called the *universal covering* of N and is unique up to homeomorphism. By a *deck transformation* associated with a covering map $p : M \to N$ we mean a continuous map $\gamma : M \to M$ which satisfies the identity $p \circ \gamma = p$, so that the diagram

$$\begin{tikzcd}[column sep=small] M \arrow[rr, "\gamma"] \arrow[dr, "p"'] & & M \arrow[dl, "p"] \\ & N & \end{tikzcd}$$

is commutative. For our purposes, the *fundamental group* $\pi_1(N)$ can be defined as the group Γ consisting of all deck transformations for the universal covering $\widetilde{N} \to N$. Note that this universal covering is always a *normal covering* of N. That is, given two points $x, x' \in M = \widetilde{N}$ with $p(x) = p(x')$, there exists one and only one deck transformation which maps x to x'. It follows that N can be identified with the quotient $\widetilde{N}/\Gamma$ of $\widetilde{N}$ by this action of Γ. A given group Γ of homeomorphisms of a connected manifold M gives rise in this way to a normal covering $M \to M/\Gamma$ if and only if

(1) Γ acts *properly discontinuously*; that is, any compact set $K \subset M$ intersects only finitely many of its translates $\gamma(K)$ under the action of Γ; and

(2) Γ acts *freely*; that is, every nonidentity element of Γ acts without fixed points on M.

Now let S be a Riemann surface. Then the universal covering manifold $\widetilde{S}$ inherits the structure of a Riemann surface, and every deck transformation is a conformal automorphism of $\widetilde{S}$. According to the Uniformization Theorem 1.1, since this universal covering surface $\widetilde{S}$ is simply connected, it must be conformally isomorphic to one of the three model surfaces. Thus we have the following.

Theorem 2.1 (Uniformization for Arbitrary Riemann Surfaces). *Every Riemann surface S is conformally isomorphic to a quotient of the form $\widetilde{S}/\Gamma$, where $\widetilde{S}$ is a simply connected Riemann surface, which is necessarily isomorphic to either $\mathbb{D}$, $\mathbb{C}$, or $\widehat{\mathbb{C}}$, and where $\Gamma \cong \pi_1(S)$ is a group of conformal automorphisms which acts freely and properly discontinuously on $\widetilde{S}$.*

The group $\mathcal{G}(\widetilde{S})$ consisting of *all* conformal automorphisms of $\widetilde{S}$ has been studied in §1. It is a Lie group, and in particular has a natural topology. Since the action of Γ on $\widetilde{S}$ is properly discontinuous, it is not difficult to check that Γ must be a *discrete* subgroup of $\mathcal{G}(\widetilde{S})$; that is, there exists a neighborhood of the identity element in $\mathcal{G}(\widetilde{S})$ which intersects Γ only in the identity element. (Compare Problem 2-a.)

As a curious consequence, we obtain a remarkable property of complex 1-manifolds, which was first proved by Radó. (Compare Ahlfors and Sario [1960].) By definition, a topological space is *σ-compact* if it can be expressed as a countable union of compact subsets.

Corollary 2.2 (σ-Compactness). *Every Riemann surface can be expressed as a countable union of compact subsets.*

(It can be shown that a connected manifold is σ-compact if and only if it is paracompact, or metrizable, or has a countable basis for the open subsets. However, in general, manifolds need not satisfy any of these conditions.)

Proof of Corollary 2.2. This follows from Theorem 2.1, since the corresponding property is clearly true for each of the three simply connected surfaces. $\square$

We can now give a very rough catalogue of all possible Riemann surfaces. The discussion will be divided into two easy cases and one hard case.

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

Euclidean Case. By Corollary 1.6, the group $\mathcal{G}(\mathbb{C})$ 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 $S \cong \mathbb{C}/\Gamma$ is a surface with universal covering $\widetilde{S} \cong \mathbb{C}$, then Γ must be a discrete group of translations $z \mapsto z + c$ of the complex plane $\mathbb{C}$. There are three subcases:

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

- If Γ has just one generator, then S is isomorphic to the infinite *cylinder* $\mathbb{C}/\mathbb{Z}$, where $\mathbb{Z} \subset \mathbb{C}$ is the additive subgroup of integers. Note that this cylinder is isomorphic to the *punctured plane* $\mathbb{C} \smallsetminus \{0\}$ under the isomorphism

$$z \mapsto \exp(2\pi iz) \in \mathbb{C} \smallsetminus \{0\}.$$

- If Γ has two generators, then it can be described as a 2-dimensional *lattice* $\Lambda \subset \mathbb{C}$, that is, an additive group generated by two complex numbers which are linearly independent over $\mathbb{R}$. (Two generators, such as 1 and $\sqrt{2}$, which are dependent over $\mathbb{R}$ would not generate a discrete group.) The quotient $\mathbb{T} = \mathbb{C}/\Lambda$ is called a *torus*.

In all three subcases, note that our surface inherits a locally Euclidean geometry from the Euclidean metric $|dz|$ on its universal covering surface. As an example, the punctured plane $\mathbb{C} \smallsetminus \{0\}$, consisting of points $\exp(2\pi iz) = w$, has a complete locally Euclidean metric $2\pi|dz| = |dw/w|$. (Such a metric is well defined only up to multiplication by a positive constant, since we could equally well use a coordinate of the form $z' = \lambda z + c$ in the universal covering, with $|dz'| = |\lambda\, dz|$. Compare Corollary 1.6.) It will be convenient to use the term *Euclidean surface* for these Riemann surfaces, which admit a complete locally Euclidean metric. The term *parabolic surface* is also commonly used in the literature.

Hyperbolic Case. In all other cases, the universal covering $\widetilde{S}$ must be conformally isomorphic to the unit disk. Such Riemann surfaces are said to be *hyperbolic*. It follows from the discussion above that $\widehat{\mathbb{C}}$, $\mathbb{C}$, $\mathbb{C}/\mathbb{Z}$, and the various tori $\mathbb{C}/(\mathbb{Z} \oplus \tau\mathbb{Z})$ are the only nonhyperbolic Riemann surfaces, up to conformal isomorphism. In particular:

> *Any Riemann surface which is not homeomorphic to the sphere or torus* (*in the compact case*), *or homeomorphic to the plane or cylinder* (*in the noncompact case*), *must necessarily be hyperbolic, with universal covering surface conformally isomorphic to the unit disk.*

For example, any Riemann surface of genus ≥ 2, or more generally any Riemann surface with nonabelian fundamental group, is hyperbolic. (Compare Problem 2-g.)

Remark. Here the word "hyperbolic" is a reference to hyperbolic Geometry, that is, the non-Euclidean geometry of Lobachevsky and Bolyai. (Compare Corollary 2.10 below.) Unfortunately the term "hyperbolic" has at least three quite distinct well-established meanings in holomorphic dynamics. We may refer to a *hyperbolic periodic orbit* (with multiplier off the unit circle), or to a *hyperbolic map* (see §19), or to a *hyperbolic surface*, as here. In order to avoid confusion, I will sometimes use the more explicit phrase *conformally hyperbolic* when the word is used with this geometric meaning, and reserve the phrase *dynamically hyperbolic* for the other two meanings.

The inclusions $\mathbb{D} \to \mathbb{C} \to \widehat{\mathbb{C}}$ provide examples of nonconstant holomorphic maps from the hyperbolic surface $\mathbb{D}$ to the Euclidean surface $\mathbb{C}$ and then to the Riemann sphere $\widehat{\mathbb{C}}$. However, no maps in the other direction are possible:

Lemma 2.3 (Maps between Surfaces of Different Type). *Every holomorphic map from a Euclidean Riemann surface to a hyperbolic one is necessarily constant. Similarly, every holomorphic map from the Riemann sphere to a Euclidean or hyperbolic surface is necessarily constant.*

Proof. Any holomorphic map $f : S \to S'$ can be lifted to a holomorphic map $\widetilde{f} : \widetilde{S} \to \widetilde{S}'$ between universal covering surfaces. (Compare Problem 2-b at the end of this section.) However, as noted following Theorem 1.3, any holomorphic map $\hat{\mathbb{C}} \to \mathbb{C}$ or $\mathbb{C} \to \mathbb{D}$ must be constant, by the Maximum Modulus Principle and by Liouville's Theorem. □

Example 2.4. The Annulus and the Punctured Disk. We have seen that all Euclidean Riemann surfaces have abelian fundamental group, either trivial or isomorphic to $\mathbb{Z}$ or $\mathbb{Z} \oplus \mathbb{Z}$. However, there also exist hyperbolic surfaces with fundamental group $\mathbb{Z}$. The *punctured disk* $\mathbb{D} \smallsetminus \{0\}$ provides one example, and each *annulus*

$$\mathbb{A}_r = \{ z \in \mathbb{C} \, ; \, 1 < |z| < r \}$$

provides another example. It follows immediately from Lemma 2.3 that these surfaces are indeed hyperbolic. We can also see this by a more explicit construction as follows. The exponential map $z \mapsto e^z$ carries the left half-plane $\{x + iy \, ; \, x < 0\}$ onto the punctured disk by a universal covering map. Hence the fundamental group $\pi_1(\mathbb{D} \smallsetminus \{0\})$ can be identified with the group of deck transformations for the exponential map, which is free cyclic, generated by the translation $z \mapsto z + 2\pi i$. Similarly, the exponential map carries the vertical strip $\{x + iy \, ; \, 0 < x < \log r\}$ onto the annulus $\mathbb{A}_r$ by a universal covering map. But it is easy to see that such a strip is conformally isomorphic to the upper half-plane $\mathbb{H}$, since the exponential map carries the analogous horizontal strip $\{x + iy \, ; \, 0 < y < \pi\}$ diffeomorphically onto $\mathbb{H}$. Again the group of deck transformations is isomorphic to $\mathbb{Z}$.

In fact annuli and the punctured disk are the only hyperbolic surfaces with abelian fundamental group, other than the disk itself. A closely related property is that annulus, punctured disk, and disk all have a nontrivial Lie group of automorphisms. This is not possible for any other hyperbolic surface. (See Problems 2-f and 2-g.)

The next example will play a fundamental role in later sections. (Compare Theorem 3.7.)

Lemma 2.5 (The Triply Punctured Sphere). *If we remove three or more points from the Riemann sphere, then the resulting Riemann surface S is hyperbolic, with universal covering $\widetilde{S}$ conformally isomorphic to the disk.*

Proof. This follows immediately from the discussion above, since S is clearly not homeomorphic to the plane or cylinder, for example, since its fundamental group is nonabelian. A more elementary argument, not involving the fundamental group, can be given as follows. For any sufficiently large compact set $K \subset S$, note that the complement $S \smallsetminus K$ has at least three connected components. We say that the thrice punctured sphere has three *ends*, while any nonhyperbolic surface has at most two ends. □

For an explicit description of the universal covering map $\mathbb{H} \to \mathbb{C} \smallsetminus \{0,1\}$ for a thrice punctured sphere, see, for example, Ahlfors [1966], and compare Problem 7-g.

Closely related is the statement that any Riemann surface can be made hyperbolic by removing at most three points. (In the case of a torus, it suffices to remove just one point, since that will correspond to removing infinitely many points in the universal covering of the torus.)

Theorem 2.6 (Picard's Theorem). *Every holomorphic map $f : \mathbb{C} \to \mathbb{C}$ which omits two different values must necessarily be constant.*

This follows immediately from Lemmas 2.3 and 2.5. For if f omits two values a, b, then it can be considered as a map from the Euclidean surface $\mathbb{C}$ to the hyperbolic surface $\mathbb{C} \smallsetminus \{a, b\}$. □

The Poincaré Metric. Every hyperbolic surface has a preferred Riemannian metric, constructed as follows. We first consider the simply connected case.

Lemma 2.7 (The Poincaré Metric on $\mathbb{D}$). *There exists one and, up to multiplication by a positive constant, only one Riemannian metric on the disk $\mathbb{D}$ which is invariant under every conformal automorphism of $\mathbb{D}$.*

As an immediate corollary, we get exactly the same statement for the upper half-plane $\mathbb{H}$, or for any other surface which is conformally isomorphic to $\mathbb{D}$.

Proof of Lemma 2.7. Geometrically, we can prove this statement as follows. To define a Riemannian metric on a smooth manifold M, we must assign a length $\|v\|$ to every tangent vector v at every point of M. Consider then a tangent vector v to the open disk $\mathbb{D}$ at some point $z_0 \in \mathbb{D}$. Choose an automorphism $g \in \mathcal{G}(\mathbb{D})$ which maps z_0 to the origin. Then the first derivative of g at z_0 yields a linear map Dg_{z_0} from the tangent space of $\mathbb{D}$ at z_0 onto the tangent space at the origin. We define $\|v\|$ to be *twice* the Euclidean length of the image vector $Dg_{z_0}(v)$. (The factor of 2 is inserted for convenience; compare formula (2:3) below.) Since g is unique up to composition with a rotation of the disk, this length $\|v\|$ is well defined and is clearly invariant under all automorphisms of $\mathbb{D}$. Finally, since the correspondence $v \mapsto \|v\|^2$, for tangent vectors at a specified point of $\mathbb{D}$, is clearly a homogeneous quadratic function, this construction does indeed yield a Riemannian metric.

Alternatively, using classical notations, we can prove Lemma 2.7 more explicitly as follows. A *Riemannian metric* on an open subset of $\mathbb{C}$ can be described as an expression of the form

$$ds^2 \;=\; g_{11}dx^2 + 2g_{12}dxdy + g_{22}dy^2,$$

where $[g_{jk}]$ is a positive definite matrix which depends smoothly on the point $z = x + iy$. Such a metric 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 positive multiple of the identity matrix. In other words, a conformal metric is one which 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 $w = f(z)$ if and only if it satisfies the identity $\gamma(w)|dw| = \gamma(z)|dz|$, or in other words,

$$\gamma(f(z)) = \gamma(z)/|f'(z)|. \tag{2:1}$$

Equivalently, an f satisfying this condition is called 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 $f(w) = aw + b$ (where $a > 0$). Since $f(i) = ai + b$, equation (2:1) takes the form $\gamma(ai+b) = \gamma(i)/a$. After multiplying the metric by a positive constant, we may assume that $\gamma(i) = 1$. Thus we are led to the formula $\gamma(u+iv) = 1/v$, or in other words

$$ds = |dw|/v \qquad \text{for} \quad w = u + iv \in \mathbb{H}. \tag{2:2}$$

In fact, the metric defined by this formula is invariant under *every* conformal automorphism g of $\mathbb{H}$. For, if we select some arbitrary point $w_1 \in \mathbb{H}$ and set $g(w_1) = w_2$, then g can be expressed as the composition of a linear automorphism of the form $g_1(w) = aw + b$ which maps w_1 to w_2 and an automorphism g_2 which fixes w_2. We have constructed the metric (2:2) so that g_1 is an isometry, and it follows from Lemmas 1.2 and 1.8 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 Lemma 2.7, we must show that a metric which is invariant under all automorphisms of $\mathbb{D}$ or $\mathbb{H}$ is necessarily conformal. For this purpose, given any point $w_0 \in \mathbb{H}$ choose the unique automorphism f which fixes the point w_0 and has derivative $f'(w_0) = \sqrt{-1}$. A brief computation shows that the induced map on Riemannian metrics takes the expression $g_{11}du^2 + 2g_{12}du\,dv + g_{22}dv^2$ at the point w_0 to the expression $g_{22}dv^2 - 2g_{12}du\,dv + g_{11}dv^2$ at w_0. Thus invariance implies that $g_{11} = g_{22}$ and $g_{12} = 0$ at the arbitrary point w_0, as required. $\square$

Definition. This metric $ds = |dw|/v$ is called the *Poincaré metric* on the upper half-plane $\mathbb{H}$. The corresponding expression on the unit disk $\mathbb{D}$ is

$$ds = 2|dz|/(1 - |z|^2) \qquad \text{for} \qquad z \in \mathbb{D}, \tag{2:3}$$

as can be verified by a brief computation using Lemma 1.8 and (2:1).

Remark. The most basic invariant for a Riemannian metric on a surface S is the *Gaussian curvature* function $K : S \to \mathbb{R}$. Since there is an isometry carrying any point of $\mathbb{D}$ (or of $\mathbb{H}$) to any other point, it follows that the Poincaré metric has *constant* Gaussian curvature. In fact this metric, as defined above, has Gaussian curvature $K \equiv -1$. (Compare Problem 2-h.)

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

Thus there is a preferred Riemannian metric ds on $\mathbb{D}$ or on $\mathbb{H}$. More generally, if S is any hyperbolic surface, then the universal covering $\widetilde{S}$ is conformally isomorphic to $\mathbb{D}$, and hence has a preferred metric which is invariant under all conformal isomorphisms of $\widetilde{S}$. In particular, it is invariant under deck transformations. It follows that there is one and only one Riemannian metric on S so that the projection $\widetilde{S} \to S$ is a *local* *isometry,* mapping any sufficiently small open subset of $\widetilde{S}$ isometrically onto its image in S. By definition, the metric ds constructed in this way is called the *Poincaré metric* on the hyperbolic surface S.

Example 2.8. The Punctured Disk. The universal covering surface for the punctured disk $\mathbb{D} \smallsetminus \{0\}$ can be identified with the left half-plane $\{w = u + iv\,;\, u < 0\}$ under the exponential map

$$w \mapsto z = e^w \in \mathbb{D} \smallsetminus \{0\},$$

with $dz/z = dw$. Evidently the Poincaré metric $|dw/u|$ on the left half-plane corresponds to the metric $|dz/r \log r|$ on the punctured disk, where $r = |z|$ and $u = \log r$. (Thus the circle $|z| = r$ has length $2\pi/|\log r|$, which tends to zero as $r \to 0$, although this circle has infinite Poincaré distance from the boundary point $z = 0$.) A neighborhood of zero, intersected with $\mathbb{D} \smallsetminus \{0\}$, can be embedded isometrically as a surface of revolution in Euclidean 3-space. (The generating curve is known as a "tractrix.")

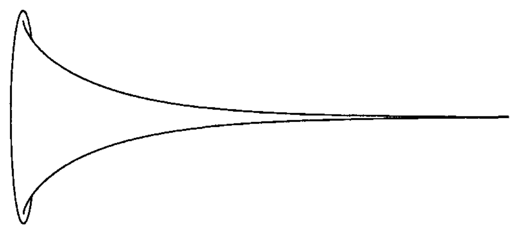


Figure 2. A surface of revolution of constant negative curvature.

Definition. Let S be a hyperbolic surface with Poincaré metric ds. The integral $\int_P ds$ along any piecewise smooth path $P : [0,1] \to S$ is called the *Poincaré length* of this path. For any two points z_1 and z_2 in S, the *Poincaré distance* $\mathrm{dist}(z_1, z_2) = \mathrm{dist}_S(z_1, z_2)$ is defined to be the infimum, over all piecewise smooth paths P joining z_1 to z_2, of the Poincaré length $\int_P ds$. In fact we will see that there always exists a path of minimal length.

Lemma 2.9 (Completeness Lemma). *Every hyperbolic surface S is complete with respect to its Poincaré metric. That is:*

(*a*) *every Cauchy sequence with respect to the metric* $\mathrm{dist_S}$ *converges, or equivalently:*

(*b*) *every closed neighborhood*

$$N_r(z_0, \mathrm{dist}_S) = \{z \in S \,;\, \mathrm{dist}_S(z, z_0) \le r\}$$

is a compact subset of S. Furthermore:

(*c*) *any two points of S are joined by at least one minimal geodesic.*

(In the simply connected case, there is exactly one geodesic between any two points.)

Proof of Lemma 2.9. First consider the special case $S = \mathbb{D}$. Given any two points of $\mathbb{D}$ we can first choose a conformal automorphism which moves the first point to the origin and the second to some point r on the positive real axis. For any path P between 0 and r within $\mathbb{D}$ we have

$$\int_P ds = \int_P \frac{2|dz|}{1-|z|^2} \ge \int_P \frac{2|dx|}{1-x^2} \ge \int_0^r \frac{2dx}{1-x^2} = \log\frac{1+r}{1-r},$$

with equality if and only if P is the straight line segment $[0,r]$. For any $z \in \mathbb{D}$, it follows that the Poincaré distance from 0 to z is given by

$$\delta = \mathrm{dist}_{\mathbb{D}}(0,z) = \log\frac{1+|z|}{1-|z|}.$$

(Compare Problem 2-c. Equivalently, we can write $|z| = \tanh(\delta/2)$.) Furthermore, the straight line segment from 0 to z is the unique minimal Poincaré geodesic. This proves (b) and (c) in the simply connected case. The general case follows immediately, and assertion (a) then follows easily. (Compare Willmore [1959].) □

Corollary 2.10 (Constant Curvature Metrics). *Every Riemann surface admits a complete conformal metric with constant curvature which is either positive, negative, or zero according to whether the surface is spherical, hyperbolic, or Euclidean.*

In fact, in the hyperbolic case there is one and only one conformal metric which is complete, with constant Gaussian curvature equal to -1. (Compare Problem 2-i.) In the Euclidean case, the corresponding metric is unique only up to multiplication by a positive constant. In the spherical case, identifying the Riemann sphere $\widehat{\mathbb{C}}$ with the unit sphere in $\mathbb{R}^3$ under stereographic projection, we obtain the standard *spherical metric*

$$ds = 2|dz|/(1+|z|^2), \tag{2:4}$$

with constant Gaussian curvature $+1$. (See Problem 2-h.) This spherical metric is smooth and well behaved, even in a neighborhood of the point at infinity. In fact the inversion map $z \mapsto 1/z$ is an isometry. However, the spherical metric is far from unique, since it is not preserved by most Möbius transformations. Its group of orientation-preserving isometries $\mathbf{SO}(3)$ is much smaller than the full group $\mathcal{G}(\widehat{\mathbb{C}})$ of all conformal automorphisms. □

Remark. For computer calculations, a more convenient metric for $\widehat{\mathbb{C}}$ is given by the chordal distance formula

$$\operatorname{dist}'(z_1\,,\,z_2) \;=\; \frac{2\,|z_1 - z_2|}{\sqrt{(1+|z_1|^2)\,(1+|z_2|^2)}} \;=\; 2\sin(s/2), \qquad (2{:}5)$$

where $s = \operatorname{dist}(z_1, z_2)$ is the usual spherical distance. As an example, using (2:5) the distance between z and the "antipodal" point $-1/\bar{z}$ is always equal to $+2$.

These nonhyperbolic metrics of curvature ≥ 0 are certainly of interest. However, in the hyperbolic case, the Poincaré metric with curvature -1 is of fundamental importance because of its marvelous property of never increasing under holomorphic maps.

Theorem 2.11 (Pick Theorem). *If $f : S \to S'$ is a holomorphic map between hyperbolic surfaces, then exactly one of the following three statements is valid:*

- *f is a conformal isomorphism from S onto S', and it maps S with its Poincaré metric isometrically onto S' with its Poincaré metric.*

- *f is a covering map but is not one-to-one. In this case, it is locally but not globally a Poincaré isometry. Every smooth path $P : [0,1] \to S$ of arclength ℓ in S maps to a smooth path $f \circ P$ of the same length ℓ in S', and it follows that*

$$\operatorname{dist}_{S'}(f(p), f(q)) \;\leq\; \operatorname{dist}_S(p, q)$$

for every $p, q \in S$. Here equality holds whenever p is sufficiently close to q, but strict inequality will hold, for example, if $f(p) = f(q)$ with $p \neq q$.

- *In all other cases, f strictly decreases all nonzero distances. In fact, for any compact set $K \subset S$ there is a constant $c_K < 1$ so that*

$$\operatorname{dist}_{S'}(f(p), f(q)) \;\leq\; c_K \operatorname{dist}_S(p, q)$$

for every $p, q \in K$ and so that every smooth path in K with arclength ℓ (using the Poincaré metric for S) maps to a path of Poincaré arclength $\leq c_K\ell$ in S'.

Here is an example to illustrate Theorem 2.11. The map $f(z) = z^2$ on the disk $\mathbb{D}$ is certainly not a covering map or a conformal automorphism. Hence it is distance decreasing for the Poincaré metric on $\mathbb{D}$. On the other hand, we can also consider f as a map from the *punctured disk* $\mathbb{D} \smallsetminus \{0\}$ to itself. In this case, f is a two-to-one covering map. Hence f is a local isometry for the Poincaré metric on $\mathbb{D} \smallsetminus \{0\}$. In fact, the universal covering of $\mathbb{D} \smallsetminus \{0\}$ can be identified with the left half-plane, mapped onto $\mathbb{D} \smallsetminus \{0\}$ by the exponential map. (Compare Example 2.8.) Then f lifts to the automorphism $F : w \mapsto 2w$ of this half-plane, which evidently preserves the Poincaré metric.

Proof of Theorem 2.11. Let TS_p be the tangent space of S at p. This is a complex 1-dimensional vector space. We will think of the Poincaré metric on S as specifying a *norm* $\|v\|$ for each vector $v \in TS_p$, with $\|v\| > 0$ for $v \neq 0$. The holomorphic map $f : S \to S'$ induces a linear first derivative map $Df_p : TS_p \to TS'_{f(p)}$. Let us compare the Poincaré norm $\|v\|$ of a vector $v \in TS_p$ with the Poincaré norm of its image in $TS'_{f(p)}$. Evidently the ratio

$$\|Df_p(v)\|/\|v\|$$

is independent of the choice of nonzero vector v and can be described as the *norm* $\|Df_p\|$ of the first derivative at p. In the special case of a fixed point $z = f(z)$ of a map on a hyperbolic open subset of $\mathbb{C}$, note that $\|Df_z\|$ can be identified with the absolute value of the classical first derivative $f'(z) = df/dz$. Therefore, for a holomorphic map $f : \mathbb{D} \to \mathbb{D}$ with $f(0) = 0$, the Schwarz Lemma asserts that $\|Df_0\| \leq 1$, with equality if and only if f is a conformal automorphism. More generally, if $f : S \to S'$ is a holomorphic map between simply connected hyperbolic surfaces, and if $p \in S$, it follows immediately that $\|Df_p\| \leq 1$, with equality if and only if f is a conformal isomorphism. Now consider the case where S and S' are not necessarily simply connected. Choose some lifting $F : \widetilde{S} \to \widetilde{S}'$ to the universal covering surfaces and some point $\tilde{p}$ over p. From the commutative diagram

$$\begin{array}{ccc} T\widetilde{S}_{\tilde{p}} & \longrightarrow & T\widetilde{S}'_{F(\tilde{p})} \\ \downarrow & & \downarrow \\ TS_p & \longrightarrow & TS'_{f(p)} \end{array}$$

where the vertical arrows preserve the Poincaré norm and where both $\widetilde{S}$ and $\widetilde{S}'$ are conformally isomorphic to $\mathbb{D}$, we see that $\|Df_p\| \leq 1$, with equality if and only if F is a conformal isomorphism from $\widetilde{S}$ onto $\widetilde{S}'$, or in other words if and only if $f : S \to S'$ is a covering map. (Compare Problem 2-b.)

In particular, if f is not a covering map, then F cannot be a conformal isomorphism, and hence $\|Df_p\| < 1$ for all $p \in S$. If K is a compact subset of S, it follows by continuity that $\|Df_p\|$ attains some maximum value $c < 1$ as p varies over K. Now for any smooth path $P : [0, 1] \to S$

the derivative DP_t carries the unit tangent vector at $t \in [0,1]$ to a vector in $TS_{P(t)}$ which is called the *velocity vector* $P'(t)$ for the path P at $P(t)$. By definition the Poincaré arclength of P is the integral

$$\mathrm{length}_S(P) \;=\; \int_0^1 \|P'(t)\|\, dt.$$

Similarly

$$\mathrm{length}_{S'}(f \circ P) \;=\; \int_0^1 \|Df_{P(t)}(P'(t))\|\, dt,$$

so if $\|Df_p\| \le c$ throughout K, it follows that

$$\mathrm{length}_{S'}(f \circ P) \;\le\; c\ \mathrm{length}_S(P)$$

for every smooth path within K. In order to compare distances within K, it is necessary to choose some larger compact set $K' \subset S$ so that any two points p and q of K can be joined by a geodesic of length $\mathrm{dist}_S(p,q)$ within K'. If $c_K < 1$ is the maximum of $\|Df_p\|$ for $p \in K'$, then it follows that

$$\mathrm{dist}_{S'}(f(p), f(q)) \;\le\; c_K\ \mathrm{dist}_S(p,q),$$

as required. $\square$

Remark. In the distance-reducing case, it may happen that there is a uniform constant $c < 1$ so that

$$\mathrm{dist}_{S'}(f(p), f(q)) \;\le\; c\, \mathrm{dist}_S(p,q)$$

for *all* p and q in S. In the special case of a map from S to itself, a standard argument then shows that f has a (necessarily unique) fixed point. (See Problem 2-j.) However, the example of the map $f(w) = w + i$ from the upper half-plane into itself shows that a distance-reducing map need not have a fixed point. Even if f does have a fixed point, it does not follow that such a constant $c < 1$ exists. For example, if $f(z) = z^2$, mapping the unit disk onto itself, then a brief computation shows that $\|Df_z\| = 2|z|/(1+|z|^2)$, taking values arbitrarily close to $+1$.

One important application of Theorem 2.11 is to the inclusion $\iota : S \to S'$ where S' is a hyperbolic Riemann surface and S is a connected open subset. If $S \ne S'$ then it follows from Theorem 2.11 that

$$\mathrm{dist}_{S'}(p,q) \;<\; \mathrm{dist}_S(p,q) \tag{2:6}$$

for every $p \ne q$ in S. *Thus distances measured relative to a larger Riemann surface are always smaller.* For sharper forms of this inequality see Theorem 3.4, as well as Corollary A.8 in the appendix.

Concluding Problems

Problem 2-a. Properly discontinuous groups. (1) Let S be a simply connected Riemann surface, and let $\Gamma \subset \mathcal{G}(S)$ be a discrete group of automorphisms; that is, suppose that the identity element is an isolated point of Γ within the Lie group $\mathcal{G}(S)$. If every nonidentity 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 γ 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. (More generally, analogous statements are true for any discrete group of isometries of a Riemannian manifold.) **(2)** On the other hand, show that the free cyclic group consisting of all transformations $z \mapsto 2^n z$ of $\mathbb{C}$, with $n \in \mathbb{Z}$, forms a discrete subgroup of $\mathcal{G}(\mathbb{C})$, but is not properly discontinuous.

Problem 2-b. Lifting to the universal covering. (1) If $S \cong \mathbb{D}/\Gamma$ and $S' \cong \mathbb{D}/\Gamma'$ are hyperbolic surfaces, show that any holomorphic $f : S \to S'$ lifts to a holomorphic map $\tilde{f} : \mathbb{D} \to \mathbb{D}$, unique up to composition with an element of Γ'. **(2)** Show that $\tilde{f}$ induces a group homomorphism $\gamma \mapsto \gamma'$ from Γ to Γ', satisfying the identity

$$\tilde{f} \circ \gamma = \gamma' \circ \tilde{f}$$

for every $\gamma \in \Gamma$. **(3)** Show that f is a covering map if and only if $\tilde{f}$ is a conformal automorphism.

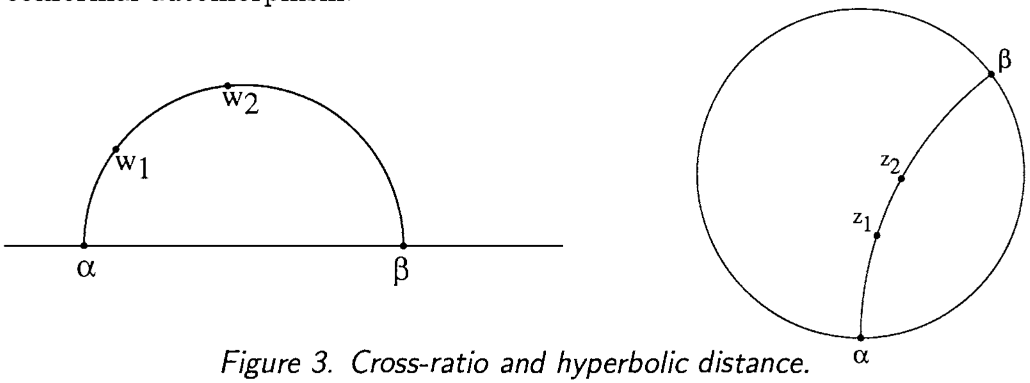


Figure 3. Cross-ratio and hyperbolic distance.

Problem 2-c. Poincaré geodesics. (1) Show that each geodesic for the Poincaré metric in the upper half-plane is a straight line or semicircle which meets the real axis orthogonally. (Compare Problem 1-f. A related statement, for any Riemannian manifold, is that a curve which is the fixed point set of an isometric involution must necessarily be a geodesic.)

(2) If the geodesic through w_1 and w_2 meets $\partial\mathbb{H} = \mathbb{R} \cup \infty$ 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 $\chi(\alpha\,,\,w_1\,,\,w_2\,,\,\beta)$, as defined in Problem 1-c. **(3)** Show that each Poincaré neighborhood $N_r(w_0\,,\,\mathrm{dist}_{\mathbb{H}})$ in the upper half-plane is bounded by a Euclidean circle, but that w_0 is not its Euclidean center. Prove corresponding statements for the unit disk.

Problem 2-d. The action of $\mathcal{G}(\mathbb{D})$. **(1)** Show that the action of the automorphism group $\mathcal{G}(\mathbb{D})$ carries two points of $\mathbb{D}$ into two other specified points if and only if they have the same Poincaré distance. **(2)** Show that the action of $\mathcal{G}(\mathbb{D})$ on the boundary circle $\partial\mathbb{D}$ carries three specified points into three other specified points if and only if they have the same cyclic order.

Problem 2-e. Classifying automorphisms of $\mathbb{D}$. Show that an automorphism of $\mathbb{H}$ or $\mathbb{D}$ is hyperbolic (Problem 1-d) if and only if it carries some Poincaré geodesic into itself without fixed points.

Problem 2-f. Infinite strip, cylinder, and annulus. Define the *infinite strip* $B \subset \mathbb{C}$ of height π to be the set of all $z = x + iy$ with $|y| < \pi/2$. **(1)** Show that the exponential map carries B isomorphically onto the right half-plane. **(2)** Show that the Poincaré metric on B takes the form

$$ds = |dz|/\cos y. \tag{2:7}$$

(3) 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. **(4)** For any $c > 0$, form the quotient cylinder $S_c = B/(c\mathbb{Z})$ by identifying each $z \in B$ with $z + c$. By definition, the *modulus* $\mathrm{mod}(S_c)$ of the resulting cylinder is the ratio π/c of height to circumference. Show that this cylinder, with its Poincaré metric, has a unique simple closed geodesic, with

$$\text{length} \;=\; c \;=\; \pi/\mathrm{mod}(S_c).$$

(5) Show that S_c is conformally isomorphic to the annulus

$$\mathbb{A}_r \;=\; \{\, z \in \mathbb{C}\,;\; 1 < |z| < r \,\}$$

where $\log r = 2\pi^2/c$, and conclude that

$$\mathrm{mod}(\mathbb{A}_r) \;=\; \frac{\log r}{2\pi}$$

is a conformal invariant.

Problem 2-g. Abelian fundamental groups. **(1)** Show that every hyperbolic surface with abelian fundamental group is conformally isomorphic to either the disk $\mathbb{D}$, to the punctured disk $\mathbb{D} \setminus \{0\}$, or to the annulus $\mathbb{A}_r$ for some uniquely defined $r > 1$. (Compare Theorem 1.12 and Problems 1-e, 2-f.) **(2)** Show that this annulus has a unique simple closed geodesic, which has length $\ell = 2\pi^2/\log r$. On the other hand, show that the punctured disk $\mathbb{D} \setminus \{0\}$ has no closed geodesic. (Either the punctured disk or the punctured plane $\mathbb{C} \setminus \{0\}$ might reasonably be considered as the limiting case of an annulus, as the modulus tends to infinity.) **(3)** Show that the conformal automorphism group $\mathcal{G}(\mathbb{D} \setminus \{0\})$ of a punctured disk is isomorphic to the circle group $\mathbf{SO}(2)$, while the conformal automorphism group of an annulus is isomorphic to the nonabelian group $\mathbf{O}(2)$. What is the automorphism group for $\mathbb{C} \setminus \{0\}$? **(4)** Using Lemma 1.10, Theorem 1.12, and Problem 2-b, show that a Riemann surface admits a one-parameter group of conformal automorphisms if and only if its fundamental group is abelian.

Problem 2-h. Gaussian curvature. The Gaussian curvature of a conformal metric $ds = \gamma(w)|dw|$ with $w = u + iv$ is given by the formula

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

where the subscripts stand for partial derivatives. (Compare Willmore [1959, p. 79].) Check that the Poincaré metrics (2:2), (2:3) and (2:7) have curvature $K \equiv -1$ and that the spherical metric (2:4) has curvature $K \equiv +1$.

Problem 2-i. Metrics of constant curvature. A theorem of Heinz Hopf 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 [1959, p. 162].) **(1)** Using this result, show that any nonspherical Riemann surface has one and, up to a multiplicative constant, only one conformal Riemann metric which is complete, with constant Gaussian curvature. **(2)** On the other hand, show that the Riemann sphere $\hat{\mathbb{C}}$ has a 3-dimensional family of distinct conformal metrics with curvature $+1$.

Problem 2-j. Fixed points and contracting maps. **(1)** If S is hyperbolic, show that a holomorphic map $f : S \to S$ can have at most one fixed point, unless some iterate $f^{\circ k}$ is the identity map. (The case of a covering map from S to itself requires special care.) On the other hand, show that any nonhyperbolic surface has a nonidentity holomorphic map with more than one fixed point.

We will say that a map $f : X \to X$ from a metric space to itself is *strictly contracting* if there is a constant $0 < c < 1$ so that

$$\mathrm{dist}(f(x), f(y)) \;\le\; c\, \mathrm{dist}(x, y) \qquad (2:8)$$

for every $x, y \in X$. **(2)** If X is a *complete* metric space, show that all orbits under a strictly contracting map must converge towards a unique fixed point. In particular, this statement applies to a self-map of a hyperbolic surface, whenever (2:8) is satisfied. (However, the example $z \mapsto z^2$ on the unit disk shows that a map with a unique fixed point need not satisfy (2:8), and the example $w \mapsto w + i$ on the upper half-plane shows that a map which simply reduces Poincaré distance need not have any fixed point.)

Problem 2-k. No nontrivial holomorphic attractors. In real dynamics, one often encounters extremely complicated *attractors*, that is, compact sets K with $f(K) = K$ such that for any orbit $x_0 \mapsto x_1 \mapsto x_2 \mapsto \cdots$ in some neighborhood of K the distance $\mathrm{dist}(x_n, K)$ converges uniformly to zero. **(1)** Show that no such behavior can occur for a holomorphic $f : S \to S$. If $K \subset S$ is compact with $f(K) = K$, and if f maps some connected hyperbolic neighborhood U of K into a proper subset of itself, show that f must be strictly contracting on K with respect to the metric dist_U, and hence that K must consist of a single point.

Problem 2-ℓ. Functions on the punctured disk. Every holomorphic function $f : \mathbb{D} \smallsetminus \{0\} \to \mathbb{C}$ can be expressed as the sum of a Laurent series

$$f(z) \;=\; \sum_{n\in\mathbb{Z}} a_n z^n, \qquad \text{where} \qquad a_n \;=\; \frac{1}{2\pi i}\oint_{|z|=r} \frac{f(z)\,dz}{z^{n+1}}.$$

Here the sum of $a_n z^n$ for $n > 0$ defines a holomorphic function throughout the unit disk $\mathbb{D}$, while the sum for $n < 0$ defines a function which is holomorphic throughout the punctured plane $\mathbb{C} \smallsetminus \{0\}$. **(1)** Show that the maximum of $|f(z)|$ on the circle of radius r satisfies

$$\max_{|z|=r} |f(z)| \;\ge\; |a_n|\, r^n$$

for every $n \in \mathbb{Z}$ and for every r with $0 < r < 1$. **(2)** If f takes values in $\mathbb{D}$, conclude that f extends to a holomorphic function from $\mathbb{D}$ to $\mathbb{D}$.

Problem 2-m. The Picard Theorem near infinity. Prove the following statement in two steps, as indicated.

> *Any holomorphic map $f : \mathbb{D} \smallsetminus \{0\} \to \widehat{\mathbb{C}} \smallsetminus \{a, b, c\}$ to the triply punctured sphere extends to a holomorphic map from $\mathbb{D}$ to $\widehat{\mathbb{C}}$.*

(1) In the special case where $f(z)$ converges to a as $z \to 0$, use Problem 2-ℓ to prove this statement. **(2)** On the other hand, suppose that $f(z)$ does

not converge to a, b, or c as $z \to 0$. Show then that there must exist some point $p \in \widehat{\mathbb{C}} \setminus \{a, b, c\}$ which is an accumulation point of images $f(z)$ as $z \to 0$. Using the Poincaré metric as described in Example 2.8, show that the image of some small circle $|z| = r$ lies in a small neighborhood of p. Conclude that f restricted to this circle lifts to the universal covering space of $\widehat{\mathbb{C}} \setminus \{a, b, c\}$, and hence that f on the entire punctured disk lifts to this covering space. Again, complete the proof using Problem 2-ℓ.
(3) Now apply this result for a disk centered at infinity to prove the following.

> **Strong Picard Theorem.** *If $f : \mathbb{C} \to \mathbb{C}$ is holomorphic but not a polynomial, then for every neighborhood $\mathbb{C} \setminus \overline{\mathbb{D}}_r$ of infinity the image $f(\mathbb{C} \setminus \overline{\mathbb{D}}_r)$ omits at most a single point of $\mathbb{C}$. In fact, f takes on every value in $\mathbb{C}$, with at most one exception, infinitely often.*

§3. Normal Families: Montel's Theorem

Let S and T be Riemann surfaces. This section will study compactness in the function space $\mathrm{Hol}(S,T)$ consisting of all holomorphic maps with source S and target T. We first define a topology on this space, and on the larger space $\mathrm{Map}(S,T)$ consisting of all continuous maps from S to T. This topology is known to complex analysts as the *topology of uniform convergence on compact subsets*, or more succinctly as the *topology of locally uniform convergence*. It is known to topologists as the *compact-open topology* (Problem 3-a), or when dealing with smooth manifolds as the C^0-*topology.*

Definition. Let X be a locally compact space and let Y be a metric space. For any f in the space $\mathrm{Map}(X,Y)$ of continuous maps from X to Y we define a family $N_{K,\epsilon}(f)$ of *basic neighborhoods* of f as follows. For any compact subset $K \subset X$ and any $\epsilon > 0$, let $N_{K,\epsilon}(f)$ be the set of all $g \in \mathrm{Map}(X,Y)$ satisfying the condition that

$$\mathrm{dist}(f(x), g(x)) < \epsilon \qquad \text{for all} \qquad x \in K.$$

A subset $\mathcal{U} \subset \mathrm{Map}(X,Y)$ is defined to be *open* if and only if, for every $f \in \mathcal{U}$, there exist K and ϵ as above so that the basic neighborhood $N_{K,\epsilon}(f)$ is contained in $\mathcal{U}$.

Lemma 3.1 (The Topology of Locally Uniform Convergence). *With these definitions,* $\mathrm{Map}(X,Y)$ *is a well defined Hausdorff space. A sequence of maps* $f_i \in \mathrm{Map}(X,Y)$ *converges to the limit* g *in this topology if and only if*

(*a*) *for every compact* $K \subset X$, *the sequence of maps* $f_i|_K : K \to Y$ *converges uniformly to* $g|_K$,

or equivalently if and only if

(*b*) *every point of* X *has a neighborhood* N *so that the sequence* $\{f_i|_N\}$ *converges uniformly to* $g|_N$.

This topology on $\mathrm{Map}(X,Y)$ *depends only on the topologies of* X *and* Y *and not on the particular choice of metric for* Y. *Furthermore, if* X *is* σ*-compact, then* $\mathrm{Map}(X,Y)$ *is itself a metrizable topological space.*

Proof. The first two statements follow immediately from the definitions. To prove that the topology is independent of the metric on Y, we describe a slightly different form of the definition, which depends only on the topology of Y. Let U be any neighborhood of the diagonal in the product space $Y \times Y$. For any compact $K \subset X$ and any $f \in \mathrm{Map}(X, Y)$, let

$$N_{K,U}(f) = \{ g \in \mathrm{Map}(X, Y) ; (f(x), g(x)) \in U \quad \text{for all} \quad x \in K \}.$$

Given K and U, it is not difficult to construct an $\epsilon > 0$ so that every pair $(f(x), y)$ with $x \in K$ and $\mathrm{dist}(f(x), y) < \epsilon$ belongs to this set U, and it then follows that $N_{K,\epsilon}(f) \subset N_{K,U}(f)$. On the other hand, if $U(\epsilon)$ is the set of all pairs (y, y') with $\mathrm{dist}(y, y') < \epsilon$, then $N_{K,\epsilon}(f) = N_{K,U(\epsilon)}(f)$. Thus, if we take $\{N_{K,U}(f)\}$ as our collection of "basic neighborhoods," then we obtain the same topology, without mentioning any particular choice of metric.

Now suppose that X is σ-compact; that is, suppose that X is a countable union of compact subsets. Since X is also assumed to be locally compact, we can choose compact sets $K_1 \subset K_2 \subset \ldots$ with union X, so that each K_n is contained in the interior of K_{n+1}. It will be convenient to replace the given metric $\mathrm{dist}(y, y')$ by the bounded metric

$$\mu(y, y') = \mathrm{Min}(\mathrm{dist}(y, y'), 1) \leq 1,$$

which evidently gives rise to the same topology. Define the "locally uniform distance" between two maps from X to Y by the formula

$$\mu'(f, g) = \sum_n \frac{1}{2^n} \mathrm{Max}\{ \mu(f(x), g(x)) ; x \in K_n\}.$$

We must show that this metric gives rise to the required topology on $\mathrm{Map}(X, Y)$. Let $N'_\epsilon(f)$ be the ϵ-neighborhood of f in this metric μ'. Given ϵ, we can choose n so that $1/2^n < \epsilon/2$, and set $K = K_n$. It is then easy to check that $N_{K,\epsilon/2}(f) \subset N'_\epsilon(f)$. Conversely, given K and ϵ we can choose n so that $K \subset K_n$ and check that $N'_{\epsilon/2^n}(f) \subset N_{K,\epsilon}(f)$. Thus the two topologies are indeed the same. $\square$

Now we can specialize to maps between two Riemann surfaces S and T. Since every Riemann surface is σ-compact by Corollary 2.2 and metrizable, for example, by Corollary 2.10, we obtain a well-defined metrizable topological space $\mathrm{Map}(S, T)$. It follows easily from the Weierstrass Theorem 1.4 that the space $\mathrm{Hol}(S, T)$ of holomorphic maps is a closed subset of $\mathrm{Map}(S, T)$.

Theorem 3.2 (Hyperbolic Compactness). *If S and T are hyperbolic Riemann surfaces, then the space $\mathrm{Hol}(S, T)$ of holomorphic maps is locally compact and σ-compact. Furthermore, if $K \subset S$ and $K' \subset T$ are nonvacuous compact subsets,* *then the set of all holomorphic maps* $f : S \to T$ *satisfying* $f(K) \subset K'$ *forms a compact subset of* $\mathrm{Hol}(S,T)$.

In particular, if T itself is compact, it follows that the entire space $\mathrm{Hol}(S,T)$ is compact. More generally, if $k_0 \in S$ is any base point, it follows that the evaluation map $f \mapsto f(k_0)$ is a *proper map*

$$\mathrm{Hol}(S,T) \to T.$$

That is, the preimage of any compact set $K' \subset T$ is a compact subset of $\mathrm{Hol}(S,T)$.

Note that these statements are clearly false in the nonhyperbolic case. For example, if $S = T$ is either the Riemann sphere $\widehat{\mathbb{C}}$ or the complex numbers $\mathbb{C}$ or a quotient $\mathbb{C}/\mathbb{Z}$ or $\mathbb{C}/\Lambda$, then the sequence of maps $z \mapsto nz$ in $\mathrm{Hol}(S,S)$ takes the compact set $K = K' = \{0\}$ into itself, and yet has no convergent subsequence since the first derivatives at zero do not converge. The space $\mathrm{Hol}(\widehat{\mathbb{C}},\widehat{\mathbb{C}})$ of all rational maps is actually locally compact but not compact, while the space $\mathrm{Hol}(\mathbb{C},\mathbb{C})$ is not even locally compact (Problem 3-c).

Proof of Theorem 3.2. The proof will be based on the *Bolzano-Weierstrass Theorem*, which asserts that a metric space is compact if and only if every infinite sequence of points in the space possesses a convergent subsequence (Problem 3-d). Thus it will suffice to show that every sequence of holomorphic maps $f_n : S \to T$ with $f_n(K) \subset K'$ contains a convergent subsequence.

It follows easily from Corollary 2.2 that the Riemann surface S possesses a countable dense subset $\{s_j\}$, where we may assume that $s_1 \in K$. Since K' is compact, the sequence of image points $f_n(s_1) \in K'$ certainly contains a convergent subsequence. Thus we can first choose an infinite set $Q_1 = \{n_p\}$ of positive integers so that the images $f_{n_p}(s_1)$ converge to some point $t_1 \in K' \subset T$ as $p \to \infty$. It then follows from Lemma 2.9 and Theorem 2.11 that the images $f_{n_p}(s_2)$ of the point s_2 lie within some compact subset of T. Hence we can choose an infinite subset $Q_2 \subset Q_1$ so that the points $f_n(s_2)$ with $n \in Q_2$ converge to some limit $t_2 \in T$. Continuing inductively, we can find infinite sets $Q_1 \supset Q_2 \supset Q_3 \supset \cdots$ so that the $f_n(s_k)$ with $n \in Q_k$ converge to a limit $t_k \in T$. Using a diagonal procedure, let $\hat{Q} = \{q_j\}$ consist of the first element in Q_1, the second element of Q_2, and so on. Then $\lim_{j\to\infty} f_{q_j}(s_k) = t_k$ for every s_k. We claim that this new sequence of maps $g_j = f_{q_j}$ converges, uniformly on every compact subset of S, to a holomorphic map $g : S \to T$.

Given any compact set $L \subset S$ and any $\epsilon > 0$, we can cover L by finitely many open balls of radius ϵ, centered at points s_j. In other words, we can choose finitely many points from $\{s_j\}$ so that every $z \in L$ has Poincaré distance $\mathrm{dist}_S(z, s_j) < \epsilon$ from one of these s_j. Further, we can choose n_0 so that $\mathrm{dist}_T(g_m(s_j),\, g_n(s_j)) < \epsilon$ for each of these finitely many s_j, whenever $m, n > n_0$. For any $z \in L$ it then follows using Theorem 2.11 and the triangle inequality that

$$\mathrm{dist}_T(g_m(z),\, g_n(z)) \;<\; 3\epsilon$$

whenever $m, n > n_0$. Thus the $g_m(z)$ form a Cauchy sequence. It follows that the sequence of functions $\{g_m|_L\}$ converges uniformly to a limit. Since L is an arbitrary compact set in S, this proves that $\{g_m\}$ converges locally uniformly to a limit function, which must belong to $\mathrm{Hol}(S,T)$. Therefore, by Bolzano-Weierstrass, the set of all $f \in \mathrm{Hol}(S,T)$ with $f(K) \subset K'$ is compact.

In particular, it follows that the evaluation map $f \mapsto f(k_0)$ from $\mathrm{Hol}(S,T)$ to T is proper. Since T is locally compact and σ-compact, this implies that $\mathrm{Hol}(S,T)$ is also locally compact and σ-compact. □

Normal Families. Here is a preliminary definition: A collection $\mathcal{F}$ of holomorphic functions from a Riemann surface S to a *compact* Riemann surface T is called a *normal* family if its closure $\bar{\mathcal{F}} \subset \mathrm{Hol}(S,T)$ is a compact set, or equivalently if every infinite sequence of functions $f_n \in \mathcal{F}$ contains a subsequence which converges locally uniformly to some limit function $g : S \to T$.

We will also need to consider the case of a noncompact target surface T. For this purpose, we need the following definition: A sequence of points $\{t_n\}$ in the noncompact surface T *diverges* from T if for every compact set $K \subset T$ we have $t_n \neq K$ for n sufficiently large. (Here the qualification "from T" is essential. For example, the sequence of points i, $i/2$, $i/3$, ..., where $i = \sqrt{-1}$, diverges from the upper half-plane $\mathbb{H}$, but converges to 0 within its closure $\overline{\mathbb{H}} \subset \mathbb{C}$.) Similarly, we will say that a sequence of maps $f_n : S \to T$ *diverges locally uniformly* from T if, for every compact $K \subset S$ and $K' \subset T$, we have $f_n(K) \cap K' = \emptyset$ for n sufficiently large. (Of course this can never happen if T itself is compact.)

Definition. A collection $\mathcal{F}$ of maps from a Riemann surface S to a (possibly noncompact) Riemann surface T will be called *normal* if every infinite sequence of maps from $\mathcal{F}$ contains either a subsequence which converges locally uniformly or a subsequence which diverges locally uniformly from T. We can now restate Theorem 3.2 as follows.

Corollary 3.3. *If S and T are hyperbolic, then every family $\mathcal{F}$ of holomorphic maps from S to T is normal.*

Proof. Choose base points $s_0 \in S$ and $t_0 \in T$. If the set of images $\{f(s_0)\ ;\ f \in \mathcal{F}\}$ lies in some compact subset $K' \subset T$, then it follows immediately from Theorem 3.2 that $\overline{\mathcal{F}}$ is compact. Otherwise, we can choose an infinite sequence of maps $f_n \in \mathcal{F}$ so that the Poincaré distance $\mathrm{dist}_T(t_0\,,\, f_n(s_0))$ tends to infinity. Using Pick's Theorem 2.11, it then follows easily that this sequence of maps f_n diverges locally uniformly from T. □

As an application of this result, we can compare the Poincaré metrics in a pair of Riemann surfaces $S \subset S'$. (Compare (2 : 6).) We will use the notation $N_r(p) \subset S$ for the open neighborhood of radius r consisting of all $q \in S$ with $\mathrm{dist}_S(p, q) < r$, using the S Poincaré metric.

Theorem 3.4 (Poincaré Metric near the Boundary). *Suppose that $S \subset S'$ are Riemann surfaces with S hyperbolic, and let $p_1\,, p_2\,, \ldots$ be a sequence of points in S which converges (in the topology of S') to a boundary point $\hat{p} \in \partial S \subset S'$. Then for any fixed r the entire neighborhood $N_r(p_j)$ converges uniformly to $\hat{p}$ as $j \to \infty$. If S has compact closure in S', then choosing some metric on S' compatible with its topology, we can make the following sharper statement. The diameter in this S'-metric of the neighborhood $N_r(p_j)$ converges uniformly to zero as p_j converges towards ∂S.*

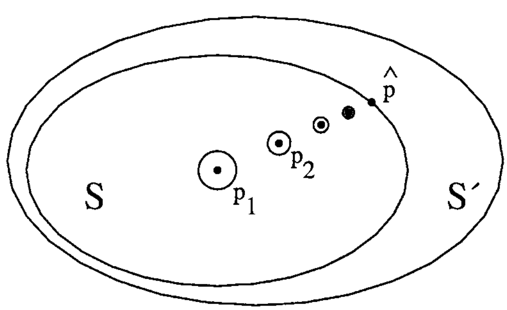


Figure 4. Poincaré neighborhoods with respect to a subset $S \subset S'$.

(See Corollary A.8 in Appendix A for a more quantitative estimate when S is a simply connected open subset of $\mathbb{C}$.)

Proof of Theorem 3.4. First note the following preliminary statement: *If K is any compact subset of S and if $\{p_j\}$ converges to ∂S, then $N_r(p_j) \cap K = \emptyset$ for j sufficiently large.* To see this, let k_0 be a basepoint in K and let r_K be the diameter of K. Then $N_{r+r_K}(k_0)$ is compact by Lemma 2.9. For j sufficiently large, p_j will be outside of this compact set, and hence $N_r(p_j)$ will be disjoint from K.

Let $N_r^0 \subset \mathbb{D}$ be the disk of Poincaré radius r about the central point of the unit disk. If $\widetilde{S}$ is the universal covering of S, then composing a suitable isomorphism $\mathbb{D} \cong \widetilde{S}$ with the projection $\widetilde{S} \to S$, we can construct a covering map $f_j : \mathbb{D} \to S$ with $f_j(0) = p_j$. Evidently $N_r(p_j)$ can be identified with the image $f_j(N_r^0)$ of this standard disk.

For any sufficiently large compact set $K \subset S$, note that each component of $S' \setminus K$ will be a hyperbolic Riemann surface. The maps $f_j|_{N_r^0}$, for j sufficiently large, take values in $S' \setminus K$ and hence form a normal family. If the p_j all lie in some compact subset of S' (for example, if $\{p_j\}$ converges to some point of S'), then we can choose a subsequence so that $f_j|_{N_r^0}$ converges locally uniformly to a holomorphic limit map $f : N_r^0 \to S' \setminus K$. (In fact, since we can apply the same argument to a disk of radius $r+1$, it follows that this subsequence converges uniformly on the closed disk $\overline{N}_r^0$.)

We claim that f must map the entire disk N_r^0 to a single point of $\partial S \subset S'$. For if this limit map were not constant, then its image $f(N_r^0)$ would be an open subset of $S' \setminus K$. Hence this image would have to intersect S. But this is impossible since S can be exhausted by a sequence of compact subsets $K_1 \subset K_2 \subset \cdots$, and the argument above shows that $f(N_r^0)$ must be disjoint from each K_n.

The above discussion applied only to some subsequence of the $N_r(p_j)$. Now we must deal with the entire sequence. Choosing some metric $\mathrm{dist}'(p, q)$ on the space S', let d'_j be the diameter of the set $N_r(p_j) \subset S \subset S'$ with respect to this dist' metric. We must show that the sequence d'_j converges uniformly to zero. Otherwise we could choose a subsequence with $d'_j \geq \epsilon > 0$ and then choose a subsequence of this so that $f_j|_{\overline{N}_r^0}$ converges uniformly to a constant map. Evidently this is impossible. □

Lemma 3.5. *Given Riemann surfaces S and $U \subset T$, let $f_j : S \to U$ be a sequence of holomorphic maps which diverges locally uniformly from U but not from T. Then there exists a subsequence which converges locally uniformly to a constant map from S to a single point of $\partial U \subset T$.*

Proof. Since $\{f_j\}$ does not diverge locally uniformly from T, we can choose compact sets $K \subset S$ and $L \subset T$ so that $f_j(K) \cap L \neq \emptyset$ for infinitely many j. After passing to a subsequence, we may choose points $k_{j_i} \in K$ so that the images $f_{j_i}(k_{j_i})$ converge to a limit $\ell \in L$.

First suppose that S and U are hyperbolic. Since K has finite diameter in the S-Poincaré metric, it follows from Theorem 3.4 and Pick's Theorem that the entire image $f_{j_i}(K)$ converges to ℓ. Again using Theorem 3.4, it follows easily that this sequence of maps $f_{j_i} : S \to U \subset T$ converges locally uniformly to the constant map $K \mapsto \ell \in \partial U \subset T$, as required.

If S and U are not hyperbolic, then we can choose a hyperbolic neighborhood S_0 of $K \subset S$ with compact closure, and a hyperbolic set U_0 of the form $U_0 = U \smallsetminus \text{compact}$. The above argument shows that a subsequence of the f_j restricted to S_0 converges locally uniformly to a constant map. Since S_0 can be arbitrarily large, this completes the proof. □

Corollary 3.6. *Given Riemann surfaces S and $U \subset T$, a family of maps from S to U is normal if and only if it is normal considered as a family of maps from S into the larger surface T.*

The proof is immediate. Combining Corollaries 3.6 and 3.3 with the fact that the thrice punctured sphere is hyperbolic (see Lemma 2.5), we obtain the following important consequence. (See, for example, Montel [1927].)

Theorem 3.7 (Montel). *Let S be a Riemann surface and let $\mathcal{F}$ be a collection of holomorphic maps $f : S \to \widehat{\mathbb{C}}$ which omit three different values. That is, assume that there are distinct points $a, b, c \in \widehat{\mathbb{C}}$ so that $f(S) \subset \widehat{\mathbb{C}} \smallsetminus \{a, b, c\}$ for every $f \in \mathcal{F}$. Then $\mathcal{F}$ is a normal family; that is, the closure $\overline{\mathcal{F}} \subset \mathrm{Hol}(S, \widehat{\mathbb{C}})$ is a compact set.*

Concluding Problems

Problem 3-a. The compact-open topology. If X and Y are locally compact spaces, the *compact-open topology* on the space $\mathrm{Map}(X, Y)$ of all maps is defined to be the *smallest* topology (that is, the topology with fewest open sets), such that, for every compact $K \subset X$ and every open $U \subset Y$, the set of $f : X \to Y$ with $f(K) \subset U$ forms an open subset of $\mathrm{Map}(X, Y)$. **(1)** If Y is metrizable, show that this coincides with the topology of locally uniform convergence, as described above. **(2)** Show that the composition operation

$$(f, g) \mapsto g \circ f$$

is continuous as a mapping from $\mathrm{Map}(X, Y) \times \mathrm{Map}(Y, Z)$ to $\mathrm{Map}(X, Z)$.

(3) If U is an open subset of Y, show that $\mathrm{Map}(X,U)$ embeds homeomorphically as a subset of $\mathrm{Map}(X,Y)$, but that this subset need not be either open or closed.

(4) Now suppose that S and T are Riemann surfaces and that U is a connected open subset of T. Show that the topological boundary of $\mathrm{Hol}(S,U)$ in $\mathrm{Hol}(S,T)$ consists completely of constant maps from S to ∂U.

Problem 3-b. Uniform convergence or divergence? Consider the family of maps $f_n(z) = z+n$ from $\mathbb{C}$ or $\widehat{\mathbb{C}}$ to itself. **(1)** Show that this sequence diverges locally uniformly from $\mathbb{C}$. However, in $\widehat{\mathbb{C}}$ show that this sequence neither converges nor diverges locally uniformly, although it does converge pointwise to the constant function which maps all of $\widehat{\mathbb{C}}$ to the single point $\infty \in \widehat{\mathbb{C}}$. **(2)** Similarly, show that the sequence of degree 1 rational functions $g_n(z) = 1/(n^2z-n)$ converges pointwise, but not locally uniformly, to the constant function $g(z)=0$.

Problem 3-c. Local compactness? **(1)** Show that $\mathrm{Hol}(\mathbb{C},\mathbb{C})$ is not locally compact, since every neighborhood of the zero map contains a sequence of polynomial maps of the form

$$f_n(z) \;=\; \epsilon(1+\epsilon z+\epsilon^2z^2+\cdots+\epsilon^nz^n)$$

with no limit point in $\mathrm{Hol}(\mathbb{C},\mathbb{C})$. Similarly show that $\mathrm{Hol}(\mathbb{C},\widehat{\mathbb{C}})$ is not locally compact. **(2)** However, if S and T are compact, show that $\mathrm{Hol}(S,T)$ is always locally compact.*

Problem 3-d. The Bolzano-Weierstrass Theorem. Let X be a metric space with the property that every infinite sequence in X possesses an accumulation point (or equivalently a convergent subsequence). **(1)** For each $n>0$, show that X can be covered by finitely many open balls $B_{n,j}$ of radius $1/n$. **(2)** Given any collection of open subsets U_α with union X, show that the collection of all of these balls $B_{n,j}$ which are contained in some U_α forms a covering of X by *countably many* open sets. **(3)** Given a sequence $V_1, V_2, \ldots$ of open sets with union X, show

---

*The space $\mathrm{Hol}(S,\widehat{\mathbb{C}})$ of meromorphic functions on S is of particular interest. As an example, for each $d\ge 1$ the space $\mathrm{Rat}_d = \mathrm{Hol}_d(\widehat{\mathbb{C}},\widehat{\mathbb{C}})$ of degree d rational maps forms a complex $(2d+1)$-dimensional manifold which is noncompact. In fact it can be identified with an open subset of $(2d+1)$-dimensional projective space. (Compare Segal [1979] or Milnor [1993].) On the other hand, there exists a surface S of genus 5 so that the space $\mathrm{Hol}(S,\widehat{\mathbb{C}})$ has singularities. (Private communication from J. Harris.) If the target space T is a torus, still assuming that S is compact, then each connected component of $\mathrm{Hol}(S,T)$ is a copy of T, while if T has genus ≥ 2 then $\mathrm{Hol}(S,T)$ is a finite set.

that $V_1 \cup \cdots \cup V_k = X$ provided that k is sufficiently large. Combining these statements, show that finitely many of the U_α suffice to cover X, thus proving that X is compact.

Problem 3-e. Locally normal families. Show that normality is a local property. More precisely, let S and T be any Riemann surfaces, and let $\{f_\alpha\}$ be a family of holomorphic maps from S to T. If every point of S has a neighborhood U such that the collection $\{f_\alpha|_U\}$ of restricted maps is a normal family in $\mathrm{Hol}(U,T)$, show by a diagonal argument, similar to the proof of Theorem 3.2, that the family $\{f_\alpha\}$ itself is normal.

Problem 3-f. Normality and derivatives. Let $f : S \to T$ be holomorphic. Given Riemannian metrics on the Riemann surfaces S and T, we can define the *norm* of the derivative of f at a point $s \in S$ to be the real number $\|f'(s)\| \geq 0$ such that the induced linear mapping from the tangent space of S at s to the tangent space of T at $f(s)$ carries vectors of length 1 to vectors of length $\|f'(s)\|$. If T is compact, show that a family $\mathcal{F}$ of maps $f : S \to T$ is normal if and only if the collection of norms $\|f'(s)\|$ is uniformly bounded as f varies over $\mathcal{F}$ and s varies over any compact subset of S.

Problem 3-g. The one-point compactification. For any locally compact space X, let $X \cup \infty$ be the topological space which is obtained by adjoining a single point ∞ to X, defining the basic neighborhoods of ∞ to be the complements of compact subsets of X. **(1)** Show that $X \cup \infty$ is a compact Hausdorff space, which is metrizable if X is metrizable and σ-compact. **(2)** Now let S and T be noncompact Riemann surfaces. Show that the closure of $\mathrm{Hol}(S,T)$ in the larger space $\mathrm{Map}(S, T \cup \infty)$ consists of $\mathrm{Hol}(S,T)$ together with the constant map $[\infty]$ which carries all of S to the point ∞. **(3)** If S and T are hyperbolic, show that this closure

$$\mathrm{Hol}(S,T) \cup [\infty] \quad \subset \quad \mathrm{Map}(S, T \cup \infty)$$

is compact and can be identified with the one point compactification of $\mathrm{Hol}(S,T)$.

ITERATED HOLOMORPHIC MAPS

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

4.1. A Brief History.* The study of iterated holomorphic mappings began in the 19th century but only came to flower in the 20th. One primary focus in the 19th century was the study of functional equations relating different known or unknown functions. Charles Babbage [1815], "*An essay on the calculus of functions,*" set the stage with an attempt to understand many such functional equations. In particular (on page 412), he implicitly described what we now call a *semiconjugacy* φ between two different functions f and g, that is a map satisfying $\varphi \circ f = g \circ \varphi$ so that the following diagram is commutative:

$$\begin{array}{ccc} X & \xrightarrow{f} & X \\ \varphi \downarrow & & \varphi \downarrow \\ Y & \xrightarrow{g} & Y. \end{array}$$

It follows that any orbit $x_0 \mapsto x_1 \mapsto \cdots$ under f maps to an orbit $\varphi(x_0) \mapsto \varphi(x_1) \mapsto \cdots$ under g. In the special case where $g(y) = y + 1$, such a semiconjugacy reduces to the *Abel functional equation*

$$\varphi(f(x)) = \varphi(x) + 1,$$

described by Niels Abel in an 1824 paper which was published only posthumously. In the special case $g(y) = \lambda y$, it reduces to the *Schröder functional equation*

$$\varphi(f(x)) = \lambda\,\varphi(x),$$

introduced in Ernst Schröder's 1871 paper "*Ueber iterirte Functionen*". Although Schröder didn't think in dynamic terms, his paper was far ahead of its time, and contained a wealth of ideas. In particular, he used elliptic function theory to construct a one-parameter family of what we now call *Lattès maps* of the Riemann sphere, with explicitly understandable chaotic dynamics. (Compare §7.) Another contribution, in 1879, was Cayley's analysis of Newton's method for degree two polynomial equations. (See Problem 7-a.) In 1884, Gabriel Kœnigs proved that this Schröder equation had an essentially unique solution in the neighborhood of a fixed point $z_0 = f(z_0)$, provided that the derivative (or "*multiplier*") $\lambda = f'(z_0)$ satisfies $|\lambda| \neq 0, 1$. Thus the dynamics in a neighborhood of such a fixed point

*Compare Alexander [1994], as well as Ghys [1999] and Milnor [2004b].

can be completely understood in terms of the dynamics of the linear map $w \mapsto \lambda w$. Leau in 1897 studied the more complicated case where the multiplier λ is a root of unit, and Böttcher in 1904 treated the case $\lambda = 0$. (See §§8–10 below.) Böttcher was perhaps the first to try to understand the global dynamics of iterated holomorphic functions and to try to distinguish between predictable orderly behavior and chaotic behavior.

The case where $|\lambda| = 1$ but λ is not a root of unity is much more difficult and was not understood until the work of Hubert Cremer in 1927 and Carl Ludwig Siegel in 1942. (Compare §11.)

Meanwhile, the global study of iterated holomorphic maps flowered dramatically. In 1906, Pierre Fatou described a 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 (Problems 4-e, 4-f). This aroused great interest. After a hiatus during the First World War, the subject was taken up in depth by Fatou and by Gaston Julia (as well as others such as Samuel Lattès [1918] and Joseph Fels Ritt [1920]). The most fundamental and incisive contributions were those of Fatou himself. However, Julia was a determined competitor and tended to get more credit because of his status as a wounded war hero—in 1918, he was awarded the "Grand Prix des Sciences Mathématiques" by the Paris Academy of Sciences.

For many years after this spate of activity, the subject seemed to be forgotten. Interest began to revive with a 1965 paper by Hans Brolin, but it was not until the 1980s that the subject really came back to life with the work of Douady, Hubbard, and Sullivan, as well as Thurston (compare Douady and Hubbard [1982]). The possibility of computer experimentation and computer illustration for the complicated geometry which is involved was also a dramatic influence.

Definition 4.2. The Fatou and Julia Sets. Let S be a compact Riemann surface, let $f : S \to S$ be a nonconstant holomorphic mapping, and let $f^{\circ n} : S \to S$ be its n-fold iterate. The domain of normality for the collection of iterates $\{f^{\circ n}\}$ is called the ***Fatou set**** for f, and its complement is called the ***Julia set***. I will use the notation $J = J(f)$ for the Julia set and write the Fatou set simply as $S \smallsetminus J$.

Thus, for any point $p_0 \in S$, we have the following basic dichotomy: *If*

*For the definition in the noncompact case, see §5. The choice as to which of these two sets should be named after Julia and which after Fatou is rather arbitrary, but the term "Julia set" now seems firmly established. (Note, however, that this form of the definition of J is actually due to Fatou. For Julia's definition, see §14.) The Fatou set $S \smallsetminus J$ is sometimes called by other names, such as "stable set" or "normal set."

there exists some neighborhood U of p_0 so that the sequence of iterates $\{f^{\circ n}\}$ restricted to U forms a normal family of maps from U to S, then we say that p_0 belongs to the Fatou set of f. Otherwise, if no such neighborhood exists, we say that p_0 belongs to the Julia set.

By its very definition, the Julia set J is a closed subset of S, while the complementary Fatou set $S \setminus J$ is an open subset. We will see that a point p_0 belongs to the Julia set if and only if dynamics in a neighborhood of p_0 displays ***sensitive dependence on initial conditions***, so that nearby initial conditions lead to wildly different behavior after a large (or sometimes not so large) number of iterations. (Compare Problem 4-h, as well as Corollary 14.2.)

The classical example, and the one which we will emphasize, is the case where S is the Riemann sphere $\widehat{\mathbb{C}} = \mathbb{C} \cup \infty$. Any holomorphic map $f : \widehat{\mathbb{C}} \to \widehat{\mathbb{C}}$ 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. For all but finitely many choices of constant $c \in \widehat{\mathbb{C}}$, this degree can be described as the number of distinct solutions to the equation $f(z) = c$. We will usually assume that $d \geq 2$ and always that $d \geq 1$ so that f is a nonconstant map from $\widehat{\mathbb{C}}$ onto itself.

As a simple example, consider the squaring map $s : z \mapsto z^2$ on $\widehat{\mathbb{C}}$. The entire open disk $\mathbb{D}$ is contained in the Fatou set of s, since successive iterates on any compact subset converge uniformly to zero. Similarly the exterior $\widehat{\mathbb{C}} \setminus \overline{\mathbb{D}}$ is contained in the Fatou set, since the iterates of s converge to the constant function $z \mapsto \infty$ outside of $\overline{\mathbb{D}}$. 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. Therefore the Julia set $J(s)$ is precisely equal to the unit circle.

Such smooth Julia sets are rather exceptional. (See §7.) Figures 5 and 6 illustrate more typical rational maps. In each case, the Julia set is black and the Fatou set is white. Figure 5 illustrates five different quadratic polynomial maps. Figure 5a shows a rather wild Jordan curve, Figure 5b a ***dendrite*** (that is, a compact, connected set without interior which does not separate the plane), Figure 5c a rather thick totally disconnected set (compare Problem 4-e), and the last two show Julia sets whose complement has infinitely many connected components. In both cases, the critical orbit is superattracting of period 3 (compare Definition 4.5). The arrows in Figure 5e give a rough indication of what maps to what. In each of these pictures, since $f(z)$ is an even function, the Julia set is centrally symmetric.

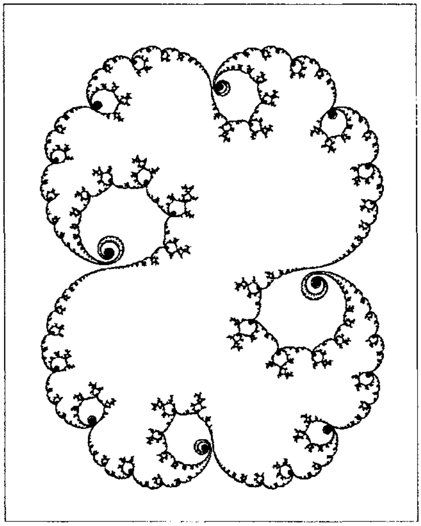


Figure 5a. A simple closed curve,
$z \mapsto z^2 + (.99 + .14i)z.$

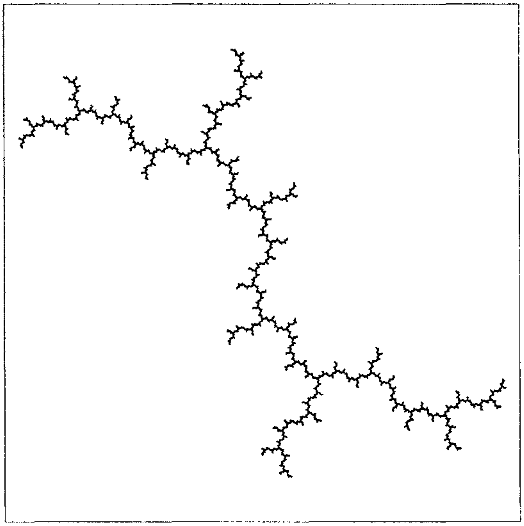


Figure 5b. A "dendrite,"
$z \mapsto z^2 + i.$

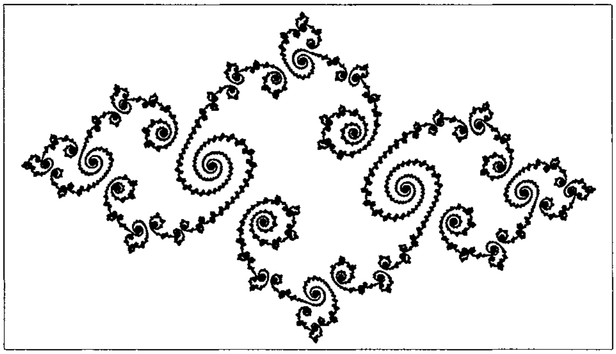


Figure 5c. A Cantor set,
$z \mapsto z^2 + (-.765 + .12i).$

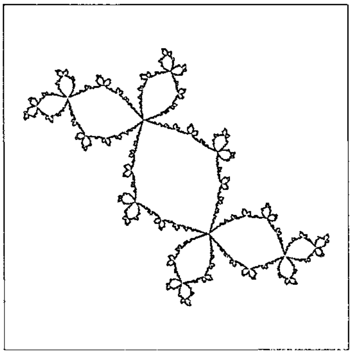


Figure 5d. The Douady rabbit
$z \mapsto z^2 + (-.122 + .745\,i).$

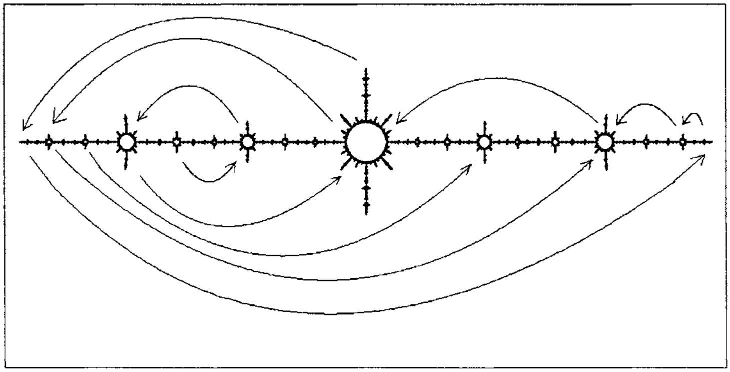
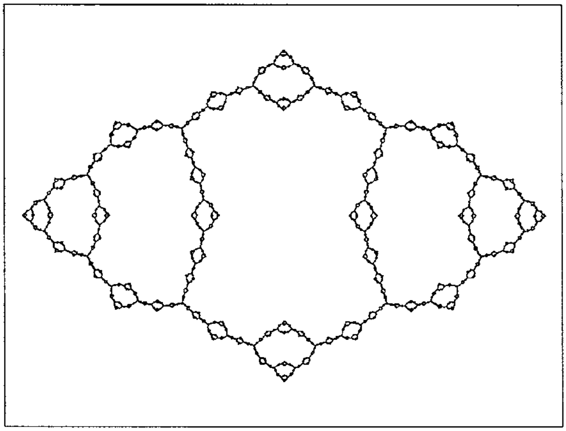
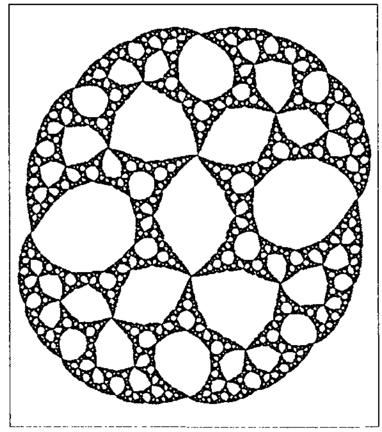


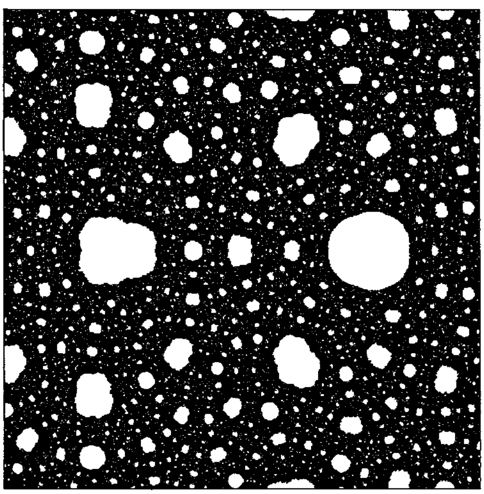
Figure 5e. The "airplane": $z \mapsto z^2 - 1.75488\ldots.$



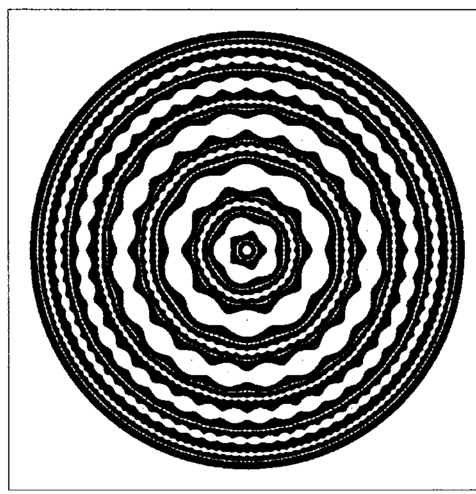
6a. $f(z) = 1 - 1/z^2$



6b. $f(z) = \frac{c+z^2}{1-z^2}$ *with* $c = \frac{1+i\sqrt{3}}{2}$



6c. $f(z) = -.138(z + 1/z) - .303$



6d. $f(z) = (z^5 - .00001)/z^3$

Figure 6. Julia sets for four rational maps.

Nonpolynomial Julia sets can be even more diverse, as illustrated in Figure 6. The first example, on the upper left, has a superattracting cycle $\{1\,,\,0\,,\,\infty\}$ containing both critical points 0 and ∞. The second has two superattracting cycles $\{-1\,,\,\infty\}$ and $\{0\,,\,\omega\,,\,\omega^2\}$, where $\omega = e^{\pi i/3}$. (This example can be constructed by "mating" two quadratic polynomials—see, for example, Milnor [2004a]. Note the distorted copy of the "rabbit" in the middle.) The third example is a *Sierpinsky carpet*. That is, the Fatou set is everywhere dense, and its various components are bounded by disjoint simple closed curves. (Compare Devaney [2004]. These degree 2 examples are all taken from Milnor [1993].) The last example, due to McMullen [1988], has uncountably many connected components which are topological circles.

Here are some basic properties of the Julia set.

Lemma 4.3 (Invariance Lemma). *The Julia set $J = J(f)$ of a holomorphic map $f : S \to S$ is fully invariant under f. That is, z belongs to J if and only if the image $f(z)$ belongs to J.*

A completely equivalent statement is that the Fatou set is fully invariant. In fact, for any open set $U \subset S$, some sequence of iterates $f^{\circ n_j}$ converges uniformly on compact subsets of U if and only if the corresponding sequence of iterates $f^{\circ n_j+1}$ converges uniformly on compact subsets of the open set $f^{-1}(U)$. Further details 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 4-d.)

Lemma 4.4 (Iteration Lemma). *For any $k > 0$, the Julia set $J(f^{\circ k})$ of the k-fold iterate coincides with the Julia set $J(f)$.*

Proof outline. Again we can equally well work with the Fatou set $S \smallsetminus J$. Suppose, for example, that z belongs to the Fatou set of $f \circ f$. This means that, for some neighborhood U of z, the collection of all even iterates $f^{\circ 2n}|_U$ is contained in a compact subset $K \subset \mathrm{Hol}(U, S)$. It follows that every iterate of f, restricted to U, belongs to the compact set $K \cup (f \circ K) \subset \mathrm{Hol}(U, S)$, hence z belongs to the Fatou set of f. Further details will be left to the reader. □

Definition 4.5. Consider a *periodic orbit* or *cycle*

$$f \,:\, z_0 \mapsto z_1 \mapsto \cdots \mapsto z_{m-1} \mapsto z_m = z_0$$

for a holomorphic map $f : S \to S$. (Here S can be any Riemann surface, compact or not.) If the points $z_1, \ldots, z_m$ are all distinct, then the integer $m \geq 1$ is called the *period*. The first derivative of the m-fold iterate $f^{\circ m}$ at a point of the orbit is a well-defined complex number called the *multiplier* of the orbit. If the Riemann surface S is an open subset of $\mathbb{C}$, then we have the product formula

$$\lambda \;=\; (f^{\circ m})'(z_i) \;=\; f'(z_1) \cdot f'(z_2) \cdots f'(z_m).$$

In particular, $\lambda = 0$ if and only if some point z_j of the orbit is a *critical point* of f, that is, a point at which the first derivative f' vanishes. More generally, for self-maps of an arbitrary Riemann surface we have a corresponding product formula, using a local uniformizing parameter (that is, a local coordinate chart) around each point of the orbit. The product λ is independent of the choice of uniformizing parameters. By definition, a periodic orbit is either *attracting* or *repelling* or *indifferent* (also called *neutral*) according as, its multiplier satisfies $|\lambda| < 1$ or $|\lambda| > 1$ or $|\lambda| = 1$. (Compare §8.) The orbit will be called *superattracting* if $\lambda = 0$ and *geometrically attracting* if $0 < |\lambda| < 1$. As examples, the maps illustrated in Figures 5d, 5e, 6a, 6b, and 6c all have superattracting orbits of period three.

Caution: In the special case where the point at infinity is periodic under a rational map, $f^{\circ m}(\infty) = \infty$, this definition may be confusing. The multiplier λ is *not* equal to the limit as $z \to \infty$ of the derivative of $f^{\circ m}(z)$, but rather turns out to be equal to the reciprocal of this number (Problem 4-c). As examples, if $f(z) = 2z$ then ∞ is an attracting fixed point with multiplier $\lambda = 1/2$, while if f is a polynomial of degree $d \geq 2$ then ∞ is a superattracting fixed point, with $\lambda = 0$.

Definition. If $\mathcal{O}$ is an attracting periodic orbit of period m, we define the *basin of attraction* to be the open set $\mathcal{A} \subset S$ consisting of all points $z \in S$ for which the successive iterates $f^{\circ m}(z)$, $f^{\circ 2m}(z)$, $\ldots$ converge towards some point of $\mathcal{O}$. Assuming once more that S is compact, we have the following.

> **Lemma 4.6 (Basins and Repelling Points).** *Every attracting periodic orbit is contained in the Fatou set of f. In fact the entire basin of attraction $\mathcal{A}$ for an attracting periodic orbit is contained in the Fatou set. However, every repelling periodic orbit is contained in the Julia set.*

Proof. First consider a fixed point $z_0 = f(z_0)$ with multiplier λ. If $|\lambda| > 1$, then no sequence of iterates of f can converge uniformly near z_0, for the first derivative of $f^{\circ n}$ at z_0 is λ^n, which diverges to infinity as $n \to \infty$. (Compare the Weierstrass Uniform Convergence Theorem 1.4.) On the other hand, if $|\lambda| < 1$, then choosing $|\lambda| < c < 1$ it follows from Taylor's Theorem that $|f(z) - z_0| \leq c|z - z_0|$ for z sufficiently close to z_0, hence the successive iterates of f, restricted to a small neighborhood, converge uniformly to the constant function $z \mapsto z_0$. (See Lemma 8.1 for details.) The corresponding statement for any compact subset of the basin $\mathcal{A}$ then follows easily. These statements for fixed points generalize immediately to periodic points, making use of Lemma 4.4, since a periodic point of f is just a fixed point of some iterate $f^{\circ m}$. $\square$

The case of an indifferent periodic point is much more difficult. (Compare §§10 and 11.) One particularly important case is the following.

Definition. A periodic point $z_0 = f^{\circ n}(z_0)$ 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 rational map $f(z) = z/(z-1)$ has two fixed points, both with multiplier equal to -1. However, these do not count as parabolic points since $f \circ f(z)$ is identically equal to z. We must exclude such cases so that the following will be true. (In any case, such examples cannot occur for maps of degree 2 or more.)

Lemma 4.7 (Parabolic Points). *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 $w \mapsto w + a_q w^q + a_{q+1} w^{q+1} + \cdots$, where $q \geq 2$, $a_q \neq 0$. It follows that $f^{\circ mk}$ corresponds to a power series $w \mapsto w + k a_q w^q + \cdots$. Thus the qth derivative of $f^{\circ mk}$ at 0 is equal to $q!\, k\, a_q$, which diverges to infinity as $k \to \infty$. It follows from Theorem 1.4 that no subsequence $\{f^{\circ m k_j}\}$ can converge locally uniformly as $k_j \to \infty$. □

Now and for the rest of §4, let us specialize to the case of a rational map $f : \hat{\mathbb{C}} \to \hat{\mathbb{C}}$ of degree $d \geq 2$.

Lemma 4.8 (J Is not Empty). *If f is rational of degree 2 or more, then the Julia set $J(f)$ is nonvacuous.*

Proof. If $J(f)$ were vacuous, then some sequence of iterates $f^{\circ n_j}$ would converge, uniformly over the entire sphere $\hat{\mathbb{C}}$, to a holomorphic limit $g : \hat{\mathbb{C}} \to \hat{\mathbb{C}}$. Here we are using the fact that normality is a local property (Problem 3-e). A standard topological argument would then show that the degree of the map $f^{\circ n_j}$ is equal to the degree of g for large j. (In fact, if two maps f_j and g are sufficiently close that the spherical distance $\sigma(f_j(z)\,,\, g(z))$ is uniformly less than the distance π between antipodal points, then we can deform $f_j(z)$ to $g(z)$ along the unique shortest geodesic; hence these two maps are homotopic and have the same degree.) But the degree of $f^{\circ n}$ cannot equal the degree of g for large n, since 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 Corollary 12.8.

We will also need the following concepts.

Definition. By the *grand orbit* of a point z under $f : S \to S$ we mean the set $\mathrm{GO}(z, f)$ 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 $n \geq 0$. A point $z \in S$ will be called *grand orbit finite* or (to use the classical terminology) *exceptional* under f if its grand orbit $\mathrm{GO}(z, f) \subset S$ is a finite set. Using Montel's Theorem 3.7, we prove the following.

Lemma 4.9 (Finite Grand Orbits). *If $f : \hat{\mathbb{C}} \to \hat{\mathbb{C}}$ is rational of degree $d \geq 2$, 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 superattracting periodic points of f and hence must belong to the Fatou set.*

Proof. (Compare Problem 4-b.) Since f maps $\hat{\mathbb{C}}$ onto itself, it must map any grand orbit $\mathrm{GO}(z, f)$ onto itself. Hence, if this grand orbit is finite, it must map bijectively onto itself, and hence constitute a single periodic orbit $a_0 \mapsto a_1 \mapsto \cdots \mapsto a_m = a_0$. Now note that an arbitrary point $\hat{z} \in \hat{\mathbb{C}}$ has exactly d preimages under f, counted with multiplicity, where the *multiplicity* of $z_j \in f^{-1}(\hat{z})$ as a preimage is greater than 1 if and only if z_j is a critical point of f. Setting $f(z) = p(z)/q(z)$, and assuming that $\hat{z} \neq \infty, f(\infty)$, this is just a matter of counting roots of the degree d polynomial equation $p(z) - \hat{z}q(z) = 0$, checking that the derivative of f vanishes at any multiple root. It follows that every a_j in this finite periodic orbit must be a critical point of f. This proves that any finite grand orbit is superattracting and hence contained in the Fatou set.

(Caution: This argument makes strong use of the compactness of $\hat{\mathbb{C}}$. An entire map from $\mathbb{C}$ to itself, for example $z \mapsto 2ze^z$, may well have a repelling point which is grand orbit finite. See Problem 6-c.)

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 $\hat{\mathbb{C}}$ would be hyperbolic, with $f(U) = U$. Therefore, the set of iterates of f restricted to U would be normal by Montel's Theorem. Hence both U and its complement would be contained in the Fatou set, contradicting Lemma 4.8. $\square$

Theorem 4.10 (Transitivity). *Let z_1 be an arbitrary point of the Julia set $J(f) \subset \hat{\mathbb{C}}$ and let N be an arbitrary neighborhood of z_1. Then the union U of the forward images $f^{\circ n}(N)$ contains the entire Julia set and contains all but at most two points* *of $\hat{\mathbb{C}}$. More precisely, if N is sufficiently small, then U is the complement $\hat{\mathbb{C}} \setminus \mathcal{E}(f)$ of the set of grand orbit finite points.*

In §14 we will prove the much sharper statement that the single forward image $f^{\circ n}(N)$ actually contains the entire Julia set (or the entire Riemann sphere in the special case where there are no grand orbit finite points), provided that n is sufficiently large.

Proof of Theorem 4.10. First note that the complementary set $\hat{\mathbb{C}} \setminus U$ can contain at most two points. For otherwise, since $f(U) \subset U$, it would follow from Montel's Theorem that U must be contained in the Fatou set, which is impossible since $z_1 \in U \cap J$. Again making use of the fact that $f(U) \subset U$, we see that any preimage of a point $z \in \hat{\mathbb{C}} \setminus U$ must itself belong to the finite set $\hat{\mathbb{C}} \setminus U$. It follows by a counting argument that some iterated preimage of z is periodic; hence z itself is periodic and grand orbit finite. Since the set $\mathcal{E}(f)$ of grand orbit finite points is disjoint from J, it follows that $J \subset U$. Finally, if N is small enough so that $N \subset \hat{\mathbb{C}} \setminus \mathcal{E}(f)$, it follows easily that $U = \hat{\mathbb{C}} \setminus \mathcal{E}(f)$. □

Corollary 4.11 (Julia Set with Interior). *If the Julia set contains an interior point, then it must be equal to the entire Riemann sphere.*

For if $J = J(f)$ has an interior point z_1, then choosing a neighborhood $N \subset J$ of z_1, the union $U \subset J$ of forward images of N is everywhere dense, $\overline{U} = \hat{\mathbb{C}}$. Since J is a closed set, it follows that $J = \hat{\mathbb{C}}$. (For examples, see §7.) □

Corollary 4.12 (Basin Boundary = Julia Set). *If $\mathcal{A} \subset \hat{\mathbb{C}}$ is the basin of attraction for some attracting periodic orbit, then the topological boundary $\partial\mathcal{A} = \overline{\mathcal{A}} \setminus \mathcal{A}$ is equal to the entire Julia set. Every connected component of the Fatou set $\hat{\mathbb{C}} \setminus J$ either coincides with some connected component of this basin $\mathcal{A}$ or else is disjoint from $\mathcal{A}$.*

Proof. If N is any neighborhood of a point of the Julia set, then Theorem 4.10 implies that some $f^{\circ n}(N)$ intersects $\mathcal{A}$, hence N itself intersects $\mathcal{A}$. This proves that $J \subset \overline{\mathcal{A}}$. But J is disjoint from $\mathcal{A}$, so it follows that $J \subset \partial\mathcal{A}$. On the other hand, if N is a neighborhood of a point of $\partial\mathcal{A}$, then any limit of iterates $f^{\circ n}|_N$ must have a jump discontinuity between $\mathcal{A}$ and $\partial\mathcal{A}$, hence $\partial\mathcal{A} \subset J$. Finally, note that any connected Fatou component which intersects $\mathcal{A}$, since it cannot intersect the boundary of $\mathcal{A}$, must coincide with some component of $\mathcal{A}$. □

Caution. $\partial\mathcal{A}$ is not the same thing as the union of the boundaries of the connected components of $\mathcal{A}$, which tends to be much smaller when $\mathcal{A}$ is not connected (for example in Figure 6). It may be instructive to compare a Cantor set in the line, which is uncountably infinite although the union of the boundaries of its complementary intervals is countable.

Corollary 4.13 (Iterated Preimages are Dense). *If z_0 is any point of the Julia set $J(f)$, then the set of all iterated preimages*

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

is everywhere dense in $J(f)$.

Because $z_0 \notin \mathcal{E}(f)$, Theorem 4.10 shows that every point $z_1 \in J(f)$ can be approximated arbitrarily closely by points z whose forward orbits contain z_0. □

Remark on computer graphics. (Compare Appendix H.) This corollary suggests an algorithm for computing pictures of the Julia set: Starting with any $z_0 \in J(f)$, first compute all z_1 with $f(z_1) = z_0$; then for each such z_1, compute all z_2 with $f(z_2) = z_1$, 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 and since the number d^n of n-fold iterated preimages is smallest when the degree d is 2. 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 Problems 4-e, 4-f, as well as §19.) However, it does have disadvantages: the number d^n grows very rapidly with n, yet it may take an enormous number of iterated preimages to get close to certain points of J.

Corollary 4.14 (No Isolated Points). *If f has degree 2 or more, then $J(f)$ has no isolated points.*

Proof. First note that $J(f)$ must be an infinite set. For if $J(f)$ were finite it would consist of grand orbit finite points, contradicting Lemma 4.9. Hence $J(f)$ contains at least one limit point z_0. Now the iterated preimages of z_0 form a dense set of nonisolated points in $J(f)$. □

Corollary 4.15 (Julia Components). *For any rational map of degree 2 or more, the Julia set J is either connected or else has uncountably many connected components.*

Proof. If J is not connected, then it can be expressed as the union $J_0 \cup J_1$ of two disjoint, nonvacuous compact subsets. Note that both of these subsets must be infinite, since J has no isolated points. To each $z \in J$

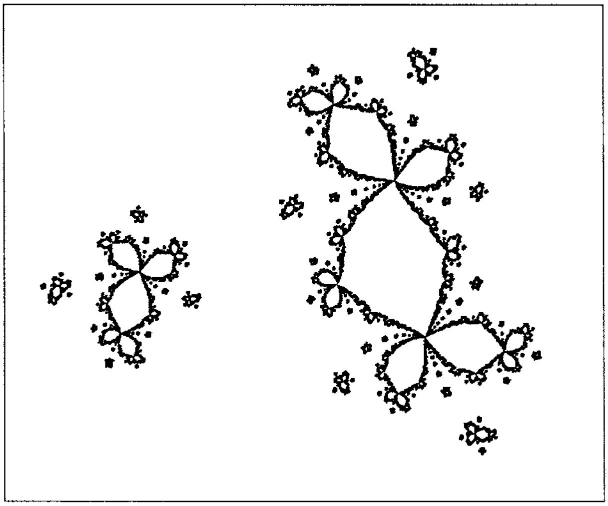


Figure 7. A family of "rabbits" (compare Figure 5d). The Julia set for the cubic polynomial,

$$f(z) \;=\; z^3 - .48z + (.706260 + .502896i),$$

has infinitely many nontrivial connected components.

we can assign an infinite sequence $(\beta_0, \beta_1, \beta_2, \ldots)$ of *bits* $\beta_n = \beta_n(z) \in \{0,1\}$ by the requirement that $f^{\circ n}(z) \in J_{\beta_n}$. Evidently the points in any connected component of J must all have the same bit sequence. Thus, to complete the proof, it suffices to show that uncountably many distinct bit sequences can actually be realized by points of J. (In some cases it may be possible to choose the partition $J = J_0 \cup J_1$ very carefully so that *all* bit sequences will be realized. However, this is certainly not true for most such partitions.)

Let $(\beta_0, \beta_1, \ldots, \beta_k)$ be a finite bit sequence which can actually be realized by some point $z' \in J$. *We will first show that there is another point $z'' \in J$ with the same initial bit sequence, but with $\beta_n(z') \neq \beta_n(z'')$ for some $n > k$.*

The union of either J_β with the Fatou set forms an open hyperbolic neighborhood $U_\beta = \hat{\mathbb{C}} \smallsetminus J_{1-\beta}$. Let $J_{\beta_0,\ldots,\beta_k}$ be the compact set consisting of all $z \in J$ with $f^{\circ n}(z) \in J_{\beta_n}$ for $0 \le n \le k$. and let $U_{\beta_0,\ldots,\beta_k}$ be the open neighborhood consisting of all points $z \in \hat{\mathbb{C}}$ with $f^{\circ n}(z) \in U_{\beta_n}$ for $0 \le n \le k$. Assuming that all points of $J_{\beta_0,\ldots,\beta_k}$ have the same infinite bit sequence $\{\beta_n\}$, we will prove that $U_{\beta_0,\ldots,\beta_k}$ must be contained in the Fatou set, which is impossible.

The sequence $\{\beta_n\}$ must contain either infinitely many zeros or infinitely many ones or both. Hence any infinite sequence of iterates of f must either contain an infinite subsequence $\{f^{\circ n_j}\}$ such that $f^{\circ n_j}(U_{\beta_0,\ldots,\beta_k}) \subset U_0$ for all j, or an infinite subsequence such that $f^{\circ n_j}(U_{\beta_0,\ldots,\beta_k}) \subset U_1$ for all j. Since U_0 and U_1 are hyperbolic, this would contradict the hypothesis that $J_{\beta_0,\ldots,\beta_k} \subset U_{\beta_0,\ldots,\beta_k}$ is in the Julia set. Thus, every finite bit sequence which is realized by a point of J can be extended in two or more different ways. It then follows easily that it can be extended in uncountably many different ways, as required. □

Remark. In the polynomial case, it seems likely that all but countably many of these connected components must be single points. (Compare Branner and Hubbard [1992].) However, this is certainly not true for arbitrary rational maps. (See McMullen's example, Figure 6d. For more on disconnected Julia sets, see Blanchard [1986].)

For the last corollary, we will need some definitions. A topological space X is called a ***Baire space*** if every countable intersection of dense open subsets of X is again dense. We will make use of ***Baire's Theorem***, which asserts that every complete metric space is a Baire space, and also that every locally compact space is a Baire space. (Compare Problem 4-j.) It will be convenient to say that a property of points in the Baire space X is true for ***generic*** $x \in X$ if it is true for all points in some countable intersection of dense open subsets of X. We apply this concept to the topological space $J(f)$.

Corollary 4.16 (Topological Transitivity). *For a generic choice of the point $z \in J = J(f)$, the forward orbit*

$$\{z\,,\ f(z)\,,\ f^{\circ 2}(z)\,,\ \ldots\}$$

is everywhere dense in J.

Proof. (Compare Problem 4-j.) For each integer $j > 0$, we can cover the Julia set $J = J(f)$ by finitely many open sets N_{jk} of diameter less than $1/j$, using the spherical metric. For each such N_{jk}, let U_{jk} be the union of the iterated preimages $f^{-n}(N_{jk})$. It follows from Corollary 4.13 that the closure $\overline{U_{jk} \cap J}$ is equal to the entire Julia set J. In other words, $U_{jk} \cap J$ is a dense open subset of the Julia set. Now if z belongs to the intersection of these dense open sets, then the forward orbit of z intersects every one of the N_{jk} and hence is everywhere dense in J. □

Concluding Problems

Problem 4-a. Degree 1. If $f : \hat{\mathbb{C}} \to \hat{\mathbb{C}}$ is rational of degree $d = 1$, show that the Julia set $J(f)$ is either vacuous or consists of a single repelling or parabolic fixed point.

Problem 4-b. Maps with grand orbit finite points. Now suppose that f is rational of degree $d \geq 2$. **(1)** Show that f is actually a polynomial if and only if $f^{-1}(\infty) = \{\infty\}$, so that the point at infinity is a grand orbit finite fixed point for f. **(2)** Show that f has both zero and infinity as grand orbit finite points if and only if $f(z) = \alpha z^n$, where $n = \pm d$ and $\alpha \neq 0$. **(3)** Conclude that f 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 $z \mapsto 1/z^d$.

Problem 4-c. Fixed point at infinity. If f is a rational function with a fixed point at infinity, show that the multiplier λ at infinity is equal to $\lim_{z\to\infty} 1/f'(z)$. In particular, this fixed point is superattracting if and only if $f'(z) \to \infty$ as $z \to \infty$. (Take $\zeta = 1/z$ and use the series expansion $1/f(1/\zeta) = \lambda\zeta + a_2\zeta^2 + a_3\zeta^3 + \cdots$ in some neighborhood of $\zeta = 0$.)

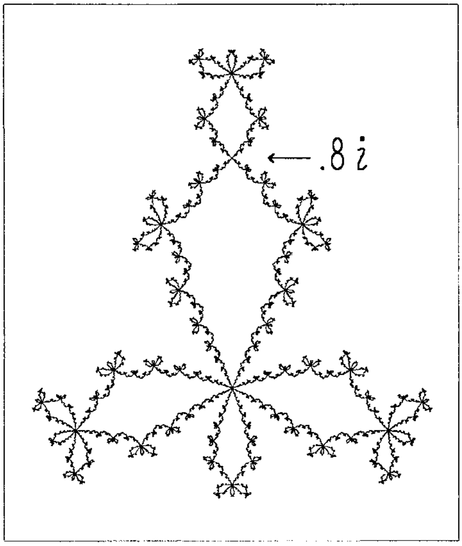


Figure 8. Julia set for $f(z) = z^3 + \frac{12}{25}z + \frac{116}{125}i$.

Problem 4-d. 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 $J \cap N'$. **(1)** Using Corollary 4.13, show that the set of z for which (J, z) is locally conformally isomorphic to (J, z_0) is everywhere dense in J unless the following very exceptional condition is satisfied: *For every backward orbit* $z_0 \leftarrow\!\dashv z_1 \leftarrow\!\dashv z_2 \leftarrow\!\dashv \cdots$ *under* f *which terminates at* z_0*, some* z_j *with* $j > 0$ *must be a critical point of* f. **(2)** As an example, for the map $f(z) = z^2 - 2$ studied in §7, show that this condition is satisfied for the endpoints $z_0 = \pm 2$. Similarly, show that it is satisfied for the point $z_0 = .8i$ of Figure 8. **(3)** For any f, show that there can be only finitely many such exceptional points z_0.

Problem 4-e. A Cantor Julia set. By definition, a topological space is called a *Cantor set* if it is homeomorphic to the standard middle third Cantor set $K \subset [0,1]$ consisting of all infinite sums $\sum_1^\infty 2a_k/3^k$ with coefficients $a_k \in \{0,1\}$. If X is a compact metric space, then a standard theorem asserts that X is a Cantor set if and only if it is totally disconnected (no connected subsets other than points), with no isolated points. (Compare Hocking and Young [1961].) 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 $\epsilon_j = z_j/|z_j| = \pm 1$. In fact

$$z_0 \;=\; \epsilon_0\sqrt{6 + \epsilon_1\sqrt{6 + \epsilon_2\sqrt{6 + \cdots}}}.$$

(Use Corollary 4.13 and the fact that the branch $z \mapsto \sqrt{6+z}$ of f^{-1} is a strictly contracting map which carries the interval $[-3\,,\,3]$ onto $[\sqrt{3}\,,\,3]$. Compare Problem 2-j.) Using Lemma 4.6, show that every orbit outside of this Cantor set must escape to infinity.

Problem 4-f. Fatou [1906]. Similarly, let $f(z) = z^2 + c$ where $c > 1/4$ is real. **(1)** Show that J is a Cantor set disjoint from the real axis and that each orbit $z_0 \mapsto z_1 \mapsto \cdots$ is uniquely determined by the sequence of signs $\epsilon_n = \operatorname{sgn}(\operatorname{Im}(z_n))$. In fact

$$z_0 \;=\; \lim_{n\to\infty} \epsilon_0 g(\epsilon_1 g(\epsilon_2 \cdots \epsilon_{n-1} g(\epsilon_n \hat{z}) \cdots)),$$

where $g(z) = \sqrt{z-c}$ is the branch of f^{-1} which maps the slit plane $U = \mathbb{C} \smallsetminus [c, +\infty)$ onto the upper half-plane, and $\hat{z}$ is some fixed basepoint in J, for example, the fixed point in the upper half-plane. (Use the fact that g restricted to the compact set $J \subset U$ is a strictly contracting map for the Poincaré metric ρ_U.) **(2)** Show that every orbit outside of J must escape to infinity. Using the substitution $w = c/z$, prove corresponding statements for Fatou's 1906 example

$$w \;\mapsto\; \frac{w^2}{c + w^2}.$$

Problem 4-g. Newton's method. (See also Problem 7-a.) Let U be an open subset of $\mathbb{C}$ and let $F : U \to \mathbb{C}$ be a holomorphic map with derivative F'. Newton's method of searching for solutions to the equation $F(\hat{z}) = 0$ can be described as follows.* Consider the auxiliary function $N : U \to \widehat{\mathbb{C}}$, where

$$N(z) = z - F(z)/F'(z).$$

For example, if F is a polynomial of degree d, then N is a rational function of degree $\leq d$. Starting with any initial guess z_0 one can form the successive images $z_{k+1} = N(z_k)$. With luck, these will converge towards a fixed point $\hat{z} = N(\hat{z})$ and it follows immediately that $F(\hat{z}) = 0$.

By definition, $\hat{z}$ is a root of F of *multiplicity* m if the Taylor expansion of F about $\hat{z}$ has the form

$$F(z) = a(z - \hat{z})^m + \text{(higher terms)},$$

with $a \neq 0$ and $m \geq 1$. **(1)** Show that the fixed points of N in the finite region U are precisely the roots of F, and in fact that every root of F is an attracting fixed point for N with multiplier $\lambda = 1 - 1/m$ where m is the multiplicity. Thus every *simple* root of F, with $m = 1$, is a superattracting fixed point of N, while every root of higher multiplicity is a geometrically attracting fixed point. **(2)** Now suppose that F is a polynomial of degree $d > 1$, so that N is a rational map. Show that ∞ is the unique repelling fixed point of N, the multiplier at infinity being $d/(d-1) > 1$. Show that $N(z) = \infty$ for $z \in \mathbb{C}$ only if $F'(z) = 0$, $F(z) \neq 0$. Show that N has derivative

$$N'(z) = \frac{F(z)F''(z)}{F'(z)^2}.$$

Problem 4-h. Lyapunov stability. A point $z_0 \in \widehat{\mathbb{C}}$ is *stable in the sense of Lyapunov* for a rational map f if the orbit of any point which is sufficiently close to z_0 remains uniformly close to the orbit of z_0 for all time. More precisely, for every $\epsilon > 0$ there should exist $\delta > 0$ so that if z has spherical distance $\sigma(z_0, z) < \delta$ then $\sigma(f^{\circ n}z_0, f^{\circ n}z) < \epsilon$ for all n. Show that a point is Lyapunov stable if and only if it belongs to the Fatou set.

Problem 4-i. Fatou components. If Ω is a connected component of the Fatou set of f, show that $f(\Omega)$ is also a connected component of

---

*Ever since the work of Cayley [1879] and Schröder [1871] in the 19th century, the problem of understanding Newton's method has been a primary inspiration for the study of iterated rational functions. For recent work, see, for example, Tan Lei [1997], Roesch [1998], and compare Keen [1989].

Fatou(f).

Problem 4-j. Baire's Theorem and transitivity. For any locally compact space X, prove Baire's Theorem that any countable intersection $U_1 \cap U_2 \cap \cdots$ of dense open subsets of X is again dense. (Within any nonvacuous open set $V \subset X$ choose a nested sequence

$$K_1 \supset K_2 \supset K_3 \supset \cdots$$

of compact sets $K_j \subset U_j$ with nonvacuous interior, and take the intersection.)

A map $f : X \to X$ is called *topologically transitive* if for every pair U and V of nonvacuous open subsets there exists an integer $n \geq 0$ so that $f^{\circ n}(U) \cap V$ is nonvacuous. (Compare Theorem 4.10.) If this condition is satisfied and if there is a countable basis for the open subsets of the locally compact space X, show that a generic orbit under f is dense. (Compare Corollary 4.16.)

§5. Dynamics on Hyperbolic Surfaces

This section will begin the discussion of dynamics on Riemann surfaces other than the Riemann sphere. It turns out that the possibilities for dynamics on a hyperbolic surface are rather limited. Let us first restate the definition in greater generality, allowing Riemann surfaces which may be noncompact.

Definition. For a holomorphic map $f : S \to S$ of an arbitrary Riemann surface, the *Fatou set* of f is the union of all open sets $U \subset S$ such that every sequence of iterates $f^{\circ n_j}|_U$ either

(1) contains a locally uniformly convergent subsequence, or

(2) contains a subsequence which diverges locally uniformly from S, so that the images of a compact subset of U eventually leave any compact subset of S.

(If S is compact, then no sequence can diverge from S, so that case (2) can never occur.) As usual, the complement of the Fatou set is called the *Julia set*.

Remark. In the special case of a surface S which can be described as an open subset with compact closure within a larger Riemann surface T, a completely equivalent condition would be the following. *A point $z \in S$ belongs to the Fatou set of f if and only if, for some neighborhood U of z, any sequence of iterates $f^{\circ n}|_U$ considered as maps from U to T contains a subsequence which converges locally uniformly to a map $U \to T$. Furthermore, if the limit map does not take values in S, then it is necessarily constant.* (Compare Lemma 3.5 and Corollary 3.6.)

Lemma 5.1 (No Julia Set). *For any map $f : S \to S$ of a hyperbolic surface, the Julia set $J(f)$ is vacuous. In particular, f can have no repelling points, parabolic points, or basin boundaries.*

Proof. This follows immediately from Corollary 3.3, together with the proofs of Lemmas 4.6, 4.7, and Corollary 4.12. □

In fact we can give a much more precise statement, essentially due to Fatou. (Compare §16.)

Theorem 5.2 (Classification). *For any holomorphic map $f : S \to S$ of a hyperbolic Riemann surface, exactly one of the following four possibilities holds:*

• **Attracting Case.** *If f has an attracting fixed point, then it follows from Lemma 5.1 (or from Problem 1-h) that all orbits under f converge towards this fixed point. The convergence is uniform on compact subsets of S.*

• **Escape.** *If some orbit under f has no accumulation point in S, then no orbit has an accumulation point. In fact, for any compact set $K \subset S$ there exists an integer n_K so that $K \cap f^{\circ n}(K) = \emptyset$ for $n \geq n_K$.*

• **Finite Order.** *If f has two distinct periodic points, then some iterate $f^{\circ n}$ is the identity map and every point of S is periodic.*

• **Irrational Rotation.** *In all other cases, (S, f) is a rotation domain. That is, S is isomorphic either to a disk $\mathbb{D}$, to a punctured disk $\mathbb{D} \smallsetminus \{0\}$, or to an annulus*

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

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

Much later, in §16, we will apply this theorem to the case where S is an open subset of the Riemann sphere and f is a rational map carrying this set into itself. Here is an important example.

Corollary 5.3 (Siegel Disks). *Let f be a rational map of degree $d \geq 2$. If a connected component U of the Fatou set $\widehat{\mathbb{C}} \smallsetminus J$ contains an indifferent fixed point,*

$$f(z_0) = z_0, \quad |f'(z_0)| = 1,$$

then U is conformally isomorphic to the unit disk $\mathbb{D}$ in such a way that $f|_U$ corresponds to an irrational rotation of the disk.

The proof, assuming Theorem 5.2, is immediate, since U is clearly hyperbolic and maps into itself under f. (No iterate $f^{\circ n}$ can be the identity map on any open set, since f has degree ≥ 2.) The actual existence of nonlinear rational maps which possess such a rotation domain U is highly nontrivial and will be discussed in §11.

In the special case where S is equal to the open unit disk $\mathbb{D}$, the following more precise version of Theorem 5.2 was proved in Denjoy [1926], refining an earlier result by Wolff.

Theorem 5.4 (Denjoy–Wolff). *Let $f : \mathbb{D} \to \mathbb{D}$ be any holomorphic map. Then either:*

(a) f is a "rotation" (with respect to the Poincaré metric) about some fixed point $z_0 \in \mathbb{D}$, or

(b) the successive iterates $f^{\circ n}$ converge, uniformly on compact subsets of $\mathbb{D}$, to a constant function $z \mapsto c_0$, where c_0 may belong either to the open disk $\mathbb{D}$ or to the boundary circle $\partial\mathbb{D}$.

(This is sharper than Theorem 5.2 only in the Escape Case: If some orbit has no accumulation point in $\mathbb{D}$, then every orbit must converge to a single boundary point of $\mathbb{D}$.) According to Lemma 1.8, there is an automorphism of $\hat{\mathbb{C}}$ carrying $\mathbb{D}$ to the upper half-plane $\mathbb{H}$, so the analog of Theorem 5.4 is true for $\mathbb{H}$ also. Here are three examples: If $f : \mathbb{H} \to \mathbb{H}$ is either the parabolic automorphism $z \mapsto z + 1$ or the hyperbolic automorphism $z \mapsto 2z$ or the embedding $z \mapsto z + i$, then all orbits in $\mathbb{H}$ converge within $\hat{\mathbb{C}}$ to the single boundary point ∞.

Unfortunately, Theorem 5.4 is no longer true if we replace $\mathbb{D}$ or $\mathbb{H}$ by an arbitrary hyperbolic open subset of $\hat{\mathbb{C}}$. (See Problem 5-a.) However, we can make the following statement, which will be important in §16.

Lemma 5.5 (Convergence to a Boundary Fixed Point). *Suppose that U is a hyperbolic open subset of a compact Riemann surface and that the map $f : U \to U$ extends continuously to the boundary ∂U, with at most a finite number of fixed points in ∂U. If some orbit of f in U has no accumulation point within U, then all orbits in U must converge within the closure $\overline{U}$ to a single boundary fixed point*

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

This convergence is uniform on compact subsets of U.

The proofs follow.

Proof of Theorem 5.2. Choose some base point $p_0 \in S$ and consider the orbit $p_0 \mapsto p_1 \mapsto p_2 \mapsto \cdots$ under f. It may happen that this orbit diverges from S, so that the Poincaré distance satisfies

$$\lim_{n\to\infty} \operatorname{dist}(p_n\,,\, p_0) \;=\; \infty.$$

If this happens, then for any point q_0 within the ball of radius r about p_0, the corresponding orbit $q_0 \mapsto q_1 \mapsto \cdots$ satisfies $\operatorname{dist}(q_n\,,\, p_n) \le r$, and hence

$$\operatorname{dist}(q_n\,,\, p_0) \;\;\ge\;\; \operatorname{dist}(p_n\,,\, p_0) - r \;\;\to\;\; \infty.$$

Thus all orbits diverge from S, and this divergence is uniform on compact subsets of S.

Otherwise, if $\operatorname{dist}(p_n, p_0)$ does not tend to infinity, then we can find infinitely many p_n within some bounded neighborhood of p_0. These must have some accumulation point $\hat{p} \in S$. Choose integers $n(1) < n(2) < \cdots$ so that the sequence $\{p_{n(j)}\}$ converges to $\hat{p}$, and consider the sequence of maps

$$g_j = f^{\circ(n(j+1)-n(j))}.$$

Then g_j maps $p_{n(j)}$ to $p_{n(j+1)}$. If r_j is the Poincaré distance between $\hat{p}$ and $p_{n(j)}$, it follows that $\operatorname{dist}(g_j(\hat{p}), p_{n(j+1)}) \le r_j$, hence

$$\operatorname{dist}(g_j(\hat{p}), \hat{p}) \;\le\; r_j + r_{j+1} \tag{5:1}$$

by the triangle inequality. Let r be the maximum of $\{r_j\}$. (This exists since the r_j converge to zero.) It follows that the points $g_j(\hat{p})$ all lie within some compact ball $B_{2r} \subset S$. Hence by the Hyperbolic Compactness Theorem 3.7, it follows that the maps g_j all lie within a compact subset of the space $\operatorname{Hol}(S, S)$. Therefore we can choose an accumulation point g of $\{g_j\}$ within $\operatorname{Hol}(S, S)$. Furthermore, since $r_j + r_{j+1} \to 0$ as $j \to \infty$, it follows from $(5:1)$ that $g(\hat{p}) = \hat{p}$. There are now two cases.

Distance-Decreasing Case. If f decreases Poincaré distances, then every iterate of f must satisfy

$$\operatorname{dist}(f^{\circ n}(p), f^{\circ n}(q)) \;\le\; \operatorname{dist}(f(p), f(q)) \;<\; \operatorname{dist}(p, q)$$

for $p \neq q$, and hence the limit g must also decrease Poincaré distances. The maps f and g commute, since g is a limit of iterates of f, and hence f must map the fixed point $\hat{p} = g(\hat{p})$ to a fixed point $f(\hat{p}) = f(g(\hat{p})) = g(f(\hat{p}))$ of g. But g cannot have two distinct fixed points, since it decreases the Poincaré distance. This proves that $\hat{p} = f(\hat{p})$ is also a fixed point for f. It is not hard to see that it must be an attracting fixed point, so that all orbits under f converge to $\hat{p}$.

Distance-Preserving Case. Now suppose that f is a local isometry for the Poincaré metric. *Then we will show that some sequence of iterates of f converges locally uniformly to the identity map of S.* Proceeding as above, some sequence of iterates g_j of f converges to a map g which has a fixed point $\hat{p}$. The multiplier at this fixed point must have absolute value equal to 1, say $g'(\hat{p}) = e^{2\pi i\alpha}$. Whether or not the angle α is rational we can choose some multiple $m\alpha$ which is arbitrarily close to an integer, and conclude that $g^{\circ m}$ has multiplier arbitrarily close to $+1$ at $\hat{p}$. On the other hand, these iterates of g belong to a normal family by Theorem 3.7, so we can choose a subsequence $\{g^{m(j)}\}$ which converges locally uniformly throughout S, with multiplier at $\hat{p}$ converging to $+1$. The limit function has multiplier equal to $+1$. Lifting to the universal covering and applying the Schwarz Lemma, we see that this limit is indeed the identity map of S. Finally, it is not difficult to reduce this double limit of iterates of f to a single limit.

To complete the proof of Theorem 5.2, we must prove the following.

Lemma 5.6 (Iterates near the Identity Map). *If $f : S \to S$ is a map of a hyperbolic surface, with the property that some sequence of iterates $f^{\circ m(i)}$ converges locally uniformly to the identity map, then either f has finite order, or else S is isomorphic to $\mathbb{D}$ or $\mathbb{D} \smallsetminus \{0\}$ or to an annulus $\mathbb{A}_r$, and f corresponds to an irrational rotation.*

(A similar assertion holds for nonhyperbolic surfaces: Compare Problem 6-d.)

Proof of Lemma 5.6. First note that f must be one-to-one, for if $f(p) = f(q)$ with $p \neq q$, then any limit of iterates of f must also map p and q to the same point, so no such limit can be the identity map. Similarly, note that f must be onto. Suppose to the contrary that $f(S)$ is a proper subset of S with say $p \notin f(S)$. If B is a closed disk neighborhood of p, then any map g sufficiently close to the identity map of S must map B to a set $g(B)$ containing p. Hence no such g can be an iterate of f. Combining these two statements, we see that f must be a conformal automorphism of the surface S.

In the simply connected case, the automorphisms of $S \cong \mathbb{D}$ have been described in Theorem 1.7. (See also Problems 1-d and 2-e.) Evidently the "hyperbolic" and "parabolic" automorphisms, with no interior fixed point, behave as in the Escape Case, with no iterate close to the identity map. Thus the only automorphisms satisfying the hypothesis of Lemma 5.6 are the rotations about some fixed point.

For the non-simply connected case, we have to work a bit harder. Suppose that the sequence of maps $f^{\circ m(j)}$ converges, uniformly on compact sets, to the identity map of S. Lifting to the universal covering surface, we obtain a sequence of automorphisms $F^{\circ m(j)} : \tilde{S} \to \tilde{S}$ which converge to the identity modulo the action of the group Γ of deck transformations. In other words, given a compact set $K \subset \tilde{S}$, for j sufficiently large, we can find a deck transformation γ_j so that the composition $F_j = \gamma_j \circ F^{\circ m(j)}$ is uniformly close to the identity throughout K.

Now note that each F_j induces a group homomorphism $\gamma \mapsto \gamma'$ from Γ to itself satisfying the identity $F_j \circ \gamma = \gamma' \circ F_j$. (See Problem 2-b. Since both $F_j \circ \gamma$ and F_j cover the same map from S to itself, it follows that there is some deck transformation γ' carrying $F_j \circ \gamma(\tilde{p})$ to $F_j(\tilde{p})$, and it is easy to check that γ' does not depend on the choice of $\tilde{p} \in \tilde{S}$.) Therefore

$$\gamma' = F_j \circ \gamma \circ F_j^{-1}.$$

If F_j is very close to the identity, then γ' will be very close to γ throughout some large compact set. But Γ is a discrete group, so this implies that $\gamma' = \gamma$ or, in other words,

$$F_j \circ \gamma = \gamma \circ F_j,$$

provided that j is sufficiently large. If some F_j is actually equal to the identity map, then some iterate of f is the identity map of S. Let us assume that this is not the case, so that no F_j is the identity map.

Recall from Theorem 1.12 that each nonidentity element g in the automorphism group $\mathcal{G}(\tilde{S}) \cong \mathcal{G}(\mathbb{D})$ belongs to a unique maximal commutative subgroup, which we will denote by $\mathcal{C}(g)$. Thus two nonidentity elements g_1 and g_2 in $\mathcal{G}(\tilde{S})$ commute if and only if $\mathcal{C}(g_1) = \mathcal{C}(g_2)$. In particular, any nonidentity $\gamma \in \Gamma \subset \mathcal{G}(\tilde{S})$ determines such a group $\mathcal{C}(\gamma)$, and any F_j which is sufficiently close to the identity map must satisfy $\mathcal{C}(F_j) = \mathcal{C}(\gamma)$. But the same is true for any other nonidentity element of Γ. This proves that the commutative group $\mathcal{C}(\gamma) \subset \mathcal{G}(\tilde{S})$ is independent of the particular choice of γ. We will denote this group briefly by $\mathcal{C}(\Gamma)$.

In particular, it follows that Γ is a commutative group. Since we have assumed that S is not simply connected, this implies that S must be either an annulus or a punctured disk. (Problem 2-g.) Furthermore, if j is large then $F_j = \gamma_j \circ F^{\circ m(j)}$ belongs to $\mathcal{C}(\Gamma)$, and hence $F^{\circ m(j)}$ does also. But F commutes with $F^{\circ m(j)}$, so F also belongs to $\mathcal{C}(\Gamma)$.

If the one-parameter group $\mathcal{C}(\Gamma) \subset \mathcal{G}(\tilde{S})$ consists of parabolic transformation, then it will be convenient to use the upper half-plane model $\tilde{S} \cong \mathbb{H}$ and to identify $\mathcal{C}(\Gamma)$ with the group of real translations $w \mapsto w + c$. On the other hand, if $\mathcal{C}(\Gamma)$ consists of hyperbolic transformations, then it is convenient to use the infinite strip model, as in Problem 2-f. In this case also, we can identify $\mathcal{C}(\Gamma)$ with the group of real translations $w \mapsto w + c$. In either case, the nontrivial discrete subgroup Γ must be cyclic, generated by some translation $w \mapsto w + c_0$, and the map F must correspond to some other translation $w \mapsto w + c'$. Now setting $z = e^{2\pi i w/c_0}$ we see that F corresponds to a rotation of an annulus or punctured disk, as required. This completes the proof of Lemma 5.6 and Theorem 5.2 $\square$

Proof of the Denjoy–Wolff Theorem 5.4. Let $f : \mathbb{D} \to \mathbb{D}$ be any holomorphic map. 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 $f_\epsilon(z) = (1-\epsilon)f(z)$ from $\mathbb{D}$ into a proper subset of itself. Each map f_ϵ has a unique fixed point z_ϵ. (Compare Problem 2-j, or note that existence of a fixed point follows from the Brouwer Fixed Point Theorem applied to the disk $(1-\epsilon)\overline{\mathbb{D}}$ and that uniqueness is clear since f_ϵ decreases Poincaré distances.) Since the closed disk $\overline{\mathbb{D}}$ is compact, we can choose a sequence $\{\epsilon_i\}$ tending to zero so that the corresponding fixed points z_{ϵ_i} converge to some limit $\hat{z} \in \overline{\mathbb{D}}$. If $|\hat{z}| < 1$, then $\hat{z}$ is a fixed point of f, and the conclusion follows easily from Theorem 5.2. Assume then that $\hat{z} \in \partial\mathbb{D}$. Choose some arbitrary basepoint $z_0 \in \mathbb{D}$, and let r_i be the Poincaré distance between z_0 and z_{ϵ_i}. Let B_i be the closed neighborhood of Poincaré radius r_i which is centered at z_{ϵ_i} and has the basepoint z_0 on its boundary. Since the map f_{ϵ_i} reduces Poincaré distances, it necessarily carries B_i into itself. These neighborhoods B_i are actually round disks with respect to the Euclidean metric also. (However the Euclidean center is usually different from the Poincaré center. Compare Problem 2-c.) As $i \to \infty$, the round disks B_i must tend to a limit B_∞, which can only be the round disk which is tangent to the unit circle at $\hat{z}$ and whose boundary passes through z_0. (By definition, such a disk tangent to the circle at infinity of $\mathbb{D}$ is called a "horodisk" in $\mathbb{D}$.) It follows by continuity that f maps $\mathbb{D} \cap B_\infty$ into itself. In particular, it follows that the entire orbit of z_0 under f must be contained in B_∞. On the other hand, we know by Theorem 5.2 that the orbit of z_0 must tend to the boundary of $\mathbb{D}$. But a sequence in B_∞ which tends to the boundary of $\mathbb{D}$ can only tend to the point of tangency $\hat{z}$. It follows easily that all orbits in $\mathbb{D}$ tend to the same limiting point $\hat{z}$, as required. □

Proof of Lemma 5.5. Let U be a hyperbolic open subset of the Riemann surface S. Suppose that $f : \overline{U} \to \overline{U}$ is continuous on the compact set $\overline{U}$ and maps U holomorphically into itself, and suppose that some orbit $p_0 \mapsto p_1 \mapsto p_2 \mapsto \cdots$ in U has no accumulation point in U. It follows that the Poincaré distance $\mathrm{dist}_U(p_0, p_n)$ must tend to infinity as $n \to \infty$. Choose some continuous path $p : [0,1] \to U$ from the point $p_0 = p(0)$ to $f(p_0) = p(1)$, and continue this path inductively for all $t \geq 0$ by setting $p(t+1) = f(p(t))$. 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 $\leq \delta$. It follows that $\mathrm{dist}_U(p_0, p(t))$ also tends to infinity as $t \to \infty$.

Let $\hat{p}$ be any accumulation point of $\{p(t)\}$ in ∂U as $t \to \infty$. It follows from Theorem 3.4 that, for any neighborhood V of $\hat{p}$, we can find a smaller neighborhood W, so that any set of Poincaré diameter δ which intersects $U \cap W$ must be contained in V. Hence, for any such V, we can find images $p[n, n+1]$ which are contained in V. Since f maps $p(n)$ to $p(n+1)$, it follows by continuity that $f(\hat{p}) = \hat{p}$. *Thus every accumulation point of the path* $p : [0, \infty) \to U$ *in* ∂U *must be a fixed point of* f. On the other hand, it is not difficult to show that the set of all accumulation points of $p(t)$ as $t \to \infty$ is a connected set (Problem 5-b).

Now assume that f has only finitely many fixed points in ∂U. Since a finite connected set can only be a single point, it follows that $p(t)$ converges to a single point $\hat{p} \in \partial U$ as $t \to \infty$. In particular, the orbit $p_0 \mapsto p_1 \mapsto \cdots$ converges to $\hat{p}$. Now consider an arbitrary orbit $q_0 \mapsto q_1 \mapsto \cdots$ under f: If $\mathrm{dist}_U(p_0, q_0) = r$, then $\mathrm{dist}_U(p_n, q_n) \le r$. Using Theorem 3.4, it follows that the sequence $\{q_n\}$ also converges to $\hat{p}$, and it is easy to check that this convergence is uniform on compact subsets of U. □

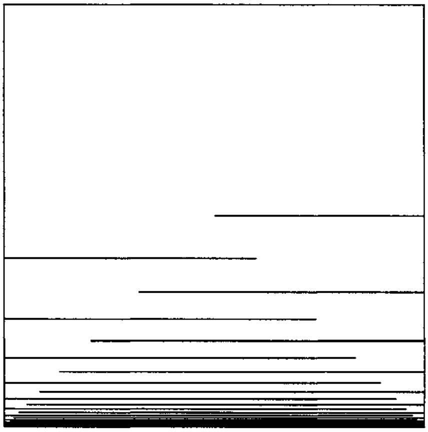


Figure 9. Region with boundary which is not locally connected.

Concluding Problems

Problem 5-a. A badly behaved example. (Compare §17.) Given a sequence of numbers $1 > a_1 > a_2 > \cdots$ converging to 0, let $U \subset \mathbb{C}$ be obtained from the open unit square $(0,1) \times (0,1)$ by removing the line

$$[a_n, \quad 1] \times \{a_n\} \quad \text{for each odd value of } n, \text{ and removing}$$
$$[0,\ 1 - a_n] \times \{a_n\} \quad \text{for each even value of } n,$$

as illustrated in Figure 9. Give U the Poincaré metric. Given a basepoint z_0 in the open set U, for each unit vector v in the tangent space at z_0 there is a unique geodesic ray $g_v : [0, \infty) \to U$ which starts at z_0 with initial velocity vector $g'_v(0)$. Evidently $g_v(t)$ tends towards the boundary of U as $t \to \infty$. Now consider the sequence of line segments

$$L_n = U \cap \Big([0,1] \times \{a_n\}\Big),$$

each of which cuts U into two components. Let V_n be the set of all unit vectors v at z_0 such that $g_v[0,\infty)$ intersects the line segment L_n. **(1)** If z_0 lies near the top of U, show that each V_n contains the closure of V_{n+1} and show that there exists a vector $\hat{v}$ which belongs to the intersection of the V_n. **(2)** Show that every point on the bottom edge of the unit square is an accumulation point for the ray $g_{\hat{v}}(t)$ as $t \to \infty$. **(3)** Now consider the conformal automorphism $f : U \to U$ which maps this geodesic to itself with $f(g_{\hat{v}}(t)) = g_{\hat{v}}(t+1)$. (Compare Problem 2-e.) Show that every point of the bottom edge is also an accumulation point for the orbit $f : z_0 \mapsto z_1 \mapsto \cdots$.

Problem 5-b. Some compact connected sets. **(1)** In any Hausdorff space X, show that the closure of a connected set is connected and show that the intersection of any nested sequence $K_1 \supset K_2 \supset \cdots$ of compact connected sets is again connected. **(2)** Now consider an infinite path $p : [0,\infty) \to X$ in a compact Hausdorff space. Show that the set of all accumulation points of $p(t)$ as $t \to \infty$ can be identified with the intersection of closures

$$\bigcap_t \overline{p[t,\infty)}$$

and therefore is a nonvacuous compact connected set. **(3)** By a similar argument, show that the topological boundary of any simply connected region in $\widehat{\mathbb{C}}$ is connected.

§6. Dynamics on Euclidean Surfaces

This section considers surfaces S such that the universal covering surface $\tilde{S}$ is conformally isomorphic to the complex numbers $\mathbb{C}$. Thus S can be either $\mathbb{C}$ itself or the punctured plane $\mathbb{C} \smallsetminus \{0\} \cong \mathbb{C}/\mathbb{Z}$ or a torus $\mathbb{T} = \mathbb{C}/\Lambda$. (Compare §2.) It turns out that the torus case is interesting but quite easy to understand, while the remaining two cases are extremely difficult.

In the case of a torus $\mathbb{T} = \mathbb{C}/\Lambda$, where Λ is a lattice in $\mathbb{C}$, we will prove the following.

> **Theorem 6.1.** *Every holomorphic map $f : \mathbb{T} \to \mathbb{T}$ is an affine map, $f(z) \equiv \alpha z + c \pmod{\Lambda}$, with degree $d = |\alpha|^2$. The corresponding Julia set $J(f)$ is either the empty set or the entire torus according to whether $d \le 1$ or $d > 1$.*

If $\alpha \neq 1$, note that f has a fixed point $z_0 = c/(1-\alpha)$ and hence is conjugate to the linear map

$$z \mapsto f(z + z_0) - z_0 = \alpha z$$

with multiplier α. For the description of which multipliers can occur, see Problem 6-a.

Proof of Theorem 6.1. To fix our ideas, suppose that $\mathbb{T} = \mathbb{C}/\Lambda$ where the lattice $\Lambda \subset \mathbb{C}$ is spanned by the two numbers 1 and τ and where $\tau \notin \mathbb{R}$. Any holomorphic map $f : \mathbb{T} \to \mathbb{T}$ lifts to a holomorphic map $F : \mathbb{C} \to \mathbb{C}$ on the universal covering surface. Note first that there exists a lattice element $\alpha \in \Lambda$ so that

$$F(z+1) = F(z) + \alpha \qquad \text{for all} \qquad z \in \mathbb{C},$$

for 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 $\mathbb{C}$ is connected and the target space Λ is discrete. Similarly, there exists β so that $F(z+\tau) = F(z) + \beta$ for all z. Now let $g(z) = F(z) - \alpha z$, so that $g(z+1) = g(z)$. Then

$$g(z+\tau) = F(z+\tau) - \alpha(z+\tau) = g(z) + (\beta - \alpha\tau).$$

We claim that g must be constant, say $g(z) = c$ for all z. In fact, g gives rise to a map from $\mathbb{T}$ to the quotient space $\mathbb{C}/(\beta - \alpha\tau)\mathbb{Z}$, which is either $\mathbb{C}$ itself or an infinite cylinder according to whether $\beta - \alpha\tau$ is zero or nonzero. In either case, this quotient is a noncompact Riemann surface, while $\mathbb{T}$ is compact, so such a map must be constant by the Maximum

Modulus Principle. Now $F(z) = g(z) + \alpha z = \alpha z + c$, as required. Since this map multiplies areas by $|\alpha|^2$, it follows easily that f has degree equal to $|\alpha|^2$.

Further properties of f depend on the multiplier α. If $|\alpha| \le 1$, then the derivatives $|d f^{\circ n}(z)/dz| = |\alpha^n|$ are uniformly bounded, so the domain of normality for $\{f^{\circ n}\}$ is the entire torus $\mathbb{T}$. In other words, the Fatou set of f is equal to $\mathbb{T}$. On the other hand, if $|\alpha| > 1$, then $|d f^{\circ n}(z)/dz| = |\alpha^n| \to \infty$ as $n \to \infty$, and it follows that the Julia set of f is the entire torus. $\square$

For further information, see Problems 6-a and 6-b.

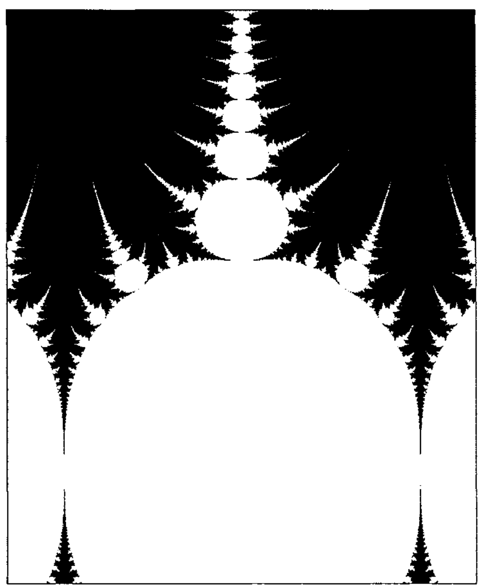


Figure 10. The function $z \mapsto \sin(z)$ *can be considered as a holomorphic map from the cylinder* $\mathbb{C}/2\pi\mathbb{Z}$ *to itself. In this case the Julia set, shown in black, has infinite area (McMullen* [1987]). *The Fatou set* $(\mathbb{C}/2\pi\mathbb{Z}) \smallsetminus J$ *is dense, but has finite area (Schubert* [*to appear*]). *The region shown is* $[-.5, \pi + .5] \times [-1, 4]$.

The Noncompact Euclidean Surfaces. If S is noncompact and Euclidean, then it must be isomorphic to the plane $\mathbb{C}$ or the cylinder $\mathbb{C}/\mathbb{Z}$. First suppose that S is the complex plane $\mathbb{C}$ itself. We can distinguish two different classes of holomorphic maps $\mathbb{C} \to \mathbb{C}$. A *polynomial map* of $\mathbb{C}$ extends uniquely over the Riemann sphere $\widehat{\mathbb{C}}$. Hence the theory of polynomial mappings can be subsumed as a special case of the theory of rational maps of $\widehat{\mathbb{C}}$. (Compare §§9 and 18.) On the other hand, *transcendental* *mappings* from $\mathbb{C}$ to itself form an essentially distinct and more difficult subject of study. Such mappings have been studied for almost eighty years by many authors, starting with Fatou [1926]. The contributions of Baker [1968, 1976] are especially noteworthy.* Even iteration of the exponential map $\exp : \mathbb{C} \to \mathbb{C}$ provides a number of quite challenging problems. For example, according to Lyubich [1987] and Rees [1986b], for Lebesgue almost every starting point $z \in \mathbb{C}$, the set of accumulation points for the orbit of z is equal to the orbit $\{0, 1, e, e^e, \ldots\}$ of zero. (This assertion is an amusing subject for computer experimentation: Random empirical orbits seem to land exactly at 0 after relatively few iterations, unless they first encounter an overflow error.) However, according to Misiurewicz [1981] the Julia set of the exponential map is the entire complex plane, hence a generic orbit is everywhere dense in the plane. (Compare Corollary 4.16. A proof that $J(\exp) = \mathbb{C}$ is included in Devaney [1989, Theorem 9.5]. For a polynomial map of the interval with the analogous property that a generic orbit is dense but almost every orbit is not, see Bruin, Keller, Nowicki, and van Strien [1996].)

Further information about iterated transcendental functions may be found, for example, in Devaney [1986], Goldberg and Keen [1986], and Eremenko and Lyubich [1990, 1992]. The study of iterated maps from the *cylinder* $\mathbb{C}/\mathbb{Z} \cong \mathbb{C} \smallsetminus \{0\}$ to itself is closely related and is also a difficult and interesting subject. See, for example, Keen [1988]. Note that any periodic function from $\mathbb{C}$ to itself can also be considered as a function from the cylinder to itself. (Compare Figure 10.)

Remark. The study of iterated *meromorphic* functions $\mathbb{C} \to \widehat{\mathbb{C}}$, although surely of great interest, does not fit into the framework we have described, since compositions are not everywhere defined. Compare Bergweiler [1993].

Concluding Problems

Problem 6-a. The derivative of a torus map. Consider the torus $\mathbb{T} = \mathbb{C}/\Lambda$, where we may assume that $\Lambda = \mathbb{Z} \oplus \tau\mathbb{Z}$ with $\tau \notin \mathbb{R}$. **(1)** Given $\alpha \in \mathbb{C}$, show that there exists a holomorphic map $f(z) \equiv \alpha z + c$ from $\mathbb{T}$ to itself with derivative α if and only if $\alpha\Lambda \subset \Lambda$, or in other words if

*See Remarks 14.7 and 16.6, and also Lemma F.2. Noel Baker was born in Adelaide and did his first mathematical work there. He then obtained his doctorate under Helmut Kneser in Tübingen. After two years in Edmonton, Canada, he moved to England where he was a Lecturer and later Professor at Imperial College London from 1959 to 1997.

and only if both α and $\alpha\tau$ belong to Λ. Show that an arbitrary integer $\alpha \in \mathbb{Z}$ will satisfy this condition. **(2)** On the other hand, for $\alpha \notin \mathbb{Z}$, show that there exists such a map with derivative α if and only if α satisfies a quadratic equation of the form

$$\alpha^2 + p\alpha + d = 0,$$

where $d = |\alpha|^2$ is the degree and where p is an integer with $p^2 < 4d$. Thus for each choice of degree there are only finitely many possible choices for

$$\alpha = \frac{-p \pm \sqrt{p^2 - 4d}}{2}.$$

(Such a torus is said to admit *complex multiplications.*) **(3)** For a map of degree $d = |\alpha|^2 = 1$ show that α must be an mth root of unity with $m = 1, 2, 3, 4$, or 6. If $m \neq 1$, conclude that $f^{\circ m}$ must be the identity map. Show that the cases $m = 3, 4, 6$ occur for suitably chosen lattices and that the cases $m = 1, 2$ occur for an arbitrary lattice. **(4)** In the special case $\alpha = 1$, show that the closure of every orbit under f is either a finite set, a finite union of parallel circles, or the full torus $\mathbb{T}$.

Problem 6-b. Periodic points of torus maps. **(1)** If $\alpha \neq 0$, show that any equation of the form $f(z) = z_0$ has exactly $d = |\alpha|^2$ solutions $z \in \mathbb{T}$. 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.) **(2)** More generally, if $|\alpha| > 1$ show that the equation $f^{\circ n}(z) = z$ has exactly $|\alpha^n - 1|^2$ solutions in $\mathbb{T}$, all repelling with multiplier α^n. Show that the periodic points of f are everywhere dense in $\mathbb{T}$ whenever $\alpha \neq 0, 1$.

Problem 6-c. Grand orbit finite points. **(1)** Show that a nonlinear holomorphic map $f : \mathbb{C} \to \mathbb{C}$ has at most one grand orbit finite point. **(2)** Show by examples such as $f(z) = \lambda z e^z$ and $f(z) = z^2 e^z$ that this fixed point need not be attracting, and in fact can have arbitrary multiplier.

Problem 6-d. Nonhyperbolic rotation domains. Prove the following analog of Lemma 5.6. If $f : S \to S$ is a self-map of the nonhyperbolic surface S such that some sequence of iterates of f converges locally uniformly to the identity map but no iterate is actually equal to the identity, show that, up to conformal isomorphism, f is either a rotation of $\hat{\mathbb{C}}$, or $\mathbb{C}$, or $\mathbb{C} \smallsetminus \{0\}$, or a translation of a torus.

§7. Smooth Julia Sets

Most Julia sets tend to be complicated fractal subsets of $\hat{\mathbb{C}}$, but there are three exceptions: D. H. Hamilton has shown that every Julia set which is a 1-dimensional topological manifold must be either a circle or closed line segment up to Möbius automorphism, or must have Hausdorff dimension strictly greater than 1. (See Hamilton [1995].) If we count the entire Riemann sphere as another smooth example, it follows that, up to automorphism, there are only three possible smooth subsets of $\hat{\mathbb{C}}$ which can be Julia sets of rational functions of degree ≥ 2. (Compare Corollary 4.11.) However, each of these can appear as a Julia set for many different rational functions, a property which is itself exceptional. This section will discuss these examples.

Example 1: The Circle. The unit circle appears as a Julia set for the mapping $z \mapsto z^{\pm n}$ for any $n \geq 2$. (Compare the discussion of the squaring map in Definition 4.2.) Other rational maps with this same Julia set are described in Problem 7-b. Similarly, the real axis $\mathbb{R} \cup \infty$, as the image of the unit circle under a conformal automorphism, can appear as a Julia set (Problem 7-a).

Example 2: The Interval. Consider the map $f(z) = z^2 - 2$, which carries the closed interval $I = [-2, 2]$ onto itself. (An equivalent example was perhaps first studied as a dynamical system by Ulam and Von Neumann [1947].) This map, and its generalizations to higher degree, are known as *Chebyshev polynomials*. (See Problem 7-c and compare 7-d.)

> **Lemma 7.1.** *The Julia set J for $f(z) = z^2 - 2$ is equal to the interval $I = [-2, 2]$, and every point outside of I belongs to the attractive basin $\mathcal{A}(\infty)$ of the point at infinity.*

First Proof. 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 Corollary 4.13 that I contains the entire Julia set $J(f)$. On the other hand, the basin $\mathcal{A}(\infty)$ is a neighborhood of infinity whose boundary $\partial\mathcal{A}(\infty)$ is equal to $J(f) \subset I$ by Corollary 4.12. It follows that everything outside of I belongs to $\mathcal{A}(\infty)$. Since every point of I has bounded orbit, this proves that $\mathcal{A}(\infty) = \hat{\mathbb{C}} \smallsetminus I$, and it follows that $J(f) = I$. □

Alternative Proof. 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

$z_0 \notin I$, 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 $\mathbb{C} \smallsetminus I$. Since the squaring map in the w-plane is related to f by the identity

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

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 Corollary 4.12, it follows that $J(f) = I$. $\square$

Example 3: All of $\hat{\mathbb{C}}$. The rest of this section will describe a family of examples discovered by Ernst Schröder [1871, p. 307], and rediscovered in greater generality by Samuel Lattès [1918]. Given any lattice $\Lambda \subset \mathbb{C}$ we can form the quotient torus $\mathbb{T} = \mathbb{C}/\Lambda$, as in §2 or §6. Thus $\mathbb{T}$ 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 $\Lambda = \mathbb{Z} + \tau\mathbb{Z}$ is the lattice with basis 1 and τ, where $\tau \notin \mathbb{R}$, 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 $\mathbb{T}$ by identifying each $z \in \mathbb{T}$ 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 $\mathbb{T} \to S$ is two-to-one, except at the four ramification points. To compute the genus of S, we use the following.

Theorem 7.2 (Riemann–Hurwitz Formula). *Let $T \to S$ 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 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 T. However, if v is a critical value, then there are too few preimages of v. The number of missing preimages is precisely the number of ramification points over v, each counted with an appropriate multiplicity. The conclusion follows. $\square$

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

In our example, since T is a torus $\mathbb{T}$, with Euler characteristic $\chi(\mathbb{T}) =$

0, and since there are exactly four simple branch points, we conclude that $2\chi(S) - \chi(\mathbb{T}) = 4$ 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.* (Note: The projection map from $\mathbb{T}$ to the sphere $\widehat{\mathbb{C}}$, suitably normalized, is known as the *Weierstrass* $\wp$*-function*.)

Now consider the doubling map $z \mapsto 2z$ on $\mathbb{T}$. This commutes with multiplication by -1, and hence induces a map $f : S \to S$. Since the doubling map has degree 4, it follows that f is a rational map of degree 4. (More generally, in place of the doubling map, we could use any linear map which carries the lattice Λ into itself, as in Problem 6-a.)

Theorem 7.3 (Lattès). *The Julia set for this rational map* f *is the entire sphere* S.

Proof. Evidently the doubling map on $\mathbb{T}$ 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 2. Evidently f inherits the same property, and the conclusion follows by Lemma 4.6. (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 for n 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 $\mathbb{T}$ map to four "ramification points" on S, which will play a special role. Let us choose a conformal isomorphism from S onto $\widehat{\mathbb{C}}$ which maps the first three of these points to ∞, 0, 1, respectively. The fourth ramification point must then map to some $a \in \mathbb{C} \setminus \{0, 1\}$. In this way we construct a projection map $\wp : T \to \widehat{\mathbb{C}}$ of degree 2, which satisfies $\wp(-z) = \wp(z)$ and which has critical values

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

(Note: This $\wp$ is a linear function of the usual Weierstrass $\wp$-function.) Here a can be any number distinct from 0, 1, ∞. In fact, given $a \in \mathbb{C} \setminus \{0, 1\}$, it is not difficult to show that there is one and only one branched covering $\mathbb{T}' \to \widehat{\mathbb{C}}$ of degree 2 with precisely $\{\infty, 0, 1, a\}$ as ramification points. (Compare Appendix E.) The Riemann-Hurwitz formula shows that this branched covering space $\mathbb{T}'$ is a torus, necessarily isomorphic to $\mathbb{C}/(\mathbb{Z} + \tau\mathbb{Z})$ for some $\tau \notin \mathbb{R}$. The unique deck transformation which interchanges the two preimages of any point must preserve the linear structure, and hence must be multiplication by -1.

Now the doubling map on $\mathbb{T}$ corresponds under $\wp$ to a specific rational map $f_a : \widehat{\mathbb{C}} \to \widehat{\mathbb{C}}$, where

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

with $J(f_a) = \widehat{\mathbb{C}}$ by Theorem 7.3. (For more about f_a, see Problem 7-g.)

Definition 7.4. It is convenient to call the rational map f a *Lattès map* whenever there is an affine map $L : \mathbb{T} \to \mathbb{T}$ of a torus and a holomorphic map $\Theta : \mathbb{T} \to \widehat{\mathbb{C}}$ so that $f = \Theta \circ L \circ \Theta^{-1}$. In fact one can always choose Θ to be the projection map from some torus $\mathbb{T}$ to a quotient surface isomorphic to $\mathbb{T}/G_k$, where G_k is a cyclic group of rotations of the torus about a point, with order* k equal to 2, 3, 4, or 6. (Compare Milnor [2004b].) For a quite different characterization of this same class of maps, see Theorem 19.9 (Case 0) in §19.

Remark. Mary Rees has proved the existence of many more rational maps with $J(f) = \widehat{\mathbb{C}}$. (See Rees [1984, 1986a], and also Herman [1984].) For any degree $d \geq 2$, let Rat_d be the complex manifold consisting of all rational maps of degree d. Rees shows that there is a subset of Rat_d of positive measure consisting of maps f which are "ergodic." By definition, this means that any measurable subset of $\widehat{\mathbb{C}}$ which is fully invariant under f must have either full measure or measure zero. Using Theorem 16.1, it is not hard to see that any ergodic map must necessarily have $J(f) = \widehat{\mathbb{C}}$.

Concluding Problems

Problem 7-a. A Newton's method example. (See Schröder [1871], Cayley [1879].) Let $f(z) = z^2 + 1$. Trying to solve the equation $f(z) = 0$ by Newton's method (Problem 4-g), we are led to the rational map

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

from $\widehat{\mathbb{C}} = \mathbb{C} \cup \infty$ to itself. **(1)** Show that every orbit of N in the upper half-plane converges to $+i$ and that every orbit in the lower half-plane converges to $-i$. Conclude that the Julia set $J(N)$ is equal to $\mathbb{R} \cup \infty$. (Alternatively, note that N is conjugate to $z \mapsto z^2$ under a holomorphic change of coordinates.) **(2)** More generally, for any quadratic polynomial equation with distinct roots, show that $J(N)$ is a straight line together with the point ∞. **(3)** What happens for a quadratic equation with double root?

---

* Caution: A more restrictive definition was used in previous editions, allowing only the case $k = 2$.

Problem 7-b. Blaschke products. For any $a \in \mathbb{D}$ the map

$$\phi_a(z) = (z-a)/(1-\bar{a}z)$$

carries the unit disk $\mathbb{D}$ isomorphically onto itself. (Compare Theorem 1.7.) 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. **(1)** Show that every such f is a rational map which carries $\mathbb{D}$ onto $\mathbb{D}$ and $\hat{\mathbb{C}} \smallsetminus \overline{\mathbb{D}}$ onto $\hat{\mathbb{C}} \smallsetminus \overline{\mathbb{D}}$. Conclude that the Julia set $J(f)$ is contained in the unit circle. **(2)** If $g(z) = 1/f(z)$, interchanging the interior and exterior of the unit circle, show that $J(g)$ is also contained in the unit circle. **(3)** If $n \geq 2$ and if one of the factors is $\phi_0(z) = z$, show that f has attracting fixed points at zero and infinity and show that $J(f)$ is the entire unit circle.

Problem 7-c. Chebyshev polynomials. Define monic polynomials

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

inductively by the formula $P_{n+1}(z) + P_{n-1}(z) = zP_n(z)$. **(1)** 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 $P_m \circ P_n = P_{mn}$. **(2)** For $n \geq 2$ show that the Julia set of $\pm P_n$ is the interval $[-2, 2]$. **(3)** For $n \geq 3$ show that P_n has $n-1$ distinct critical points in the finite plane, but only two critical values, namely ± 2.

Problem 7-d. More maps with interval Julia set. Now suppose that f is a Blaschke product with real coefficients and with an attracting fixed point at the origin. (Compare Problem 7-b.) Show that there is one and only one rational map F of the same degree so that the following diagram is commutative:

$$\begin{array}{ccc} \hat{\mathbb{C}} & \xrightarrow{f} & \hat{\mathbb{C}} \\ \downarrow z+1/z & & \downarrow z+1/z \\ \hat{\mathbb{C}} & \xrightarrow{F} & \hat{\mathbb{C}}, \end{array}$$

and show that $J(F) = [-2, 2]$. (In the special case $f(z) = z^n$, F will be a Chebyshev polynomial.) In this way, construct a 1-real-parameter family of conformally distinct quadratic rational maps with Julia set $[-2, 2]$.

Problem 7-e. Periodic orbits. Show that the Julia sets for Chebyshev maps and Lattès maps, and also for the power map $z \mapsto z^{\pm d}$, have the following extraordinary property. For all but finitely many periodic orbits $z_0 \mapsto z_1 \mapsto \cdots \mapsto z_n = z_0$, show that 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 $J = \hat{\mathbb{C}}$, where d is the degree. (Compare Problem 19-d.)

Problem 7-f. A quadratic Lattès map. Let $\mathbb{T}$ be the torus $\mathbb{C}/\mathbb{Z}[i]$, where $Z[i] = \mathbb{Z} \oplus i\mathbb{Z}$ is the lattice of Gaussian integers, and let $L : \mathbb{T} \to \mathbb{T}$ be the linear map $L(z) = (1+i)z$ of degree $|1+i|^2 = 2$. (Compare Theorem 6.1 and Problem 6-a.) Let $\wp : \mathbb{T} \to \widehat{\mathbb{C}}$ be the associated Weierstrass map, with $\wp(-z) = \wp(z)$, and let $F = \wp \circ L \circ \wp^{-1}$ be the associated quadratic rational map. **(1)** Show that F has critical orbits

$$\wp((1+i)/4) \;\mapsto\; \wp(i/2) \;\mapsto \wp((1 \pm i)/2) \;\mapsto \wp(0)$$

and

$$\wp((1-i)/4) \;\mapsto\; \wp(1/2) \;\mapsto \wp((1 \pm i)/2) \;\mapsto \wp(0).$$

Show that the multiplier at the fixed point $\wp(0)$ is equal to $(1+i)^2 = 2i$. **(2)** After conjugating F by a Möbius automorphism, we may assume that the critical points are ± 1 and the postcritical fixed point is at ∞. Show that the most general quadratic map with critical points ± 1 and a fixed point at ∞ has the form $f(z) = a(z + z^{-1}) + b$ and show that the required critical orbit relations are satisfied if and only if $a^2 = -1/2$ and $b = 0$. More precisely, by computing the multiplier of the fixed point at infinity, show that $a = 1/2i$. (Compare Milnor [2004a].)

Problem 7-g. The family of degree 4 Lattès maps. For the torus $\mathbb{T} = \mathbb{C}/(\mathbb{Z} + \tau\mathbb{Z})$ of Example 3 and Theorem 7.3, show that the involution $z \mapsto z + 1/2$ of $\mathbb{T}$ corresponds under $\wp$ to an involution of the form $w \mapsto a/w$ of $\widehat{\mathbb{C}}$, with fixed points $w = \pm\sqrt{a}$. **(1)** Show that the rational map $f = f_a$ has poles at ∞, 0, 1, a and double zeros at $\pm\sqrt{a}$. **(2)** 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

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

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

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

and also $a(-\overline{\tau}) = \overline{a}(\tau)$. 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 $\mathbb{H}$ as a universal covering of the thrice-punctured sphere $\mathbb{C} \smallsetminus \{0, 1\}$. Compare Ahlfors [1966, pp. 269–274].)

Problem 7-h. Postcritical finiteness. For each of the six critical points ω of this map f, show that $f(f(\omega))$ is the repelling fixed point at infinity. (According to Corollary 16.5, the fact that each critical orbit terminates on a repelling cycle is already enough to imply that $J(f) = \widehat{\mathbb{C}}$. Rational maps satisfying this condition are much more common than Lattès maps.)

LOCAL FIXED POINT THEORY

§8. Geometrically Attracting or Repelling Fixed Points

The next four sections will study the dynamics of a holomorphic map in some small neighborhood of a fixed point. This local theory is a fundamental tool in understanding more global dynamics. It has been studied for well over a hundred years by mathematicians such as Ernst Schröder, Gabriel Kœnigs, Léopold Leau, Lucjan Böttcher, Pierre Fatou, Gaston Julia, Hubert Cremer, Carl Ludwig Siegel, Thomas Cherry, Alexander Bryuno, Jean Écalle, Serguei Voronin, Michel Herman, Jean-Christophe Yoccoz, and Ricardo Perez-Marco. In most cases it is now well understood, but a few cases still present extremely difficult problems.

We start by expressing our map in terms of a local uniformizing parameter z, which can be chosen so that the fixed point corresponds to $z = 0$. We can then describe the map by a power series of the form

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

which converges for $|z|$ sufficiently small. Recall that the initial coefficient $\lambda = f'(0)$ is called the *multiplier* of the fixed point: It is given a special name since it plays a dominant role in the discussion.

Attracting Points. By definition, a fixed point p of a map f is *topologically attracting* if it has a neighborhood U so that the successive iterates $f^{\circ n}$ are all defined throughout U and so that this sequence $\{f^{\circ n}|_U\}$ converges uniformly to the constant map $U \to p$.

> **Lemma 8.1 (Topological Characterization of Attracting Points).** *A fixed point for a holomorphic map of a Riemann surface is topologically attracting if and only if its multiplier satisfies* $|\lambda| < 1$.

Proof. In one direction, this follows from elementary calculus. We can assume as above that the fixed point is $0 = f(0) \in \mathbb{C}$, with Taylor expansion $f(z) = \lambda z + O(z^2)$ as $z \to 0$. In other words there are constants $r_0 > 0$ and C so that

$$|f(z) - \lambda z| \;\le\; C\,|z^2| \qquad \text{for} \qquad |z| < r_0. \tag{8:1}$$

Choose c so that $|\lambda| < c < 1$ and choose $0 < r \le r_0$ so that $|\lambda| + Cr < c$. For all $|z| < r$, it follows that

$$|f(z)| \;\le\; |\lambda z| + C\,|z^2| \;\le\; c|z|,$$

and hence

$$|f^{\circ n}(z)| \;\le\; c^n|z| \;<\; c^n r.$$

As $n \to \infty$, this tends uniformly to zero, as required.

Conversely, if f is topologically attracting, then for any sufficiently small disk $\mathbb{D}_\epsilon$ about the origin there exists an iterate $f^{\circ n}$ which maps $\mathbb{D}_\epsilon$ onto a proper subset of itself. By the Schwarz Lemma 1.2, this implies that the multiplier of $f^{\circ n}$ satisfies $|\lambda^n| < 1$, and hence $|\lambda| < 1$ as required. $\square$

Definition. An attracting fixed point will be called either *superattracting* or *geometrically attracting*, according to whether its multiplier is zero, or satisfies $0 < |\lambda| < 1$.

In either case, we will show that f can be reduced to a simple normal form by a suitable change of coordinates. This section considers only the geometrically attracting case $\lambda \ne 0$. In other words, we assume that the origin is not a critical point. The following was proved in Kœnigs [1884].*

Theorem 8.2 (Kœnigs Linearization). *If the multiplier λ satisfies $|\lambda| \ne 0, 1$, 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 nonzero constant.*

In other words, the following diagram is commutative,

$$\begin{array}{ccc} U & \xrightarrow{f} & f(U) \\ \phi\downarrow & & \phi\downarrow \\ \mathbb{C} & \xrightarrow{\lambda\cdot} & \mathbb{C}, \end{array}$$

where ϕ is univalent on the neighborhood $U \cup f(U)$ of zero. The usefulness of this functional equation

$$\phi \circ f \circ \phi^{-1}(w) \;=\; \lambda w \tag{8 : 2}$$

had been pointed out some years earlier by E. Schröder. (Compare §4.1.) However, Schröder had been able to find solutions only in very special cases.

Proof of Uniqueness. If there were two such maps ϕ and ψ, then the composition $\psi \circ \phi^{-1}$ would commute with the map $w \mapsto \lambda w$. Expanding

*Much later, in 1920, Kœnigs became the first Secretary General of the International Mathematical Union, where his insistence on strictly excluding German mathematicians caused a great deal of dissension. The 1924 International Congress of Mathematicians was initially supposed to take place in New York, but the American Mathematical Society withdrew its invitation because of this discriminatory policy. (See Lehto [1998].)

as a power series,

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

and then composing on the left or right with multiplication by λ, we see by comparing coefficients 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 $|\lambda| < 1$. Choose a constant $c < 1$ so that $c^2 < |\lambda| < c$. As in the proof of Lemma 8.1, we can choose a neighborhood $\mathbb{D}_r$ of the origin so that $|f(z)| \le c|z|$ for $z \in \mathbb{D}_r$. Thus for any starting point $z_0 \in \mathbb{D}_r$, the orbit $z_0 \mapsto z_1 \mapsto \cdots$ under f converges geometrically towards the origin, with $|z_n| \le rc^n$. By Taylor's Theorem $(8:1)$ we have $|f(z) - \lambda z| \le C|z^2|$ for $z \in \mathbb{D}_r$, and hence

$$|z_{n+1} - \lambda z_n| \le C|z_n|^2 \le Cr^2 c^{2n}.$$

Setting $k = Cr^2/|\lambda|$, it follows that the numbers $w_n = z_n/\lambda^n$ satisfy

$$|w_{n+1} - w_n| \le k(c^2/|\lambda|)^n.$$

These differences converge uniformly and geometrically to zero. Thus the holomorphic functions $z_0 \mapsto w_n(z_0)$ converge, uniformly throughout $\mathbb{D}_r$, to a holomorphic limit $\phi(z_0) = \lim_{n\to\infty} z_n/\lambda^n$. (Compare Theorem 1.4.) The required identity $\phi(f(z)) = \lambda\phi(z)$ follows immediately. Furthermore, since each correspondence $z_0 \mapsto w_n = z_n/\lambda^n$ has derivative 1 at the origin, it follows that the limit function ϕ has derivative $\phi'(0) = 1$ and hence is a local conformal isomorphism.

Proof when $|\lambda| > 1$. The statement in this case follows immediately by applying the argument above to the map f^{-1}, which can be defined as a single-valued holomorphic function in some neighborhood of zero, with multiplier satisfying $0 < |\lambda^{-1}| < 1$. This completes the proof of Theorem 8.2. □

Remark 8.3. More generally, suppose that we consider a family of maps f_α of the form

$$f_\alpha(z) = \lambda(\alpha)z + b_2(\alpha)z^2 + \cdots$$

which depend holomorphically on one (or more) complex parameters α and have multiplier satisfying $|\lambda(\alpha)| \ne 0, 1$. Then a similar argument shows that *the Kœnigs function* $\phi(z) = \phi_\alpha(z)$ *depends holomorphically on* α. (This fact will be important in the proof of Lemma 11.15.) To prove this statement, first fix some $0 < c < 1$ and suppose that $|\lambda(\alpha)|$ varies through some compact subset of the interval (c^2, c). Then the convergence in the proof of Theorem 8.2 is uniform in α. Since we are free to choose c, the general case follows. $\square$

In the attracting case $0 < |\lambda| < 1$, we can restate Theorem 8.2 in a more global form as follows. Suppose that $f : S \to S$ is a holomorphic map from a Riemann surface into itself with an attracting fixed point $\hat{p} = f(\hat{p})$ of multiplier $\lambda \neq 0$. Recall from §4 that the total ***basin of attraction*** $\mathcal{A} = \mathcal{A}(\hat{p}) \subset S$ consists of all $p \in S$ for which $\lim_{n\to\infty} f^{\circ n}(p)$ exists and is equal to $\hat{p}$. The ***immediate basin*** $\mathcal{A}_0$ is defined to be the connected component of $\mathcal{A}$ which contains $\hat{p}$. (Equivalently, $\mathcal{A}_0$ is the connected component of the Fatou set $S \smallsetminus J$ which contains $\hat{p}$. Compare Lemma 4.6 and Corollary 4.12.)

Corollary 8.4 (Global Linearization). *With $\hat{p} = f(\hat{p})$ as above, there is a holomorphic map ϕ from $\mathcal{A}$ to $\mathbb{C}$, with $\phi(\hat{p}) = 0$, so that the diagram*

$$\begin{array}{ccc} \mathcal{A} & \xrightarrow{f} & \mathcal{A} \\ \downarrow \phi & & \downarrow \phi \\ \mathbb{C} & \xrightarrow{\lambda \cdot} & \mathbb{C} \end{array} \qquad (8:3)$$

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

In fact, to compute $\phi(p_0)$ at an arbitrary point of $\mathcal{A}$ we must simply follow the orbit of p_0 until we reach some point p_k which is very close to $\hat{p}$, then evaluate the Kœnigs coordinate $\phi(p_k)$ and multiply by λ^{-k}. Alternatively, in terms of a local uniformizing coordinate z with $z(\hat{p}) = 0$ we can simply set $\phi(p) = \lim_{n\to\infty} z(f^{\circ n}(p))/\lambda^n$. $\square$

Now let us specialize to the case of the Riemann sphere. Suppose that $f : \widehat{\mathbb{C}} \to \widehat{\mathbb{C}}$ is a rational function of degree $d \geq 2$. Let $\hat{z} \in \widehat{\mathbb{C}}$ be a geometrically attracting fixed point with basin of attraction $\mathcal{A} \subset \widehat{\mathbb{C}}$. In some small neighborhood $\mathbb{D}_\epsilon$ of $0 \in \mathbb{C}$, note that there is a well-defined holomorphic map $\psi_\epsilon : \mathbb{D}_\epsilon \to \mathcal{A}_0$ which is inverse to the map $\phi : \mathcal{A} \to \mathbb{C}$ of Corollary 8.4 in the sense that $\phi \circ \psi_\epsilon$ is equal to the identity map of $\mathbb{D}_\epsilon$, and which satisfies $\psi_\epsilon(0) = \hat{z}$.

Lemma 8.5 (Finding a Critical Point). *This local inverse $\psi_\epsilon : \mathbb{D}_\epsilon \to \mathcal{A}_0$ extends, by analytic continuation, to some maximal open disk $\mathbb{D}_r$ about the origin in $\mathbb{C}$. This yields a uniquely defined holomorphic map $\psi : \mathbb{D}_r \to \mathcal{A}_0$ with $\psi(0) = \hat{z}$ and $\phi(\psi(w)) \equiv w$. Furthermore, ψ extends homeomorphically over the boundary circle $\partial\mathbb{D}_r$, and the image $\psi(\partial\mathbb{D}_r) \subset \mathcal{A}_0$ necessarily contains a critical point of f.*

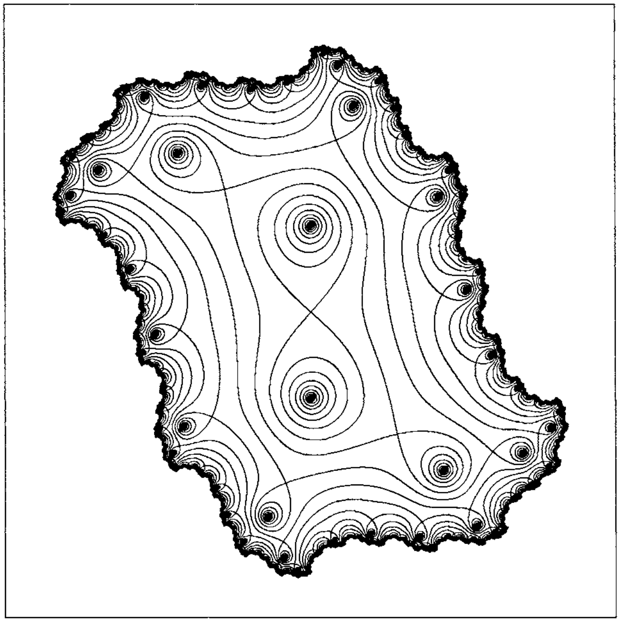


Figure 11. Julia set for $z \mapsto z^2 + .7iz$, *with curves* $|\phi| = \text{constant}$.

As an example, Figure 11 illustrates the map $f(z) = z^2 + 0.7iz$. Here the Julia set J is the outer Jordan curve, bounding the basin $\mathcal{A}$ of the attracting fixed point $\hat{z} = 0$. The critical point $c = -0.35i$ is the center of symmetry, and the fixed point $\hat{z} = 0$ is at the center of the nested circles directly above it, while the preimage $-.7i$ of the fixed point is directly below it. The curves $|\phi(z)| = \text{constant} = |\phi(c)/\lambda^n|$ have been drawn in. Thus the region $\psi(\mathbb{D}_r)$ of Lemma 8.5 is bounded by the top half of the figure eight curve through the critical point. Note that ϕ has zeros at all iterated preimages of $\hat{z}$, and has critical points (crossing points in the figure) at all iterated preimages of the critical point c. The function $z \mapsto \phi(z)$ is unbounded and oscillates wildly as z tends to $J = \partial\mathcal{A}$.

Proof of Lemma 8.5. Let us try to extend ψ_ϵ by analytic continuation along radial lines through the origin. It cannot be possible to extend indefinitely far in every direction, for that would yield a holomorphic map ψ from the entire complex plane onto an open set $\psi(\mathbb{C}) \subset \mathcal{A}_0 \subset \widehat{\mathbb{C}}$, with $\phi(\psi(w)) = w$. This would only be possible if the complement $\widehat{\mathbb{C}} \smallsetminus \psi(\mathbb{C})$ consisted of a single point. But the map $f|_{\psi(\mathbb{C})}$ is one-to-one, so this would imply that f is one-to-one, contradicting our hypothesis that f has degree $d \geq 2$.

Thus there must exist some largest radius r so that ψ_ϵ extends analytically throughout the open disk $\mathbb{D}_r$. Let U be the image $\psi(\mathbb{D}_r) \subset \mathcal{A}_0$.

Thus we obtain a commutative diagram of conformal isomorphisms

$$\begin{array}{ccc} U & \xrightarrow{f} & f(U) \\ \psi \uparrow\downarrow \phi & & \psi \uparrow\downarrow \phi \\ \mathbb{D}_r & \xrightarrow{\lambda\cdot} & \lambda\mathbb{D}_r . \end{array}$$

Note that the closure $\overline{U} \subset \widehat{\mathbb{C}}$ must be contained in the attracting basin $\mathcal{A}$. In fact, since the image of $\mathbb{D}_r$ under multiplication by λ is contained in a compact subset $\lambda\overline{\mathbb{D}}_r \subset \mathbb{D}_r$, it follows that the image $f(U)$ is contained in a corresponding compact subset $K \subset U$. It then follows by continuity that $f\left(\overline{U}\right) \subset K \subset U \subset \mathcal{A}$, which implies that $\overline{U} \subset \mathcal{A}$. In particular, it follows that ϕ is defined and holomorphic throughout a neighborhood of $\overline{U}$.

We will next show that the topological boundary ∂U contains a critical point of f, for otherwise we could analytically continue the map $\psi : \mathbb{D}_r \to \mathcal{A}$ over a strictly larger disk, as follows. For any boundary point $w_0 \in \partial\mathbb{D}_r$ choose some accumulation point z_0 in ∂U for the curve $t \mapsto \psi(tw_0)$ as $t \to 1$. If z_0 is not a critical point of f then we can choose a holomorphic branch g of f^{-1} in some neighborhood of $f(z_0)$, so that $g(f(z_0)) = z_0$, and then extend ψ holomorphically throughout a neighborhood of w_0 by the formula

$$w \mapsto g(\psi(\lambda w)).$$

If there were no critical points at all in ∂U, then evidently these local extensions would piece together to yield a holomorphic extension of ψ through a disk which is strictly larger than $\mathbb{D}_r$.

Finally we will show that ϕ maps the compact set $\overline{U}$ homeomorphically onto the closed disk $\overline{\mathbb{D}}_r$. It suffices to show that two distinct points $z \neq z'$ in the boundary ∂U must have distinct images $\phi(z) \neq \phi(z')$ in $\partial\mathbb{D}_r$. Suppose to the contrary that $\phi(z) = \phi(z') = w \in \partial\mathbb{D}_r$. Choose a sequence of points $z_j \in U$ converging to z and a sequence of points $z'_j \in U$ converging to z'. Then the sequences $\{\phi(z_j)\}$ and $\{\phi(z'_j)\}$ converge to the same limit in $\partial\mathbb{D}_r$. Let L_j be the straight line segment from $\phi(z_j)$ to $\phi(z'_j)$ in $\mathbb{D}_r$, and let $X \subset \partial U$ be the set of all accumulation points for the curves $\psi(L_j)$ as $j \to \infty$. Then it is not difficult to show that X is a compact connected set containing both z and z', but that $f(X)$ consists of a single point in U. Evidently this is impossible. $\square$

More generally, if $\mathcal{O} = \{z_1, \ldots, z_m\}$ is an attracting periodic orbit of period m, so that each z_j is an attracting fixed point for the m-fold composition $f^{\circ m}$, then the *immediate basin* $\mathcal{A}_0 = \mathcal{A}_0(\mathcal{O}, f)$ is defined to be the union of the immediate attractive basins $\mathcal{A}_0(z_j)$ of the m fixed points $z_j = f^{\circ m}(z_j)$ under the map $f^{\circ m}$. The following fundamental result is due to Fatou and Julia.

Theorem 8.6 (Finding Periodic Attractors). *If f is a rational map of degree $d \geq 2$, then the immediate basin of every attracting periodic orbit contains at least one critical point. Hence the number of attracting periodic orbits is finite, less than or equal to the number of critical points.*

Proof. In the case of a geometrically attracting fixed point, the first statement follows immediately from Lemma 8.5, while a superattracting fixed point is itself the required critical point in its basin. Now consider a period m attracting orbit $\{z_j\}$ with $f(z_j) = z_{j+1}$, taking the subscripts j to be integers modulo m. Evidently $f(\mathcal{A}_0(z_j)) \subset \mathcal{A}_0(z_{j+1})$. If none of the $\mathcal{A}_0(z_j)$ contained a critical point, then, by the chain rule, the m-fold composition mapping each $\mathcal{A}_0(z_j)$ into itself would not have any critical point, which is impossible.

The conclusion now follows since the attractive basins of the various periodic attractors are clearly pairwise disjoint and since a nonconstant rational map can have only finitely many critical points. □

(For another proof, see Problem 8-g.)

As an example, for a polynomial map of degree $d \geq 2$, there are at most $d-1$ finite critical points, and hence at most $d-1$ periodic attractors (not counting the fixed point at infinity). In the case of a rational map, there are $2d-2$ critical points, counted with multiplicity. This follows from the Riemann-Hurwitz Formula 7.2, or by simply inspecting the required polynomial equation $p'q - q'p = 0$ where $f(z) = p(z)/q(z)$, taking particular care with the possibility of a critical point at infinity. Hence if $d \geq 2$ there are at most $2d-2$ periodic attractors. (Compare Corollary 10.16 and Lemma 13.2.)

Remark 8.7. The situation with two complex variables is quite different. A holomorphic map from the complex projective plane to itself can have infinitely many periodic attractors. Similarly, a polynomial automorphism of $\mathbb{C}^2$ can have infinitely many periodic attractors. In this last case, there are evidently no critical points to work with. See Appendix D for further discussion.

Remark 8.8. Theorem 8.6 gives rise to a simple algorithm for trying to locate the attracting periodic points, if they exist, for any nonlinear rational map. Starting at each one of the critical points, simply iterate the map many times and then test for (approximate) periodicity. Of course if the period is very large, then this becomes impractical. As an explicit example, it is easy to check that the quadratic map $f(z) = z^2 - 1.5$ of Figure 15 (page

97) has no attracting orbits of reasonable period. However, I know no way of deciding whether it has an attracting orbit of some very high period.

Theorem 8.9 (Topology of $\mathcal{A}_0$). *Let $\mathcal{A}_0$ be the immediate basin of an attracting fixed point (either geometrically attracting or superattracting). Then the complement $\widehat{\mathbb{C}} \smallsetminus \mathcal{A}_0$ is either connected or else has uncountably many connected components.*

(Compare Corollary 4.15.) It follows that $\mathcal{A}_0$ itself is either simply connected or infinitely connected. For infinitely connected examples, see Figures 5c, 7, or 16 (pages 42, 50, 97). Note that the analogous statement for an attracting periodic point of period p follows by applying Theorem 8.9 to the iterate $f^{\circ p}$.

Proof of Theorem 8.9. Choose a small open disk N_0 about the attracting point $\hat{z}$ so that $f(\overline{N_0}) \subset N_0$ and so that the boundary $\Gamma = \partial N_0$ is a simple closed curve containing no iterated forward images of critical points. *Setting N_k equal to the connected component of $f^{-k}(N_0)$ which contains $\hat{z}$, we have*

$$N_0 \subset N_1 \subset N_2 \subset \cdots$$

with union equal to the entire immediate basin $\mathcal{A}_0$. In fact any point of $\mathcal{A}_0$ can be joined to $\hat{z}$ by a path $P \subset \mathcal{A}_0$. Since P is compact, some forward image $f^{\circ k}(P)$ must be contained in N_0; and since P is a connected set containing $\hat{z}$, this implies that $P \subset N_k$. Thus the union of the N_k is all of $\mathcal{A}_0$.

Evidently each N_k is bounded by some finite number of simple closed curves, while the complement $\widehat{\mathbb{C}} \smallsetminus N_k$ is a disjoint union of the same number of closed topological disks. There are now two possibilities.

Case 1. If each N_k is bounded by one simple closed curve, then $\widehat{\mathbb{C}} \smallsetminus N_k$ is connected, and $\widehat{\mathbb{C}} \smallsetminus \mathcal{A}_0$, being the intersection of a nested sequence of connected sets, is itself connected.

Case 2. Otherwise, there is a unique smallest integer m such that N_m has more than one boundary component. Without loss of generality, we may assume that $m = 1$. (Simply discard all N_i with $i < m - 1$ and then renumber each N_{m-1+j} as N_j.) Thus we can assume that ∂N_0 is connected but that ∂N_1 is bounded by a collection of simple closed curves $\Gamma_1, \ldots, \Gamma_n$ with $n \geq 2$. Each Γ_i is the boundary of a corresponding component D_i of $\widehat{\mathbb{C}} \smallsetminus N_1$. *Now for each finite sequence $(i_1, \ldots, i_k)$ of numbers between 1 and n, we will show inductively that the set N_k has at least one boundary component $\Gamma_{i_1 \cdots i_k}$ which is contained in D_{i_1} and which has image $f(\Gamma_{i_1 \cdots i_k}) = \Gamma_{i_2 \cdots i_k}$.* (If there is more than one such component, simply choose one of them.)

To prove this statement, note first that each $\overline{N}_k$ is a branched covering of $\overline{N}_{k-1}$ under the map f. (Compare Appendix E.) It follows that each connected component of $\overline{N}_k \smallsetminus f^{-1}(N_0)$ is a branched covering of $\overline{N}_{k-1} \smallsetminus N_0$. It is not hard to see that each such component is contained in just one of the disks D_i, and that each D_i must contain at least one such component. Therefore it follows that each of the curves $\Gamma_{i_2 \cdots i_k}$ in ∂N_{k-1} is covered by at least one curve $\Gamma_{i_1 i_2 \cdots i_k} \subset D_{i_1} \cap \partial N_k$, as required. The corresponding components $D_{i_1 \cdots i_k}$ of $\widehat{\mathbb{C}} \smallsetminus N_k$ are disjoint, and it is easy to check that

$$D_{i_1} \supset D_{i_1 i_2} \supset D_{i_1 i_2 i_3} \supset \cdots .$$

Thus $\widehat{\mathbb{C}} \smallsetminus \mathcal{A}_0$ has one complementary component for each infinite sequence of numbers between 1 and n, and hence has uncountably many such components. $\square$

Repelling Points. For most purposes we can simply define a "repelling" fixed point to be one with multiplier satisfying $|\lambda| > 1$. However, it is more satisfying to have a topologically invariant characterization.

Definition. A fixed point $\hat{p} = f(\hat{p})$ of a continuous map will be called *topologically repelling* if there is a neighborhood U of $\hat{p}$ so that for every $p \neq \hat{p}$ in U there exists some $n \geq 1$ so that the nth forward image $f^{\circ n}(p)$ lies outside of U. In other words, the only infinite orbit $p_0 \mapsto p_1 \mapsto p_2 \mapsto \cdots$ which is completely contained in U must be the orbit of the fixed point itself. Such a U is called a *forward isolating neighborhood* of $\hat{p}$.

Now suppose that U is an open subset of a Riemann surface and that f is holomorphic.

Lemma 8.10 (Topologically Repelling Points). *A fixed point of such a holomorphic map is topologically repelling if and only if its multiplier satisfies* $|\lambda| > 1$.

Proof. If $|\lambda| > 1$, then it follows from Theorem 8.2 (or from a much more elementary exercise in calculus) that the point is topologically repelling. I am indebted to S. Zakeri for the following proof of the converse statement. If $\hat{p}$ is a topologically repelling point for f, note first that $\lambda \neq 0$ (and in fact $|\lambda| \geq 1$), since $\hat{p}$ clearly cannot be both attracting and repelling. Thus we can choose a compact forward isolating neighborhood N, which is small enough so that f maps N homeomorphically onto some compact neighborhood $f(N)$ of $\hat{p}$. Let

$$N_k = N \cap f^{-1}(N) \cap \cdots \cap f^{-k}(N)$$

be the compact neighborhood consisting of points for which the first k forward images all belong to N. Thus $N = N_0 \supset N_1 \supset N_2 \supset \cdots$, with intersection the single point $\hat{p}$ since N is an isolating neighborhood. By compactness, it follows that the diameter of N_k tends to zero as $k \to \infty$. But it follows immediately from the construction that

$$f(N_k) = N_{k-1} \cap f(N),$$

where $N_{k-1} \subset f(N)$ for k large since the diameters tend to zero. Thus $f(N_k) = N_{k-1}$ for k large; in fact f maps N_k homeomorphically onto N_{k-1}. Now let U_k be the connected component of the interior of N_k which contains $\hat{p}$. Then it follows that f^{-1} maps U_{k-1} biholomorphically onto the strictly smaller set U_k. By the Schwarz Lemma, its multiplier must satisfy $|\lambda^{-1}| < 1$, which proves that $|\lambda| > 1$ as required. □

Remark. Lemmas 8.1 and 8.10 work only over the complex numbers. Over the real numbers, examples such as $f(x) = x \pm x^3$ show that a fixed point with multiplier $\lambda = 1$ may perfectly well be topologically attracting or topologically repelling.

The Kœnigs Linearization Theorem 8.2, in the repelling case, helps us to understand why the Julia set $J(f)$ is so often a complicated "fractal" set.

Corollary 8.11. *Suppose that the rational function f has a 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 $\hat{\mathbb{C}}$.*

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 $z_0 \leftarrow\!\shortmid z_1 \leftarrow\!\shortmid z_2 \leftarrow\!\shortmid \ldots$ with Kœnigs coordinates $\phi(z_n) = w_0/\lambda^n$ which lie along a logarithmic spiral and converge to zero. Evidently such a set can not lie in any smooth 1-dimensional real submanifold of $\mathbb{C}$. □

In fact, if we recall that the iterated preimages of our periodic point are everywhere dense in $J(f)$, then we see that such sequences lying on logarithmic spirals are extremely pervasive. Compare Figure 12, which shows two examples of such spiral structures, associated with repelling points of periods 1 and 2, respectively. (For more on fractals, see, for example, Falconer [1990].)

The global form of the Linearization Theorem in the repelling case, due to Poincaré, is rather different from the statement in the geometrically attracting case (Corollary 8.4). In particular, there is no analogous concept of a "repelling basin," and no analogous extension to a map from the entire basin $\mathcal{A} \subset S$ to $\mathbb{C}$. However, we can extend ϕ^{-1} to a map $\mathbb{C} \to S$.

Corollary 8.12 (Global Extension of ϕ^{-1}). *If $\hat{p}$ is a repelling fixed point for the holomorphic map $f : S \to S$, then there is a holomorphic map $\psi : \mathbb{C} \to S$, with $\psi(0) = \hat{p}$, so that the diagram*

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

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

Proof. For ϵ sufficiently small, let $\psi_\epsilon : \mathbb{D}_\epsilon \to S$ be that branch of ϕ^{-1} which maps zero to $\hat{p}$. Now to compute $\psi(w)$ for any $w \in \mathbb{C}$, we simply choose n large enough so that $w/\lambda^n \in \mathbb{D}_\epsilon$, and then set $\psi(w) = f^{\circ n}(\psi_\epsilon(w/\lambda^n))$. Details will be left to the reader. □

Concluding Problems

Problem 8-a. The identification torus of a fixed point. Suppose that f has a geometrically attracting or repelling fixed point $\hat{p}$ with multiplier λ. Let U be any neighborhood of $\hat{p}$ small enough so that f maps U biholomorphically, with $f(U) \subset U$ in the attracting case or $f(U) \supset U$ in the repelling case. Form an identification space $T = (U \smallsetminus \{\hat{p}\})/f$ by identifying p with $f(p)$ whenever both belong to U. **(1)** Show that T is a Riemann surface, independent of the choice of U, with the topology of a torus. **(2)** Show in fact that T is conformally isomorphic to the quotient $\mathbb{C}/\Lambda$, where Λ is the lattice $2\pi i\mathbb{Z} \oplus (\log\lambda)\mathbb{Z}$.

Problem 8-b. Global linearization. Let $f : S \to S$ be a holomorphic map from a Riemann surface S to itself. **(1)** Show that $p_0 \in \mathcal{A}$ is a critical point of ϕ if and only if the orbit $f : p_0 \mapsto p_1 \mapsto p_2 \mapsto \cdots$ contains some critical point of f. **(2)** If f is onto, show that the linearizing map ϕ of Corollary 8.4 maps the attracting basin $\mathcal{A}$ onto $\mathbb{C}$.

Problem 8-c. Asymptotic values. In order to extend Theorem 8.6 to a noncompact Riemann surface such as $\mathbb{C}$ or $\mathbb{C} \smallsetminus \{0\}$, we need some definitions. Let $f : S \to S'$ be a holomorphic map between Riemann surfaces. A point $v \in S'$ is a *critical value* if it is the image under f of a *critical point*, that is, a point at which the first derivative of f vanishes. It is an *asymptotic value* if there exists a continuous path $[0, 1) \to S$ which

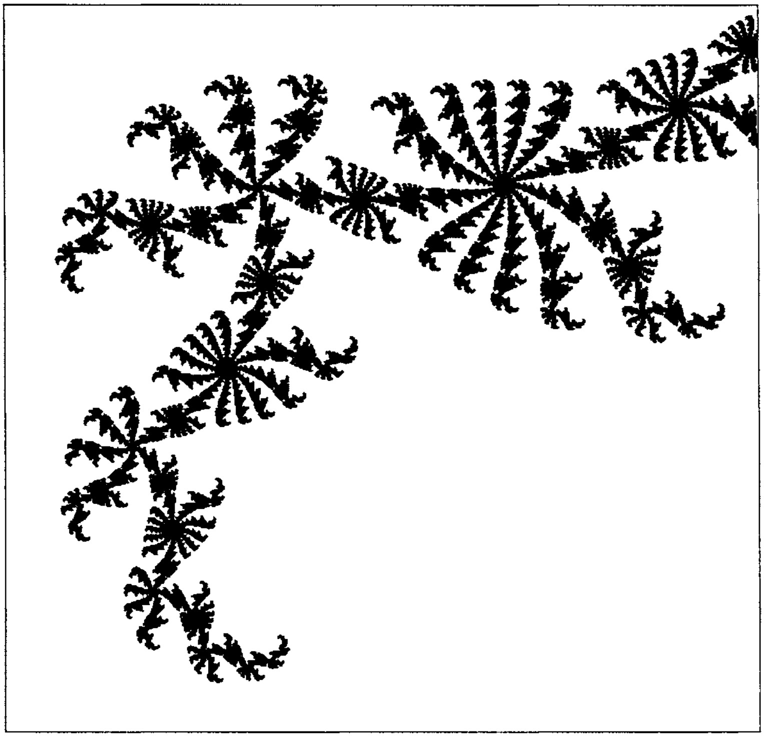


Figure 12a. Detail of Julia set for $z \mapsto z^2 + .424513 + .207530i$. *Compare Corollary 8.11.*

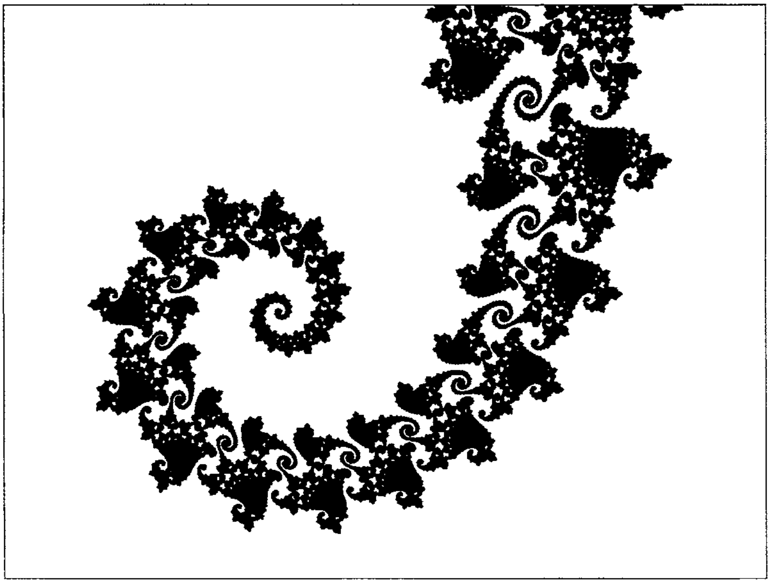


Figure 12b. Detail of Julia set for $z \mapsto z^2 - .744336 + .121198i$.

diverges from S in the sense that it eventually leaves any compact subset of S, but whose image under f converges to the point v. Recall from §2 that a connected open set $U \subset S'$ is *evenly covered* if every component of $f^{-1}(U)$ maps homeomorphically onto U and that f is a *covering map* if every point of S' has a neighborhood which is evenly covered.

(1) Show that a simply connected open subset of S' is evenly covered by f if and only if it contains no critical value or asymptotic value. (Compare Goldberg and Keen [1986].) In particular, f is a covering map if and only if S' contains no critical values and no asymptotic values.

(2) For a holomorphic self-map $f : S \to S$, show that the immediate basin of any attracting periodic orbit must contain either a critical value or an asymptotic value or both, except in the special case of a linear map from $\mathbb{C}$ or $\widehat{\mathbb{C}}$ to itself. As an example, for any $c \neq 0$, show that the transcendental map $f(z) = c\,e^z$ from $\mathbb{C}$ to itself has no critical points and just one asymptotic value, namely $z = 0$. Conclude that it has at most one periodic attractor. If $|c| < 1/e$ show that f maps the unit disk into itself and that f has an attracting fixed point in this disk.

(3) The map f is *proper* if the preimage $f^{-1}(K)$ of every compact set $K \subset S'$ is a compact subset of S. (If S is compact, then every map on S is proper.) Show that a proper map has no asymptotic values.

Problem 8-d. Topological attraction and repulsion. Suppose that X is a locally compact space and that f maps a compact neighborhood N of $\hat{x}$ homeomorphically onto a compact neighborhood N', with $f(\hat{x}) = \hat{x}$. Show that f is topologically repelling at $\hat{x}$ if and only if f^{-1} is topologically attracting at $\hat{x}$. (Here the hypothesis that f is locally one-to-one is essential. For example, the map $f(z) = z^2$ is attracting at the origin, and the nonsmooth map $g(z) = 2z^2/|z|$ is repelling at the origin, although neither one has a local inverse.)

Problem 8-e. The image $\psi(\mathbb{C}) \subset S$. If $\hat{p}$ is a repelling point for the holomorphic map $f : S \to S$, show that the image of the map $\psi : \mathbb{C} \to S$ of Corollary 8.12 is everywhere dense and in fact that the complement $S \smallsetminus \psi(\mathbb{C})$ consists of grand orbit finite points. (Compare Theorem 4.10. There are at most two such points when $S = \widehat{\mathbb{C}}$, at most one when $S = \mathbb{C}$, and none for other nonhyperbolic surfaces.)

Problem 8-f. Counting basin components. Let $\mathcal{A}$ be the attracting basin of a periodic point which may be either superattracting or geometrically attracting. **(1)** If some connected component of $\mathcal{A}$ is not periodic, show that $\mathcal{A}$ has infinitely many components. **(2)** Suppose then that $\mathcal{A}$ has only finitely many components forming a periodic cycle. If these components are simply connected, use the Riemann-Hurwitz Formula 7.2 to show that the period is at most 2. (Example: $f(z) = 1/z^2$.) **(3)** If they are infinitely connected, show that the period must be 1.

Problem 8-g. Critical points in the basin. Give another proof of Theorem 8.6 as follows. **(1)** Suppose there were an attracting periodic orbit $\mathcal{O}$ with no critical point in its immediate basin. If U is a small neighborhood of a point $p \in \mathcal{O}$, show that for each $k \geq 1$ there would be a unique branch $g_k : U \to \widehat{\mathbb{C}}$ of $f^{-k}|_U$ which maps p into $\mathcal{O}$. **(2)** Show then that the family $\{g_k\}$ would have to be normal, which is impossible since the first derivative of g_k at p must be unbounded.

§9. Böttcher's Theorem and Polynomial Dynamics

This section studies the case of a *superattracting* fixed point, with multiplier λ equal to zero. As usual, we can choose a local uniformizing parameter z with fixed point $z = 0$. Thus our map takes the form

$$f(z) \;=\; a_n z^n \,+\, a_{n+1} z^{n+1} \,+\, \cdots\,, \tag{9:1}$$

with $n \ge 2$ and $a_n \ne 0$, where the integer n is called the *local degree*.

> **Theorem 9.1 (Böttcher [1904]).*** *With f as above, there exists a local holomorphic change of coordinate $w = \phi(z)$, with $\phi(0) = 0$, which conjugates f to the nth power map $w \mapsto w^n$ throughout some neighborhood of zero. 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 $n \ge 2$. This theorem has important applications to polynomial dynamics, since any polynomial map $\mathbb{C} \to \mathbb{C}$ of degree $d \ge 2$ extends to a rational map of $\hat{\mathbb{C}}$ which has a superattracting fixed point at infinity with local degree $n = d$. (Compare Theorem 9.5.)

Proof of Existence. The proof will be quite similar to the proof of Theorem 8.2. With f as in (9 : 1), let us first choose a solution c to the equation $c^{n-1} = a_n$. Then the linearly conjugate map $cf(z/c)$ will have leading coefficient equal to $+1$. Thus we may assume, without loss of generality, that our map has the form $f(z) = z^n(1+b_1z+b_2z^2+b_3z^3+\cdots)$, or briefly

$$f(z) \;=\; z^n\,(1+\eta(z)) \qquad \text{with} \qquad \eta(z) \;=\; b_1 z + b_2 z^2 + \cdots\,. \tag{9:2}$$

Choose a radius $0 < r < 1/2$ which is small enough so that $|\eta(z)| < 1/2$ on the disk $\mathbb{D}_r$ of radius r. Then clearly f maps this disk into itself, with $|f(z)| \le \frac{3}{2}|z|^n \le \frac{3}{4}|z|$ and with $f(z) \ne 0$ for $z \in \mathbb{D}_r \smallsetminus \{0\}$. The k-fold iterate $f^{\circ k}$ also maps $\mathbb{D}_r$ into itself, and we see inductively that it has the form $f^{\circ k}(z) = z^{n^k}(1 + n^{k-1}b_1\, z + \text{(higher terms)})$. The idea of the proof is

*L. E. Böttcher was born in Warsaw in 1878. He took his doctorate in Leipzig in 1898, working in Iteration Theory, and then moved to Lvov, where he retired in 1935. He published in Polish and Russian. (The Russian form of his name is Бётхер.)

to set

$$\phi_k(z) = \sqrt[n^k]{f^{\circ k}(z)} = z\,(1 + n^{k-1} b_1\, z + \cdots)^{1/n^k} = z\Big(1 + \frac{b_1}{n}\, z + \cdots\Big),$$

where the expression on the right provides an explicit choice of n^kth root. Clearly $\phi_k(f(z)) = \phi_{k+1}(z)^n$. We will show that the functions ϕ_k converge uniformly to a limit function $\phi : \mathbb{D}_r \to \mathbb{D}$. In order to prove convergence, let us make the substitution $z = e^Z$, where Z ranges over the left half-plane $\mathbb{H}_r$ defined by the inequality $\mathrm{Re}(Z) < \log r$. Then the map f from the disk $\mathbb{D}_r$ into itself corresponds to a map $F(Z) = \log f(e^Z)$ from $\mathbb{H}_r$ into itself. Setting $\eta = \eta(e^Z) = b_1 e^Z + b_2 e^{2Z} + \cdots$ as in (9 : 2), with $|\eta| < 1/2$, we see that this can be written more precisely as

$$\begin{aligned} F(Z) &= \log\big(e^{nZ}(1+\eta)\big) = nZ + \log(1+\eta) \\ &= nZ + (\eta - \eta^2/2 + \eta^3/3 - + \cdots), \end{aligned}$$

where now the final expression provides an explicit choice as to which branch of the logarithm function we are using.

Evidently $F : \mathbb{H}_r \to \mathbb{H}_r$ is a well-defined holomorphic function. Since $|\eta| < 1/2$, we have

$$|F(Z) - nZ| = |\log(1+\eta)| < \log 2 < 1 \qquad (9:3)$$

for all Z in this half-plane. Similarly, the map $\phi_k(z) = f^{\circ k}(z)^{1/n^k}$ for $|z| < r$ corresponds to a map

$$\Phi_k(Z) = \log \phi_k(e^Z) = F^{\circ k}(Z)/n^k,$$

which is defined and holomorphic throughout $\mathbb{H}_r$. By (9 : 3), we have

$$|\Phi_{k+1}(Z) - \Phi_k(Z)| = |F^{\circ k+1}(Z) - n\, F^{\circ k}(Z)|/n^{k+1} < 1/n^{k+1}.$$

Since the exponential map from the left half-plane onto $\mathbb{D}$ reduces distances, it follows that

$$|\phi_{k+1}(z) - \phi_k(z)| < 1/n^{k+1}$$

for $|z| < r$. Therefore, as $k \to \infty$ the sequence of holomorphic functions $z \mapsto \phi_k(z)$ on the disk $|z| < r$ converges uniformly to a holomorphic limit $\phi(z)$. It is easy to check that ϕ satisfies the required identity $\phi(f(z)) = \phi(z)^n$.

Proof of Uniqueness. It suffices to study the special case $f(z) = z^n$. If a map of the form $\phi(z) = c_1\, z + c_k\, z^k +$(higher terms) conjugates $z \mapsto z^n$ to itself, then the series

$$\phi(z^n) = c_1\, z^n + c_k\, z^{nk} + \cdots$$

must be equal to

$$\phi(z)^n \;=\; c_1^n\, z^n + nc_1^{n-1} c_k\, z^{n+k-1} + \cdots ,$$

with $nk > n + k - 1$. Comparing coefficients, we find that $c_1^{n-1} = 1$ and that all higher coefficients are zero. □

Remark. Given a global holomorphic map $f : S \to S$ with a superattracting point $\hat{p}$, we can choose a local uniformizing parameter $z = z(p)$ with $z(\hat{p}) = 0$ and construct the Böttcher coordinate $w = \phi(z(p))$ as above. *To simplify the notation, we will henceforth forget the intermediate parameter z and simply write $w = \phi(p)$.*

In analogy with Corollary 8.4, one might hope that the local mapping $p \mapsto \phi(p)$ could be extended throughout the entire basin of attraction of $\hat{p}$ as a holomorphic mapping $\mathcal{A} \to \mathbb{D}$. However, this is not always possible. Such an extension would involve computing expressions of the form

$$p \mapsto \sqrt[n]{\phi(f^{\circ n}(p))},$$

and this may not work, since the nth root cannot always 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. However, if we consider only the absolute value of ϕ, then there is no problem.

Corollary 9.2 (Extension of $|\phi|$). *If $f : S \to S$ has a superattracting fixed point $\hat{p}$ with basin $\mathcal{A}$, then the function $p \mapsto |\phi(p)|$ of Theorem 9.1 extends uniquely to a continuous map $|\phi| : \mathcal{A} \to [0, 1)$ which satisfies the identity $|\phi(f(p))| = |\phi(p)|^n$. Furthermore, $|\phi|$ is real analytic, except at the iterated preimages of $\hat{p}$ where it takes the value zero.*

Proof. Set $|\phi(p)|$ equal to $|\phi(f^{\circ k}(p))|^{1/n^k}$ for large k. □

Now let $f : \widehat{\mathbb{C}} \to \widehat{\mathbb{C}}$ be a rational function with a superattracting fixed point $\hat{p}$. Then the associated Böttcher map ϕ, which carries a neighborhood of $\hat{p}$ biholomorphically onto a neighborhood of zero, has a local inverse ψ_ϵ mapping the ϵ-disk around zero to a neighborhood of $\hat{p}$. Using an argument similar to the proof of Lemma 8.5, we have the following statement.

Theorem 9.3 (Critical Points in the Basin). *There exists a unique open disk $\mathbb{D}_r$ of maximal radius $0 < r \le 1$ such that ψ_ϵ extends holomorphically to a map ψ from the disk $\mathbb{D}_r$ into the immediate basin $\mathcal{A}_0$ of $\hat{p}$. If $r = 1$, then ψ maps the unit disk $\mathbb{D}_1$ biholomorphically onto $\mathcal{A}_0$ and $\hat{p}$ is the only critical* *point in this basin. On the other hand, if* $r < 1$ *then there is at least one other critical point in* $\mathcal{A}_0$, *lying on the boundary of* $\psi(\mathbb{D}_r)$.

Compare Figure 13, which illustrates the case $r = 1$ and Figure 14, which illustrates the case $r < 1$.

Proof of Theorem 9.3. As in the proof of Lemma 8.5, we can extend ψ_ϵ by analytic continuation either to the unit disk $\mathbb{D}$ of radius $r = 1$ or to some maximal disk $\mathbb{D}_r$ of radius $r < 1$. In either case, we will show that the extended map ψ carries $\mathbb{D}_r$ biholomorphically onto its image $U = \psi(\mathbb{D}_r) \subset \mathcal{A}_0$. First note that ψ_r has no critical points, for if there were a point $w \in \mathbb{D}_r$ with $\psi'(w) = 0$ then certainly $w \neq 0$. Hence the equation

$$\psi(w^n) \;=\; f(\psi(w)) \tag{9:4}$$

would imply that w^n is also a critical point. This would yield a sequence of critical points $w, w^n, w^{n^2}, \ldots$ tending to zero, which is impossible. Thus ψ is locally one-to-one, and the set of all pairs $w_1 \neq w_2$ with $\psi(w_1) = \psi(w_2)$ forms a closed subset of $\mathbb{D}_r \times \mathbb{D}_r$.

We must show that ψ is actually one-to-one on $\mathbb{D}_r$. The proof will be based on the observation that the map $|\phi|$ of Corollary 9.2 satisfies the identity $|\phi(\psi(w))| = |w|$ for w close to zero, and hence for all $w \in \mathbb{D}_r$ by analytic continuation. Suppose that $\psi(w_1) = \psi(w_2)$ with $w_1 \neq w_2$. Applying the map $|\phi|$ to both sides, it follows that $|w_1| = |w_2|$. Choose such a pair with $|w_1| = |w_2|$ minimal. Since ψ is an open mapping, if we choose any w_1' sufficiently close to w_1 then we can find w_2' close to w_2 with $\psi(w_1') = \psi(w_2')$. Taking $|w_1'| < |w_1|$, this yields a contradiction.

In the case $r = 1$, the image $U = \psi(\mathbb{D})$ must be the entire immediate basin $\mathcal{A}_0$. For otherwise U would have some boundary point $z_0 \in \mathcal{A}_0$. Approximating z_0 by points $\psi(w_j)$, we must see that the sequence of numbers $|\phi(\psi(w_j))| = |w_j|$ must converge to 1. But this implies that $\psi(z_0) = 1$, which is impossible.

Now suppose that $r < 1$. Then the proof that $\partial U \subset \mathcal{A}_0$, and that there exists a critical point in ∂U is completely analogous to the corresponding argument in Lemma 8.5. Details will be left to the reader. □

Caution: Examining Figures 13 and 14 and in analogy with Lemma 8.5, one might expect that ϕ always extends to a homeomorphism between the closure $\overline{U}$ and the closed disk $\overline{\mathbb{D}}_r$; however, this is false. Compare Figure 15 for the case $r = 1$ and Figure 16 for the case $r < 1$. (In both cases $\hat{p}$ is the point at infinity.)

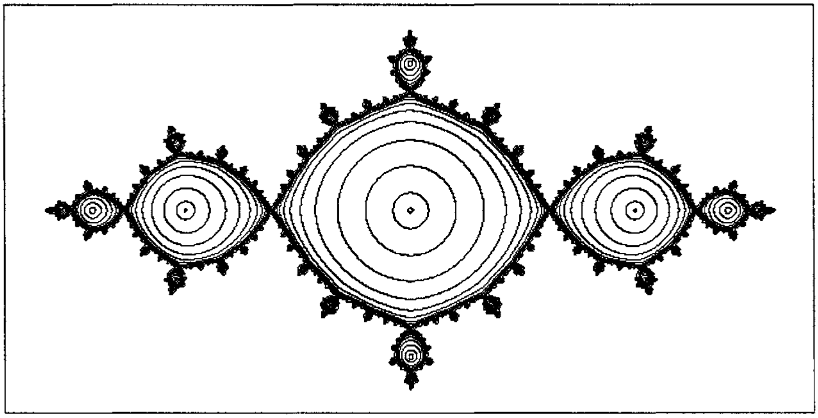


Figure 13. Julia set for $f(z) = z^2 - 1$. The fourth degree map $f \circ f$ has two superattracting fixed points at $z = 0$ and $z = -1$, with no other critical points in the immediate basins. The grand orbit of a representative curve $|\phi| = \text{constant}$ has been drawn in for both attracting basins. Note that each such curve in an immediate basin $\mathcal{A}_0$ maps to the next smaller curve in $\mathcal{A}_0$ by a twofold covering.

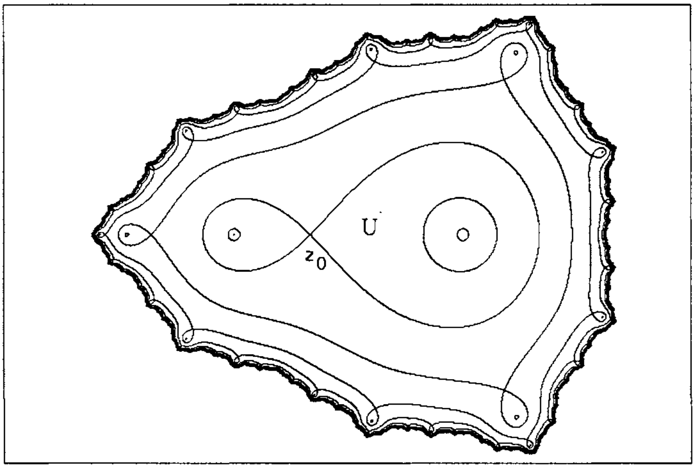


Figure 14. Julia set for the map $f(z) = z^3 + z^2$, which has a critical point $z_0 = -2/3$ in the immediate basin of the superattracting point $z = 0$. The grand orbit of the curve $|\phi| = \text{constant}$ through z_0 has been drawn in.

Application to Polynomial Dynamics. Let

$$f(z) = a_d z^d + a_{d-1} z^{d-1} + \cdots + a_1 z + a_0$$

be a polynomial of degree $d \geq 2$. We always assume that the leading coefficient a_d is nonzero. In fact, it is often convenient to assume that f is *monic*, with $a_d = 1$. This does not involve any loss of generality: Since $a_d \neq 0$ and $d \geq 2$, we can always choose c with $c^{d-1} = a_d$ and note that the linearly conjugate polynomial $c\,f(z/c)$ is monic.

We will see that such a map f has a superattracting fixed point at infinity, so that we can apply Böttcher's Theorem. But first a more elementary construction.

By definition, the *filled Julia set* $K = K(f)$ is the set of all $z \in \mathbb{C}$ for which the orbit of z under f is bounded.

Lemma 9.4 (The Filled Julia Set). *For any polynomial f of degree at least 2, this filled Julia set $K \subset \mathbb{C}$ is compact, with connected complement, with topological boundary ∂K equal to the Julia set $J = J(f)$ and with interior equal to the union of all bounded components U of the Fatou set $\mathbb{C} \smallsetminus J$. Thus K is equal to the union of all such U, together with J itself. Any such bounded component U is necessarily simply connected.*

(More generally, a compact subset of $\mathbb{C}$ is called *full* if its complement is connected. Any compact subset of $\mathbb{C}$ can be *filled* by adjoining all bounded components of its complement.)

Proof of Lemma 9.4. Evidently the ratio $f(z)/z^d$ converges to the limit a_d as $|z| \to \infty$. Assuming for convenience that $a_d = 1$, we can choose a constant $r_0 \geq 2$ so that $|f(z)/z^d - 1| < 1/2$ for $|z| > r_0$, and it follows that

$$|f(z)| > |z^d|/2 > 2|z| \qquad \text{for} \qquad |z| > r_0.$$

It follows that any z with $|z| > r_0$ belongs to the attracting basin $\mathcal{A} = \mathcal{A}(\infty)$ of the point at infinity. Evidently K can be identified with the complement $\widehat{\mathbb{C}} \smallsetminus \mathcal{A}$. Hence K is compact, and it follows from Corollary 4.12 that $\partial K = \partial \mathcal{A}$ is equal to the Julia set of f.

We must show that $\mathcal{A}$ is connected. Let U be any bounded component of the Fatou set $\mathbb{C} \smallsetminus J$. Then $|f^{\circ d}(z)| \leq r_0$ for every $z \in U$ and every $n \geq 0$, for otherwise, by the Maximum Modulus Principle, there would be some $\hat{z} \in \partial U \subset J$ with $|f^{\circ d}(\hat{z})| > r_0$. But this would imply that $\hat{z} \in \mathcal{A}$, which is impossible. Thus every bounded component of $\mathbb{C} \smallsetminus J$ is contained in the filled Julia set K, and the unique unbounded component can be identified with $\mathbb{C} \smallsetminus K = \mathbb{C} \cap \mathcal{A}(\infty)$.

Similarly, if Γ is a simple closed curve lying in a bounded component U, and if V is the bounded component of $\mathbb{C} \smallsetminus \Gamma$, then it follows from the Maximum Modulus Principle that $V \subset K$. In particular, V cannot contain any points of $J = \partial K$, so it follows that $V \subset U$. This proves that U is simply connected. □

To better understand this filled Julia set K, we consider the dichotomy of Theorem 9.3 for the complementary domain $\mathcal{A}(\infty) = \widehat{\mathbb{C}} \smallsetminus K$. This yields the following.

Theorem 9.5 (Connected $K \Leftrightarrow$ Bounded Critical Orbits). *Let f be a polynomial of degree $d \geq 2$. If the filled Julia set $K = K(f)$ contains all of the finite critical points of f, then both K and $J = \partial K$ are connected and the complement of K is conformally isomorphic to the exterior of the closed unit disk $\overline{\mathbb{D}}$ under an isomorphism*

$$\widehat{\phi} : \mathbb{C} \smallsetminus K \longrightarrow \mathbb{C} \smallsetminus \overline{\mathbb{D}},$$

which conjugates f on $\mathbb{C} \smallsetminus K$ to the dth power map $w \mapsto w^d$. On the other hand, if at least one critical point of f belongs to $\mathbb{C} \smallsetminus K$, then both K and J have uncountably many connected components.

Compare Figure 15, which illustrates the first possibility, and Figure 16, which illustrates the second. The proof of Theorem 9.5 will be based on the following. To study the behavior of f near infinity, we make the usual substitution $\zeta = 1/z$ and consider the rational function

$$F(\zeta) = \frac{1}{f(1/\zeta)}.$$

Again we may assume that f is monic. From the asymptotic equality $f(z) \sim z^d$ as $z \to \infty$, it follows that $F(\zeta) \sim \zeta^d$ as $\zeta \to 0$. Thus F has a superattracting fixed point at $\zeta = 0$. (More explicitly, it is not difficult to derive a power series expansion of the form

$$F(\zeta) = \zeta^d - a_{d-1}\zeta^{d+1} + (a_{d-1}^2 - a_{d-2})\zeta^{d+2} + \cdots$$

for $|\zeta|$ small.) There is an associated Böttcher map

$$\phi(\zeta) = \lim_{k\to\infty} F^{\circ k}(\zeta)^{1/d^k} \in \mathbb{D},$$

which is defined and biholomorphic for $|\zeta|$ sufficiently small, with $\phi'(0) = 1$ since f is assumed to be monic. In practice, it is more convenient to work

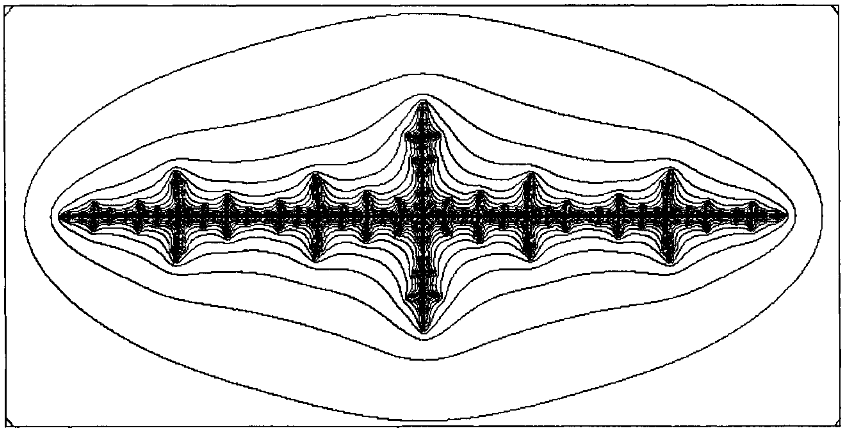


Figure 15. Julia set for $f(z) = z^2 - 3/2$*, showing an equipotential*

$$G = \log|\widehat{\phi}| = \text{constant}$$

and its iterated forward and backward images. (Definition 9.6.) Each such curve maps to the next larger curve by a twofold covering.

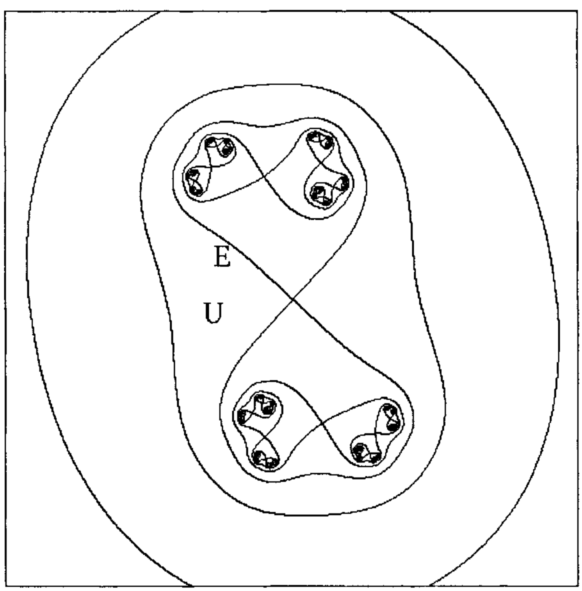


Figure 16. The Julia set for $f(z) = z^2 + (1 + i/2)$ *is totally disconnected (a Cantor set). The neighborhood of infinity* $U = \widehat{\psi}(\mathbb{C} \setminus \overline{\mathbb{D}}_r)$ *is the complement of the region bounded by the figure eight equipotential curve* E *through the critical point* $z = 0$*. Note that the two components of* $f^{-1}(E)$ *are also figure eight curves, as are the four components of* $f^{-2}(E)$*, and so on. However, the iterated forward images of* E *are all smooth topological circles.*

with the reciprocal

$$\widehat{\phi}(z) \;=\; \frac{1}{\phi(1/z)} \;=\; \lim_{k\to\infty} f^{\circ k}(z)^{1/d^k} \;\in\; \mathbb{C}\smallsetminus\overline{\mathbb{D}}.$$

Thus $\widehat{\phi}$ maps some neighborhood of infinity biholomorphically onto a neighborhood of infinity, with $\widehat{\phi}(z)\sim z$ as $|z|\to\infty$, and $\widehat{\phi}$ conjugates the degree d polynomial map f to the dth power map, so that

$$\widehat{\phi}(f(z)) \;=\; \widehat{\phi}(z)^d. \qquad (9:5)$$

Proof of Theorem 9.5. (Compare Problem 9-e.) Suppose first that there are no critical points other than ∞ in the attracting basin $\mathcal{A}=\mathcal{A}(\infty)$. Then by Theorem 9.3 the Böttcher map extends to a conformal isomorphism $\mathcal{A}\xrightarrow{\cong}\mathbb{D}$. It follows that the function $\widehat{\phi}$ extends to a conformal isomorphism $\mathbb{C}\smallsetminus K\xrightarrow{\cong}\mathbb{C}\smallsetminus\overline{\mathbb{D}}$. Now each annulus

$$\mathbb{A}_{1+\epsilon} \;=\; \{z\in\mathbb{C}\;;\;1<|z|<1+\epsilon\}$$

maps under $\widehat{\psi}=\widehat{\phi}^{-1}$ to a connected set $\widehat{\psi}(\mathbb{A}_{1+\epsilon})\subset\mathbb{C}\smallsetminus K$. The closure $\overline{\widehat{\psi}(\mathbb{A}_{1+\epsilon})}$ is a compact connected set which evidently contains the Julia set $J=\partial\mathcal{A}$. It follows that the intersection

$$J \;=\; \bigcap_{\epsilon>0}\overline{\widehat{\psi}(\mathbb{A}_{1+\epsilon})}$$

is also connected, and it then follows easily from Lemma 9.4 that K is connected.

Now suppose that there is at least one critical point in $\mathbb{C}\smallsetminus K$. Then the conclusion of Theorem 9.3 translates as follows: *There is a smallest number $r>1$ so that the inverse of $\widehat{\phi}$ near infinity extends to a conformal isomorphism*

$$\widehat{\psi}\;:\;\mathbb{C}\smallsetminus\overline{\mathbb{D}}_r\;\xrightarrow{\cong}\;U\;\subset\;\mathbb{C}\smallsetminus K.$$

Furthermore the boundary ∂U of this open set $U=\widehat{\psi}(\mathbb{C}\smallsetminus\overline{\mathbb{D}}_r)$ is a compact subset of $\mathbb{C}\smallsetminus K$ which contains at least one critical point of f.

We will show that the closure $\overline{U}$ separates the plane into two or more bounded open sets, each of which contains uncountably many points of the Julia set. Let c be a critical point in ∂U. Then the corresponding *critical value* $v=f(c)$ clearly belongs to U, with $|\widehat{\phi}(v)|=r^d>r$. Consider the infinite *ray* $R\subset\mathbb{C}\smallsetminus\mathbb{D}_r$ consisting of all products $t\widehat{\phi}(v)$ with $t\geq 1$. The image $R'=\widehat{\psi}(R)\subset U$ is called an *external ray* to the point v, associated with the compact set $K\subset\mathbb{C}$.

Now consider the full inverse image $f^{-1}(R')\subset\overline{U}$. Clearly the intersec-

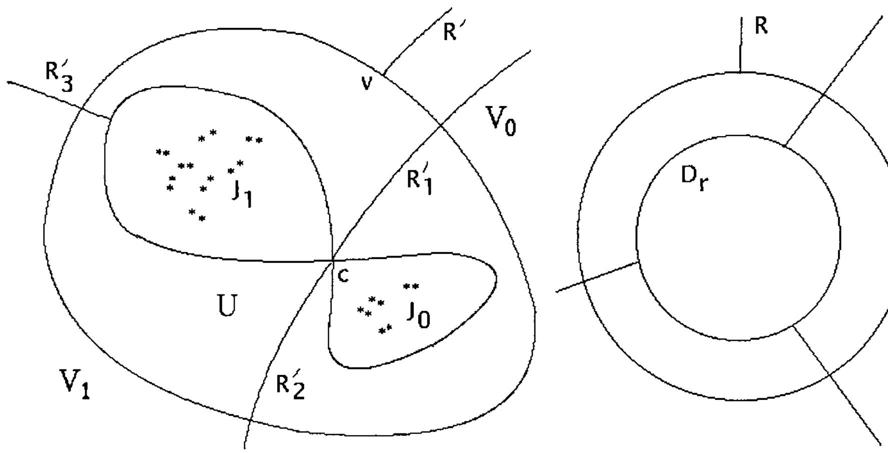


Figure 17. Sketch illustrating the proof of Theorem 9.5 in degree $d = 3$, with the z-plane on the left and the $w = \widehat{\phi}(z)$-plane on the right. The open set $U = \widehat{\psi}(\mathbb{C} \smallsetminus \overline{\mathbb{D}}_r)$ is the exterior of the region bounded by the figure eight through the critical point.

tion $U \cap f^{-1}(R')$ consists of d distinct external rays, corresponding to the d distinct components of the set $\sqrt[d]{R} \subset \mathbb{C} \smallsetminus \mathbb{D}_r$. Each of these d external rays R'_j will end at some solution z to the equation $f(z) = v$. But this equation has at least a double solution at the critical point c, so at least two of these external rays, say R'_1 and R'_2, will land at c. Evidently the union $R'_1 \cup R'_2 \subset \overline{U}$ will cut the plane into two connected open sets, which we will call V_0 and V_1.

Next note that each of the images $f(V_0)$ and $f(V_1)$ contains all points of the complex plane, except possibly for the points of R'. In fact each $f(V_k)$ is an open set. If $\hat{z} \in \mathbb{C}$ is a boundary point of $f(V_k)$, then we can choose a sequence of points $z_j \in V_k$ so that the images $f(z_j)$ converge to $\hat{z}$. The z_j must certainly be bounded, so we can choose a subsequence which converges to some point $z' \in \mathbb{C}$. Now $z' \notin V_k$ since $f(z') = \hat{z}$ is a boundary point and f is an open map, so it follows that $z' \in \partial V_k = R'_1 \cup R'_2$ and hence $\hat{z} \in R'$. Since $\mathbb{C} \smallsetminus R'$ is connected, this implies that

$$f(V_k) \supset \mathbb{C} \smallsetminus R' \supset K.$$

Now let $J_0 = J \cap V_0$ and $J_1 = J \cap V_1$. Then it follows that

$$f(J_0) = f(J_1) = J.$$

Note that J_0 and J_1 are disjoint compact sets with $J_0 \cup J_1 = J$. Similarly, we can split each J_k into two disjoint compact subsets $J_{k0} = J_k \cap f^{-1}(J_0)$

and $J_{k1} = J_k \cap f^{-1}(J_1)$, with $f(J_{k\ell}) = J_\ell$. Continuing inductively, we split J into 2^{p+1} disjoint compact sets

$$J_{k_0 \cdots k_p} = J_{k_0} \cap f^{-1}(J_{k_1}) \cap \cdots \cap f^{-p}(J_{k_p}),$$

with $f(J_{k_0 \cdots k_p}) = J_{k_1 \cdots k_p}$. Similarly, for any infinite sequence $k_0\, k_1\, k_2 \ldots$ of zeros and ones, let $J_{k_0 k_1 k_2 \cdots}$ be the intersection of the nested sequence

$$J_{k_0} \supset J_{k_0 k_1} \supset J_{k_0 k_1 k_2} \supset \cdots .$$

Each such intersection is compact and nonvacuous. In this way, we obtain uncountably many disjoint nonvacuous subsets with union J. Every connected component of J must be contained in exactly one of these, so J has uncountably many components. The proof for the filled Julia set is completely analogous. □

The Green's Function of a Polynomial Map. As in Corollary 9.2, the function $z \mapsto |\widehat{\phi}(z)|$ extends continuously throughout the attracting basin $\mathbb{C} \setminus K$, taking values $|\widehat{\phi}(z)| > 1$. (This function is finite valued, since a polynomial has no poles in the finite plane.) In practice it is customary to work with the logarithm of $|\widehat{\phi}|$.

Definition 9.6. By the *Green's function* or the *canonical potential function* associated with the filled Julia set K of the monic degree d polynomial f we mean the function $G : \mathbb{C} \to [0, \infty)$ which is identically zero on K and takes the values

$$G(z) = \log |\widehat{\phi}(z)| = \lim_{k \to \infty} \frac{1}{d^k} \log |f^{\circ k}(z)| > 0$$

outside of K. It is not difficult to check that G is continuous everywhere and harmonic, that is

$$G_{xx} + G_{yy} = 0,$$

outside of the Julia set. (See Problems 9-b and 9-c. Here the subscripts denote partial derivatives, with $z = x + iy$.) The curves $G = \text{constant} > 0$ in $\mathbb{C} \setminus K$ are known as *equipotentials*. Note the equation

$$G(f(z)) = G(z)\, d,$$

which shows that f maps each equipotential to an equipotential.

Concluding Problems

Problem 9-a. Grand orbit closures. Let f be a rational function, and let $\mathcal{A}$ be the attracting basin of some attracting fixed point p. **(1)** Show that the grand orbit $\mathrm{GO}(p)$ is a discrete subset of $\mathcal{A}$. If p

is not a grand orbit finite point, show that its set of accumulation points within $\widehat{\mathbb{C}}$ is equal to the Julia set J. **(2)** In the geometrically attracting case, if $z_0 \in \mathcal{A}$ but $z_0 \notin \mathrm{GO}(p)$, show that $\mathrm{GO}(z_0)$ is a discrete subset of $\mathcal{A}$ and that its set of accumulation points is equal to $J \cup \mathrm{GO}(p)$. **(3)** If p is superattracting of local degree n and if $z_0 \in \mathcal{A} \smallsetminus \mathrm{GO}(p)$, show that the closure of $\mathrm{GO}(z_0)$ consists of all points z such that $|\phi(z)|$ is equal to some power $|\phi(z_0)|^{n^k}$, together with $J \cup \mathrm{GO}(p)$. (Compare Figures 13 through 16, pages 94 and 97.)

Problem 9-b. Harmonic functions. **(1)** If U is a simply connected open set of complex numbers $z = x + iy$, show that a smooth function $G : U \to \mathbb{R}$ is *harmonic*, $G_{xx} + G_{yy} = 0$, if and only if there is another smooth function $H : U \to \mathbb{R}$, uniquely defined up to an additive constant, satisfying

$$H_x = -G_y, \quad H_y = G_x.$$

It then follows that H is harmonic and that $G + iH$ is holomorphic. Any such H is called a *harmonic conjugate* of G. **(2)** On the other hand, for the non-simply connected set $U = \mathbb{C} \smallsetminus \{0\}$, show that the function $z \mapsto \log|z|$ is harmonic but has no globally defined harmonic conjugate. **(3)** For an arbitrary Riemann surface S, show that there is a corresponding concept of harmonic function from S to $\mathbb{R}$ which is independent of any choice of local uniformizing parameters. **(4)** Show that a harmonic function cannot have any local maximum or minimum, unless it is constant. (Use the Maximum Modulus Principle for holomorphic functions.) In particular, show that every harmonic function on a compact surface is constant. Similarly, if the harmonic function G on an open surface has the property that $\{p\ ;\ |G(p)| \ge \epsilon\}$ is compact for every $\epsilon > 0$, show that G is identically zero.

Problem 9-c. Green's function. Consider a monic polynomial f of degree $d \ge 2$. **(1)** Show that the Green's function $G(z) = \log|\widehat{\phi}(z)|$ is harmonic on $\mathbb{C} \smallsetminus K$, that it tends to zero as z approaches K, and that it satisfies

$$G(z) = \log|z| + o(1) \qquad \text{as} \qquad |z| \to \infty.$$

(In other words, $G(z) - \log|z|$ tends to zero as $|z| \to \infty$.) **(2)** Show that the function G is uniquely characterized by these properties. Hence G is completely determined by the compact set $K = K(f)$, although our construction of G depends explicitly on the polynomial f.

Problem 9-d. The punctured disk. Let $G : \mathbb{D} \smallsetminus \{0\} \to \mathbb{R}$ be a harmonic function on the punctured disk. **(1)** Show that there exists a harmonic conjugate $H : \mathbb{D} \smallsetminus \{0\} \to \mathbb{R}$ if and only if the integral

$$b = \frac{1}{2\pi} \oint (G_x dy - G_y dx)$$

around the origin is zero. (Compare Problem 9-b.) **(2)** Let $\widehat{G}(z)$ be the average of $G(e^{i\theta} z)$ for $0 \le \theta \le 2\pi$. Show that this is a harmonic function of the form

$$\widehat{G}(z) = a + b \log |z|$$

where a and b are constants, with b as above. **(3)** If G is bounded, conclude that $b = 0$, so that G is the real part of a holomorphic function $f(z) = G(z) + iH(z)$ from $\mathbb{D} \smallsetminus \{0\}$ to $\mathbb{C}$ and so that $e^{f(z)}$ is a *bounded* holomorphic function on $\mathbb{D} \smallsetminus \{0\}$. Using Problem 2-$\ell$, conclude that G extends to a harmonic function which is defined and smooth throughout the disk $\mathbb{D}$.

Problem 9-e. Cellular sets and Riemann-Hurwitz. Here is another approach to Theorem 9.5. Again let f be a polynomial of degree $d \ge 2$. For each number $g > 0$ let V_g be the bounded open set consisting of all complex numbers z with $G(z) < g$. Using the maximum modulus principle, show that each connected component of V_g is simply connected. Hence the Euler characteristic $\chi(V_g)$ can be identified with the number of connected components of V_g. Show similarly that each component of V_g intersects the filled Julia set.

The Riemann-Hurwitz formula (Theorem 7.2) applied to the map $f : V_g \to V_{gd}$ asserts that $d\chi(V_{gd}) - \chi(V_g)$ is equal to the number of critical points of f in V_g, counted with multiplicity. Since V_g is clearly connected for g sufficiently large, conclude that V_g is connected if and only if it contains all of the $n-1$ critical points of f.

A compact subset of Euclidean n-space is said to be *cellular** if it is a nested intersection of closed topological n-cells, each containing the next in its interior. Show that the filled Julia set $K = \bigcap V_g$ is cellular (and hence connected) if and only if it contains all of the $n-1$ finite critical points of f. (In fact, if one of these critical points lies outside of K, and hence outside of some V_g, show that V_g and hence K are not connected.)

Problem 9-f. Quadratic polynomials. Now let $f(z) = z^2 + c$ and suppose that the critical orbit escapes to infinity. Let $V = V_{G(c)}$ be

*Compare Brown [1960], where it is shown that a subset K of the sphere S^n is cellular if and only if its complement $S^n \smallsetminus K$ is an open n-cell. This concept is of more interest in higher dimensions. In fact it is not hard to see that a compact subset of $\mathbb{C}$ is cellular if and only if it is connected with connected complement.

the open set consisting of all $z \in \mathbb{C}$ with $|\widehat{\phi}(z)| < |\widehat{\phi}(c)|$. Show that V is conformally isomorphic to $\mathbb{D}$ and that $f^{-1}(V)$ has two connected components. Conclude that $f^{-1}|_V$ has two holomorphic branches g_0 and g_1 mapping V into disjoint open subsets, each having compact closure in V. Show that each g_j strictly contracts the Poincaré metric of V. Proceeding as in Problem 4-e, show that J is a Cantor set, canonically homeomorphic to the space of all infinite sequences of zeros and ones. Show that the map f from J to itself corresponds to the *shift map*

$$(j_0, j_1, j_2, \ldots) \mapsto (j_1, j_2, j_3, \ldots)$$

from this space of sequences to itself.

§10. 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. (Compare Lemma 4.7.) First consider the special case $\lambda = 1$. Then we can write our map as

$$\begin{aligned} f(z) &= z + a\, z^{n+1} + \text{(higher terms)} \\ &= z\,(1 + a\, z^n + \text{(higher terms)}), \end{aligned} \tag{10 : 1}$$

with $n \geq 1$ and $a \neq 0$. The integer $n+1$ is called the *multiplicity* of the fixed point. (Compare Lemma 12.1.) We are concerned here with fixed points of multiplicity $n + 1 \geq 2$. (Thus, for the moment, we exclude *simple* fixed points, that is, those with multiplicity equal to 1, or equivalently those with multiplier $\lambda \neq 1$, so that the graph of f intersects the diagonal transversally.)

Definition. A complex number $\mathbf{v}$ will be called a *repulsion vector* for f at the origin if the product $na\mathbf{v}^n$ is equal to $+1$, and an *attraction vector* if $na\mathbf{v}^n = -1$. Here we use a boldface letter, and use the term "vector," to indicate that $\mathbf{v}$ should be thought of intuitively as a tangent vector to $\mathbb{C}$ at the origin (for example, as the tangent vector to the curve $t \mapsto t\mathbf{v}$ at $t = 0$). Thus there are n equally spaced attraction vectors at the origin, separated by n equally spaced repulsion vectors. I will number these distinguished vectors as $\mathbf{v}_0, \mathbf{v}_1, \ldots, \mathbf{v}_{2n-1}$, where $\mathbf{v}_0$ is repelling and where

$$\mathbf{v}_j = e^{\pi i j/n}\mathbf{v}_0 \qquad \text{so that} \qquad na\mathbf{v}_j^n = (-1)^j.$$

Thus $\mathbf{v}_j$ is attracting or repelling according to whether j is odd or even. Note that the inverse map f^{-1} is also well-defined and holomorphic in some neighborhood of the origin, and that the repulsion vectors for f are just the attraction vectors for f^{-1}.

Here is a preliminary description of the local dynamics. Consider some orbit $z_0 \mapsto z_1 \mapsto \cdots$ for the map f of equation $(10:1)$. We will say that this orbit converges to zero *nontrivially* if $z_k \to 0$ as $k \to \infty$, but no z_k is actually equal to zero.

Lemma 10.1. *If an orbit $f : z_0 \mapsto z_1 \mapsto \cdots$ converges to zero nontrivially, then z_k is asymptotic to $\mathbf{v}_j / \sqrt[n]{k}$ as $k \to +\infty$*

for one of the n attraction vectors $\mathbf{v}_j$. In other words, the limit $\lim_{k\to\infty} \sqrt[n]{k}\, z_k$ exists and is equal to one of the $\mathbf{v}_j$ with j odd. Similarly, if an orbit $f^{-1} : z'_0 \mapsto z'_1 \mapsto \cdots$ under f^{-1} converges to zero nontrivially, then z'_k is asymptotic to $\mathbf{v}_j / \sqrt[n]{k}$, where $\mathbf{v}_j$ is now one of the n repulsion vectors, with j even. Any attraction or repulsion vector can occur.

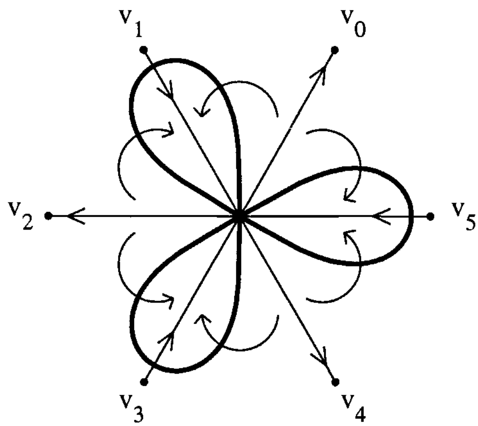


Figure 18. Schematic picture of a parabolic point of multiplicity $n+1 = 4$. (Here $a = -1$.) Each arrow indicates roughly how points are moved by f. The three attraction vectors are indicated by arrows pointing towards the origin, and the three repulsion vectors by arrows pointing away from the origin.

Definition. If an orbit $z_0 \mapsto z_1 \mapsto \cdots$ under f converges to zero, with $z_k \sim \mathbf{v}_j / \sqrt[n]{k}$ (where j is necessarily odd), then we will say that this orbit $\{z_k\}$ tends to zero *from the direction* $\mathbf{v}_j$.

Remark 10.2. The array of attraction-repulsion vectors at a fixed point transforms naturally under a holomorphic change of coordinate. (This array can be thought of as a geometric representation for the leading terms of the power series for f at the fixed point.) More generally, consider a Riemann surface S and a map $p \mapsto f(p) \in S$ which is defined and holomorphic in the neighborhood of a fixed point $\hat{p}$ of multiplicity $n + 1 \geq 2$. Then there is a corresponding uniquely defined array of attraction-repulsion vectors in the tangent space at $\hat{p}$ with completely analogous properties. Details will be left to the reader.

Proof of Lemma 10.1. The proof will be based on the substitution

$$w = \varphi(z) = c/z^n$$

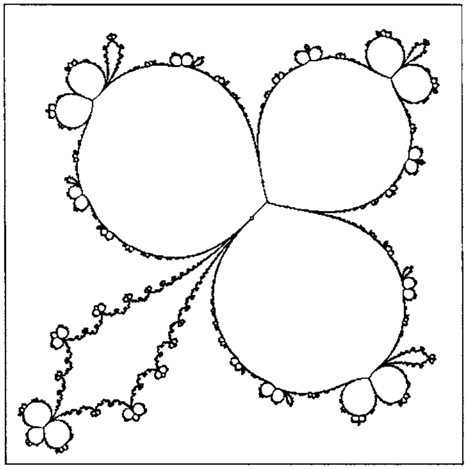


Figure 19. Julia set for $f(z) = z^5+(.8+.8i)z^4+z$. This map has a parabolic fixed point of rotation number zero and petal number three at $z=0$ (and also an attracting fixed point at $z=-.8-.8i$). The immediate basins for the three attraction vectors resemble balloons, pulled together at the parabolic point and separated by the three repulsion vectors.

where $c = -1/(na)$, and particularly on the real part of w

$$\mathrm{Re}(w) \;=\; \mathrm{Re}\,\varphi(z) = \mathrm{Re}(c/z^n).$$

Note that in the special case of an attraction or repulsion vector we have

$$\varphi(\mathbf{v}_j) \;=\; \mathrm{Re}\,\varphi(\mathbf{v}_j) \;=\; (-1)^{j+1}.$$

However, we are primarily interested in behavior when $|z|$ is very small, or in other words when $|w|$ is very large.

Let $\mathbb{R}_+ = [0,\infty)$ be the positive real axis and let $\mathbb{R}_- = (-\infty\,,0]$ be the negative real axis. The half-line $\mathbb{R}_+\mathbf{v}_j$ will be called either a repelling ray or an attracting ray according to whether j is even or odd. In order to label the various branches of the many-valued function $\varphi^{-1}(w) = \sqrt[n]{c/w}$ let us cover the punctured plane somewhat redundantly by $2n$ open sectors with angle $2\pi/n$, bounded either by two consecutive repelling rays or by two consecutive attracting rays. More explicitly, for each attraction or repulsion vector $\mathbf{v}_j$ let Δ_j be the corresponding open sector consisting of all $re^{i\theta}\mathbf{v}_j$ with $r>0$ and $|\theta| < \pi/n$. (Figure 20.) Then φ maps Δ_j biholomorphically onto a slit plane, with

$$\varphi(\Delta_j) \;=\; \begin{cases} \mathbb{C}\smallsetminus\mathbb{R}_+ & \text{for } j \text{ even (so that } \mathbf{v}_j \text{ is repelling),} \\ \mathbb{C}\smallsetminus\mathbb{R}_- & \text{for } j \text{ odd (so that } \mathbf{v}_j \text{ is attracting).} \end{cases}$$

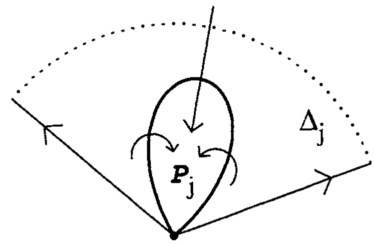


Figure 20. Sector Δ_j with j odd, $n = 3$, and an enclosed attracting petal $\mathcal{P}_j$.

Hence there is a uniquely defined branch ψ_j of φ^{-1} with

$$\psi_j \,:\, \mathbb{C} \smallsetminus \mathbb{R}_{(-1)^j} \xrightarrow{\cong} \Delta_j \,.$$

Note that each $\Delta_j \cap \Delta_{j+1}$ is a sector of angle π/n bounded by the rays $\mathbb{R}_+\mathbf{v}_j$ and $\mathbb{R}_+\mathbf{v}_{j+1}$. The image $\varphi(\Delta_j \cap \Delta_{j+1})$ is the upper half-plane if j is even or the lower half-plane if j is odd.

We can write

$$f(z) \;=\; z\,(1 + az^n + o(z^n)) \quad \text{as} \quad z \to 0,$$

where the notation $o(z^n)$ stands for a remainder term, depending on z, which tends to zero faster than z^n so that $o(z^n)/z^n \to 0$ as $z \to 0$. To understand the behavior of this map for z close to zero in the sector Δ_j, we look at at the corresponding transformation

$$w \;\mapsto\; F_j(w) \;=\; \varphi \circ f \circ \psi_j(w)$$

which is defined outside a large disk in the slit w-plane and which takes values in the full w-plane. Note that

$$f \circ \psi_j(w) \;=\; \sqrt[n]{c/w}\left(1 + a\,\frac{c}{w} + o\Big(\frac{1}{w}\Big)\right) \quad \text{as} \quad |w| \to \infty.$$

Composing with the function $\varphi(z) = c/z^n$, we obtain

$$F_j(w) \;=\; w\left(1 + a\,\frac{c}{w} + o\Big(\frac{1}{w}\Big)\right)^{-n} \;=\; w\left(1 + \frac{-nac}{w} + o\Big(\frac{1}{w}\Big)\right).$$

Since $nac = -1$, this can be written briefly as

$$F_j(w) \;=\; w + 1 + o(1) \qquad \text{as} \qquad |w| \to \infty. \tag{10:2}$$

(With just a little more work, we obtain the more precise statement that

$$F_j(w) \;=\; w + 1 + O\Big(1/\sqrt[n]{|w|}\Big) \quad \text{as} \quad |w| \to \infty. \tag{10:3}$$

In other words, the remainder term not only tends to zero, but has absolute value bounded by some constant times $1/\sqrt[n]{|w|}$ for large $|w|$.)

It will be convenient to choose a number $R > 0$ so that

$$|F(w) - w - 1| < 1/2 \qquad \text{whenever} \qquad |w| > R. \tag{10:4}$$

In particular, it follows easily from (10 : 4) that

$$\mathrm{Re}(F_j(w)) > \mathrm{Re}(w) + 1/2$$

whenever $|w| > R$, and hence that the related function $\mathrm{Re}\,\varphi(z) = \mathrm{Re}(c/z^n)$ satisfies

$$\mathrm{Re}\,\varphi(f(z)) > \mathrm{Re}\,\varphi(z) + 1/2 \tag{10:5}$$

whenever $|z|$ is sufficiently small.

Remark 10.3. As an immediate consequence of (10 : 5), we see the following. *There are no small cycles near a parabolic fixed point.* The fixed point itself is the only periodic orbit contained in a small neighborhood.

Another consequence of (10 : 4) is that the slope of the line segment from w to $F_j(w)$ is bounded. In particular,

$$\left|\mathrm{Im}\Big(F_j(w) - w\Big)\right| < \mathrm{Re}\Big(F_j(w) - w\Big) \quad \text{when} \quad |w| > R \tag{10:6}$$

(or more precisely the slope is bounded by $\sqrt{3}/3$). The proof is easily supplied.

Choosing R sufficiently large, as in (10 : 4), let $\mathbb{H}_R$ be the right half-plane consisting of all w with $\mathrm{Re}(w) > R$, and let $\mathcal{P}_j(R)$ be its image $\psi_j(\mathbb{H}_R)$, consisting of all points $z \in \Delta_j$ with $\mathrm{Re}\,\varphi(z) > R$. Then clearly F_j maps $\mathbb{H}_R$ into itself, and it is not hard to check that f maps the image $\mathcal{P}_j(R)$ into itself. Furthermore, the successive iterates of f restricted to $\mathcal{P}_j(R)$ converge uniformly to the constant map $\mathcal{P}_j(R) \to 0$. We will refer to this set $\mathcal{P}_j(R)$ as an ***attracting petal***. (Compare Definition 10.6.)

Now consider any orbit $z_0 \mapsto z_1 \mapsto \cdots$ under f which converges to zero nontrivially. Then the inequality $\mathrm{Re}\,\varphi(z_{k+1}) > \mathrm{Re}\,\varphi(z_k) + 1/2$ will be satisfied whenever k is sufficiently large. In particular, it follows that there exists an m such that $\mathrm{Re}\,\varphi(z_m) > R$. This z_m must belong to one of the n attracting petals $\mathcal{P}_j(R) \subset \Delta_j$. Since $f(\mathcal{P}_j(R)) \subset \mathcal{P}_j(R)$, this implies that z_k belongs to this same petal $\mathcal{P}_j(R)$ for all $k \geq m$.

Next consider the sequence $w_0 \mapsto w_1 \mapsto \cdots$ where $w_k = \varphi(z_k) \in \mathbb{C}$. Then $w_k \in \mathbb{H}_{\mathbb{R}}$ and $w_{k+1} = F_j(w_k)$ for $k \geq m$. Since $\mathrm{Re}(w_k) \to \infty$ hence $|w_k| \to \infty$, it follows from (10 : 2) that the difference $w_{k+1} - w_k$ converges to $+1$ as $k \to \infty$. Therefore the average

$$\frac{w_k - w_0}{k} = \frac{1}{k}\sum_0^{k-1}(w_{h+1} - w_h)$$

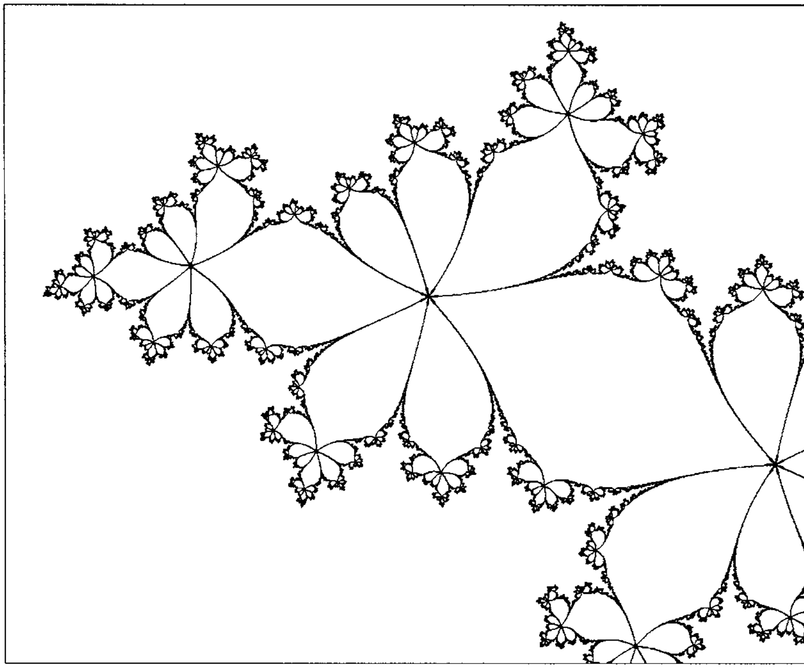


Figure 21. Julia set for $z \mapsto z^2 + e^{2\pi i t} z$ *with* $t = 3/7$.

also converges to $+1$. This implies that the ratio w_k/k converges to $+1$, or in other words $w_k \sim k$ as $k \to \infty$. Since $1/w_k = -naz_k^n$, it follows that naz_k^n is asymptotic to $-1/k$, and since $na\mathbf{v}_j^n = -1$ we can write this as $z_k^n \sim \mathbf{v}_j^n/k$. Now extracting the n-th root, since z_k is known to be in the petal $\mathcal{P}_j$, it follows that $z_k \sim \mathbf{v}_k/\sqrt[n]{k}$, thus proving Lemma 10.1. $\square$

Now suppose that the multiplier λ at a fixed point is a qth root of unity, say $\lambda = \exp(2\pi i p/q)$, where p/q is a fraction in lowest terms.

Lemma 10.4. *If the multiplier λ at a fixed point $f(\hat{z}) = \hat{z}$ is a primitive qth root of unity, then the number n of attraction vectors at $\hat{z}$ must be a multiple of q. In other words, the multiplicity $n+1$ of $\hat{z}$ as a fixed point of $f^{\circ q}$ must be congruent to 1 modulo q.*

As an example, Figure 21 shows part of the Julia set for a quadratic map f having a fixed point of multiplier $\lambda = e^{2\pi i(3/7)}$ at the origin, near the center of the picture. In this case, the sevenfold iterate $f^{\circ 7}$ is a map of degree 128 with a fixed point of multiplicity $7+1=8$ at the origin. The seven immediate attracting basins are clearly visible in the figure.

Proof of Lemma 10.4. If $\mathbf{v}$ is any attraction vector for $f^{\circ q}$ at $\hat{z}$, then we can choose an orbit $z_0 \mapsto z_q \mapsto z_{2q} \cdots$ under $f^{\circ q}$ which converges to $\hat{z}$ from the direction $\mathbf{v}$. Evidently the image $z_1 \mapsto z_{q+1} \mapsto z_{2q+1} \mapsto \cdots$ under f will be an orbit which converges to $\hat{z}$ from the direction $\lambda\mathbf{v}$. Thus multiplication by $\lambda = e^{2\pi ip/q}$ permutes the n attraction vectors, and the conclusion follows easily. □

Remark. If we replace $f = f_0$ by a nearby map f_t, so as to change λ slightly, then the $(n+1)$-fold fixed point $\hat{z}$ of $f^{\circ q}$ will split up into $n+1$ simple fixed points of $f_t^{\circ q}$. Since $\hat{z}$ is a simple fixed point of $f^{\circ k}$ for $k < q$, it follows that only one of these $n+1$ points will be fixed by f_t or by any $f_t^{\circ k}$ with $0 < k < q$. The remaining n will partition into n/q orbits, each of period exactly q.

Definition. Now consider a holomorphic map f from a Riemann surface S to itself with a fixed point $\hat{p}$ of multiplier $+1$. (Compare Remark 10.2.) Given an attraction vector $\mathbf{v}_j$ in the tangent space of S at $\hat{p}$, the associated ***parabolic basin of attraction*** $\mathcal{A}_j = \mathcal{A}(\hat{p}, \mathbf{v}_j)$ is defined to be the set consisting of all $p_0 \in S$ for which the orbit $p_0 \mapsto p_1 \mapsto \cdots$ converges to $\hat{p}$ from the direction $\mathbf{v}_j$. Evidently these basins $\mathcal{A}_1, \ldots, \mathcal{A}_n$ are disjoint fully invariant open sets, with the property that an orbit $p_0 \mapsto p_1 \mapsto \cdots$ under f converges to $\hat{p}$ nontrivially if and only if it belongs to one of the $\mathcal{A}_j$. The ***immediate basin*** $\mathcal{A}_j^0$ is defined to be the unique connected component of $\mathcal{A}_j$ which maps into itself under f. Equivalently, $\mathcal{A}_j$ can be described as that connected component of the Fatou set $S \smallsetminus J$ which contains p_k for large k whenever $\{p_k\}$ converges to $\hat{p}$ from the direction $\mathbf{v}_j$.

More generally, if $\hat{p}$ is a periodic point of period k with multiplier $\lambda = e^{2\pi ip/q}$ for the map $f : S \to S$, then $\hat{p}$ is a fixed point of multiplier $+1$ for the iterate $f^{\circ kq}$. By definition, the parabolic basins for $f^{\circ kq}$ at $\hat{p}$ are also called parabolic basins for f.

> **Lemma 10.5.** *For a holomorphic map $f : S \to S$, each parabolic basin $\mathcal{A}_j$ is contained in the Fatou set $S \smallsetminus J(f)$, but each basin boundary $\partial\mathcal{A}_j$ is contained in the Julia set $J(f)$.*

Proof. It suffices to consider the special case of a fixed point of multiplier $+1$. It is clear that $\mathcal{A}_j$ is contained in the Fatou set, and we already know by Lemma 4.7 that the fixed point $\hat{p}$ itself belongs to the Julia set. If an orbit $p_0 \mapsto p_1 \mapsto \cdots$ eventually lands at $\hat{p}$ or in other words converges ***trivially*** to $\hat{p}$, then it follows that p_0 also belongs to the Julia set. Thus it suffices to consider a point $p_0 \in \partial\mathcal{A}_j$ whose orbit does not contain $\hat{p}$. Since p_0 is not in any of the attractive basins $\mathcal{A}_j$, the orbit $p_0 \mapsto p_1 \mapsto \cdots$ does not converge to $\hat{p}$ either trivially or nontrivially. Hence we can extract a subsequence $p_{k(i)}$ which is bounded away from $\hat{p}$. Since the sequence of iterates $f^{\circ k}$ converges to $\hat{p}$ throughout the open set $\mathcal{A}_j$, it follows that $\{f^{\circ k}\}$ cannot be normal in any neighborhood of the boundary point p_0. □

It is often convenient to have a purely local analog for the global concept of "basin of attraction."

Definition 10.6. Let $\hat{p} \in S$ be a fixed point of multiplicity $n+1 \geq 2$ for a map f which is defined and univalent on some neighborhood $N \subset S$, and let $\mathbf{v}_j$ be an attraction vector at $\hat{p}$. An open set $\mathcal{P} \subset N$ will be called an ***attracting petal*** for f for the vector $\mathbf{v}_j$ at $\hat{p}$ if

(1) f maps $\mathcal{P}$ into itself, and

(2) an orbit $p_0 \mapsto p_1 \mapsto \cdots$ under f is eventually absorbed by $\mathcal{P}$ if and only if it converges to $\hat{p}$ from the direction $\mathbf{v}_j$.

Similarly, if $f : N \xrightarrow{\cong} N'$, then an open subset $\mathcal{P} \subset N'$ will be called a ***repelling petal*** for the repulsion vector $\mathbf{v}_k$ if $\mathcal{P}$ is an attracting petal for the map $f^{-1} : N' \to N$ and for this vector $\mathbf{v}_k$.

Caution: There doesn't seem to be any standard definition for the concept of petal. Different authors impose different restrictions. The present definition is very flexible. For example it has the property that any intersection of petals for $\mathbf{v}_j$ is itself a petal. Furthermore, any neighborhood N on which f is univalent contains a unique maximal petal for $\mathbf{v}_j$, namely the union of all forward orbits in N which converge to $\hat{p}$ from this direction. It does allow rather wild sets as petals, but we can always impose further restrictions. The petals $\mathcal{P}_j(R)$ described in the proof of Lemma 10.1 are particularly well behaved, being simply connected, with smooth boundary except at the point $\hat{p}$ itself, and with

$$\bigcap f^{\circ k}\left(\overline{\mathcal{P}_j(R)}\right) = \{\hat{p}\}$$

where we take the intersection over $k > 0$ or $k < 0$ according as the petal is attracting or repelling. (Compare Figure 18, page 105.) However, we will need petals which are a bit fatter.

The following result was proved in a preliminary form by Léopold Leau* [1897], and in increasingly satisfactory forms by Julia [1918] and Fatou [1919–1920].

*In the non-mathematical world, Leau is best known for his efforts to establish an international auxiliary language. He helped to organize an International Delegation in 1907 which advocated an improved variant of Esperanto. Unfortunately, this angered advocates of the existing form of Esperanto and led to a great deal of controversy.

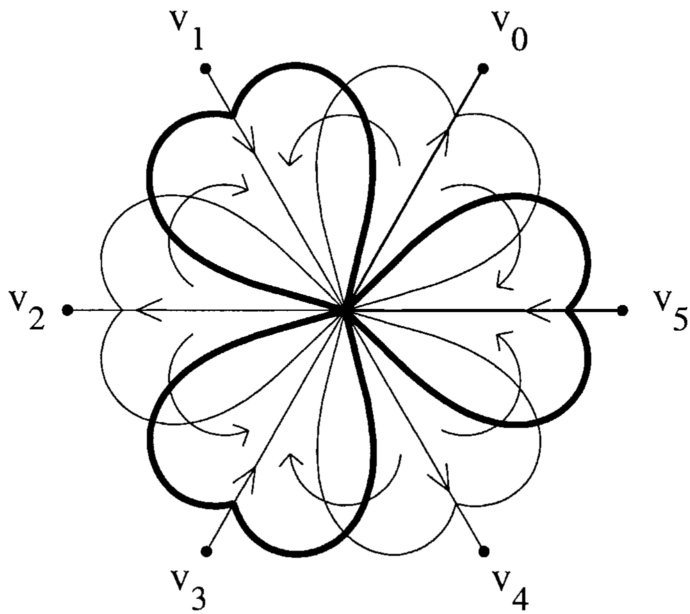


Figure 22. Flower with three attracting petals (emphasized) and three repelling petals.

Theorem 10.7 (Parabolic Flower Theorem). *If $\hat{z}$ is a fixed point of multiplicity $n+1 \geq 2$, then within any neighborhood of $\hat{z}$ there exist simply connected petals $\mathcal{P}_j$, where the subscript j ranges over the integers modulo $2n$ and where $\mathcal{P}_j$ is either repelling or attracting according to whether j is even or odd. Furthermore, these petals can be chosen so that the union*

$$\{\hat{z}\} \cup \mathcal{P}_0 \cup \cdots \cup \mathcal{P}_{2n-1}$$

is an open neighborhood of $\hat{z}$. When $n > 1$, each $\mathcal{P}_j$ intersects each of its two immediate neighbors in a simply connected region $\mathcal{P}_j \cap \mathcal{P}_{j\pm1}$ but is disjoint from the remaining $\mathcal{P}_k$.

Compare Figure 22. The situation when $n = 1$ is only slightly different. The left neighbor of a petal is then the same as its right neighbor, so that the intersection $\mathcal{P}_0 \cap \mathcal{P}_1$ actually has *two* simply connected components.

Proof of Theorem 10.7. (Compare Buff and Epstein [2002].) As in the proof of Lemma 10.1, we may assume that $\hat{z} = 0$. Choose a large number R as in (10 : 4) and (10 : 6), and let $W_R \supset \mathbb{H}_{2R}$ be the set of all $w = u + iv \in \mathbb{C}$ with $u + |v| > 2R$. Then for each odd j it is easy to check that F_j maps W_R into itself, and that the image $\mathcal{P}_j = \psi_j(W_R) \subset \Delta_j$ is an attracting petal. Similarly, let $-W_R$ be the set of all $-w$ with

$w \in W_R$. Then, for each even j, making corresponding estimates for F_j^{-1}, we see that F_j^{-1} maps $-W_R$ into itself and that $\mathcal{P}_j = \psi_j(-W_R)$ is a repelling petal. (These petals are heart shaped, as shown in Figure 22.)

The intersection $W_R \cap (-W_R)$ is a disjoint union $V_R^+ \cup V_R^-$ where V_R^+ is the V-shaped region consisting of all $u + iv$ in the upper half-plane with $v > |u| + 2R$ and where V_R^- is its reflection in the lower half-plane. The image $\psi_j(\mathcal{P}_j \cap \mathcal{P}_{j+1})$ is either W_V^+ or V_R^- according as j is even or odd. Further details will be left to the reader. □

If $f : S \to S$ is a globally defined holomorphic function and $\hat{z}$ is a fixed point of multiplicity $n + 1 \geq 2$, then each attracting petal $\mathcal{P}_j$ about $\hat{z}$ determines a corresponding parabolic basin of attraction $\mathcal{A}_j$, consisting of all z_0 for which the orbit $z_0 \mapsto z_1 \mapsto \cdots$ eventually lands in $\mathcal{P}_j$, and hence converges to the fixed point from the associated direction $\mathbf{v}_j$.

We can further describe the geometry around a parabolic fixed point as follows. As in (10 : 1), consider a local analytic map with a fixed point of multiplier $\lambda = 1$. Let $\mathcal{P}$ be either an attracting petal or a repelling petal. Form an identification space $\mathcal{P}/f$ from $\mathcal{P}$ by identifying z with $f(z)$ whenever both z and $f(z)$ belong to $\mathcal{P}$. (This means that z is identified with $f(z)$ for every $z \in \mathcal{P}$ in the case of an attracting petal and for every $z \in \mathcal{P} \cap f^{-1}(\mathcal{P})$ in the case of a repelling petal.)

Evidently this quotient manifold $\mathcal{P}/f$ depends only on the choice of attracting or repulsion vector and not on the particular choice of $\mathcal{P}$, for if $\mathcal{P}'$ is another petal for the same vector, then $\mathcal{P} \cap \mathcal{P}'$ is also a petal, and the inclusions $\mathcal{P} \cap \mathcal{P}' \subset \mathcal{P}$ and $\mathcal{P} \cap \mathcal{P}' \subset \mathcal{P}$ induce the required conformal isomorphisms.

Theorem 10.8 (Cylinder Theorem). *For each attracting or repelling petal $\mathcal{P}$, the quotient manifold $\mathcal{P}/f$ is conformally isomorphic to the infinite cylinder $\mathbb{C}/\mathbb{Z}$.*

By definition, the quotient $\mathcal{P}/f$ is called an *Écalle cylinder* for $\mathcal{P}$. This term is due to Douady, suggested by the work of Écalle on holomorphic maps tangent to the identity. (Compare Écalle [1975]. The behavior of Écalle cylinders under perturbation of the mapping f is a very important topic in holomorphic dynamics. See for example Lavaurs [1989] and Shishikura [1998].)

We will derive Theorem 10.8 as an immediate consequence of the following basic result, which was proved by Leau and Fatou. However, it would be equally possible to first prove Theorem 10.8 and then derive Theorem 10.9 from it. (Compare Problem 10-c.)

Theorem 10.9 (Parabolic Linearization Theorem). *For any attracting or repelling petal $\mathcal{P}$, there is one and, up to composition with a translation of $\mathbb{C}$, only one conformal embedding $\alpha : \mathcal{P} \to \mathbb{C}$ which satisfies the Abel functional equation*

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

for all $z \in \mathcal{P} \cap f^{-1}(\mathcal{P})$.

Here we can think of $\mathbb{C}$ the universal covering space of the cylinder $\mathbb{C}/\mathbb{Z}$. The linearizing coordinate $\alpha(z)$ is often referred to as a *Fatou coordinate* in $\mathcal{P}$. The proofs of these two theorems, loosely following Steinmetz [1993], will be based on the following lemma, The argument is completely constructive, and can be used for actual computation of Fatou coordinates, although convergence is rather slow.

Lemma 10.10. *Let $F : \mathbb{H}_R \to \mathbb{H}_R$ satisfy the inequality*

$$\mathrm{Re}(F(w)) > \mathrm{Re}(w) + 1/2$$

and also the inequality

$$|F(w+1) - w - 1| \le C/|w|^\epsilon$$

for some positive constants C and ϵ. (Compare equation (10 : 3), taking $\epsilon = 1/n$.) Let $\hat{w}$ be a base point in $\mathbb{H}_R$. Then the sequence of functions

$$\beta_k(w) = F^{\circ k}(w) - F^{\circ k}(\hat{w})$$

converges locally uniformly to a biholomorphic map

$$\beta : \mathbb{H}_R \xrightarrow{\cong} U \subset \mathbb{C}$$

which satisfies the Abel equation $\beta(F(w)) = \beta(w) + 1$. Furthermore, the ratio $\beta(w)/w$ tends to $+1$ as w tends to infinity within $\mathbb{H}_R$.

Proof. First note that the derivative F' of F tends to $+1$ as w tends to infinity within the smaller half-plane $\mathbb{H}_{2R}$. In fact the inequality

$$|F'(w) - 1| < C/S^{1+\epsilon} \quad \text{whenever} \quad |w| \ge 2S \ge 2R$$

follows from the Cauchy Derivative Estimate, Lemma 1.2′, since the function $w \mapsto F(w) - w - 1$ caries the disk of radius S centered at w into the disk of radius C/S^ϵ centered at zero. Given two points w' and w'' in the half-plane $\mathbb{H}_{2S}$, since the average of $F'(w) - 1$ over the line segment from w' to w'' is equal to $\big(F(w'') - F(w')\big)/(w'' - w') - 1$, we conclude that

$$\left|\frac{F(w'') - F(w')}{w'' - w'} - 1\right| \;\le\; \frac{C}{S^{1+\epsilon}}\,.$$

Now choose some base point $\hat{w} \in \mathbb{H}_{2R}$ and consider the sequence of functions $\beta_k(w) = F^{\circ k}(w) - F^{\circ k}(\hat{w})$. Since $\mathrm{Re}\big(F^{\circ k}(w)\big) > k/2$, we can write

$$\left|\frac{\beta_k(w)}{\beta_{k-1}(w)} - 1\right| \;\le\; \frac{C'}{k^{1+\epsilon}} \qquad (10:5)$$

for suitable choice of C' and for all $k \ge 1$. In particular, it follows that

$$1 - \frac{C'}{k^{1+\epsilon}} \;\le\; \left|\frac{\beta_k(w)}{\beta_{k-1}(w)}\right| \;\le\; 1 + \frac{C'}{k^{1+\epsilon}}\,. \qquad (10:6)$$

Since the value of the infinite product

$$P \;=\; \prod_{k\ge 1}\left(1 + C'/k^{1+\epsilon}\right)$$

is finite, it follows from (10 : 6) that

$$|\beta_k(w)| \;\le\; |w - \hat{w}|\,P$$

for every k. Substituting $k-1$ in place of k and multiplying by (10 : 5), we see that

$$|\beta_k(w) - \beta_{k-1}(w)| \;\le\; PC'|w - \hat{w}|/k^{1+\epsilon}. \qquad (10:7)$$

Since the infinite series $\sum 1/k^{1+\epsilon}$ converges, this proves that the series

$$\beta_0(w) \;+\; \sum_{k=1}^{\infty}\big(\beta_k(w) - \beta_{k-1}(w)\big)$$

is absolutely convergent, and therefore that the limit

$$\beta(w) \;=\; \lim_{k\to\infty}\beta_k(w)$$

exists for all $w \in \mathbb{H}_{2R}$. *Furthermore, if we divide by $\beta_0(w) = w - \hat{w}$, then the convergence $\beta_k(w)/\beta_0(w) \to \beta(w)/\beta_0(w)$ is uniform throughout the half-plane* $\mathbb{H}_{2R}$. This proves that the limit function β is holomorphic. Now choose k_0 large enough so that $\prod_{k\ge k_0}(1 - C'/k^{1+\epsilon}) > 0$. (For example, take $k_0 > C'$.) Using (10 : 6), since $F^{\circ k_0}$ is univalent, it follows that β is also univalent on $\mathbb{H}_{2R}$. The extension of β as a univalent function on the larger half-plane $\mathbb{H}_R$ then follows immediately, since some iterate of F embeds $\mathbb{H}_R$ into $\mathbb{H}_{2R}$.

To prove the asymptotic formula

$$\beta(w) \sim w \qquad \text{as} \qquad |w| \to \infty \quad \text{within} \quad \mathbb{H}_R,$$

that is, to prove that $\beta(w)/w \to 1$, we proceed as follows. For each fixed k, it is easy to check that

$$\beta_k(w) \sim F^{\circ k}(w) \sim w+k \sim w \sim \beta_0(w) \qquad \text{as} \quad |w| \to \infty.$$

On the other hand, since the sequence of ratios $\beta_k(w)/\beta_0(w)$ converges uniformly to $\beta(w)/\beta_0(w)$ throughout the half-plane H_{2R}, this proves that $\beta(w) \sim \beta_0(w) \sim w$ as $|w| \to \infty$ within $\mathbb{H}_{2R}$. The extension of this statement to $\mathbb{H}_R$ is then straightforward. □

We will also need the following.

Lemma 10.11. *With $\beta(\mathbb{H}_R) = U \subset \mathbb{C}$ as in the preceding lemma, the image of U under the projection from $\mathbb{C}$ to $\mathbb{C}/\mathbb{Z}$ is the entire cylinder. Equivalently, the union of all translates $U+n$ with $n \in \mathbb{Z}$ is all of $\mathbb{C}$.*

Proof. Choose S large enough so that $|\beta(w) - w| < |w|/3$ for every $w \in \mathbb{H}_R$ with $|w| > S$. For any point of $\mathbb{C}/\mathbb{Z}$ we can choose a representative $w_0 = u_0 + iv_0$ which is far enough to the right so that $|w_0| > 2S$, and so that the closed disk $\overline{D}$ of radius $|w_0|/2$ centered at w_0 is contained in $\mathbb{H}_R$. Then for every $w \in \overline{D}$, since $S < |w| \le 3|w_0|/2$, it follows that the difference $|\beta(w) - w|$ is less than the radius $|w_0|/2$. It then follows from the Argument Principle (or from Rouché's Theorem) that the image of this disk under β must contain the given w_0. □

Proof of Theorem 10.9: Existence. In the special case of the petal $\mathcal{P} = \mathcal{P}_j(R) \cong \mathbb{H}_R$, we can define $\alpha : \mathcal{P}_j(R) \to \mathbb{C}$ to be the composition $\alpha(z) = \beta(\varphi(z))$, using the function β as described in Lemma 10.10. For the case of an arbitrary petal $\mathcal{P}$ with attraction vector $\mathbf{v}_j$, note that any point $z \in \mathcal{P}$ must have some forward image $f^{\circ k}(z) \in \mathcal{P}_j(R)$. Thus we can simply define $\alpha(z)$ to be $\alpha(f^{\circ k}(z)) - k$. The necessary properties are easily verified.

Uniqueness. Consider $\mathcal{P}$ and the associated $\mathcal{P}_j(R)$, as above. Then any Fatou coordinate on $\mathcal{P}$ can be restricted to $\mathcal{P} \cap \mathcal{P}_j(R)$ and then uniquely extended to $\mathcal{P}_j(R)$. Thus it again suffices to consider the special case $\mathcal{P} = \mathcal{P}_j(R)$. Let $\alpha : \mathcal{P} \xrightarrow{\cong} U$ be the Fatou coordinate as constructed above and let $\alpha' : \mathcal{P} \xrightarrow{\cong} U'$ be some arbitrary Fatou coordinate on $\mathcal{P}$. Then $g = \alpha' \circ \alpha^{-1}$ maps U bijectively onto U' with $g(w+1) = g(w)+1$. Since the union of all integer translates of U is the entire plane by Lemma 10.11, we can extend g to a bijective map $\hat{g} : \mathbb{C} \xrightarrow{\cong} U' + \mathbb{Z} \subset \mathbb{C}$ by setting $\hat{g}(w+n) = g(w) + n$ for all $w \in U$ and $n \in \mathbb{Z}$. It follows that $\hat{g}$ is a linear map from $\mathbb{C}$ onto $\mathbb{C}$. Since $\hat{g}(u+1) = \hat{g}(u) + 1$, it can only be a translation. This completes the proof of Theorem 10.9. □

Proof of Theorem 10.8. This follows as an immediate corollary. Simply map each equivalence class $\{f^{\circ k}(z)\} \in \mathcal{P}/f$ to the residue class of $\alpha(z)$ modulo $\mathbb{Z}$. $\square$

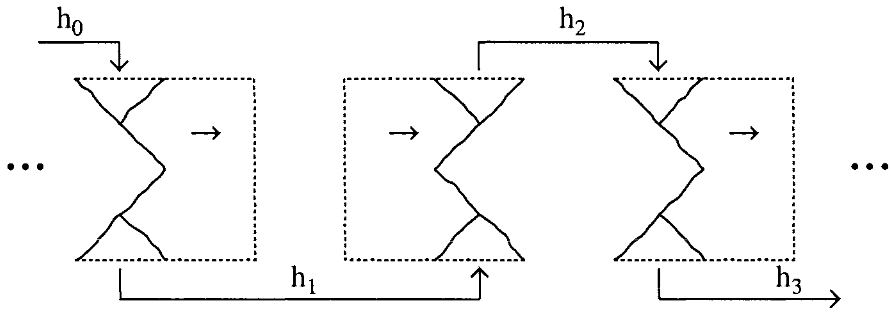


Figure 23. The Écalle-Voronin classification: A sketch of the open sets $\alpha_j(\mathcal{P}_j)$ for $j = 1, 2, 3$ and the associated pasting maps $h_j : \alpha_j(\mathcal{P}_j \cap \mathcal{P}_{j+1}) \to \alpha_{j+1}(\mathcal{P}_j \cap \mathcal{P}_{j+1})$. The arrow in each $\alpha_j(\mathcal{P}_j)$ points from α to $\alpha + 1$.

Remark 10.12. The Écalle-Voronin Classification. Note that this preferred Fatou coordinate system is defined only within one of the $2n$ attracting or repelling petals. In order to describe 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 biholomorphic maps

$$h_j : \alpha_j(\mathcal{P}_j \cap \mathcal{P}_{j+1}) \xrightarrow{\cong} \alpha_{j+1}(\mathcal{P}_j \cap \mathcal{P}_{j+1}),$$

which must satisfy the functional equation $h_j(\alpha+1) = h_j(\alpha)+1$ whenever both α and $\alpha+1$ lie in $\alpha_j(\mathcal{P}_j \cap \mathcal{P}_{j+1})$. (Compare Figure 23.) In fact, it can be shown that each h_j must have the form*

$$h_j(\alpha) = \alpha + H_j\left(e^{\pm 2\pi i \alpha}\right) \quad \text{so that} \quad \alpha_{j+1} = \alpha_j + H_j\left(e^{\pm 2\pi i \alpha_j}\right),$$

taking the plus sign if j is even or the minus sign if j is odd. Here each H_j is a function which is defined and holomorphic throughout some punctured

* An alternative procedure which treats α_j and α_{j+1} more symmetrically would be to introduce the variable $t_j(z) = t_j(f(z)) \in \mathbb{D} \smallsetminus \{0\}$ by the formula $t_j = e^{\pm 2\pi i \alpha_j}$, taking the plus sign if $\mathrm{Im}(\alpha_j) \gg 0$ or the minus sign if $\mathrm{Im}(\alpha_j) \ll 0$. Then we can express t_{j+1} as a holomorphic function

$$t_{j+1} = \eta_j(t_j) = t_j \exp(\pm 2\pi i H_j(t_j))$$

for t_j in some disk $\mathbb{D}_\epsilon$. The η_j are known as *horn maps* and h_j can be described as a *lifted horn map.*

neighborhood $\mathbb{D}_\epsilon \smallsetminus \{0\}$ of zero. In fact H_j extends holomorphically over $\mathbb{D}_\epsilon$. Furthermore, as z tends to the fixed point $\hat{z}$ within $\mathcal{P}_j \cap \mathcal{P}_{j+1}$, the number $e^{\pm 2\pi i \alpha_j(z)}$ tends, to zero, so that we get the limiting formula

$$\alpha_{j+1}(z) = \alpha_j(z) + H_j(0) + o(1) \qquad \text{as} \qquad z \to \hat{z}.$$

These germs H_j of holomorphic functions are not quite uniquely defined, since we are free to add an arbitrary constant to each Fatou coordinate. However, since each H_j depends on infinitely many parameters while there are only $2n$ Fatou coordinates, one gets the following statement.

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,

(Compare Voronin [1981], Malgrange [1981/82], Martinet and Ramis [1983], Buff and Epstein [2002].) On the other hand, if we allow a change of coordinate given by a formal power series, then there is a normal form $z \mapsto z + z^{n+1} + \beta\, z^{2n+1}$ depending on just one complex parameter (Problem 10-d), while if we allow a topological change of coordinate, then Camacho [1978] showed that the normal form $z \mapsto z + z^{n+1}$ will suffice.

It is not hard to see that the sum

$$I(f, \hat{z}) = H_0(0) + H_1(0) + \cdots + H_{2n-1}(0) \tag{10:8}$$

does not depend on these additive constants, and hence is a well defined invariant. If we replace f by some iterate $f^{\circ k}$, note that each Fatou coordinate must be divided by k, so that

$$I(f^{\circ k}, \hat{z}) = I(f, \hat{z})/k\,. \tag{10:9}$$

For further information about this invariant, see Remark 12.12.

After adding appropriate constants to the α_j, we can always normalize so that

$$H_0(0) = H_1(0) = \cdots = H_{2n-1}(0) = I(f,\, \hat{z})/2n\,.$$

The various H_j are then uniquely defined up to a simultaneous linear change of argument, replacing each $\mathbb{H}_j(t)$ by either $\mathbb{H}_j(\lambda t)$ or $H_j(t/\lambda)$ according to whether j is even or odd.

Although the coordinate α is well-defined only up to an additive constant, its differential $d\alpha$ is uniquely defined. Thus another way of describing Theorem 10.9 is to say that within each attracting or repelling petal there is a unique holomorphic 1-form $d\alpha = (d\alpha/dz)\,dz$ which is f-invariant and satisfies $\int_z^{f(z)} d\alpha = +1$. Equivalently, within each petal there is a uniquely defined holomorphic vector field $(d\alpha/dz)^{-1}\partial/\partial z$ with the following property: *The time 1 map for the flow* $t \mapsto f_t(z)$ *generated by the associated differential equation*

$$\frac{dz(t)}{dt} = (d\alpha/dz)^{-1}$$

on $\mathcal{P}_j$ *is precisely the given map,* $f_1 = f$. In particular, we can write f as an iterate $f = g \circ g$ within each petal, where $g = f_{1/2}$.

In general, these flows $\alpha \mapsto \alpha + t$ for the different petals do not match up at all. *In fact they match up if and only if the functions* H_j *are all constants,* $H_j(t) \equiv H_j(0)$. This is precisely the case when f can be described as the time one map for the local flow $t \mapsto f_t(z)$ generated by a differential equation

$$dz/dt = \mathbf{v}(z)$$

where $\mathbf{v}(z)$ is holomorphic throughout some neighborhood of $\hat{z}$, with a multiple zero at $\hat{z}$. *Thus we see that such a holomorphic differential equation is completely classified up to local conformal conjugacy by the order of this zero together with the single invariant*

$$I(f_1, \hat{z}) = t\, I(f_t, \hat{z}).$$

(We will discuss such flows further in Lemma 12.11.) In the simplest case,

$$dz/dt = z^{n+1} \quad \text{with solution} \quad f_t(z) = z/\sqrt[n]{1 - ntz^n},$$

it is not hard to check that this invariant is identically zero. For example, this follows since each f_t with $t \neq 0$ is linearly conjugate to f_1.

Global Theory. Now suppose that $f : S \to S$ is a globally defined holomorphic map with a fixed points of multiplicity $n + 1 \geq 2$. Here S can be either $\widehat{\mathbb{C}}$ or $\mathbb{C}$ or $\mathbb{C}/\mathbb{Z}$. Although attracting petals behave much like repelling petals in the local theory, they behave quite differently in the large.

Corollary 10.13. *If* $\mathcal{P} \subset S$ *is an attracting petal for* f, *then the Fatou map*

$$\alpha : \mathcal{P} \to \mathbb{C}$$

extends uniquely to a map $\mathcal{A} \to \mathbb{C}$ *which is defined and holomorphic throughout the attractive basin of* $\mathcal{P}$, *still satisfying the Abel equation* $\alpha(f(z)) = 1 + \alpha(z)$.

In the case of a repelling petal, the analogous statement is the following.

Corollary 10.14. *If* $\mathcal{P}'$ *is a repelling petal for* $f : S \to S$, *then the inverse map*

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

extends uniquely to a globally defined holomorphic map $\gamma : \mathbb{C} \to S$ *which satisfies the corresponding equation*

$$f(\gamma(w)) = \gamma(1+w).$$

The proofs of Corollaries 10.13 and 10.14 are completely analogous to the proofs of Corollaries 8.4 and 8.12. □

In the case of a rational map $f : \widehat{\mathbb{C}} \to \widehat{\mathbb{C}}$ which is *nonlinear* (that is, of degree 2 or more), the extended map of Corollary 10.13 is surjective. However, it is not univalent, but rather has critical points whenever some iterate $f \circ \cdots \circ f$ has a critical point. Note the following basic result.

Theorem 10.15. *If* $\hat{z}$ *is a parabolic fixed point with multiplier* $\lambda = 1$ *for a rational map, then each immediate basin for* $\hat{z}$ *contains at least one critical point of* f. *Furthermore, each basin contains one and only one attracting petal* $\mathcal{P}_{\max}$ *which maps univalently onto some right half-plane under* α *and which is maximal with respect to this property. This preferred petal* $\mathcal{P}_{\max}$ *always has one or more critical points on its boundary.*

More generally, for a periodic orbit which has multiplier equal to a root of unity, we can apply this result to an appropriate iterate of f and conclude that every cycle of parabolic basins contains a critical point.

The proof of Theorem 10.15 is completely analogous to the proof of Lemma 8.5. It is not difficult to show that α^{-1} can be defined throughout *some* right half-plane. If we try to extend leftwards by analytic continuation then we must run into an obstruction, which can only be a critical point of f. (For an alternative proof that every parabolic basin contains a critical point, see Milnor and Thurston [1988, pp. 512–515].) □

As an example, Figure 24 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 $\mathcal{A}$. The critical point $z = -1/2$ lies exactly at the center of symmetry. All orbits in this basin $\mathcal{A}$ converge towards $z = 0$ to the right. The curves $\mathrm{Re}(\alpha(z)) =$ constant $\in \mathbb{Z}$ have been drawn in, using the normalization $\alpha(-1/2) = 0$. Thus the preferred petal $\mathcal{P}_{\max}$, with the critical point on $\partial\mathcal{P}_{\max}$, is bounded by the right half of the central ∞ shaped curve. Note that the function $z \mapsto \mathrm{Re}(\alpha(z))$ has a saddle critical point at each iterated preimage of ω. This function $\mathrm{Re}(\alpha(z))$ oscillates wildly as z tends to $J = \partial\mathcal{A}$.

As an immediate consequence of Theorem 10.15 we have the following.

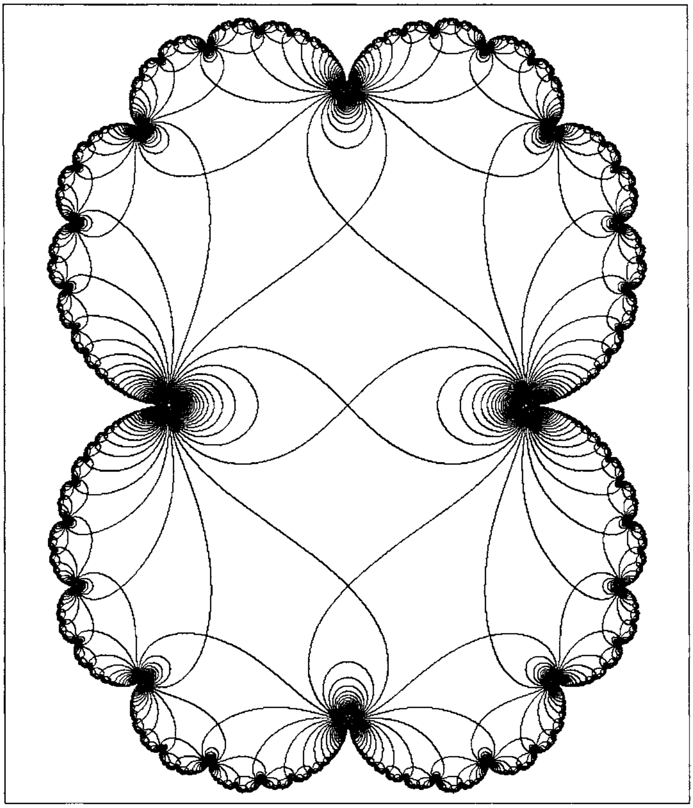


Figure 24. Julia set for $z \mapsto z^2 + z$, *with the curves* $\mathrm{Re}(\alpha(z)) \in \mathbb{Z}$ *drawn in.*

Corollary 10.16. *A rational map can have at most finitely many parabolic periodic points. In fact, for a map of degree* $d \geq 2$, *the number of parabolic cycles plus the number of attracting cycles is at most* $2d - 2$.

More precisely, the number of cycles of Fatou components which are either immediate parabolic basins or immediate attracting basins is at most equal to the number of distinct critical points. In fact, just as in the proof of Theorem 8.6, these attracting and parabolic basins must be disjoint and each cycle of basins must contain at least one critical point. □

For sharper results, see Shishikura [1987], or Buff and Epstein [2002].

Concluding Problems

Problem 10-a. Repelling petals and the Julia set. If f is a nonlinear rational function, show that every repelling petal must intersect the Julia set of f.

Problem 10-b. No small cycles. We noted in Remark 10.3 that a small neighborhood of a parabolic fixed point contains no periodic orbits other than the fixed point itself. **(1)** Give an alternative proof based on the Flower Theorem 10.7. In fact, assuming that both f and f^{-1} are defined and univalent throughout the punctured neighborhood

$$\mathcal{P}_0 \cup \mathcal{P}_1 \cup \mathcal{P}_2 \cdots \cup \mathcal{P}_{2n-1},$$

show that this union contains no periodic orbits. **(2)** On the other hand, show that any nonlinear rational function has orbits which return to every repelling petal infinitely often. (Using §14, one can show that there are *periodic* points arbitrarily close to $\hat{z}$ in each repelling petal.)

Problem 10-c. The Cylinder Theorem. (1) Give an alternate proof of Theorem 10.8 as follows. With notation as in the proof of Lemma 10.1, show that F maps the half-plane $\mathbb{H}_R$ diffeomorphically onto an open set $F(\mathbb{H}_R)$, and show that the region $U = \overline{\mathbb{H}}_R \smallsetminus F(\mathbb{H}_R)$ forms a fundamental domain for the action of F on $\mathbb{H}_R$. In particular, the quotient $\mathbb{H}_R/F$ can be identified with the space obtained from U by identifying each point $R+iv$ of its left hand boundary with the image $F(R+iv)$ on its right hand boundary. Thus the Riemann surface $\mathbb{H}_R/F$ has free cyclic fundamental group, and hence must be conformally invariant to the cylinder $\mathbb{C}/\mathbb{Z}$ or the punctured disk $\mathbb{D} \smallsetminus \{0\}$ or to an an annulus. **(2)** First suppose that $\mathbb{H}_R/F$ is conformally equivalent to an annulus of modulus μ, and hence to the finite cylinder $C_\mu \subset \mathbb{C}/\mathbb{Z}$ consisting of all $z = x + iy \pmod{\mathbb{Z}}$ with $0 < y < \mu$. Show that U would have a conformal metric $\gamma(w)|dw|$ of area

$$\int\int_U \gamma(w)^2 du dv \;=\; \mu < \infty \tag{10:10}$$

with the property that any path $w = w(t)$ from $C + iv$ to $F(C + iv)$ in U would have length $\int \gamma(w(t)|dw(t)| \geq 1$. Using the Schwarz inequality $(\int(1 \cdot \phi\, dt)^2 \leq (\int 1\, dt)(\int \phi^2 dt)$, show that the integral along the straight line segment $w(t) = (1-t)w_0 + tF(w_0)$ from $w_0 = R + iv_0$ to $F(w_0)$ satisfies

$$\int_0^1 \gamma(w)\,|dw/dt|\,dt \geq 1 \qquad \text{hence} \qquad \int_0^1 \gamma^2(w)\,|dw/dt|^2\,dt \geq 1.$$

(3) If R is sufficiently large, show that the nonconformal change of coordinates $(t, \eta) \mapsto (u, v)$ with

$$w = u + iv \;=\; (1-t)(R + i\eta) + t\,F(R + i\eta)$$

will have matrix of first derivatives arbitrarily close to the identity matrix.

Since the integral

$$\int_{-\infty}^{\infty}\left(\int_0^1 \gamma^2(w)|\partial w/\partial t|^2\,dt\right)d\eta \;\geq\; \int_{-\infty}^{\infty} d\eta$$

is infinite, conclude that the integral (10 : 10) is also infinite, yielding a contradiction. **(4)** Similarly, since the integral over each of the regions $\eta \geq 0$ and $\eta \leq 0$ is infinite, show that $\mathbb{H}_R/F$ cannot be conformally isomorphic to a punctured disk.

Problem 10-d. A formal normal form. Suppose that f is given by a power series of the form

$$f(z) \;=\; z + z^m + (\text{higher terms}) \qquad (10:9)$$

with $m \geq 2$, and let g be a local diffeomorphism of the form $g(z) = z + cz^k$ with $k \geq 2$. **(1)** Show that

$$f(g(z)) - g(f(z)) \;=\; (m-k)\,c\,z^{m+k-1} + (\text{higher terms}),$$

or equivalently that

$$g^{-1}\circ f\circ g(z) \;=\; f(z) + (m-k)\,c\,z^{m+k-1} + (\text{higher terms}).$$

(2) Conclude inductively that by such conjugations we can eliminate terms of any degree other than 1, m, and $2m-1$ from the power series for f. Thus f is locally holomorphically conjugate to a map of the form

$$g(z) \;=\; z + a\,z^m + b\,z^{2m-1} + (\text{terms of degree } > N),$$

where N can be arbitrarily large. **(3)** Conclude that conjugation by a possibly nonconvergent formal power series $\psi(z) = z + c_2\,z^2 + c_3\,z^3 + \cdots$ can transform any f of the form (10 : 9) into the normal form

$$z \;\mapsto\; z + z^m + b\,z^{2m-1}.$$

Remark. This coefficient b is a formal conjugacy class invariant, so that no further simplification is possible. (Compare Problem 12-a.) The power series ψ is definitely not convergent in general. In fact, according to Écalle or Voronin, it would take infinitely many complex parameters to specify the map f up to local holomorphic conjugacy. (See Remark 10.12.)

Problem 10-e. Examples with a parabolic point at infinity. (Compare Milnor [1993, §8].) **(1)** For $f(z) = z - 1/z$ show that the only fixed point is the point at infinity, with multiplicity $n+1=3$. Show that the two parabolic basins are the upper and lower half-planes, and that $J = \mathbb{R}\cup\{\infty\}$. (For a similar nonparabolic example, see Problem 7-a.) **(2)** For $f(z) = z - 1/z + 1$ show that the point at infinity has multiplicity 2,

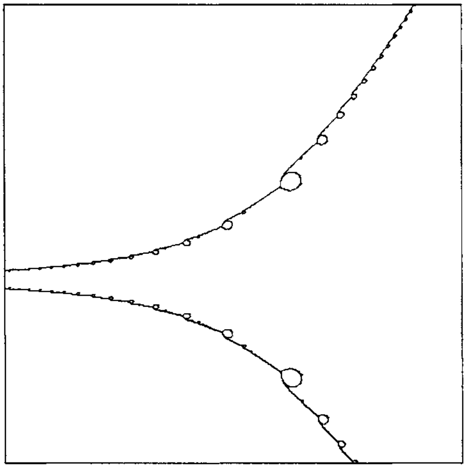


Figure 25. Julia set for $z \mapsto z + 1/(1 + z^2)$.
(See Problem 10-e (4).)

and that J is a Cantor set contained in $\mathbb{R} \cup \{\infty\}$. **(3)** For $f(z) = z + 1/z - 2$ show that ∞ again has multiplicity 2, and that J is the interval $[0, +\infty]$. **(4)** For $f(z) = z + 1/(1 + z^2)$ as illustrated in Figure 25, show that the multiplicity of ∞ is 4. Show that one of the three immediate parabolic basins contains all of $\mathbb{R}$ and hence *nearly* disconnects the Riemann sphere.

Problem 10-f. Immediate parabolic basins. By an argument similar to that of Theorem 8.9, show that the complement of an immediate parabolic basin is either connected or else has uncountably many connected components. (However, compare Problem 10-e(4).)

§11. Cremer Points and Siegel Disks

Once more we consider maps of the form

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

which are defined and holomorphic throughout some neighborhood of the origin, with a fixed point of multiplier λ at the origin. In §§8 and 9 we supposed that $|\lambda| \neq 1$, while in §10 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 $\xi \in \mathbb{R}/\mathbb{Z}$ is called the *rotation number** for the tangent space at the fixed point. The fundamental question here is whether or not f is *locally linearizable*. That is, does there exist a local holomorphic 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 Theorem 8.2.) If f is a globally defined rational function, then Theorem 5.2 implies the following statement.

Lemma 11.1. *Let f be a rational function of degree 2 or more with a fixed point z_0 which is indifferent, $|f'(z_0)| = 1$. Then the following three conditions are equivalent to each other:*

- *f is locally linearizable around z_0,*
- *z_0 belongs to the Fatou set $\widehat{\mathbb{C}} \smallsetminus J(f)$,*
- *the connected component U of the Fatou set containing z_0 is conformally isomorphic to the unit disk under an isomorphism which conjugates f on U to multiplication by λ on the disk.*

Proof. If f is locally linearizable around z_0, then the iterates of f in a suitable neighborhood of z_0 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 from Corollary 5.3 that the entire Fatou component U of z_0 must be conformally isomorphic to the unit disk, with $f|_U$ conjugate to multiplication by λ on $\mathbb{D}$. □

*A theorem of Naĭshul′ [1983] asserts that this rotation number is a local topological invariant.

Definition. We will say that an irrationally indifferent fixed point is either a *Siegel point* or a *Cremer point*, according to whether a local linearization is possible or not. A Fatou component on which f is conformally conjugate to a rotation of the unit disk is called a *Siegel disk*, with the fixed point z_0 as *center*. (In the classical literature, Siegel points were called "centers" and the question as to their existence was called the "center problem.")

This section will first survey what is known about the local linearization problem and then prove some of the easier results. Finally, it will describe the relation between Cremer points or Siegel disks and the critical points of a rational map.

Edward Kasner [1912] conjectured that such a linearization is *always* possible. However, George Pfeiffer [1917] disproved this conjecture by giving a rather complicated description of certain holomorphic functions for which no local linearization is possible. Gaston Julia [1919] claimed to settle the question completely for rational functions of degree 2 or more by showing that such a linearization is *never* possible; however, his proof was wrong. In fact the correct answer depends on a careful study of the extent to which the rotation number ξ can be very closely approximated by rational numbers.

Hubert Cremer finally put the situation in clearer perspective with a result which we can state as follows.

> **Nonlinearizability Theorem 11.2 (Cremer [1927]).** *Given λ on the unit circle and given $d \geq 2$, if the sequence of numbers $\sqrt[d^q]{1/|\lambda^q - 1|}$ is unbounded as $q \to \infty$, then no fixed point of multiplier λ for a rational function of degree d can be locally linearizable.*

This will be proved below. It is convenient to say that a property of an angle $\xi \in \mathbb{R}/\mathbb{Z}$ is true for *generic* ξ if the set of ξ for which it is true contains a countable intersection of dense open subsets of $\mathbb{R}/\mathbb{Z}$. According to Baire, such a countable intersection of dense open sets is necessarily dense and uncountably infinite. (See Problem 4-j.)

> **Corollary 11.3.** *For a generic choice of rotation number $\xi \in \mathbb{R}/\mathbb{Z}$, if z_0 is a fixed point of multiplier $e^{2\pi i \xi}$ for a completely arbitrary rational function f of degree 2 or more, then there is no local linearizing coordinate about z_0.*

(Compare Problem 11-b.) The question as to whether this statement is actually true for *all* ξ remained open for many years until Carl Ludwig

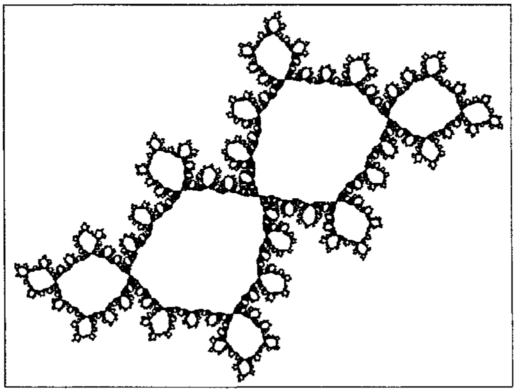


Figure 26a. Julia set for $z^2 + e^{2\pi i \xi} z$ *with* $\xi = \sqrt[3]{1/4}$ $= .62996\cdots$. *The large region on the lower left is a Siegel disk.*

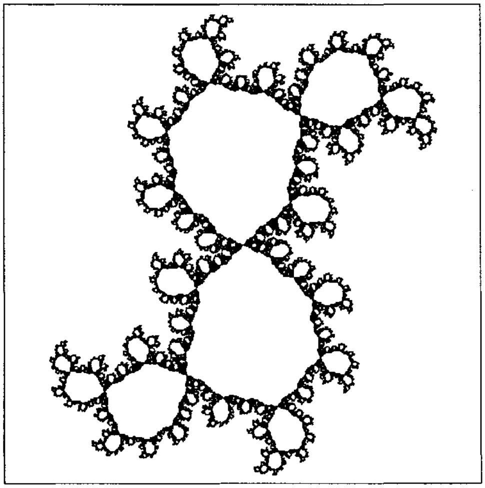


Figure 26b. Corresponding Julia set with a randomly chosen angle $\xi = .7870595\cdots$.

Siegel proved the following. Again let $\lambda = e^{2\pi i \xi}$ with $\xi \in \mathbb{R} \setminus \mathbb{Q}$.

Linearization Theorem 11.4 (Siegel [1942]). *If* $1/|\lambda^q - 1|$ *is less than some polynomial function of* q, *then every germ of a holomorphic map with fixed point of multiplier* λ *is locally linearizable.*

This will not be proved here. However, proofs may be found in Siegel [1942], Siegel and Moser [1971], Zehnder [1977], or Carleson and Gamelin [1993].

Corollary 11.5. *For every ξ outside of a set of Lebesgue measure zero, we can conclude that every holomorphic germ with a fixed point of multiplier $e^{2\pi i\xi}$ is locally linearizable.*

In other words, if the angle $\xi \in \mathbb{R}/\mathbb{Z}$ is "randomly chosen" with respect to Lebesgue measure, then with probability 1 every rational function with a fixed point of multiplier $e^{2\pi i\xi}$ will have a corresponding Siegel disk. See Figure 26b for an example. Again, this will not be proved here (except in the special case of a quadratic polynomial fixed point, as described in Theorem 11.14). However, we will see in Lemma 11.7 that Corollary 11.5 does follow from Theorem 11.4.

Remark. Comparing Corollaries 11.3 and 11.5, we see that there is a total contrast between behavior for *generic* ξ and behavior for *almost every* ξ. This contrast is quite startling, but is not uncommon in dynamics. (Compare the discussion of the iterated exponential map in §6.) In applied dynamics, it is 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.

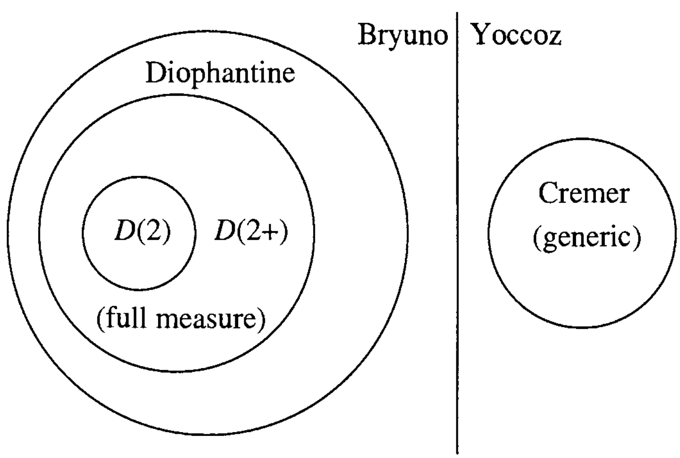


Figure 27. Schematic diagram for classes of irrational numbers.

In order to understand these statements, as well as sharper results which have been obtained more recently, it is convenient to introduce a number of different classes of irrational numbers, which are related to each other as indicated schematically in Figure 27.

Let κ be a positive real number. By definition, an irrational number ξ

is said to be *Diophantine of order* $\le \kappa$ if there exists $\epsilon > 0$ so that

$$\left|\xi - \frac{p}{q}\right| > \frac{\epsilon}{q^\kappa} \qquad \text{for every rational number} \quad p/q. \qquad (11:1)$$

The class of all such numbers will be denoted by $\mathcal{D}(\kappa)$. Evidently

$$\mathcal{D}(\kappa) \subset \mathcal{D}(\eta) \qquad \text{whenever} \qquad \kappa < \eta.$$

Setting $\lambda = e^{2\pi i\xi}$ as above, and choosing p to be the closest integer to $q\xi$ so that $|q\xi - p| \le 1/2$, note the order of magnitude estimate

$$|\lambda^q - 1| \;=\; |2\sin(\pi(q\xi - p))| \;\asymp\; 2\pi|q\xi - p|.$$

In fact, more precisely, it is not hard to see that

$$4|q\xi - p| \;\le\; |\lambda^q - 1| \;\le\; 2\pi|q\xi - p|.$$

It follows that $(11:1)$ is equivalent to the requirement that

$$|\lambda^q - 1| \;>\; \epsilon'/q^{\kappa-1} \qquad \Longleftrightarrow \qquad 1/|\lambda^q - 1| \;<\; c\,q^{\kappa-1}$$

for some $\epsilon' > 0$, with the same value of κ and with $c = 1/\epsilon'$. Thus Siegel's Theorem 11.4 can be restated as follows.

> *If the angle $\xi \in \mathbb{R}/\mathbb{Z}$ is Diophantine of any order, then any holomorphic germ with multiplier $\lambda = e^{2\pi i\xi}$ is locally linearizable.*

It turns out that the set $\mathcal{D}(\kappa)$ is vacuous for $\kappa < 2$. (See Problem 11-a.) Diophantine numbers of order 2 are said to be *of bounded type*. (Compare Corollary 11.9.) Examples are provided by quadratic irrationals. More generally, we have the following classical statement.

> **Theorem 11.6 (Liouville).** *If the irrational number ξ satisfies a polynomial equation $f(\xi) = 0$ of degree d with integer coefficients, then $\xi \in \mathcal{D}(d)$.*

Proof. We may assume that $f(p/q) \ne 0$. Clearing denominators, it follows that $|f(p/q)| \ge 1/q^d$. On the other hand, if M is an upper bound for $|f'(x)|$ in the interval of length 1 centered at ξ, then

$$|f(p/q)| \;\le\; M\,|\xi - p/q|\,.$$

Choosing $\epsilon < 1/M$, we obtain $|\xi - p/q| > \epsilon/q^d$, as required. □

Thus every algebraic number is Diophantine. It follows that any irrationally indifferent fixed point with algebraic rotation number is locally linearizable. (Compare Figure 26a, page 127.)

Remark. Irrational numbers which are not Diophantine are often called

Liouville numbers. We will see in Appendix C that the set of Liouville numbers is very small, in the sense that its Hausdorff dimension is zero.

A much sharper and more difficult version of Theorem 11.6 has been proved by Klaus Roth.

Theorem of Roth [1955]. *Every algebraic number in* $\mathbb{R} \setminus \mathbb{Q}$ *belongs to the class*

$$\mathcal{D}(2+) = \bigcap_{\kappa>2} \mathcal{D}(\kappa),$$

consisting of numbers which are Diophantine of order $\leq \kappa$ *for every* $\kappa > 2$.

We will not use Roth's result, but will make use of the following much more elementary property.

Lemma 11.7. *This set* $\mathcal{D}(2+)$ *has full measure in the circle* $\mathbb{R}/\mathbb{Z}$.

(On the other hand, we will see in Appendix C that the subset $\mathcal{D}(2)$ has measure zero.)

Proof of Lemma 11.7. Let $U(\kappa, \epsilon)$ be the open set consisting of all $\xi \in [0, 1]$ such that $|\xi - p/q| < \epsilon/q^\kappa$ for some p/q. This set has measure at most

$$\sum_{q=1}^{\infty} q \cdot 2\epsilon/q^\kappa,$$

since for each q there are q possible choices of p/q modulo 1. If $\kappa > 2$, then this sum converges, and hence tends to zero as $\epsilon \searrow 0$. Therefore the intersection $\bigcap_{\epsilon>0} U(\kappa, \epsilon)$ has measure zero, and its complement $\mathcal{D}(\kappa)$ has full measure. Taking the intersection of these complements $\mathcal{D}(\kappa)$ as $\kappa \searrow 2$, we see that the set $\mathcal{D}(2+)$ also has full measure. $\square$

Evidently Theorem 11.4 and Lemma 11.7 together imply Corollary 11.5.

For a more precise analysis of the approximation of an irrational number $\xi \in (0, 1)$ by rationals, we 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 number

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

is called the nth *convergent* to ξ. The denominators q_n will play a particularly important role. Here is a summary of some basic results. (Compare Hardy and Wright [1938] or Khintchine [1963].) Note that ξ is very close to some fraction with denominator q if and only if λ^q is very close to $+1$ (where $\lambda = e^{2\pi i\xi}$ as usual). In fact, if $\delta/q = |\xi - p/q|$ is the distance to the closest such fraction, with $0 < \delta < 1/2$, then $|\lambda^q - 1| = |e^{2\pi i\delta} - 1| = 2\sin(\pi\delta)$.

Theorem 11.8 (Best Approximations). *Each convergent p_n/q_n is the best possible approximation to ξ by fractions with denominator at most q_n. Furthermore*

$$|\lambda^h - 1| > |\lambda^{q_n} - 1| \quad \textit{for} \quad 0 < h < q_{n+1} \quad \textit{with} \quad h \neq q_n.$$

The error $|\lambda^{q_n} - 1|$ has the order of magnitude of $1/q_{n+1}$, or more precisely

$$2/q_{n+1} \le |\lambda^{q_n} - 1| \le 2\pi/q_{n+1}.$$

These denominators can be computed inductively by the formula

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

with $q_0 = 0$, $q_1 = 1$, $q_2 = a_1$.

The proof will be given in Appendix C.

Corollary 11.9. *An irrational number ξ is Diophantine if and only if the associated continued fraction denominators satisfy the condition that q_{n+1} is less than some polynomial function $f(q_n)$. More precisely, $\xi \in \mathcal{D}(\kappa)$ if and only if q_{n+1} is less than some constant times $q_n^{\kappa-1}$. In particular, ξ belongs to $\mathcal{D}(2)$ if and only if the ratios q_{n+1}/q_n are bounded, or equivalently if and only if the continued fraction coefficients $a_n = (q_{n+1} - q_{n-1})/q_n$ are bounded.*

For this reason, elements of $\mathcal{D}(2)$ are also called *numbers of bounded type* (or sometimes *numbers of constant type*). The proof of Corollary 11.9 is straightforward and will be left to the reader. $\square$

Next we state three results which give a much sharper picture of the local linearization problem. Alexander Bryuno proved the following.

Theorem 11.10 (Bryuno [1965], Rüssmann [1967]). *With λ and $\{q_n\}$ as above, if*

$$\sum_n \frac{\log(q_{n+1})}{q_n} < \infty, \tag{11:2}$$

then any holomorphic germ with a fixed point of multiplier λ is locally linearizable.

Jean-Christophe Yoccoz showed that this is a best possible result.

Theorem 11.11 (Yoccoz [1988]). *Conversely, if the sum* (11 : 2) *diverges, then the quadratic map $f(z) = z^2 + \lambda z$ has a fixed point at the origin which is not locally linearizable. Furthermore, this fixed point has the* small cycles property*: Every neighborhood of the origin contains infinitely many periodic orbits.*

(Compare Perez-Marco [1992].) Evidently such small cycles provide an obstruction to linearizability.

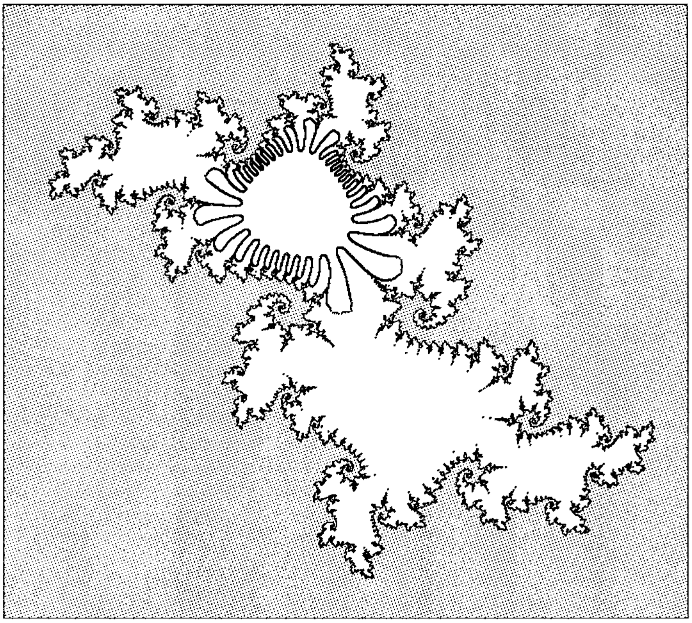


Figure 28. A rough plot of the filled Julia set for $z^2 + e^{2\pi i \xi} z$, with a Siegel disk of rotation number

$$\xi = 1/(3 + 1/(10 + 1/(200000 + 1/\cdots))) .$$

The boundary of the Siegel disk has been emphasized. Note the fjords with period $q_2 = 3$ and $q_3 = 31$ which squeeze this disk.

Without attempting to prove these theorems, we can give some intuitive idea as to what they mean in the polynomial case as follows. Whenever the summand $(\log q_{n+1})/q_n$ is large the rotation number will be extremely close to p_n/q_n, so that f will be extremely close to a parabolic map with a period q_n cycle of repelling directions. It follows that the basin of infinity for f will have a period q_n cycle of deep fjords which penetrate towards zero, squeezing the size of a possible Siegel disk. As an example, in Figure 28, the summands $\log(31)/3 = 1.144\cdots$ and $\log(6200003)/31 = 0.504\cdots$ correspond to fjords of period 3 and 31 which are visible in the figure. When the sum $(11:2)$ is infinite, such a Siegel disk can no longer exist.

Historical Remark. One early contributer to considerations of this kind was T. M. Cherry, but only part of his work had been published at the time of his death in 1966. (See Cherry [1964].) According to Love [1969]: "*Fuller details of this may possibly be written in his notebooks; it is likely that he studied this subject deeply over many years.*" These notebooks have been in France for many years. It is to be hoped that they will someday be made public or returned to Australia.

Yoccoz's theorem raises the question as to whether every Cremer point has small cycles. The answer was provided by Ricardo Perez-Marco. Suppose that $\sum \log(q_{n+1})/q_n = \infty$ so that a Cremer point can exist.

Theorem 11.12 (Perez-Marco [1990]). *If*

$$\sum \frac{\log\log(q_{n+1})}{q_n} < \infty, \qquad (11:3)$$

then every germ of a holomorphic function which has a Cremer point at the origin has the small cycles property. But if the sum $(11:3)$ *diverges (which is the generic case), there exists a germ with multiplier* λ *such that every forward orbit contained in some neighborhood of zero has zero as accumulation point. Such a germ has no small cycles but is not linearizable.*

However, it is not known whether this last possibility can occur for the germ of a rational map.

I will not try to say any more about these three big theorems. The remainder of this section will rather provide proofs for some easier results. We first show that Cremer points really exist, then prove Cremer's 1927 theorem, and finally show that Siegel disks really exist.

First a folk theorem.

Theorem 11.13 (Small Cycles). *Let* $\{f_\lambda\}$ *be a holomorphic family of nonlinear rational maps parametrized by* $\lambda \in \mathbb{C}$, *where*

$f_\lambda(0) = 0$ and $f'_\lambda(0) = \lambda$, so that $f_\lambda(z) = \lambda z +$ (higher terms). For a generic choice of λ on the unit circle, there are infinitely many periodic orbits in every neighborhood of $z = 0$, and hence zero is a Cremer point.

Proof (compare Problem 11-d). Let $\mathbb{D}_\epsilon$ be the disk of radius ϵ about the origin. For λ in some dense open subset U_ϵ of the unit circle, we will show that $\mathbb{D}_\epsilon$ contains a nonzero periodic orbit. For λ in the countable intersection $\bigcap U_{1/n}$, it will follow that f_λ has infinitely many periodic orbits converging to zero, as required.

Start with some root of unity $\lambda_0 = e^{2\pi i p/q} \neq 1$, and choose some positive $\epsilon' \leq \epsilon$ so that f_{λ_0} has no periodic points $z \neq 0$ of period $\leq q$ in the closed disk $\overline{\mathbb{D}}_{\epsilon'}$. Then the algebraic number of fixed points of $f_{\lambda_0}^{\circ k}$ in $\mathbb{D}_{\epsilon'}$ is equal to 1 for $1 \leq k < q$ (compare Problem 10-b), but is strictly greater than 1 for $k = q$. As we vary λ throughout a neighborhood of λ_0, this multiple fixed point for $f_{\lambda_0}^{\circ q}$ at the origin will split up into a collection of fixed points for $f_\lambda^{\circ q}$. Let $U(p/q)$ be a neighborhood of λ_0 which is small enough so that no periodic point of period $\leq q$ can cross through the boundary of $D_{\epsilon'}$. Then for any $\lambda \in U(p/q)$ with $\lambda \neq \lambda_0$, it follows that f_λ has an entire periodic orbit of period q contained within the neighborhood $\mathbb{D}_{\epsilon'} \subset \mathbb{D}_\epsilon$. The union U_ϵ of these open sets $U(p/q)$, for $0 < p/q < 1$, is evidently a dense and open subset of the circle, with the required property. □

Proof of Theorem 11.2 following Cremer. 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 $d^q - 1$ nonzero fixed points of $f^{\circ q}$ is equal to $\pm(\lambda^q - 1)$. Choosing q so that $|\lambda^q - 1| < 1$, it follows that at least one of these fixed points z satisfies

$$0 < |z|^{d^q} < |z|^{d^q - 1} \leq |\lambda^q - 1|.$$

Therefore, if the quantity $\liminf |\lambda^q - 1|^{1/d^q}$ is zero, it follows that there exist periodic points $z \neq 0$ in every neighborhood of zero.

In order to extend this argument to the case of a rational function f of degree $d \geq 2$, Cremer first noted 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. After a scale change, we may assume that $P(z)$ and $Q(z)$ have the form

$$P(z) = \star z^{d-1} + \cdots + \star z^2 + \lambda z, \qquad Q(z) = z^d + \cdots + 1,$$

where each $\star$ stands for a possibly nonzero coefficient. A brief computation then shows that $f^{\circ q}(z) = P_q(z)/Q_q(z)$ where $P_q(z)$ and $Q_q(z)$ have the form

$$P_q(z) = \star z^{d^q-1} + \cdots + \star z^2 + \lambda^q z, \qquad Q_q(z) = z^{d^q} + \cdots + 1.$$

Thus the equation for fixed points of $f^{\circ q}$ has the form

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

Now, if $\liminf |\lambda^q - 1|^{1/d^q} = 0$, then just as in the polynomial case we see that f has infinitely many periodic points in every neighborhood of zero, and hence that $0 \in J(f)$. This proves Theorem 11.2. $\square$

For the proof that Corollary 11.3 follows, see Problem 11-b.

Remark. As far as I know, Cremer never studied the small cycles property. His argument finds periodic points in every neighborhood of zero, but does not show that the entire periodic orbit is contained in a small neighborhood of zero. (However, compare Problem 11-d.)

Finally, let us show that Siegel disks really exist. We will describe a proof, due to Yoccoz, of the following weaker version of Siegel's Theorem. (Compare Herman [1986] or Douady [1987].) Again let $f_\lambda(z) = z^2 + \lambda z$ with $\lambda = e^{2\pi i \xi}$.

Theorem 11.14. *For Lebesgue almost every angle $\xi \in \mathbb{R}/\mathbb{Z}$, the quadratic map $f_\lambda(z)$ possesses a Siegel disk about the origin.*

The proof will depend on approximating multipliers λ on the unit circle by multipliers with $|\lambda| < 1$.

Definition. For each λ in the closed unit disk $\overline{\mathbb{D}}$, let $\sigma(\lambda)$ be the *conformal radius* of the largest linearizing neighborhood for f_λ, that is, the largest number σ such that there exists a univalent map $\psi_\lambda : \mathbb{D}_\sigma \to \mathbb{C}$, which conjugates the linear map $w \mapsto \lambda w$ on the open disk $\mathbb{D}_\sigma$ of radius σ to the map f_λ on $\mathbb{C}$, with

$$\psi_\lambda(0) = 0 \qquad \text{and} \qquad \psi'_\lambda(0) = 1,$$

taking $\sigma = 0$ if such a map cannot exist for any positive radius. Thus

$$\psi_\lambda(\lambda w) = f_\lambda(\psi_\lambda(w)) \qquad \text{for} \qquad 0 \le |w| < \sigma(\lambda). \tag{11:4}$$

Evidently, $\sigma(\lambda) > 0$ whenever f_λ has a Siegel disk about the origin, and this number does indeed measure the size of the disk in some invariant sense. Similarly $\sigma(\lambda) > 0$ for $0 < |\lambda| < 1$. However, $\sigma(\lambda) = 0$ whenever f_λ has a parabolic or Cremer point at the origin, and also when $\lambda = 0$. If f_λ has a Siegel disk, note that this conformal radius function cannot be continuous at λ, since parabolic or Cremer values for λ are everywhere dense on the unit circle.

Recall that a real-valued function σ on a topological space is said to be *upper semicontinuous* if

$$\limsup_{x\to x_0,\ x\neq x_0} \sigma(x) \ \leq\ \sigma(x_0)$$

for every x_0 in the space, or equivalently if the set of x with $\sigma(x) \geq \sigma_0$ is closed for every $\sigma_0 \in \mathbb{R}$.

Lemma 11.15. *This conformal radius function $\sigma : \overline{\mathbb{D}} \to \mathbb{R}$ is bounded and upper semicontinuous. Furthermore, for $|\lambda| < 1$ we can write $\sigma(\lambda) = |\eta(\lambda)|$, where the function $\lambda \mapsto \eta(\lambda)$ is holomorphic throughout the open unit disk.*

Proof. First note that $\sigma(\lambda) \leq 2$ for all $\lambda \in \overline{\mathbb{D}}$. In fact, if $|z| > 2$ and $|\lambda| \leq 1$, then $|f_\lambda(z)| = |z(z+\lambda)| > |z|$, and it follows easily that z lies in the basin of infinity for f_λ. Therefore any map $\mathbb{D}_\sigma \to \mathbb{C}$ satisfying $(11:4)$ must take values in $\mathbb{D}_2$ and hence must satisfy $\sigma \leq 2$ by the Schwarz Lemma.

To see that σ is upper semicontinuous, note that $\sigma(z) \geq \sigma_0$ if and only if there is a univalent map $\mathbb{D}_{\sigma_0} \to \mathbb{D}_2$ satisfying $(11:4)$. But the collection of all holomorphic maps from $\mathbb{D}_{\sigma_0}$ to $\mathbb{D}_2$ forms a normal family. Hence any sequence of such maps contains a subsequence which is locally uniformly convergent throughout $\mathbb{D}_{\sigma_0}$. In particular, given a sequence of univalent maps ψ_{λ_k} satisfying $(11:4)$, we can find a convergent subsequence, and it is not hard to check that the limit function will also be univalent and satisfy $(11:4)$. Therefore, the set of $\lambda \in \overline{\mathbb{D}}$ with $\sigma(\lambda) \geq \sigma_0$ is closed, as required.

Now let us specialize to the case $0 < |\lambda| < 1$. We can compute the conformal radius $\sigma(\lambda)$ for such values of λ as follows. Let $\mathcal{A}_\lambda$ be the attracting basin of the fixed point zero under f_λ. Evidently the critical point $-\lambda/2$ of f_λ must lie in this attracting basin. As in Theorem 8.2 and Corollary 8.4, the Kœnigs map $\phi_\lambda : \mathcal{A}_\lambda \to \mathbb{C}$ can be defined by the formula

$$\phi_\lambda(z) \ =\ \lim_{n\to\infty} f_\lambda^{\circ n}(z)/\lambda^n. \qquad (11:5)$$

Since this limit converges locally uniformly, $\phi_\lambda(z)$ depends holomorphically on both variables. (Compare Remark 8.3.) In particular, its value

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

at the critical point of f_λ is a holomorphic function of λ. It follows easily from Lemma 8.5 that the absolute value $|\eta(\lambda)|$ is precisely equal to the conformal radius $\sigma(\lambda)$, as defined above, for all $\lambda \in \mathbb{D} \smallsetminus \{0\}$. Furthermore, since $\sigma(0) = 0$, it follows from upper semicontinuity that $\eta(\lambda) \to 0$ as $\lambda \to 0$, so that η has a removable singularity at the origin. This completes the proof of Lemma 11.15. □

Remark. We can actually compute this function within the open disk by noting that $\eta(\lambda)$ can be described as the limit as $i \to \infty$ of the numbers

$$\eta_i \;=\; f_\lambda^{\circ i}(-\lambda/2)/\lambda^i .$$

These can be determined recursively by the formula

$$\eta_0 = -\lambda/2\,, \quad \eta_{i+1} \;=\; \eta_i + \lambda^{i-1}\eta_i^2 .$$

This procedure can also be used to compute the coefficients of the power series expansion of η about the origin, which takes the form

$$\eta(\lambda) \;=\; -\frac{\lambda}{4} + \frac{\lambda^2}{16} + \frac{\lambda^3}{16} + \frac{\lambda^4}{32} + \frac{9\lambda^5}{256} + \frac{\lambda^6}{256} + \frac{7\lambda^7}{256} - \frac{3\lambda^8}{512} + \cdots .$$

Details will be left to the reader.

Corollary 11.16. *For $|\lambda_0| = 1$, the map f_{λ_0} has either a Cremer point or a parabolic point at the origin if and only if the limit*

$$\lim_{\lambda \to \lambda_0,\ |\lambda|<1} \eta(\lambda)$$

is defined and equal to zero.

The proof is immediate. Now, to complete the proof of Theorem 11.14, we need only quote the following. Let c_0 be any complex constant.

Theorem of F. and M. Riesz [1916]. *Let $\eta : \mathbb{D} \to \mathbb{C}$ be a nonconstant bounded holomorphic function, and let $S(c_0)$ be the set of angles $\xi \in \mathbb{R}/\mathbb{Z}$ such that the radial limit*

$$\lim_{r \nearrow 1} \eta(r \exp(2\pi i \xi))$$

is defined and equal to c_0. Then $S(c_0)$ has Lebesgue measure zero.

(Compare Theorem 17.4.) For the proof of this result, see Theorem A.3 in Appendix A. Clearly Theorem 11.14 follows immediately. □

Remark. This argument also shows that

$$\sigma(e^{2\pi i\xi}) \;\geq\; \limsup_{r\nearrow 1} |\eta(re^{2\pi i\xi})|$$

for every $\xi \in \mathbb{R}/\mathbb{Z}$. In fact, according to Yoccoz [1995], equality always holds here.

Unsolved Problems. Although the theorems of Bryuno, Yoccoz, and Perez-Marco are very sharp, they do not answer all questions about local behavior near an irrationally indifferent fixed point. For example, there is a very complicated local structure about any Cremer point (compare Perez-Marco [1997]), yet there is not a single example which is well understood. It is not known whether any rational function can have a Cremer point without small cycles, and it is not known whether Cremer Julia sets sometimes, always, or never have positive Lebesgue measure. Similarly, it is not known whether it is possible for a nonlinear rational function to have a Siegel disk for which the Bryuno condition is not satisfied.

In Theorem 8.6 and Theorem 10.15, we found a rather direct relationship between critical points and attracting or parabolic orbits. For Siegel or Cremer orbits, the relationship is less direct.

Definition. By the *postcritical set* $P = P(f)$ of a rational map f we will mean the union of all forward images $f^{\circ k}(c)$ with $k > 0$, where c ranges over the critical points. We will be particularly interested in the topological closure $\overline{P}(f)$ of this set.

Theorem 11.17. *Every Cremer fixed point or periodic point for a rational map is contained in the postcritical closure $\overline{P}(f)$. Similarly, the boundary of any Siegel disk or cycle of Siegel disks is contained in $\overline{P}(f)$.*

Proof. (Compare §19.) We will work with the open sets $U = \widehat{\mathbb{C}} \smallsetminus \overline{P}$ and $V = f^{-1}(U)$. Since $f^{-1}(\overline{P}) \supset \overline{P}$, it follows that $V \subset U$. Since there are no critical values in U, it follows that f maps V onto U by a d-fold covering map, or more precisely that f maps each connected component of V onto some connected component of U by a covering map.

We may assume that $\overline{P}$ contains at least three distinct points, for otherwise U would be a twice punctured sphere, hence its covering space V would also be a twice punctured sphere, equal to U. It would then follow easily that f must be conjugate to a map of the form $z \mapsto z^{\pm d}$, with no Cremer points and no Siegel disks.

Thus we may assume that every connected component of V or U is conformally hyperbolic. Consider a fixed point $z_0 = f(z_0)$ which belongs to

U and hence to V. Let $V_0 \subset U_0$ be the connected components containing z_0. If $V_0 = U_0$, then V_0 maps into itself under f and hence is contained in the Fatou set. Thus in this case z_0 cannot be a Cremer point.

Now suppose that V_0 is strictly smaller than U_0. Then the inclusion $V_0 \to U_0$ strictly decreases Poincaré distances by (2 : 6), that is,

$$\mathrm{dist}_V(x,y) \;>\; \mathrm{dist}_U(x,y)$$

for every $x \neq y$ in V_0. On the other hand, by Theorem 2.11, V maps to U by a local isometry, so that

$$\mathrm{dist}_U(f(x), f(y)) \;=\; \mathrm{dist}_V(x,y)$$

whenever $x, y \in V$ are sufficiently close to each other. Therefore

$$\mathrm{dist}_U(f(x)\,,\, f(y)) \;>\; \mathrm{dist}_U(x,y) \qquad (11:6)$$

whenever $x \neq y$ in V_0 are sufficiently close to each other. It follows that the fixed point z_0 must be strictly repelling: Again, it cannot be a Cremer point.

To deal with the boundary of a Siegel disk $\Delta = f(\Delta)$, we must work just a little harder. First note that Δ, with its center point z_0 removed, is naturally foliated into f-invariant circles. The intersection $\overline{P} \cap \Delta \smallsetminus \{z_0\}$ consists of at most finitely many of these circles. Thus if a component U_0 of $\widehat{\mathbb{C}} \smallsetminus \overline{P}$ intersects the boundary $\partial\Delta$, then it must contain an entire neighborhood of $\partial\Delta$ within Δ. In particular, it must contain every invariant circle C which is sufficiently close to the boundary. Similarly, one component V_0 of $f^{-1}(U_0)$ must contain every such circle. If $V_0 = U_0$ then, arguing as above, U_0 is contained in the Fatou set and cannot intersect the boundary of Δ. On the other hand, if V_0 is strictly smaller than U_0 then, as in (11: 6), f restricted to V_0 must strictly increase the distance $\mathrm{dist}_U(x,y)$ between nearby points and similarly must map any smooth path to a path of strictly larger arclength. In particular, it must map each invariant circle $C \subset V_0$ onto a longer circle. But this is impossible since f maps C diffeomorphically onto itself. This proves that every fixed Cremer point or Siegel disk boundary must be contained in $\overline{P}$. The corresponding statement for a cycle of Cremer points or Siegel disks follows by applying the above argument to a suitable iterate $f^{\circ k}$ and noting that $P(f^{\circ k}) = P(f)$. $\square$

A different proof of Theorem 11.17 will be given in Corollary 14.4.

For further studies of the topology and geometry of Siegel disks, see for example Rogers [1998], Zakeri [1999], Yampolsky and Zakeri [2001], Buff and Cheritat [2003], as well as Petersen and Zakeri [2004].

Concluding Problems

Problem 11-a. Dirichlet. Use the "pigeon-hole principle" to show that for any irrational number x there are infinitely many fractions p/q with

$$\left|x - \frac{p}{q}\right| < \frac{1}{q^2}.$$

In fact, for any integer $Q > 1$ cut the circle $\mathbb{R}/\mathbb{Z}$ into Q half-open intervals of length $1/Q$, and consider the $Q+1$ numbers $0, x, 2x, \ldots, Qx$ reduced modulo $\mathbb{Z}$. Since at least two of these must belong to the same interval, conclude that there exist integers p and $1 \le q \le Q$ with $|qx - p| < 1/Q$, hence

$$\left|x - \frac{p}{q}\right| < \frac{1}{qQ} \le \frac{1}{q^2}.$$

Conclude the the class of Diophantine numbers $\mathcal{D}(\kappa)$ is vacuous for $0 < \kappa < 2$.

Problem 11-b. Generic angles. Given a completely arbitrary sequence of positive real numbers $\epsilon_1, \epsilon_2, \ldots \searrow 0$, let $S(q_0)$ be the set of all real numbers ξ such that

$$\left|\xi - \frac{p}{q}\right| < \epsilon_q$$

for some fraction p/q in lowest terms with $q > q_0$. **(1)** Show that $S(q_0)$ is a dense open subset of $\mathbb{R}$ and conclude that the intersection $S = \bigcap_{q_0} S(q_0)$, consisting of all ξ for which this condition is satisfied for infinitely many p/q, is a countable intersection of dense open sets. **(2)** As an example, taking $\epsilon_q = 2^{-q!}$ conclude that a generic real number belongs to the set S, and hence satisfies Cremer's condition that $\liminf |\lambda^q - 1|^{1/d^q} = 0$ for every degree d. (Compare Theorem 11.2.)

Problem 11-c. Cremer [1938]. **(1)** 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 (following Poincaré) 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. **(2)** Now suppose that we choose the a_n inductively, always equal to zero or one, so that

$|a_n + X_n| \geq 1/2$. 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. **(3)** Choosing the a_n more carefully, show that we can even choose $f(z)$ to be an entire function.

Problem 11-d. Small cycles. Suppose that

$$\limsup_{q\to\infty} \frac{\log\log(1/|\lambda^q - 1|)}{q} \;>\; \log d \;>\; 0.$$

Modify the proof of Theorem 11.2 to show that: *Any fixed point of multiplier λ for a rational function f of degree d has the small cycle property.* First choose $\epsilon > 0$ so that

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

or equivalently

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

for infinitely many q. The proof of Theorem 11.2 then constructs points z_q of period q with $|z_q| < \exp(-e^{\epsilon q})$. Now use Taylor's Theorem to find $\delta > 0$ so that $|f(z)| < e^{\epsilon}|z|$ for $|z| < \delta$, and hence $|f^{\circ q}(z)| < \delta$ for $|z| < e^{-q\epsilon}\delta$. Finally, note that $\exp(-e^{\epsilon q}) < e^{-q\epsilon}\delta$ for large q, to conclude that f has small cycles.

PERIODIC POINTS: GLOBAL THEORY

§12. The Holomorphic Fixed Point Formula

The number of fixed points of a rational map $f : \hat{\mathbb{C}} \to \hat{\mathbb{C}}$ of degree $d \geq 0$, can be counted as follows.

Lemma 12.1. *If f is not the identity map, then f has exactly $d+1$ fixed points, counted with multiplicity.*

Here the *multiplicity* of a finite fixed point $\hat{z} = f(\hat{z})$ is defined to be the unique integer $m \geq 1$ for which the power series expansion of $f(z) - z$ about $\hat{z}$ has the form

$$f(z) - z \;=\; a_m(z - \hat{z})^m + a_{m+1}(z - \hat{z})^{m+1} + \cdots$$

with $a_m \neq 0$. Thus $m \geq 2$ if and only if the multiplier λ at $\hat{z}$ is exactly 1. (Note that $\hat{z}$ is then a parabolic point with $m-1$ attracting petals, each of which maps into itself. Compare §10.) In the special case of a fixed point at infinity, we introduce the local uniformizing parameter $\zeta = \phi(z) = 1/z$ and define the multiplicity of f at infinity to be the multiplicity of the map $\phi \circ f \circ \phi^{-1}$ at the point $\phi(\infty) = 0$. As an example, any polynomial map of degree $d \geq 2$ has a fixed point at infinity with multiplier $\lambda = 0$ and hence with multiplicity $m = 1$, and therefore has d finite fixed points counted with multiplicity. On the other hand, the map $f(z) = z + 1$ has a fixed point of multiplicity $m = 2$ at infinity.

Proof of Lemma 12.1. Conjugating f by a fractional linear automorphism if necessary, we may assume that the point at infinity is *not* fixed by f. If we write f as a quotient $f(z) = p(z)/q(z)$ of two polynomials which have no common factor, this means that the degrees of $p(z)$ and $q(z)$ satisfy

$$\text{degree}(p(z)) \;\leq\; \text{degree}(q(z)) \;=\; d.$$

In this case, the equation $f(z) = z$ is equivalent to the polynomial equation $p(z) = z\,q(z)$ of degree $d + 1$, and hence has $d + 1$ solutions, counted with multiplicity, as required. $\square$

Remark. This algebraic multiplicity m is known to topologists as the *Lefschetz fixed point index*. For any map $f : M \to M$ of a compact n-dimensional manifold into itself with only finitely many fixed points, the Lefschetz index of each fixed point p_j is a uniquely defined integer $\Lambda(p_j)$, and the sum over all of the fixed points can be computed from the homology of f by the Lefschetz formula

$$\sum_j \Lambda(p_j) \;=\; \sum_{i=0}^{n} (-1)^i \operatorname{trace}\Big(f_* : H_i(M;\mathbb{R}) \to H_i(M;\mathbb{R})\Big).$$

(See, for example, Franks [1982].) In our case, with M the Riemann sphere and f rational of degree d, there is a contribution of $+1$ from the 0-dimensional homology and $+d$ from the 2-dimensional homology, so the sum of the Lefschetz indices is $d+1$.

Both Fatou and Julia made use of a "well-known" relation between the multipliers at the fixed points of a rational map. First consider an isolated fixed point $\hat{z} = f(\hat{z})$ where $f : U \to \mathbb{C}$ is a holomorphic function on a connected open set $U \subset \mathbb{C}$. The ***residue fixed point index*** of f at $\hat{z}$ is defined to be the complex number

$$\iota(f,\, \hat{z}) \;=\; \frac{1}{2\pi i}\oint \frac{dz}{z - f(z)} \qquad (12:1)$$

where we integrate in a small loop in the positive direction around $\hat{z}$.

Lemma 12.2. *If the multiplier $\lambda = f'(\hat{z})$ is not equal to $+1$, then this residue fixed point index is given by*

$$\iota(f,\, \hat{z}) \;=\; \frac{1}{1-\lambda} \;\neq\; 0\,. \qquad (12:2)$$

Proof. Without loss of generality, we may assume that $\hat{z} = 0$. Expanding f as a power series, we can write

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

Since $\lambda \neq 1$, it follows that $z - f(z) \;=\; (1-\lambda)z(1+O(z))$ and hence

$$\frac{1}{z - f(z)} \;=\; \frac{1+O(z)}{(1-\lambda)z} \;=\; \frac{1}{(1-\lambda)z} \,+\, O(1)\,.$$

Integrating this expression around a small loop $|z| = \epsilon$ and dividing by $2\pi i$, we evidently obtain a residue of $\iota = 1/(1-\lambda)$, as asserted. □

Note: This computation breaks down completely in the special case $\lambda = 1$. The residue index $\iota(f, \hat{z})$ is still well defined and finite when $\lambda = 1$, but the formula (12 : 2) no longer makes sense. For information about this case see Problem 12-a as well as Lemma 12.9.

More generally, given any isolated fixed point of a holomorphic map $F : S \to S$ from a Riemann surface to itself, we can choose some local coordinate z and then compute the index $\iota(f,\, \hat{z})$ for the associated local map $z \mapsto f(z)$.

Lemma 12.3. *This residue, computed in terms of a local coordinate near the fixed point, does not depend on any particular choice of local coordinate.*

In the generic case of a fixed point of multiplicity $m = 1$, this follows immediately from Lemma 12.2, since the multiplier $\lambda \neq 1$ clearly does not depend on the particular choice of coordinate chart. Consider then a fixed point $\hat{z} = f(\hat{z})$ with multiplicity $m \geq 2$. Assume for convenience that $\hat{z} = 0$. We can choose a family of perturbed maps $f_\alpha(z) = f(z) + \alpha$ so that the m-fold fixed point for $\alpha = 0$ splits up into m nearby fixed points for $\alpha \neq 0$. In fact, since the derivative $f'(z)$ is not identically equal to $+1$, we can choose ϵ so that $f'(z) \neq 1$ for $0 < |z| < \epsilon$, and it follows that f_α has only simple fixed points in $\mathbb{D}_\epsilon$ and no fixed points on $\partial\mathbb{D}_\epsilon$ for small $\alpha \neq 0$. Next note that the sum of residue indices over the m simple fixed points $z_j \in \mathbb{D}_\epsilon$ can be expressed as

$$\sum_j \iota(f_\alpha, z_j) \;=\; \frac{1}{2\pi i}\oint_{\partial\mathbb{D}_\epsilon} \frac{dz}{z - f_\alpha(z)}\,.$$

Evidently this integral converges to

$$\iota(f, 0) \;=\; \frac{1}{2\pi i}\oint_{\partial\mathbb{D}_\epsilon} \frac{dz}{z - f(z)}$$

as $\alpha \to 0$. Since the indices $\iota(f_\alpha, z_j)$ are invariant under a holomorphic change of coordinates, it follows that $\iota(f, 0)$ is also invariant. $\square$

Now suppose that our Riemann surface S is the Riemann sphere. (For analogous formulas on other Riemann surfaces, see Remark 12.5.)

Theorem 12.4 (Rational Fixed Point Theorem). *For any rational $f : \widehat{\mathbb{C}} \to \widehat{\mathbb{C}}$ which is not the identity map, we have the relation*

$$\sum_{z = f(z)} \iota(f, z) \;=\; 1, \qquad (12:3)$$

to be summed over all fixed points.

Proof. Conjugating f by a linear fractional automorphism if necessary, we may assume that $f(\infty) \neq 0, \infty$. Then $f(z)$ converges to $f(\infty) \in \mathbb{C} \smallsetminus \{0\}$ as $z \to \infty$, hence

$$\frac{1}{z - f(z)} - \frac{1}{z} \;=\; \frac{f(z)}{z(z - f(z))} \;\sim\; \frac{f(\infty)}{z^2}$$

as $z \to \infty$. Let $\partial\mathbb{D}_r$ be the loop $|z| = r$. It follows easily that the integral

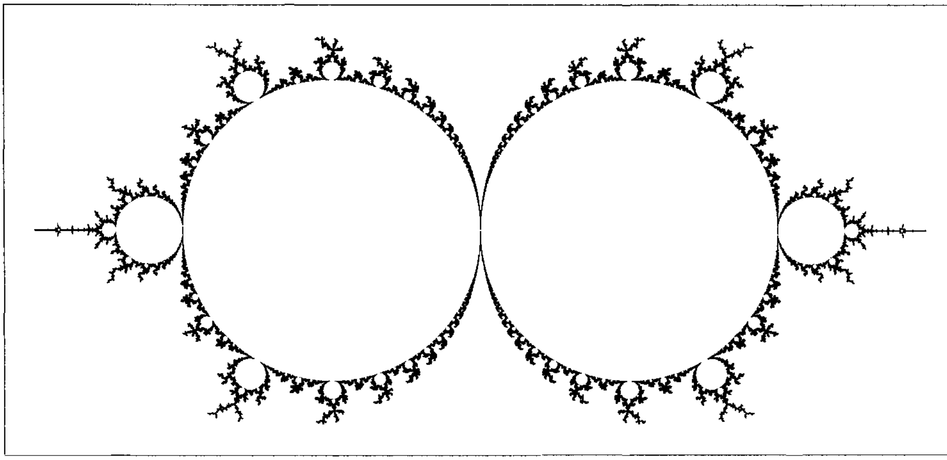


Figure 29. Double Mandelbrot set: The bifurcation locus in the λ-parameter plane for the family of quadratic maps $z \mapsto z^2 + \lambda z$. The figure is centered at $\lambda = 1$.

of this difference around $\partial\mathbb{D}_r$ converges to zero as $r \nearrow \infty$. Hence

$$\frac{1}{2\pi i}\oint_{\partial\mathbb{D}_r} \frac{dz}{z - f(z)} \;=\; \frac{1}{2\pi i}\oint_{\partial\mathbb{D}_r} \frac{dz}{z} \;=\; +1,$$

Evidently the integral on the left is equal to the sum of the residues $\iota(f, z_j)$ at the various fixed points of f, hence the required summation formula. □

Examples. If f is a polynomial map of degree $d \geq 2$, then f has a fixed point of multiplier zero at infinity. *Since $\iota(f,\infty) = +1$, it follows that the sum of $\iota(f,z)$ over all finite fixed points is equal to zero.* For a generic quadratic polynomial with two simple fixed points of multiplier λ and μ, the relation

$$\frac{1}{1-\lambda} + \frac{1}{1-\mu} \;=\; 0$$

reduces easily to $\lambda + \mu = 2$. For example, for the family of maps

$$f_\lambda(z) \;=\; z^2 + \lambda z$$

there is a fixed point of multiplier λ at $z = 0$ and a fixed point of multiplier $2 - \lambda$ at $z = 1 - \lambda$. Note that these two fixed point coalesce into a single fixed point of multiplicity 2 as $\lambda \to 1$. Evidently f_λ has an attracting fixed point if and only if λ belongs either to the unit disk $\mathbb{D}$ or to the translated unit disk $2 + \mathbb{D}$. Both disks are clearly visible in Figure 29 which shows the λ parameter plane. (Compare Appendix G.) The two disks are tangent to each other at the center of symmetry $\lambda = 1$.

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

$$\frac{1}{1-\lambda} + \frac{1}{1-\mu} = 1$$

simplifies to $\lambda\mu = 1$. The family of maps $f_\lambda(z) = \lambda z + 1$ is typical. Such a map has one fixed point of multiplier λ at $z = 1/(1-\lambda)$ and one fixed point of multiplier $1/\lambda$ at infinity. Just as in the quadratic polynomial example, as $\lambda \to 1$ the two fixed points coalesce, so that there is a single fixed point of multiplicity 2 for the limit map $f_1(z) = z + 1$.

More generally, let f be any rational map which has a fixed point with multiplier very close to 1, and hence with $|\iota|$ large. Then it follows immediately from the Fixed Point Formula that f must have at least one other fixed point with $|\iota|$ large, and hence with λ either close to 1 or equal to 1.

Remark 12.5. For a far-reaching generalization of this fixed point theorem, see Atiyah and Bott [1966]. In particular, for a holomorphic map f from a compact Riemann surface of genus g to itself, the Atiyah-Bott formula implies that the sum of the residue fixed point indices is given by

$$\sum \iota \;=\; 1 - \overline{\tau}, \qquad (12:4)$$

where τ is the trace of the induced map from the g-dimensional vector space of holomorphic 1-forms to itself, and where the overline stands for complex conjugation. In the special case of the Riemann sphere, with $g = 0$, there are no holomorphic 1-forms, so this formula reduces to $(12:3)$. For other examples, see Problem 12-e; and for a different generalization of Theorem 12.4, to higher dimensional projective spaces, see Ueda [1995].

Lemma 12.6. *A fixed point with multiplier $\lambda \neq 1$ is attracting if and only if its residue fixed point index ι has real part* $\mathrm{Re}(\iota) > \frac{1}{2}$.

Proof. Geometrically, this is proved by noting that a fixed point with multiplier λ is attracting if and only if $1-\lambda$ belongs to the disk $1+\mathbb{D}$, having the origin as boundary point. It is easy to check that the map $z \mapsto 1/z$ carries this disk $1+\mathbb{D}$ precisely onto the half-plane $\mathrm{Re}(z) > 1/2$. Computationally, this can be proved by noting that $\frac{1}{2} < \mathrm{Re}(\frac{1}{1-\lambda})$ if and only if $1 < 1/(1-\lambda) + 1/(1-\overline{\lambda})$. Multiplying both sides by $(1-\lambda)(1-\overline{\lambda}) > 0$, we easily obtain the equivalent inequality $\lambda\overline{\lambda} < 1$. □

One important consequence is the following.

Corollary 12.7. *Every rational map of degree* $d \geq 2$ *must have either a repelling fixed point or a parabolic fixed point with* $\lambda = 1$*, or both.*

Proof. If there is no fixed point of multiplier $\lambda = 1$, then there must be $d + 1$ distinct fixed points. If these were all attracting or indifferent, then each index would have real part $\mathrm{Re}(\iota) \geq \frac{1}{2}$, and hence the sum would have real part greater than or equal to $\frac{d+1}{2} > 1$, but this would contradict Theorem 12.4. □

Since repelling points and parabolic points both belong to the Julia set, this yields a constructive proof of the following. (Compare Lemma 4.8.)

Corollary 12.8. *The Julia set for a nonlinear rational map is always nonvacuous.*

The Résidu Itératif. When studying a fixed point of multiplier $\lambda = 1$, it is often convenient to use a modified form of the residue index which is due to Écalle. If $\hat{z}$ is a fixed point of multiplicity m then by definition the ***résidu itératif*** is the difference

$$\text{résit}(f, \hat{z}) = m/2 \, - \, \iota(f, \hat{z})\,.$$

(It will be convenient to use this notation even in the case $m = 1$.)

Lemma 12.9. *If* $\lambda = 1$ *or equivalently if* $m \geq 2$*, then for any integer* $k \neq 0$ *we have the formula*

$$\text{résit}(f^{\circ k}, \hat{z}) \;=\; \text{résit}(f, \hat{z})/k\,. \qquad (12:5)$$

Proof. First consider a fixed point of multiplier $\lambda = 1 - \epsilon$ where the complex number ϵ is small but non-zero. If $k > 0$ then we can write

$$1-\lambda^k \;=\; (1-\lambda)(1+\lambda+\cdots+\lambda^{k-1}) \;=\; \epsilon\Big(k-\epsilon(1+2+3+\cdots+(k-1))+O(\epsilon^2)\Big),$$

and it follows easily that

$$\frac{1}{1-\lambda^k} \;=\; \frac{1}{k\epsilon} + \frac{k-1}{2k} + o(1)$$

where the remainder term $o(1)$ tends to zero as $\epsilon \to 0$. In fact it is not hard to see that this last estimate holds also when $k < 0$. Rearranging terms and remembering that $\epsilon = 1 - \lambda$, we see that this takes the form

$$\text{résit}(f^{\circ k}, \hat{z}) \;=\; \text{résit}(f, \hat{z})/k \,+\, o(1) \qquad \text{as} \qquad \lambda \to 1\,.$$

Now consider a map f_0 with a fixed point $\hat{z} = 0$ of multiplicity $m \geq 2$. As in the proof of Lemma 12.3, see can perturb f_0 to obtain maps f_α with m nearby simple fixed points, all necessarily with multiplier close to $+1$.

Summing the résidu itératif over these m points, we get

$$\sum_j \text{résit}(f_\alpha^{\circ k}, z_j) \;=\; \sum_j \text{résit}(f_\alpha, z_j)/k \,+\, o(1)\,, \qquad (12:6)$$

where the remainder term tends to zero as $\alpha \to 0$. Since

$$\lim_{\alpha\to 0} \sum_j \text{résit}(f_\alpha, z_j) \;=\; \text{résit}(f, \hat{z})\,, \qquad (12:7)$$

the required formula $(12:5)$ is an immediate consequence of $(12:6)$. □

Definition. Following Adam Epstein [1999], a fixed point $\hat{z}$ of multiplicity $m \geq 2$ will be called

parabolic repelling if $\text{Re}(\text{résit}(f, \hat{z})) > 0$, and
parabolic attracting if $\text{Re}(\text{résit}(f, \hat{z})) < 0$.

(Buff [2003] uses the terms *virtually repelling* and *virtually attracting* for the same concepts.) In the case of a simple fixed point with $m = 1$, it follows immediately from Lemma 12.6 that the fixed point is actually

repelling if and only if $\text{Re}(\text{résit}(f, \hat{z})) > 0$, and
attracting if and only if $\text{Re}(\text{résit}(f, \hat{z})) < 0$.

Now suppose that we perturb f throughout some neighborhood of the m-fold fixed point $\hat{z}$ so that it splits up into m simple fixed points. (Compare the proofs of Lemmas 12.3 and 12.9.)

Theorem 12.10 (Buff and Epstein). *It is possible to choose perturbed maps arbitrarily close to f so that all of the m resulting simple fixed points are repelling if and only if*

$$\text{Re}(\text{résit}(f, \hat{z})) \;\geq\; 0\,.$$

Similarly, it is possible to perturb so that all of these simple fixed points are attracting if and only if $\text{Re}(\text{résit}(f, \hat{z})) \leq 0$.

For the construction of such special perturbations, the reader is referred to Buff [2003]. (For the special case of a quadratic rational map, compare Problem 12-b.) On the other hand, the proof of Theorem 12.10 in the other direction is straightforward: If we are given perturbations which yield only repelling (or attracting) fixed points, then the required inequality for $\text{Re}\text{résit}(f, \hat{z})$ follows immediately from equation $(12:7)$ together with Lemma 12.6. □

The Time One Map for a Flow. Now consider the local flow f_t generated by a holomorphic differential equation $dz/dt = v(z)$. Thus, for each fixed z_0, the map $t \mapsto z(t) = f_t(z_0)$ must satisfy $dz/dt = v(z)$ with initial condition $z(0) = z_0$. Here $f_t(z)$ is defined for $|t| < \epsilon$, where $\epsilon > 0$

depends on z. In the special case where $v(0) = 0$, we can say also that every f_t is defined for all z in *some* neighborhood of 0. The function $f_t(z)$ is holomorphic as a function of two variables.

Lemma 12.11. *If both $v(z)$ and its derivative $v'(z)$ vanish at $z = 0$, then each f_t with $t \neq 0$ has a fixed point of multiplicity $m \geq 2$ at 0, and*

$$\text{résit}(f_t, 0) \;=\; \frac{1}{2\pi i\, t}\oint \frac{dz}{v(z)}\,.$$

Proof. The identity $\text{résit}(f_t, 0) = \text{résit}(f_1, 0)/t$ for $t \neq 0$ follows easily from Lemma 12.9 when t is rational. For other values of t, it then follows by analytic continuation. Thus it suffices to compute $\text{résit}(f_1, 0)$. From the Taylor series

$$f_t(z) = z + tv(z) + \cdots + a_k(z)t^k/k! + \cdots$$

where $a_k(z) = \partial^k f_t(z)/\partial t^k = v(z)\, da_k(z)/dz$, we see that

$$f_t(z) \;=\; z + tv(z) + t^2 v(z)v'(z)/2 + \cdots \;=\; z + tv(z)(1 + O(t))$$

as $t \to 0$. Therefore the residue index ι at the origin satisfies

$$2\pi i\; \iota(f_t, 0) \;=\; \oint \frac{dz}{z - f_t(z)} \;=\; \frac{-1}{t}\oint \frac{dz}{v(z)} \;+\; O(1).$$

Changing sign and adding $m/2$, it follows that

$$2\pi i \;\text{résit}(f_t, 0) \;=\; \frac{1}{t}\oint \frac{dz}{v} \;+ O(1) \qquad \text{as} \qquad t \to 0\,.$$

Multiplying by t and then letting $t \to 0$ this proves that

$$2\pi i \;\text{résit}(f_1, 0) \;=\; \oint \frac{dz}{v}$$

as required. $\square$

Remark 12.12. Recall from Remark 10.12 that the Écalle-Voronin theory of parabolic fixed points leads to an invariant $I(f, \hat{z})$ for fixed points of multiplicity $m \geq 2$. (See equation (10 : 8)). Buff and Epstein [2002] have shown by a careful computation of Fatou coordinates that

$$I(f, \hat{z}) \;+\; 2\pi i \;\text{résit}(f, \hat{z}) \;=\; 0\,. \qquad (12:8)$$

In the special case of the time one map of a flow, there is a very easy proof of this identity (12 : 8). Given a differential equation of the form $dz/dt = v(z)$ with $v(0) = 0$, we can introduce the differential $d\alpha = dz/v(z)$ throughout some punctured disk $\mathbb{D}_\epsilon \smallsetminus \{0\}$. On each attracting or repelling petal $\mathcal{P}_j$

we can choose an integral $\alpha_j : \mathcal{P}_j \to \mathbb{C}$ of the differential $d\alpha$. Along each solution $z = z(t)$ to our differential equation within $\mathcal{P}_j$ we then have

$$\frac{d\alpha_j}{dt} = \frac{d\alpha_j}{dz}\frac{dz}{dt} = \frac{1}{v}v = +1\,.$$

Therefore α_j is a Fatou coordinate for the map f_1. On each overlap $\mathcal{P}_j \cap \mathcal{P}_{j+1}$ the difference $\alpha_{j+1} - \alpha_j$ takes a constant value c_j, and by definition

$$I(f_1, 0) = \sum_{j \in \mathbb{Z}/2n} c_j\,.$$

On the other hand, if we choose a base point z_j in each $\mathcal{P}_{j-1} \cap \mathcal{P}_j$, then

$$\oint d\alpha = \sum_j \int_{z_j}^{z_{j+1}} d\alpha_j = \sum_j \Big(\alpha_j(z_{j+1}) - \alpha_j(z_j)\Big)\,,$$

where

$$\oint d\alpha = 2\pi i\, \text{résit}(f_1, 0)$$

by Lemma 12.11. Combining the last three equations, we easily obtain the required equation $(12:8)$ for the map f_1.

Concluding Problems

Problem 12-a. An index computation. Consider a fixed point of multiplicity $n+1 \ge 2$ which has been put into the normal form

$$f(z) = z + a\,z^{n+1} + b\,z^{2n+1} + \text{(higher terms)}.$$

(Compare Problem 10-d.) Show directly from the definition $(12:1)$ that the index $\iota(f,0)$ is equal to the ratio b/a^2. In particular, if $a = 1$ so that $f(z) = z + z^{n+1} + bz^{2n+1} + \text{(higher terms)}$, show that $\iota(f,0) = b$.

Problem 12-b. Quadratic rational maps. Consider the family of rational maps

$$f(z) = z\frac{z+\mu}{\nu z + 1},$$

where $\mu\nu \ne 1$ so that f has degree 2. Thus f has a fixed point of multiplier μ at the origin, as well as a fixed point of multiplier ν at infinity. **(1)** If $\mu = 1$, show that there is a double fixed point at the origin (that is, a fixed point of multiplicity 2), with index

$$\iota(f,0) = 1 - \iota(f,\infty) = \nu/(\nu - 1).$$

(The picture in the ν-parameter plane, for this family of maps with fixed point multipliers 1 and ν looks very much like the left half of the double Mandelbrot set of Figure 29, p. 145. Compare Milnor [1993, Figure 5].)

(2) For μ close to 1 but $\mu \neq 1$, show that the single fixed point at zero splits into two fixed points, one at zero and one at $(1-\mu)/(1-\nu)$, with indices which are very large in absolute value, but with sum equal to $\nu/(\nu-1)$. Show that *any* two large indices with sum $\nu/(\nu-1)$ can occur for suitable choice of μ. (Compare the discussion following Theorem 12.10.)

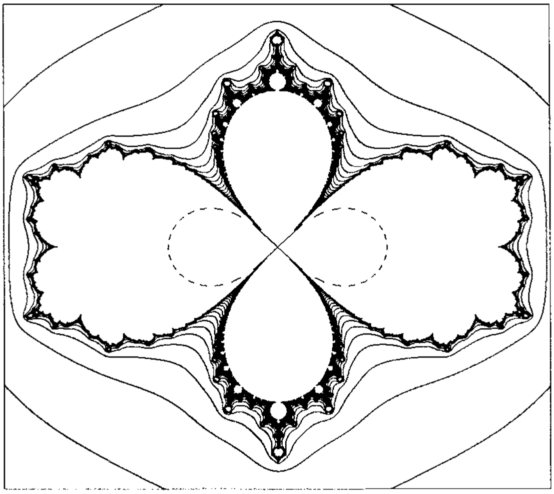


Figure 30. Parameter space picture for the family of cubic maps $z \mapsto z^3 + \alpha z^2 + z$. *In the outer region, the orbit of one critical point escapes to infinity.*

Problem 12-c. Cubic parabolic maps. Now consider the one-parameter family of cubic polynomial maps

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

with a double fixed point at the origin. **(1)** Using Theorem 12.4 or by direct calculation, show that the remaining finite fixed point $z = -\alpha$ has multiplier $\lambda = 1 + \alpha^2$ and hence 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 eight shaped region bounded by a lemniscate. This lemniscate is clearly visible as the boundary of the main upper and lower regions in Figure 30, which shows the α-parameter plane. **(2)** Show that f_α is parabolic attracting if and only if α^2 lies in the open disk of radius $1/2$ centered at $1/2$, or equivalently if and only if α lies within a corresponding region bounded by a lemniscate shaped like the symbol ∞. (This has been drawn in as a dotted line in Figure 30.) **(3)** In the parabolic attracting case, show that there are cubic polynomials arbitrarily close to f_α with two distinct attracting fixed points near the origin. (Hint: It is convenient to rescale f_α so that it has a fixed point of multiplier $\lambda = 1 + \alpha^2$ at $z = 1$, as well as the fixed point of multiplier $\mu = 1$ at the origin. Show for any $\mu \neq 2 - \lambda$ that there exists a unique cubic polynomial with multipliers μ and λ at 0 and 1 and proceed as in Problems 12-b and 12-d.)

Problem 12-d. Assigning fixed point indices. Show that a rational map with only simple fixed points is uniquely determined by its distinct fixed points $z_k \in \widehat{\mathbb{C}}$ together with the associated residue indices $\iota_k \in \mathbb{C} \smallsetminus \{0\}$, and that these can be chosen arbitrarily subject to the condition that $\sum \iota_k = +1$. For example, putting one fixed point at infinity, use the normal form

$$f(z) \;=\; z + \frac{(z - z_1) \cdots (z - z_d)}{q(z)}$$

where the polynomial $q(z)$ of degree at most $d-1$ can be chosen uniquely to realize the required ι_k.

Remark. More generally, it seems natural to conjecture the following. Suppose that we are given p distinct points $z_k \in \widehat{\mathbb{C}}$, together with integers $m_k \geq 1$ and complex numbers ι_k satisfying the requirements that

$$\sum m_k \;=\; d+1 \qquad \text{and} \qquad \sum \iota_k \;=\; 1$$

with $\iota_k \neq 0$ whenever $m_k = 1$. Then the family consisting of all rational maps of degree $d+1$ having the given z_k as fixed points of multiplicity m_k and index ι_k forms a smooth manifold of dimension $\sum(m_k - 1) = d+1-p$. Here is an example: If there is just one fixed point, necessarily of multiplicity $d+1$ and index 1, then putting the fixed point at infinity we get the family consisting of all $f(z) = z + 1/q(z)$ such that $q(z)$ is a polynomial of degree exactly $d - 1$.

Problem 12-e. The fixed point formula in higher genus. Verify the generalized fixed point formula of Remark 12.5 in the following two special cases: Let $f : \mathbb{T} \to \mathbb{T}$ be a linear torus map with derivative f' identically equal to α. (Compare Problem 6-b.) Show that the trace τ of the induced action on the 1-dimensional space of holomorphic 1-forms is equal to α. If f is not the identity map, show that there are $|1 - \alpha|^2$ fixed points, each with index $\iota = 1/(1-\alpha)$, and conclude that $\sum \iota = 1 - \overline{\tau}$, as required. Now suppose that S is a compact surface of genus g and that $f : S \to S$ is an involution with k fixed points. Use the Riemann-Hurwitz formula to conclude that the quotient S/f is a surface of genus $\hat{g} = (2 + 2g - k)/4$. For the induced action on the g-dimensional vector space of holomorphic 1-forms, show that $\hat{g}$ of the eigenvalues are equal to $+1$ and that the remaining $g - \hat{g}$ are equal to -1, so that the trace τ equals $2\hat{g} - g$. Conclude again that $\sum \iota = k/2$ is equal to $1 - \overline{\tau}$.

§13. Most Periodic Orbits Repel

This section will prove the following theorem of Fatou.

Theorem 13.1. *Let* $f : \hat{\mathbb{C}} \to \hat{\mathbb{C}}$ *be a rational map of degree* $d \geq 2$. *Then* f *has at most a finite number of cycles which are attracting or indifferent.*

By a *cycle* we will mean simply a periodic orbit of f. Recall that a cycle is called *attracting*, *indifferent*, or *repelling* according to whether its multiplier λ satisfies $|\lambda| < 1$, $|\lambda| = 1$, or $|\lambda| > 1$. We will see in §14 that there always exist infinitely many repelling cycles. Shishikura [1987] has given the sharp upper bound of $2d - 2$ for the number of attracting or indifferent cycles using methods of quasiconformal surgery. (Compare Epstein [1999], as well as Buff and Epstein [2002].) However, the classical proof, which is given here, shows only that this number is less than or equal to $6d - 6$.

Recall from Corollary 10.16 that f can have at most $2d-2$ attracting or parabolic cycles. (If $\mathcal{A}$ is any immediate attracting or parabolic basin, then some iterate $f^{\circ p}$ maps $\mathcal{A}$ into itself. By Theorems 8.6 and 10.15, $\mathcal{A}$ contains a critical point of $f^{\circ p}$. Therefore, by the chain rule, some immediate basin $f^{\circ i}(\mathcal{A})$ in the same cycle contains a critical point of f. Since the various immediate basins are all pairwise disjoint, and since f has at most $2d - 2$ critical points, it follows that f has at most $2d - 2$ cycles which are attracting or parabolic.)

Lemma 13.2. *For a rational map of degree* $d \geq 2$, *the number of indifferent cycles which have multiplier* $\lambda \neq 1$ *is at most* $4d - 4$.

Proofs. Evidently Lemma 13.2 and Corollary 10.16 together imply Theorem 13.1. Following Fatou, we prove Lemma 13.2 by perturbing the given map f in such a way that more than half of its indifferent cycles become attracting. Let $f(z) = p(z)/q(z)$ where $p(z)$ and $q(z)$ are polynomials with no common divisor, and where at least one of the two has degree d. Consider the one-parameter family of maps

$$f_t(z) = \frac{p(z) - tz^d}{q(z) - t}, \qquad (13 : 1)$$

with $f_0(z) = f(z)$ and $f_\infty(z) = z^d$. Here we must exclude the trivial special case where $f(z)$ is identically equal to z^d. For most values of the parameter t, this is a well-defined rational map of degree d which depends smoothly on t. However, we must exclude a finite number of exceptional parameter values $\hat{t} \neq 0, \infty$ such that a zero and a pole of f_t crash together as $t \to \hat{t}$ so that the degree of the limit map $f_{\hat{t}}$ is strictly less than d. It is not hard to check that the possible points $\hat{z}$ at which a zero and pole can collide are just the finitely many solutions to the equation $f(\hat{z}) = \hat{z}^d$ in $\widehat{\mathbb{C}}$. If $\hat{z}$ is finite, then the numerator and denominator of (13 : 1) must vanish simultaneously, and we can solve uniquely for $\hat{t} = q(\hat{z})$. On the other hand, for all $z \in \widehat{\mathbb{C}} \smallsetminus \{0\}$ we can solve uniquely for $\hat{t} = p(\hat{z})/\hat{z}^d$, interpreting this quotient as the leading coefficient of the polynomial $p(z)$ in the special case $\hat{z} = \infty$.

If $f = f_0$ has k distinct indifferent cycles with multipliers $\lambda_j \neq 1$, then we must prove that $k \leq 4d - 4$. Choose a representative point z_j in each of these cycles. By the Implicit Function Theorem, we can follow each of these cycles under a small deformation of f_0. Thus, for small values of $|t|$, the map f_t must have corresponding periodic points $z_j(t)$ with multipliers $\lambda_j(t)$ which depend holomorphically on t.

Sublemma 13.3. *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 $t \mapsto \lambda_j(t)$ is constant throughout a neighborhood of $t = 0$. Choose some ray

$$r \mapsto t = re^{i\theta}, \qquad 0 \leq r \leq +\infty,$$

from 0 to ∞ in $\widehat{\mathbb{C}}$ which avoids the finitely many exceptional values of t. Then we will show that it is possible to continue the function $t \mapsto z_j(t)$ analytically along a neighborhood of this ray, so that each $z_j(t)$ is a periodic point for f_t with multiplier $\lambda_j = \text{constant}$. To prove this, we will check that the set of $r_1 \in [0, \infty]$ such that we can continue for $0 \leq r \leq r_1$ is both open and closed: 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 periodic point varies smoothly with t throughout some open neighborhood in the t-plane by the Implicit Function Theorem. Now continuing analytically along the ray to $t = \infty$, we see that the map $z \mapsto z^d$ must also have an indifferent cycle with multiplier equal to λ_j. But every periodic point of this limit map is either 0 or ∞ with multiplier $\lambda = 0$, or else a root of unity with multiplier $\lambda = d^k > 1$ where k is the period. This contradicts the hypothesis that $|\lambda_j| = 1$, and hence completes the proof of Sublemma 13.3. $\square$

The proof of Lemma 13.2 continues as follows. We can express each of our k 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 $n_j \geq 1$ for $1 \leq j \leq k$. Hence

$$|\lambda_j(t)| \;=\; 1 + \operatorname{Re}(a_j t^{n_j}) + o(t^{n_j}).$$

We can divide the t-plane up into n_j sectors for which the expression $\operatorname{Re}(a_j t^{n_j})$ is positive, and n_j complementary sectors for which this expression is negative. Let

$$\sigma_j(\theta) \;=\; \operatorname{sgn}\Big(\operatorname{Re}(a_j e^{i\theta n_j})\Big),$$

so that

$$\begin{aligned}
\sigma_j(\theta) &= +1 \quad \Longrightarrow \quad |\lambda_j(re^{i\theta})| > 1 \quad \text{for small } r > 0,\\
\sigma_j(\theta) &= -1 \quad \Longrightarrow \quad |\lambda_j(re^{i\theta})| < 1 \quad \text{for small } r > 0.
\end{aligned}$$

Evidently each $\sigma_j : \mathbb{R}/2\pi\mathbb{Z} \to \{\pm 1, 0\}$ is a step function which takes the value ± 1 except at $2n_j$ jump discontinuities, and each σ_j has average

$$\frac{1}{2\pi}\int_0^{2\pi} \sigma_j(\theta)\, d\theta \;=\; 0.$$

Let ℓ be either k or $k-1$ according as k is odd or even. Then the sum $\sigma_1(\theta) + \cdots + \sigma_\ell(\theta)$ is also a well-defined step function which has average zero and which takes odd integer values almost everywhere. Hence we can choose some θ such that $\sigma_j(\theta) = -1$ for more than half, that is, for at least $(\ell+1)/2$, of the indices $j \leq \ell$. If we choose r sufficiently small and set $t = re^{i\theta}$, this means that f_t has at least $(\ell+1)/2$ distinct cycles with multiplier satisfying $|\lambda_j| < 1$. Therefore, by Theorem 8.6 or Corollary 10.16, we have $(\ell+1)/2 \leq 2d-2$. This implies that $k \leq \ell + 1 \leq 4d-4$, which completes the proof of Lemma 13.2 and Theorem 13.1. $\square$

§14. Repelling Cycles Are Dense in J

We saw in Lemma 4.6 that every repelling cycle is contained in the Julia set. A much sharper statement was proved in quite different ways by both Fatou and Julia, and both proofs are given below. Using our terminology, it reads as follows.

Theorem 14.1. *The Julia set for any rational map of degree* ≥ 2 *is equal to the closure of its set of repelling periodic points.*

Proof following Julia. According to Corollary 12.7 to the Rational Fixed Point Formula, every rational map f of degree 2 or more has either a repelling fixed point or a parabolic fixed point with $\lambda = 1$. In either case, by Lemmas 4.6 and 4.7 this fixed point belongs to the Julia set $J(f)$.

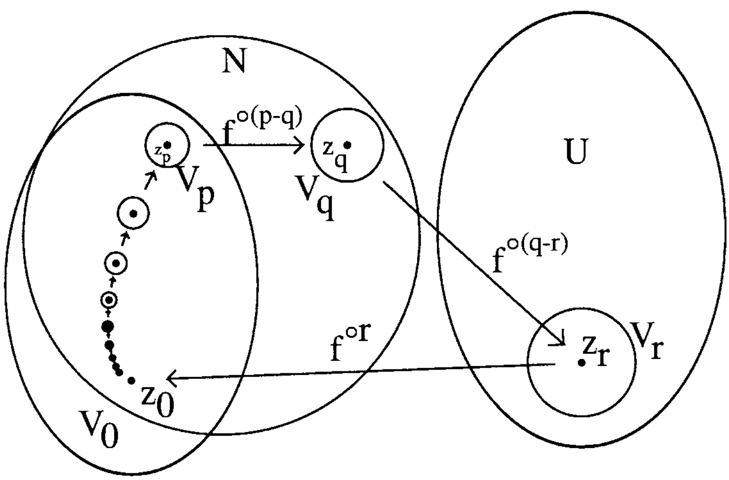


Figure 31. A homoclinic orbit with $z_j \mapsto z_{j-1}$, $\lim_{j\to\infty} z_j = z_0$, *and with* $\cdots \mapsto z_p \mapsto \cdots \mapsto z_q \mapsto \cdots \mapsto z_r \mapsto \cdots \mapsto z_0$.

Thus we can start with a fixed point z_0 in the Julia set. Let $U \subset \hat{\mathbb{C}}$ be any open set, disjoint from z_0, which intersects the Julia set. 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 backward 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 Corollary 4.13 which says that there exists an integer $r > 0$ and a point $z_r \in J(f) \cap U$ so that the rth forward image $f^{\circ r}(z_r)$ is equal to z_0. Given any neighborhood N of z_0, we can repeat this argument and conclude that there exists an integer $q > r$ and a point $z_q \in N$ so that $f^{\circ(q-r)}(z_q) = z_r$ (Figure 31).

First suppose that z_0 is a repelling fixed point. Choose N to be a linearizing neighborhood as in the Kœnigs Theorem 8.2, small enough to be disjoint from z_r. Then inductively choose preimages

$$\cdots \mapsto z_j \mapsto z_{j-1} \mapsto \cdots \mapsto z_q,$$

all inside of the neighborhood N. These preimages z_j will automatically converge to z_0 as $j \to \infty$.

If 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 $z_q \in N$ will map diffeomorphically under $f^{\circ q}$ onto a neighborhood V_0 of z_0. Furthermore, we can assume that $V_r = f^{\circ(q-r)}(V_q)$ is contained in U. 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 $\overline{V}_p \subset V_0$. Now f^{-p} maps the simply connected open set V_0 holomorphically into this compact subset $\overline{V}_p \subset V_0$. Hence it contracts the Poincaré metric on V_0 by a factor $c < 1$ and therefore must have an attracting 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.

If our homoclinic orbit contains critical points, then this argument must be modified very slightly as follows. The neighborhood V_q will no longer map diffeomorphically onto V_0; however, we can choose V_q and V_0 so that $f^{\circ q} : V_q \to V_0$ is a branched covering map, branched only at z_q. It then follows that $f^{\circ p} : V_p \to V_0$ is also a branched covering map, branched only at z_p. Choose a slit S in V_0 from the boundary to the base point z_0 so as to be disjoint from $\overline{V}_p$, 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.

The proof in the parabolic case is similar. Replacing f by some iterate if necessary, we may assume that the multiplier at z_0 is $+1$. Now take N to be a repelling petal, and proceed as above. □

Proof of Theorem 14.1 following Fatou. In this case, the main idea is an easy application of Montel's Theorem 3.7. However, we must use Theorem 13.1 to finish the argument.

To begin the proof, recall from Corollary 4.14 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

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

on N would avoid the three values $0, 1$, and ∞, and hence be a normal family. (This expression is just the cross-ratio of the four points z, $\varphi_1(z)$, $\varphi_2(z)$, $f^{\circ n}(z)$, as discussed in Problem 1-c.) 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. $\square$

There are a number of interesting corollaries.

Corollary 14.2. *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 Theorem 4.10 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. $\square$

More generally, if $K \subset \hat{\mathbb{C}}$ is any compact set which does not contain any grand orbit finite points, then $f^{\circ n}(U) \supset K$ for large n. In particular, if there are no grand orbit finite points, then $f^{\circ n}(U) = \hat{\mathbb{C}}$ for large n. (Compare Theorem 4.10 and Problem 4-b.)

As another corollary, we can make a sharper statement of the defining property for the Julia set.

Corollary 14.3. *If $U \subset \hat{\mathbb{C}}$ is any open set which intersects the Julia set $J(f)$, then no sequence of iterates $f^{\circ n(i)}$ can converge locally uniformly throughout U.*

Proof. Suppose that the sequence of functions $f^{\circ n(i)}(z)$ converged locally uniformly to $g(z)$ throughout the open set U. If z_0 is a point of $U \cap J$, then we could choose a smaller neighborhood U' of z_0 so that $|g(z) - g(z_0)| < \epsilon$ for all $z \in U'$. For large i, it would follow that $|f^{n(i)}(z) - g(z_0)| < 2\epsilon$ for all $z \in U'$, thus contradicting Corollary 14.2. □

As another consequence of Corollary 14.2, we have the following statement, which is very slightly sharper than Theorem 11.17 and provides an alternative proof for it. By a *critical orbit* we mean the forward orbit of some critical point.

Corollary 14.4. *If z_0 is a Cremer point or a boundary point of a Siegel disk, then every neighborhood of z_0 contains infinitely many distinct critical orbit points.*

Proof. First consider the Cremer case. Without loss of generality we may assume that $z_0 = 0$, and replacing f by a suitable iterate we may assume that $f(0) = 0$. Since $|f'(0)| = 1 \neq 0$, the map f is locally one-to-one. Hence, for each $n > 0$, there exists a disk $\mathbb{D}_{\epsilon(n)}$ such that there is a single valued branch of f^{-n} which is defined throughout $\mathbb{D}_{\epsilon(n)}$ and maps 0 to 0. Choosing $\epsilon(n)$ to be maximal, it follows as in the proof of Lemma 8.5 that there is a critical point of $f^{\circ n}$ on the boundary of $f^{-n}(\mathbb{D}_{\epsilon(n)})$, and hence a critical value of $f^{\circ n}$ on the boundary of $\mathbb{D}_{\epsilon(n)}$. We must prove that $\epsilon(n) \to 0$ as $n \to \infty$. But otherwise, taking $\epsilon > 0$ to be the infimum of the $\epsilon(n)$, since the images of the collection of maps $\{f^{-n}|_{\mathbb{D}_\epsilon}\}$ avoid every periodic orbit other than $\{0\}$, these maps would form a normal family. Thus some subsequence would converge to a holomorphic map $g : \mathbb{D}_\epsilon \xrightarrow{\cong} U$ with $g(0) = 0$. Now consider the subset $U' = g(\mathbb{D}_{\epsilon/2})$. Then infinitely many iterates $f^{\circ n}$ would map U' into D_ϵ. According to Corollary 14.2, this would imply that U' is contained in the Fatou set, and hence contradict the hypothesis that the origin is a Cremer point.

Now suppose that $z_0 = 0$ belongs to the boundary $\partial\Delta$ of some f-invariant Siegel disk $\Delta = f(\Delta)$. Suppose that a small disk neighborhood $\mathbb{D}_\epsilon$ of 0 contained no critical orbit points. Since each $f^{\circ n}$ is one-to-one on Δ, it would follow easily that there is a unique branch of $f^{-n}|_{\mathbb{D}_\epsilon}$ which takes $\mathbb{D}_\epsilon \cap \Delta$ into Δ. As in the argument above these would form a normal family, so some subsequence would converge to a holomorphic map $g : \mathbb{D}_\epsilon \xrightarrow{\cong} U$ with $g(\mathbb{D}_\epsilon \cap \Delta) \subset \Delta$. Just as above, it would follow that infinitely many iterates $f^{\circ n}$ would map the set $U' = g(\mathbb{D}_{\epsilon/2})$ into $\mathbb{D}_\epsilon$. By Corollary 14.2 it would follow that U' is contained in the Fatou set, which is impossible since U' must contain Siegel boundary points. This proves that critical orbit points are dense in the compact connected set $\partial\Delta$, and the conclusion follows easily. □

By definition, the rational map f is *postcritically finite* if 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 and Hubbard [1993].)

Corollary 14.5. *If f is postcritically finite, then every periodic orbit of f is either repelling or superattracting. 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 repelling or attracting; there are no parabolic cycles, Cremer cycles, or Siegel cycles.*

Proof. This follows easily from Theorem 10.15 and Corollary 14.4. □

As still another consequence, we get a simpler proof of Corollary 4.15.

Corollary 14.6. *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 nonvacuous compact subsets. After replacing f by some iterate $g = f^{\circ n}$, we may assume by Corollary 14.2 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 to whether $g^{\circ k}(z)$ 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 14.7. Arbitrary Riemann Surfaces. 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 trivial exceptional case. For transcendental maps of the plane or cylinder this result was proved by Baker [1968], and the proof for a map of a torus or hyperbolic surface is quite easy. (Compare Lemma 5.1 and Theorem 6.1.) The unique exceptional case occurs for a fractional linear transformation of $\widehat{\mathbb{C}}$ which has just one parabolic fixed point, for example, the map $f(z) = z+1$ with $J(f) = \{\infty\}$.

§15. Herman Rings

This section will be a survey, without complete proofs, describing a close relative of the Siegel disk.

Definition. A component U of the Fatou set $\widehat{\mathbb{C}} \smallsetminus J(f)$ 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 iterate of f, corresponds to an irrational rotation of this annulus. (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 [1979], is based on a careful analysis of real analytic diffeomorphisms of the circle. An alternative method, due to Shishikura [1987], uses quasiconformal surgery, starting with two copies of the Riemann sphere with a Siegel disk in each, cutting out part of the center of each disk and pasting the resulting boundaries together in order to fabricate such a ring.

The original method can be outlined as follows in the special case of a map which leaves the unit circle invariant. (Compare Sullivan [1983], Douady [1987].) First a number of definitions: If $f : \mathbb{R}/\mathbb{Z} \to \mathbb{R}/\mathbb{Z}$ is an orientation-preserving homeomorphism, then we can lift to a homeomorphism $F : \mathbb{R} \to \mathbb{R}$ which satisfies the identity $F(t+1) = F(t) + 1$ and is uniquely defined up to addition of an integer constant.

Definition. The real number

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

is independent of the choice of t_0 and will be called the *translation number* of the lifted map F. Following Poincarè, the *rotation number* $\mathrm{rot}(f) \in \mathbb{R}/\mathbb{Z}$ of the circle map f is defined to be the residue class of $\mathrm{Rot}(F)$ modulo $\mathbb{Z}$.

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 [1955] or de Melo and van Strien [1993], and see Problem 15-d.)

Lemma 15.1. *The homeomorphism f has a periodic point with period q if and only if its rotation number is rational with denominator q.*

Theorem 15.2 (Denjoy [1932]). *If f is a diffeomorphism of class C^2 and if the rotation number $\rho = \mathrm{rot}(f)$ is irrational, then f is topologically conjugate to the rotation $t \mapsto t+\rho \pmod{\mathbb{Z}}$.*

Lemma 15.3. *Consider a one-parameter family of lifted maps of the form*

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

Then the translation 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 $\mathrm{Rot}(F_\alpha)$ provided that F_0 is nonlinear.

In the real analytic case, Denjoy's Theorem has an analog which can be stated as follows. Recall from §11 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 (that is, for maps close to the identity) by Arnold [1965] and sharpened first by Herman [1979] and then by Yoccoz [2002], who also included some non-Diophantine cases.

Theorem 15.4 (Herman-Yoccoz Theorem). *If f is a real analytic diffeomorphism of $\mathbb{R}/\mathbb{Z}$ 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.

Remark: In the C^∞ case, Yoccoz [1984] proved a corresponding if and only if statement, using results of Herman: *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 Problem 7-b, as well as Theorem 1.7.) Given any constant $a \in \widehat{\mathbb{C}}$ 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 $\partial\mathbb{D}$ onto itself fixing the basepoint $z = 1$, 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 $\partial\mathbb{D}$ and maps $\mathbb{D}$ to its complement.

Lemma 15.5. *A rational map of degree d carries the unit circle into itself* if and only if it can be written as a* Blaschke product

$$f(z) = e^{2\pi i t}\beta_{a_1}(z) \cdots \beta_{a_d}(z) \qquad (15:1)$$

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

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 $\beta_a(z)\beta_b(z) = 1$. Evidently the expression in Lemma 15.5 is essentially 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 Lemma 15.5 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 7-b, 15-c.) However, we will instead be interested in the mixed case, where some of the a_j are inside the unit disk and some are outside.

Theorem 15.6. *For any odd degree $d \geq 3$ we can choose a Blaschke product f of degree d which carries the unit circle $\partial\mathbb{D}$ into itself by an orientation-preserving 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

---

* Similarly, a rational map carries $\mathbb{R} \cup \{\infty\}$ into itself if and only if it can be expressed as a quotient of polynomials with real coefficients. Thus, if we conjugate by a Möbius transformation which carries the unit circle to the real line, then such Blaschke products correspond precisely to rational maps with real coefficients.

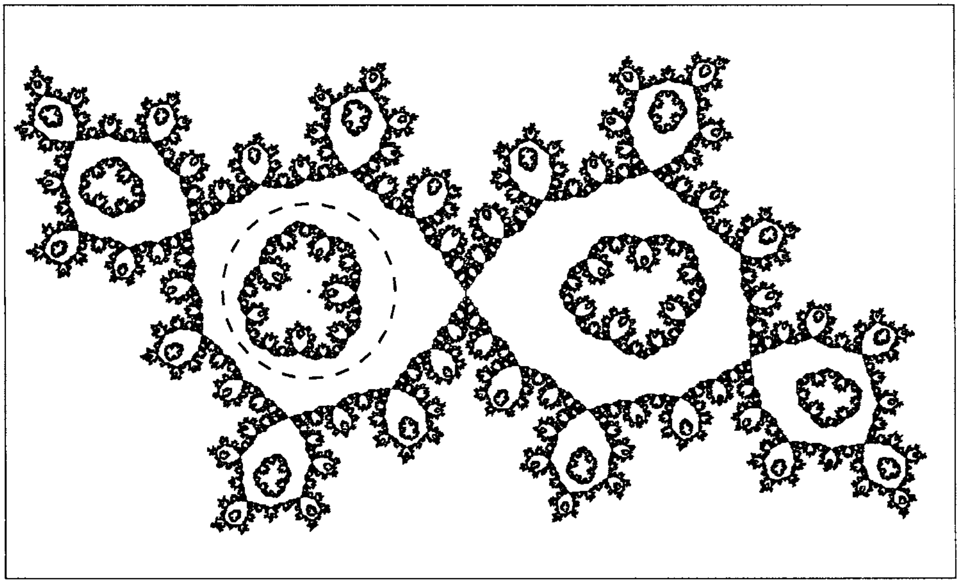


Figure 32. Julia set for a cubic rational map possessing a Herman ring.

identity map on the unit circle $\partial\mathbb{D}$. In particular, it induces an orientation-preserving diffeomorphism of $\partial\mathbb{D}$. Now multiplying by $e^{2\pi it}$ and using Lemma 15.3, 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 $\partial\mathbb{D}$ 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 $\partial\mathbb{D}$, and the conclusion follows. □

As an example, Figure 32 shows the Julia set for a cubic rational map of the form $f(z) = e^{2\pi it}z^2(z-4)/(1-4z)$, with the unit circle as an f-invariant curve. The unit circle and the superattracting fixed point at the origin have been marked. Here the constant $t = .6151732\cdots$ has been chosen so that the rotation number ρ will be equal to $(\sqrt{5}-1)/2$. (Compare Problem 15-d.) Since ρ is Diophantine, it follows that the unit circle is contained in a Herman ring. This Julia set is invariant under inversion in the unit circle (Problem 15-b). There is one critical point on each of the two components of the boundary of this Herman ring, namely one at the center of approximate symmetry and one at its image under inversion.

Just as in the case of a quadratic polynomial, the map f restricted to the basin of infinity is conjugate to the map $w \mapsto w^2$ on the unit disk. This helps to explain why the outer parts of the Julia set appear to be symmetric under $180°$ rotation, although the inner parts do not. The Julia set can be expressed as a union $J_1 \cup J_2 \cup J_3$ of three non-overlapping compact subsets where each J_α maps bijectively onto the entire Julia set. Here J_1 is everything to the right of the central critical point, J_3 is its image under inversion, and J_2 is everything else. Any two of the J_α intersect at most in a single critical point.

This is the simplest kind of example one can find, since Shishikura has shown that a Herman ring can exist only if the degree d is at least 3 (compare Milnor [2000c]), and since it is easy to prove that a polynomial map cannot have any Herman ring (Problem 15-a). The rings constructed in this way are very special in that they are symmetric under inversion in the unit circle, with $f(1/\overline{z}) = 1/\overline{f(z)}$. (Compare Problem 15-b. The original construction in Herman [1979], based on work of Helson and Sarason, was more flexible and did not require symmetry.)

Shishikura's more general construction, based on quasiconformal surgery, also avoids the need for symmetry. In the simplest case, given rational maps of degree d_1 and d_2 having invariant Siegel disks with rotation numbers $+\rho$ and $-\rho$ respectively, Shishikura's construction cuts out a small concentric disk from each, and then glues the resulting boundaries together. Making corresponding modifications at each of the infinitely many iterated preimages of each of the Siegel disks and then applying the Morrey-Ahlfors-Bers Measurable Riemann Mapping Theorem to obtain a compatible conformal structure, he obtains a rational map of degree $d_1 + d_2 - 1$ with a Herman ring of rotation number ρ. There is a converse construction, which cuts along the central circle of a Herman ring and then pastes in a pair of Siegel disks. Thus the following statement is an immediate consequence:

> *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 Theorem 11.10 can occur as the rotation number of a Herman ring.

Although Herman rings do not contain any critical points, nonetheless they are closely associated with critical points.

Lemma 15.7. *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 of the first statement is almost identical to the proof of Theorem 11.17 or Corollary 14.4, while the second follows from Problem 5-b and the Jordan curve theorem. □

Concluding Problems

Problem 15-a. No polynomial Herman rings. Using the maximum modulus principle, show that no polynomial map can have a Herman ring.

Problem 15-b. Symmetry of Blaschke products. For any Blaschke product $f : \widehat{\mathbb{C}} \to \widehat{\mathbb{C}}$ show that the Julia set is invariant under the inversion $z \mapsto 1/\bar{z}$. 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 15-c. Proper self-maps of $\mathbb{D}$. A holomorphic map $f : \mathbb{D} \to \mathbb{D}$ is said to be *proper* if the inverse image of any compact subset of $\mathbb{D}$ is compact. Show that any proper holomorphic map from $\mathbb{D}$ onto itself can be expressed uniquely as a Blaschke product $(15:1)$, with $a_j \in \mathbb{D}$.

Problem 15-d. Computing rotation numbers. **(1)** Show that the rotation number $\mathrm{rot}(f)$ can be deduced directly from the cyclic order relations on a single orbit, in a form convenient for computer calculations, as follows. Choose representatives $t_i \in [0,1)$ for the elements of the orbit of zero, so that $t_i \equiv f^{\circ i}(0) \pmod{\mathbb{Z}}$. If we exclude the trivial case $t_1 = 0$, then t_1 cuts $[0,1)$ into two disjoint intervals $I_1 = [0,t_1)$ and $I_0 = [t_1,1)$. Define a sequence of bits $(b_2, b_3, b_4, \ldots)$ by the requirement that $t_n \in I_{b_n}$. If F is the unique lift with $F(0) = t_1$, show that

$$\mathrm{Rot}(F) = \lim_{n\to\infty}(b_2 + b_3 + \cdots + b_n)/n.$$

(2) Furthermore, if a second such map f' has bit sequence $(b'_2, b'_3, \ldots)$, and if

$$(b_2, b_3, \ldots) < (b'_2, b'_3, \ldots)$$

using the lexicographical order for bit sequences, show that

$$\mathrm{Rot}(F) \le \mathrm{Rot}(F').$$

(We can then estimate $\mathrm{Rot}(F)$ rapidly by comparing its bit sequence with the bit sequences of rigid rotations.)

§16. The Sullivan Classification of Fatou Components

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

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

Theorem 16.1. *If f maps the Fatou component U onto itself, then there are just four possibilities, as follows: Either U is the immediate basin for an attracting fixed point or for one petal of a parabolic fixed point which has multiplier $\lambda = 1$ 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 a geometrically attracting fixed point, with $\lambda \neq 0$. Note that immediate attractive or parabolic basins always contain critical points by Theorems 8.6 and 10.15, while rotation domains (that is, Siegel disks and Herman rings) evidently cannot contain critical points.

Much of the proof of Theorem 16.1 has already been carried out in §5. In fact, according to Theorem 5.2 and Lemma 5.5, a priori there are just four possibilities. They are:

(a) U contains an attracting 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 2 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 U would be a subset of a Siegel disk, rather than a full Fatou component. Thus, in order to prove Theorem 16.1, we need only show that the boundary fixed point in case (b) must be parabolic with $\lambda = 1$. 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. Thus it must be indifferent, $|\lambda| = 1$. To prove Theorem 16.1, we need only show that λ is precisely equal to $+1$.

The proof will be based on the following statement, which is due to Douady and Sullivan. (Compare Sullivan [1983] or Douady and Hubbard [1984–85, p. 70]. For a more classical alternative, see Lyubich [1986, 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 V of the origin and which has a fixed point with multiplier λ at $z = 0$. By a *path in $V \setminus \{0\}$ which converges to the origin* we will mean a continuous map $p : [0, \infty) \to V \setminus \{0\}$ satisfying the condition that $p(t)$ tends to zero as $t \to \infty$. (Here $[0, \infty)$ denotes the half-open interval consisting of all real numbers $t \geq 0$.) Note that such a path p may have self-intersections.

Lemma 16.2 (Snail Lemma). *Suppose that there exists a path $p : [0, \infty) \to V \setminus \{0\}$ which is mapped into itself by f in such a way that $f(p(t)) = p(t+1)$ and which converges to the origin as $t \to \infty$. Then either $|\lambda| < 1$ or $\lambda = 1$.*

In other words, the origin must be either an attracting fixed point or a parabolic fixed point with λ precisely equal to 1.

Proof of Lemma 16.2. By hypothesis, the orbit $p(0) \mapsto p(1) \mapsto p(2) \mapsto \cdots$ in $V \setminus \{0\}$ converges towards the origin. Thus the origin is a fixed point and cannot be repelling; the multiplier must satisfy $|\lambda| \leq 1$. Let us assume that $|\lambda| = 1$ with $\lambda \neq 1$ and show that this hypothesis leads to a contradiction.

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 we have the asymptotic equality $p(t+1) \sim \lambda p(t)$ as $t \to \infty$. If the path p has no self-intersections, then the image must resemble a very tight spiral as shown in Figure 33 (left side), and we can sketch a proof as follows. Draw a radial segment E joining two turns of this spiral, as shown. Then the region W 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 $0 \in W$ must be strictly attracting, which contradicts the hypothesis that $|\lambda| = 1$.

In order to fill in the details of this argument and to allow for the possibility of self-intersections, let us introduce polar coordinates (r, θ) on $\mathbb{C} \setminus \{0\}$, setting $z = re^{i\theta}$ with $r > 0$ and $\theta \in \mathbb{R}/2\pi\mathbb{Z}$. Lift p to a path $\tilde{p}(t) = (r(t), \tilde{\theta}(t))$ in the universal covering of $\mathbb{C} \setminus \{0\}$, where $\tilde{\theta}(t)$ is a real

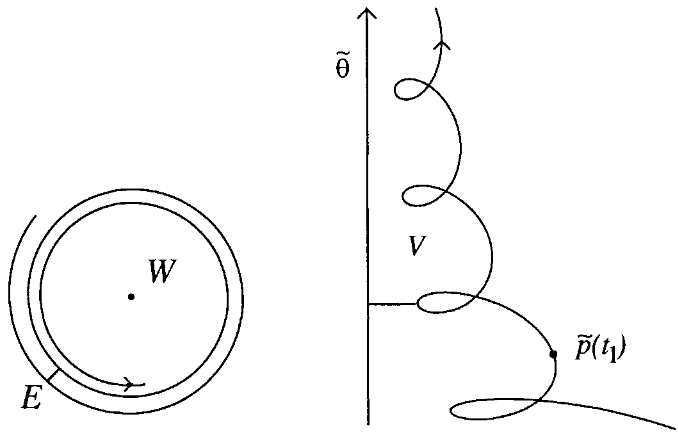


Figure 33. Simple curve in $\mathbb{C} \smallsetminus \{0\}$ on the left, and a nonsimple curve lifted to the universal covering on the right.

number. As t tends to infinity, note that

$$r(t) \searrow 0 \qquad \text{and} \qquad \tilde{\theta}(t+1) \;=\; \tilde{\theta}(t) + c + o(1), \tag{16 : 1}$$

where c is a uniquely defined real constant with $e^{ic} = \lambda$. Similarly, choosing $r_0 > 0$ so that f is univalent on the disk of radius r_0, we can lift f to a map $\tilde{f}(r, \tilde{\theta}) = (r', \tilde{\theta}')$ on the universal covering, where $\tilde{f}$ is defined and univalent for $0 < r < r_0$ and for all $\tilde{\theta} \in \mathbb{R}$. Note that

$$r' \sim r \qquad \text{and} \qquad \theta' \;=\; \theta + c + o(1) \qquad \text{as} \qquad r \searrow 0,$$

where c is the same constant which occurs in $(16:1)$ provided that we choose the correct lift of f. (It follows that we can extend $\tilde{f}$ continuously over $[0, r_0) \times \mathbb{R}$ so that it is a translation, $\tilde{f}(0, \tilde{\theta}) \;=\; (0, \tilde{\theta} + c)$, when $r = 0$.)

We must prove that $c = 0$. Suppose, for example, that c is strictly positive. Then we could derive a contradiction as follows. Choose a constant $r_1 < r_0$ so that the map $\tilde{f}(r, \tilde{\theta}) = (r', \tilde{\theta}')$ satisfies $\tilde{\theta}' > \tilde{\theta} + c/2$ whenever $r \le r_1$. Then choose t_1 so that $r(t) \le r_1$ and hence $\tilde{\theta}(t+1) > \tilde{\theta}(t) + c/2$ whenever $t \ge t_1$. Furthermore, choose $\tilde{\theta}_1$ so that $\tilde{\theta}(t) < \tilde{\theta}_1$ for $t \le t_1$. Now in the $(r, \tilde{\theta})$ plane, consider the connected region V which lies above the line $\tilde{\theta} = \tilde{\theta}_1$, to the right of the line $r = 0$, and to the left of the curve $\tilde{p}[t_1, \infty)$. It follows easily that $\tilde{f}$ maps V univalently into itself and that the r-coordinate tends to zero under iteration of $\tilde{f}$ restricted to V. Hence, if W is the image of V under projection to the z-plane, it follows that W is a neighborhood of the origin and that all orbits in W converge to the origin. Therefore, the origin is an attracting fixed point, and $|\lambda| < 1$. □

Here is a completely equivalent statement, which will be useful in §18. Again let f be a holomorphic map of the form $f(z) = \lambda z + a_2 z^2 + \cdots$ near the origin.

Corollary 16.3. *Now suppose that* $p : [0, \infty) \to V \smallsetminus \{0\}$ *is a path which converges to the origin, with* $f(p(t)) = p(t-1)$ *for* $t \ge 1$ *(so that points on this path are pushed away from the origin). Then the multiplier* λ *must satisfy either* $|\lambda| > 1$ *or* $\lambda = 1$.

Proof. Since the orbit

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

is repelled by the origin, the multiplier λ cannot be zero. Hence f^{-1} is defined and holomorphic near the origin. Applying Lemma 16.2 to the map $g = f^{-1}$, the conclusion follows. □

Proof of Theorem 16.1. Recall that we have already discussed all of the cases except (b) above. Thus we need only consider a Fatou component U which is mapped into itself by f in such a way that all orbits converge to a boundary fixed point w_0. Choose any basepoint z_0 in U, and choose any path $p : [0, 1] \to U$ from $z_0 = p(0)$ to $f(z_0) = p(1)$. Extending for all $t \ge 0$ 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 Lemma 16.2, the fixed point w_0 must be either parabolic with $\lambda = 1$ 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. Each one is either:

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

(2) the immediate 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 uniquely specified by this description. In cases (1) and (2), as noted in Theorem 8.9 and Problem 10-f, U must be either simply connected or infinitely connected.

By §10, there can be at most a finite number of attracting basins and Siegel disks. In fact, according to Shishikura [1987], there can be at most

$2d-2$ distinct cycles of periodic Fatou components, where d is the degree. (In particular, there can be at most $2d-2$ cycles of Herman rings.)

In order to complete the picture, we need the following fundamental theorem, which asserts that there are no *wandering* Fatou components.

Theorem 16.4 (Sullivan Nonwandering Theorem). *Every Fatou component U for a rational map is eventually periodic. That is, there necessarily exist integers $n \ge 0$ and $p \ge 1$ so that the nth forward image $f^{\circ n}(U)$ is mapped onto itself by $f^{\circ p}$. In particular, it follows that every Fatou component is either a branched covering or a biholomorphic copy of some periodic Fatou component, which necessarily belongs to one of the four types described above.*

The proof, by quasiconformal deformation, will be outlined in Appendix F. (Compare Sullivan [1985], Carleson and Gamelin [1993].) However, the idea of the proof can be described very briefly as follows: If a wandering Fatou component were to exist, with all forward images pairwise disjoint, then using the Measurable Riemann Mapping Theorem of Morrey, Ahlfors, and Bers one could construct an infinite-dimensional space of nonisomorphic deformations of f, 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 §14 that f is *postcritically finite* if every critical orbit is finite. (Compare Corollary 14.5.)

Corollary 16.5. *If a postcritically finite rational map has no superattractive periodic orbit, then its Julia set is the entire sphere $\widehat{\mathbb{C}}$.*

By Theorems 8.6 and 10.15 it cannot have any attracting or parabolic basins, and by Theorem 11.17 and Lemma 15.7 it cannot have any rotation domains. □

We will give a more direct proof for this statement in Corollary 19.8.

Remark 16.6. Transcendental Maps. The analogs of both Theorems 16.1 and 16.4 fail for the iterates of a transcendental map $f : \mathbb{C} \to \mathbb{C}$. In fact there are two new kinds of Fatou component, which cannot occur for rational maps. There may be wandering domains (Problem 16-c), and there may be invariant domains $U = f(U)$ such that no orbit in U has any accumulation point in the finite plane $\mathbb{C}$. These are now known as *Baker domains*. (Problem 16-d). Of course every orbit in a Baker domain must converge to the point at infinity within $\widehat{\mathbb{C}}$, but the point at

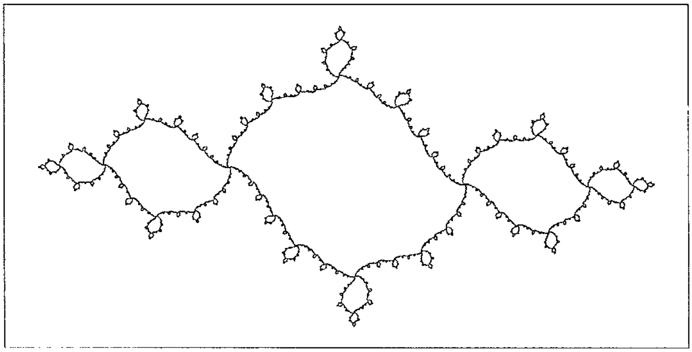


Figure 34. Julia set for $z \mapsto z^2 - 1 + 0.1i$ *(Problem 16-b).*

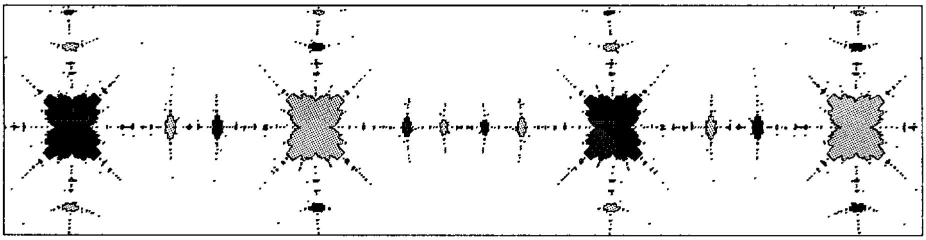


Figure 35. Julia set for $z \mapsto z + \sin(2\pi z)$ *(Problem 16-c). Here, unlike all other Julia set pictures in these notes, the Julia set has been colored white.*

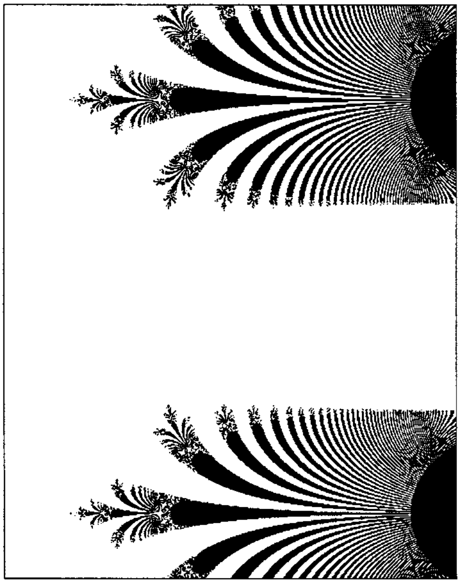


Figure 36. Julia set for $z \mapsto z + e^z - 1$ *(Problem 16-d).*

infinity is an essential singularity of f and hence looks very different from a parabolic point.

Concluding Problems

Problem 16-a. Limits of iterates. Give a sharper formulation of the defining property of the Fatou set $\widehat{\mathbb{C}} \smallsetminus J$ for a rational function as follows. If V is a connected open subset of $\widehat{\mathbb{C}} \smallsetminus J$, 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 compact one-parameter family of maps, consisting of all compositions $R_\theta \circ f^{\circ k}|_V$, with $k_0 \le k < k_0 + p$. Here $f^{\circ k_0}$ is to be some fixed iterate with values in a rotation domain belonging to a cycle of rotation domains of period p, and R_θ is the rotation of this domain through angle θ.

Problem 16-b. Counting components. **(1)** If a quadratic polynomial map has either an attracting fixed point or a parabolic fixed point of multiplier $\lambda = 1$, show that there is only one bounded Fatou component. Compare Figures 5a, 11, 24 (pp. 42, 80, 121). **(2)** If it has an attracting cycle of period 2, show that there are three bounded components which map according to the pattern $U_1 \leftrightarrow U_0 \leftarrow U_1'$ and that the remaining bounded components are iterated preimages of U_1' where each set $f^{-n}(U_1')$ is made up of 2^n distinct components. Identify nine of these components in Figure 34. **(3)** What is the corresponding description for a cycle of attracting or parabolic basins with period p, as in Figures 5d, 21 (pp. 42, 109), or for the case of a Siegel fixed point as in Figures 26, 28 (pp. 127, 132)?

Problem 16-c. Wandering domains. Show that the transcendental map

$$f(z) \;=\; z + \sin(2\pi z)$$

has one family of wandering domains $\{U_n\}$ with $f(U_n) = U_n + 1$ and one family $\{V_n\}$ with $f(V_n) = V_n - 1$ (Figure 35). Describe the Fatou set for the corresponding map of the cylinder $\mathbb{C}/\mathbb{Z}$.

Problem 16-d. A Baker domain. Show that the map

$$f(z) \;=\; z + e^z - 1$$

has a fully invariant Baker domain $U = f^{-1}(U)$ (Figure 36). In particular, show that all critical values belong to the half-plane $\mathrm{Re}(z) < 0$ and that all orbits $\{z_j\}$ in this half-plane satisfy $\lim_{j\to\infty} \mathrm{Re}(z_j) = -\infty$. Show that there is an associated map of the cylinder $\mathbb{C}/2\pi i\mathbb{Z}$.

USING THE FATOU SET TO STUDY THE JULIA SET

§17. Prime Ends and Local Connectivity

Carathéodory's theory of "prime ends" is the basic tool for relating an open set of complex numbers to its complementary closed set. Let U be a simply connected subset of $\widehat{\mathbb{C}}$ such that the complement $\widehat{\mathbb{C}} \smallsetminus U$ is infinite. The Riemann Mapping Theorem asserts that there exists a conformal isomorphism

$$\psi : \mathbb{D} \xrightarrow{\cong} U.$$

In some cases, ψ will extend to a homeomorphism from the closed disk $\overline{\mathbb{D}}$ onto the closure $\overline{U}$, so that the topological boundary ∂U is homeomorphic to the circle $\partial\mathbb{D}$. (See Figures 5a (p. 42) and 37a, together with Theorem 17.16.) However, this is not true in general, since the boundary ∂U may be an extremely complicated object. As an example, Figure 37b shows a region U such that one point of ∂U (with countably many short spikes sticking out from it) corresponds to a Cantor set of distinct points of the circle $\partial\mathbb{D}$. Figures 37c, 37d show examples for which an entire interval of points of ∂U corresponds to a single point of the circle. (See also Problem 5-a.) An effective analysis of the relationship between the compact set ∂U and the boundary circle $\partial\mathbb{D}$ was carried out by Carathéodory [1913] and will be described here.

Finding Short Arcs. The main construction will be purely topological, but we first use analytic methods to prove several lemmas about the existence of short arcs. Let $I = (0, \delta)$ be an open interval of real numbers, and let $I^2 \subset \mathbb{C}$ be the open square, consisting of all $z = x + iy$ with $x,\, y \in I$. Suppose that we are given some conformal metric on I^2 of the form $\rho(z)|dz|$ where $\rho : I^2 \to (0, \infty)$ is a continuous strictly positive real-valued function. (We do not assume that $\rho(z)$ is bounded.) By definition, the *area* of I^2 in this metric is the integral

$$\mathcal{A} \;=\; \iint_{I^2} \rho(x+iy)^2\, dx\, dy,$$

and the *length* of each horizontal line segment $\;y = \text{constant}\;$ is the integral

$$L(y) \;=\; \int_{x \in I} \rho(x+iy)\, dx.$$

We will need the following. (Compare Appendix B.)

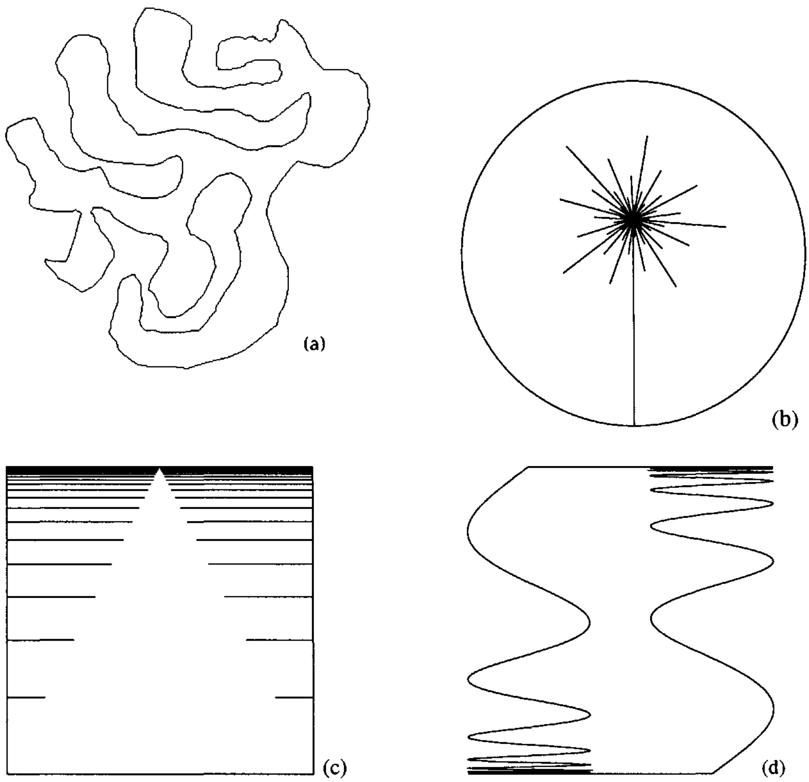


Figure 37. The boundaries of four simply connected regions in $\mathbb{C}$.

Lemma 17.1 (Length-Area Inequality). *If the area $\mathcal{A}$ is finite, then the length $L(y)$ is finite for almost every height $y \in I$, and the average of $L(y)^2$ satisfies*

$$\frac{1}{\delta}\int_I L(y)^2\, dy \;\leq\; \mathcal{A}. \qquad (17:1)$$

Proof. We will use the Schwarz Inequality in the form*

$$\left(\int_I f(x)g(x)\,dx\right)^2 \;\leq\; \left(\int_I f(x)^2\,dx\right)\left(\int_I g(x)^2\,dx\right), \qquad (17:2)$$

where f and g are square-integrable real-valued functions on $I = (0, \delta)$. Taking $f(x) = 1$ and $g(x) = \rho(x + iy)$, this yields

$$L(y)^2 \;\leq\; \delta \cdot \int_I \rho(x+iy)^2\, dx.$$

*If $\int f^2 dx = t^2 \int g^2 dx$ where $t > 0$, then $(17:2)$ can be proved by manipulating the inequalities $\int (f \pm tg)^2 dx \geq 0$. The case $\int f^2 dx = 0$ can be proved in a similar manner by letting t tend to zero.

Integrating this inequality over y and dividing by δ, we obtain the required inequality (17 : 1). If $\mathcal{A}$ is finite, it evidently follows that $L(y)$ is finite for y outside of a set of Lebesgue measure zero. □

For a "majority" of values of y, we can give a more precise upper bound as follows.

Corollary 17.2. *The set S consisting of all $y \in I$ with $L(y) \le \sqrt{2\mathcal{A}}$ has Lebesgue measure $\ell(S) > \ell(I)/2$.*

Proof. Evidently

$$\delta\,\mathcal{A} \;\ge\; \int_I L(y)^2 dy \;>\; \int_{I \smallsetminus S} \left(\sqrt{2\mathcal{A}}\right)^2 dy + \int_S 0 \;=\; 2\mathcal{A}\,\ell(I \smallsetminus S),$$

and the conclusion follows since $\ell(I) = \delta$. □

In the application, consider some univalent embedding

$$\eta : I^2 \xrightarrow{\cong} U \subset \hat{\mathbb{C}}.$$

Pulling the spherical metric from U back to I^2, we obtain a conformal metric of the form $\rho(z)|dz|$ on I^2. (Compare (2 : 4).) Evidently the area $\mathcal{A}$ of I^2 in this metric is at most equal to the area 4π of $\hat{\mathbb{C}}$.

Corollary 17.3. *Given such a univalent embedding of I^2 onto $U \subset \hat{\mathbb{C}}$, almost every horizontal line segment $y =$ constant in I^2 maps to a curve of finite spherical length; and more than half (in the sense of Lebesgue measure) of these horizontal line segments have spherical length at most $\sqrt{2\mathcal{A}}$, where $\mathcal{A}$ is the spherical area of U. Similar statements hold for vertical line segments $x =$ constant.*

The proof is immediate. □

Now consider a simply connected open set $U \subset \hat{\mathbb{C}}$ with infinite complement and some choice of conformal isomorphism $\psi : \mathbb{D} \to U$.

Theorem 17.4 (Fatou; Riesz and Riesz). *For almost every point $e^{i\theta}$ of the circle $\partial\mathbb{D}$ 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*

$$\lim_{r \nearrow 1} \psi(re^{i\theta}) \in \partial U$$

exists for Lebesgue almost every θ. However, if we fix any particular point $u_0 \in \partial U$, then the set of θ such that this radial limit is equal to u_0 has Lebesgue measure zero.

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. Fatou, in his thesis, showed that any bounded holomorphic function on $\mathbb{D}$ has radial limits in almost all directions, whether or not it is univalent. (See, for example, Hoffman [1962, p. 38].) However the univalent case is all that we will need and is easier to prove than the general theorem.

Proof of Theorem 17.4 (using results from Appendix A). The first half of Theorem 17.4 follows easily from Corollary 17.3, as follows. Let $\mathbb{H}^-$ be the left half-plane, consisting of all points $x + iy \in \mathbb{C}$ with $x < 0$. Map $\mathbb{H}^-$ onto $\mathbb{D} \smallsetminus \{0\}$ by the exponential map $x + iy \mapsto e^x e^{iy}$. Then the square

$$-2\pi < x < 0, \qquad 0 \le y < 2\pi$$

in $\mathbb{H}^-$ maps under $\psi \circ \exp$ onto a neighborhood of the boundary in U. Almost every line $y = \text{constant}$ in $\mathbb{H}^-$ maps onto a curve of finite spherical length, which therefore tends to a well-defined limit as $x \nearrow 1$.

If $U = \psi(\mathbb{D})$ is a bounded subset of $\mathbb{C}$, then a theorem of F. and M. Riesz, as stated in §11 and proved in Theorem A.3 of Appendix A, asserts that any given radial limit can occur only for a set of directions $e^{i\theta}$ of measure zero. For any univalent ψ, we can reduce to the bounded case in two steps, as follows. First suppose that the image $\psi(\mathbb{D}) = U$ omits an entire neighborhood of some point z_0 of $\widehat{\mathbb{C}}$. Then by composing ψ with a fractional linear transformation which carries z_0 to ∞, we reduce to the bounded case. In general, $\psi(\mathbb{D})$ 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. Since the squaring function $\sqrt{\psi} \mapsto \psi$ takes curves of finite spherical length to curves of finite spherical length, we are reduced to the previous case. □

Here is a topological complement. (Compare Figure 41 (p. 196), taking U to be the basin of infinity $\widehat{\mathbb{C}} \smallsetminus K$.) Recall that ∂U is connected by Problem 5-b.

> **Lemma 17.5.** *If two different curves $r \mapsto \psi(re^{i\theta_1})$ and $r \mapsto \psi(re^{i\theta_2})$ land at the same point $u_0 \in \partial U$, then this point u_0 disconnects the boundary of U.*

Proof. These two curves, together with their landing point, form a Jordan curve Γ, which separates the sphere $\widehat{\mathbb{C}}$ into two open sets V_1 and V_2. Similarly, the angles θ_1 and θ_2 separate the circle $\mathbb{R}/2\pi\mathbb{Z}$ into two open intervals I_1 and I_2, numbered so that the rays corresponding to angles $\theta_j \in I_j$ are contained in the corresponding open set V_j. By Theorem 17.4, each of these intervals must contain at least one angle θ_j corresponding to a ray which lands at a point $u_j \neq u_0$ in ∂U. Since the separating curve Γ intersects ∂U only at u_0, this proves that the point u_0 separates ∂U. □

Prime Ends. Next we describe some constructions which depend only on the topology of the pair $(\overline{U}, \partial U)$ and not on conformal structure. We continue to assume that U is a simply connected open subset of the sphere $\widehat{\mathbb{C}}$ and that ∂U has more than one element.

Definition. By a *crosscut* (or "transverse arc") for the pair $(\overline{U}, \partial U)$, we will mean a subset $A \subset U$ which is homeomorphic to the open interval $(0,1)$, such that the closure $\overline{A}$ is homeomorphic to a closed interval with only the two endpoints in ∂U.

Note that it is very easy to construct examples of crosscuts. For example, if U is a bounded subset of $\mathbb{C}$, then we can start with any short line segment inside U and extend in both directions until it first hits the boundary.

Lemma 17.6. *Any crosscut A divides U into two connected components.*

Proof. The quotient space $\overline{U}/\partial U$, in which the boundary is identified to a point, is evidently homeomorphic to the 2-sphere. Since $\overline{A}$ corresponds to a Jordan curve in this quotient 2-sphere, the conclusion follows from the Jordan Curve Theorem. (See, for example, Munkres [1975].) □

Either of the two connected components of $U \smallsetminus A$ will be called briefly a *crosscut neighborhood* $N \subset U$. Note that we can recover the crosscut A from such a crosscut neighborhood since $A = U \cap \partial N$.

Main Definition. By a *fundamental chain* $\mathcal{N} = \{N_j\}$ in U, we will mean a nested sequence

$$N_1 \supset N_2 \supset N_3 \supset \cdots$$

of crosscut neighborhoods $N_j \subset U$ such that the closures $\overline{A}_j$ of the corresponding crosscuts $A_j = U \cap \partial N_j$ are disjoint and such that the diameter of $\overline{A}_j$ tends to zero as $j \to \infty$, using the spherical metric. (Since $\overline{U}$ is compact, this condition depends only on the topology and not on the particular choice of metric.) Two fundamental chains $\{N_j\}$ and $\{N'_k\}$ are *equivalent* if every N_j contains some N'_k, and conversely every N'_k contains some N_j. An equivalence class $\mathcal{E}$ of fundamental chains is called a *prime end* for the pair $(\overline{U}, \partial U)$.

There are a number of possible minor variations on these basic definitions. (Compare Ahlfors [1973], Epstein [1981], Mather [1982], Ohtsuka [1970].) The present version is close to Carathéodory's original construction.

Note that only the crosscuts $\overline{A}_j$ are required to become small as j tends to infinity. In examples such as Figures 37c and 37d (p. 175), the crosscut neighborhood N_j may well have diameter bounded away from zero. Note also that the $\overline{A}_j$ need not converge to a single point. (Compare Figure 9, p. 63.)

Definition. The intersection of the closures $\overline{N}_j \subset \overline{U}$ is called the *impression* of the fundamental chain $\{N_j\}$ or of the corresponding end $\mathcal{E}$.

> **Lemma 17.7.** *For any fundamental chain $\{N_j\}$, the intersection of the open sets N_j is vacuous. However, the impression $\bigcap \overline{N}_j$ is a nonvacuous compact connected subset of ∂U.*

Proof. For any $z \in U$ we will find a j with $z \notin N_j$. Choose a point $z_0 \in U \smallsetminus N_1$ and a path $P \subset U$ joining z_0 to z. If δ is the distance from the compact set P to ∂U and if j is large enough so that the diameter of $\overline{A}_j$ is less than δ, then evidently $\overline{A}_j \cap P = \emptyset$, so A_j cannot disconnect z_0 from z. Since $z_0 \notin N_j$, it follows that $z \notin N_j$. Hence $\bigcap \overline{N}_j$ is a subset of ∂U. This set is clearly compact and nonvacuous. For the proof that it is connected, see Problem 5-b. □

This impression may consist of a single point $z_0 \in \partial U$, as in Figures 37a and 37b. In this case we say that $\{N_j\}$ or $\mathcal{E}$ *converges* to the point z_0. Evidently the impression consists of a single point if and only if the diameter of N_j tends to zero as $j \to \infty$. However, in examples such as Figures 37c and 37d the impression may well be a nontrivial continuum. (See also Figures 9, 38, 40; pp. 63, 187, 192.) Note that two different prime ends may converge to the same point (Figures 37b, 37c), or more generally have impressions which intersect each other.

We will say that two fundamental chains $\{N_j\}$ and $\{N'_k\}$ are *eventually disjoint* if $N_j \cap N'_k = \emptyset$ whenever both j and k are sufficiently large.

> **Lemma 17.8.** *Any two fundamental chains $\{N_j\}$ and $\{N'_k\}$ in U are either equivalent or eventually disjoint.*

Proof. If $N_j \cap N'_k \neq \emptyset$ for all j and k, then we will show that for each j there is a k so that $N_j \supset N'_k$. We first show that every crosscut A'_k with k sufficiently large must intersect the neighborhood N_{j+1}. In fact we have assumed that every N'_k intersects N_{j+1}. Since $\bigcap N'_k = \emptyset$ by Lemma 17.7, it follows that the complement $U \smallsetminus N'_k$ must also intersect N_{j+1} for large k. Since N_{j+1} is connected, this implies that the common boundary A'_k must intersect N_{j+1}.

If no N'_k were contained in N_j, then every N'_k would intersect the complement $U \smallsetminus N_j$. An argument just like that above would then show that A'_k must intersect $U \smallsetminus N_j$ whenever k is large. But if A'_k intersects both $U \smallsetminus N_j$ and N_{j+1}, then it must cross both A_j and A_{j+1}. Hence its diameter must be greater than or equal to the distance between $\overline{A}_j$ and $\overline{A}_{j+1}$. This completes the proof, since it contradicts the hypothesis that the diameter of A'_k tends to zero as $k \to \infty$. □

Now we will combine the topological and analytic arguments. We return to the study of a conformal isomorphism $\psi : \mathbb{D} \xrightarrow{\cong} U \subset \widehat{\mathbb{C}}$.

Lemma 17.9 (Main Lemma). *Given any point $e^{i\theta} \in \partial\mathbb{D}$, there exists a fundamental chain $\{N_j\}$ in $\mathbb{D}$ which converges to $e^{i\theta}$ and which maps under ψ to a fundamental chain $\{\psi(N_j)\}$ in U.*

Proof. We must construct the $N_1 \supset N_2 \supset \cdots$ in $\mathbb{D}$, converging to $e^{i\theta}$, so that the associated crosscuts A_j map to crosscuts in U which have disjoint closures and which have diameters tending to zero. As in the proof of Theorem 17.4, we will make use of the exponential map $\exp : \mathbb{H}^- \to \mathbb{D}\smallsetminus\{0\}$, where $\mathbb{H}^-$ is the left half-plane. In fact, we will actually construct crosscut neighborhoods N'_j in $\mathbb{H}^-$, converging to the boundary point $i\theta$, and then map to $\mathbb{D}$ by the exponential map. Each N'_j will be an open rectangle

$$-\epsilon < x < 0, \quad c_1 < y < c_2$$

in $\mathbb{H}^-$. Thus the corresponding crosscut $A'_j \subset \mathbb{H}^-$ will be made up of three of the four edges of this rectangle and will have endpoints $0+ic_1$ and $0+ic_2$ in $\partial\mathbb{H}^-$. The construction will be inductive. Given $N'_1, \ldots, N'_{j-1}$ we first choose $\delta < 1/j$, which is small enough so that the square S_δ defined by the inequalities

$$-2\delta \le x < 0, \quad \theta - \delta \le y \le \theta + \delta$$

is contained in N'_{j-1}. Mapping S_δ into U by $\psi \circ \exp$, let $\mathcal{A}_\delta$ be the spherical area of its image. Evidently this area tends to zero as $\delta \to 0$. Using Corollary 17.2, we can choose constants c_1 and c_2 so that

$$\theta - \delta < c_1 < \theta < c_2 < \theta + \delta$$

and so that the horizontal line segments $y = c_k$ in S_δ map to curves of length at most $\sqrt{2\mathcal{A}_\delta}$ in $U \subset \widehat{\mathbb{C}}$. This will guarantee that the images of these line segments in U land at well-defined points of ∂U as $x \nearrow 0$. We must also take care to see that these landing points are distinct from each other, and distinct from the endpoints of the crosscuts $\psi \circ \exp(A_h)$ with $h < j$. However, this does not pose any additional difficulty, in view of

Theorem 17.4.

Finally, we must choose a vertical line segment $x = -\epsilon$ inside S_δ which also maps to a curve of length $\le \sqrt{2\mathcal{A}_\delta}$ in U. Setting

$$N'_j \;=\; (-\epsilon,\, 0) \times (c_1,\, c_2) \;\subset \mathbb{H}^-,$$

the inductive construction is complete. Mapping into $\mathbb{D}$, we obtain the required crosscut neighborhoods $N_j = \exp(N'_j) \subset \mathbb{D}$. □

The inverse isomorphism $\psi^{-1} : U \to \mathbb{D}$ is much better behaved.

Corollary 17.10. *Any path $p : [0,1) \to U$ which lands at a well-defined point of ∂U maps under ψ^{-1} to a path in $\mathbb{D}$ which lands at a well-defined point of $\partial\mathbb{D}$. Furthermore, paths which land at distinct points of ∂U map to paths which land at distinct points of $\mathbb{D}$.*

Proof. Let $e^{i\theta} \in \partial\mathbb{D}$ be any accumulation point of the path $t \mapsto \psi^{-1} \circ p(t)$ as $t \nearrow 1$. Choose some fundamental chain $\{N_j\}$ converging to $e^{i\theta}$ as in Lemma 17.9, so that the image under ψ is a fundamental chain $\{\psi(N_j)\}$ in U. For each j, we will prove that $\psi^{-1} \circ p(t) \in N_j$ for all t which are sufficiently close to 1. Otherwise, for some j_0 we could find a sequence of points t_j converging to 1 so that $\psi^{-1} \circ p(t_j) \not\in N_{j_0}$. Since $e^{i\theta}$ is an accumulation point of the path $\psi^{-1} \circ p$ in $\mathbb{D}$, this would imply that this path must pass through both of the crosscuts A_{j_0} and A_{j_0+1} infinitely often as $t \nearrow 1$. Hence the image path $p : [0,1) \to U$ must pass through both $\psi(A_{j_0})$ and $\psi(A_{j_0+1})$ infinitely often. Since there is some positive distance between these crosscuts in U, this contradicts the hypothesis that $p(t)$ converges as $t \nearrow 1$.

If paths $p : [0,1) \to U$ and $q : [0,1) \to U$ landing at two distinct points of ∂U pulled back to paths $\psi^{-1} \circ p$ and $\psi^{-1} \circ q$ landing at a single point of $\partial\mathbb{D}$, then, choosing $\{N_j\}$ as above, each crosscut A_j with j large would cut both $\psi^{-1} \circ p$ and $\psi^{-1} \circ q$. Hence the image crosscut $\psi(A_j) \subset U$ would cut both p and q. As $j \to \infty$, the diameter of $\psi(A_j)$ tends to zero, while its intersection points with p and q tend to distinct points of ∂U. Evidently this is impossible. □

Corollary 17.11. *Every fundamental chain $\{N'_k\}$ in U maps under $\psi^{-1} : U \to \mathbb{D}$ to a fundamental chain $\{\psi^{-1}(N'_k)\}$ in $\mathbb{D}$.*

Proof. Let A'_k be the crosscut which bounds N'_k. It follows from Corollary 17.10 that each $\psi^{-1}(A'_k)$ is a crosscut in $\mathbb{D}$, so that each $\psi^{-1}(N'_k)$ is a crosscut neighborhood in $\mathbb{D}$. It also follows that the closures of all of these crosscuts $\psi^{-1}(A'_k)$ are disjoint. We must prove that the diameter of $\psi^{-1}(A'_k)$ tends to zero as $k \to \infty$. Choose some accumulation point $e^{i\theta} \in \partial\mathbb{D}$ for these sets $\psi^{-1}(A'_k)$ and choose a fundamental chain $\{N_j\}$ in $\mathbb{D}$ which converges to $e^{i\theta}$ and which maps to a fundamental chain $\{\psi(N_j)\}$ in U. Then $N_j \cap \psi^{-1}(N'_k) \neq \emptyset$, and hence $\psi(N_j) \cap N'_k \neq \emptyset$ for all j and k. Therefore, by Lemma 17.8, the two fundamental chains $\{\psi(N_j)\}$ and $\{N'_k\}$ in U are equivalent. It follows that every N_j contains some $\psi^{-1}(N'_k)$. Since the diameter of the entire neighborhood N_j clearly tends to zero as $j \to \infty$, the conclusion follows. □

It follows easily from Lemma 17.9 and Corollary 17.11 that ψ induces a one-to-one correspondence between prime ends of $\mathbb{D}$ and prime ends of U. Furthermore the impression of any prime end of $\mathbb{D}$ is a single point of $\partial\mathbb{D}$, and each point of $\partial\mathbb{D}$ is the impression of one and only one prime end. We can express these facts in clearer form as follows.

Define the *Carathéodory compactification* $\widehat{U}$ of U to be the disjoint union of U and the set consisting of all prime ends of U, with the following topology. For any crosscut neighborhood $N \subset U$ let $\tilde{N} \subset \widehat{U}$ be the union of the set N itself, and the collection of all prime ends $\mathcal{E}$ which are represented by fundamental chains $\{N_j\}$ with $N_j \subset N$. These neighborhoods $\tilde{N}$, together with the open subsets of U, form a basis for the required topology.

Theorem 17.12. *The Carathéodory compactification $\widehat{\mathbb{D}}$ of the open disk is canonically homeomorphic to the closed disk $\overline{\mathbb{D}}$. Furthermore, any conformal isomorphism $\psi : \mathbb{D} \to U \subset \widehat{\mathbb{C}}$ extends uniquely to a homeomorphism from $\overline{\mathbb{D}} \cong \widehat{\mathbb{D}}$ onto $\widehat{U}$.*

The proof is straightforward and will be left to the reader. □

Local Connectivity. A Hausdorff space X is said to be *locally connected* if the following condition is satisfied (see, for example, Kuratowski [1968]):

(a) *Every point $x \in X$ has arbitrarily small connected (but not necessarily open) neighborhoods.*

Other equivalent conditions can be described as follows.

Lemma 17.13. *X is locally connected if and only if:*

(b) *every $x \in X$ has arbitrarily small connected* open *neighborhoods, or*

(c) *every open subset of X is a union of connected open subsets.*

If X is compact metric, then an equivalent condition is that:

(d) *for every $\epsilon > 0$ there exists $\delta > 0$ so that any two points of distance $< \delta$ are contained in a connected subset of X of diameter $< \epsilon$.*

Proof. It is easy to see that (d) $\Rightarrow$ (a) $\Rightarrow$ (c) $\Rightarrow$ (b) $\Rightarrow$ (a). To show that (b) $\Rightarrow$ (d), let $\{Y_\alpha\}$ be the collection of all connected open sets of diameter $< \epsilon$, and let δ be the minimum of $\mathrm{dist}(x, y)$ as (x, y) varies over the compact set $(X \times X) \smallsetminus \bigcup(Y_\alpha \times Y_\alpha)$, where $\delta > 0$ by (b). □

Remark. Sometimes it is important to study the situation around a single point $x \in X$. There is no universally accepted usage, but it seems reasonable to say that X is *locally connected at* x if (a) is satisfied at the single point x, and *openly locally connected at* x if (b) is satisfied. For the difference between these two requirements, see Problem 17-b.

Theorem 17.14 (Carathéodory). *A conformal isomorphism $\psi : \mathbb{D} \xrightarrow{\cong} U \subset \widehat{\mathbb{C}}$ extends to a continuous map from the closed disk $\overline{\mathbb{D}}$ onto $\overline{U}$ if and only if the boundary ∂U is locally connected, or if and only if the complement $\widehat{\mathbb{C}} \smallsetminus U$ is locally connected.*

Proof. If either ∂U or $\widehat{\mathbb{C}} \smallsetminus U$ is locally connected, then we will show that for any fundamental sequence $\{N_j\}$ in U the impression $\bigcap \overline{N}_j$ consists of a single point. It will then follow easily that ψ extends continuously over the boundary of $\mathbb{D}$.

With ϵ and δ as in Lemma 17.13(d), choose j large enough so that the crosscut $A_j = U \cap \partial N_j$ has diameter less than δ. It follows that the two endpoints of $\overline{A}_j$ have distance less than δ and hence are contained in a compact connected set $Y \subset \widehat{\mathbb{C}} \smallsetminus U$ of diameter less than ϵ. *Then the compact set $Y \cup \overline{A}_j \subset \widehat{\mathbb{C}}$ separates N_j from $U \smallsetminus \overline{N}_j$*, for otherwise we could choose some smooth embedded arc $A' \subset \widehat{\mathbb{C}}$ which is disjoint from $Y \cup \overline{A}_j$ and joins some point $x \in N_j$ to a point $y \in U \smallsetminus \overline{N}_j$. Taking A' together with a suitably chosen arc $A'' \subset U$ from x to y which cuts once across the crosscut A_j, we could construct a Jordan curve $A' \cup A''$ which separates the two endpoints of $\overline{A}_j$. Hence it would separate Y, which is impossible since Y was assumed connected.

This compact set $Y \cup \overline{A}_j$ has diameter less than $\epsilon + \delta$. If $\epsilon + \delta < \pi/2$, then one of the connected components of the complement $\widehat{\mathbb{C}} \smallsetminus (Y \cup \overline{A}_j)$ contains an entire hemisphere, while all of the other connected components must have diameter less than $\epsilon + \delta$. If $\epsilon + \delta$ is also smaller than the diameter of $U \smallsetminus \overline{N}_1$, then it follows that $U \smallsetminus \overline{N}_j$ must be contained in the large component of $\hat{\mathbb{C}} \smallsetminus (Y \cup \overline{A}_j)$, and hence N_j must have diameter less than $\epsilon + \delta$. Since ϵ and δ can be arbitrarily small, this proves that the impression $\bigcap \overline{N}_j$ can only be a single point. Using Lemma 17.9, it follows easily that ψ extends continuously over $\partial\mathbb{D}$.

To prove the converse statement, we need the following.

Lemma 17.15. *If f is a continuous map from a compact locally connected space X onto a Hausdorff space Y, then Y is also compact and locally connected.*

In fact, it will be convenient to adopt the convention that all topological spaces are to be Hausdorff.

Proof of Lemma 17.15. This image $f(X) = Y$ is certainly compact. Given any point $y \in Y$ and open neighborhood $N \subset Y$, we can consider the compact set $f^{-1}(y) \subset X$ with open neighborhood $f^{-1}(N)$. Let V_α range over all connected open subsets of $f^{-1}(N)$ which intersect $f^{-1}(y)$. Then the union $\bigcup f(V_\alpha)$ is a connected subset of N. It is also a neighborhood of y, since it contains the open neighborhood $Y \smallsetminus f\big(X \smallsetminus \bigcup V_\alpha\big)$ of y. □

The proof of Theorem 17.14 continues as follows. If ψ extends continuously to $\overline{\psi} : \overline{\mathbb{D}} \to \overline{U}$, then $\overline{\psi}$ maps the circle $\partial\mathbb{D}$ onto ∂U, so ∂U is locally connected. We must show that $\hat{\mathbb{C}} \smallsetminus U$ is also locally connected. Here it is only necessary to consider the situation about a point $z_0 \in \partial U$, since $\hat{\mathbb{C}} \smallsetminus U$ is clearly locally connected away from ∂U. Choose an arbitrarily small connected neighborhood N of z_0 within ∂U, and then choose ϵ so that the ball of radius ϵ about z intersected with ∂U is contained in N. The union of N and the ball of radius ϵ about z within $\hat{\mathbb{C}} \smallsetminus U$ is then the required small connected neighborhood. □

Combining this theorem with Lemma 17.5, we obtain the following.

Theorem 17.16 (Carathéodory). *If the boundary of U is a Jordan curve, then $\psi : \mathbb{D} \xrightarrow{\cong} U$ extends to a homeomorphism from the closed disk $\overline{\mathbb{D}}$ onto the closure $\overline{U}$.*

Proof. If ∂U is a Jordan curve, that is, a homeomorphic image of the circle, then we certainly have a continuous extension $\psi : \overline{\mathbb{D}} \to \overline{U}$ by Theorem 17.14. Since a Jordan curve cannot be separated by any single point, it follows from Lemma 17.5 that this extension is one-to-one, and hence is a homeomorphism. □

Definitions. 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, and *arcwise connected* if there is a topological embedding of $[0,1]$ into X which joins any two given distinct points. It is *locally path-connected* if every point has arbitrarily small path-connected neighborhoods.

As one example, note that the boundary of a simply connected set is always connected by Problem 5-b. However, it need not be path-connected (Figure 37d, p. 175).

To conclude this section, we prove two well-known results.

Lemma 17.17. *If a compact metric space X is locally connected, then it is locally path-connected.*

It follows easily that every connected component is path-connected. Here is a further statement.

Lemma 17.18. *If a Hausdorff space is path-connected, then it is necessarily arcwise connected.*

Proof of Lemma 17.17. Let X be compact metric and locally connected. Given $\epsilon > 0$, it follows from Lemma 17.13 that we can choose a sequence of numbers $\delta_n > 0$ so that any two points with distance $\mathrm{dist}(x, x') < \delta_n$ 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 $\mathrm{dist}(x(0), x(1)) < \delta_0$ 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, by induction. Also, for each fraction of the form i/k_n between 0 and 1 we will choose an intermediate point $x(i/k_n)$ satisfying the following condition: *If $|i/k_n - j/k_{n+1}| \le 1/k_n$ then the distance between $x(i/k_n)$ and $x(j/k_{n+1})$ must be less than $\epsilon/2^n$.* Furthermore, the distance between $x(i/k_n)$ and $x((i+1)/k_n)$ will be less than δ_n. The inductive construction follows. Given $x(i/k_n)$ and $x((i+1)/k_n)$, choose some connected set C which contains both and has diameter less than $\epsilon/2^n$. The points $x(i/k_n)$ and $x((i+1)/k_n)$ can be joined within C by a finite chain of points so that two consecutive points have distance less than δ_{n+1}. Taking k_{n+1} to be a suitably large multiple of k_n, we can evidently choose the required points $x(j/k_{n+1})$ for $i/k_n < j/k_{n+1} < (i+1)/k_n$ from this chain, allowing duplications if necessary. Thus we may assume that $x(r)$ has been defined inductively for a dense set of rational numbers $r = i/k_n$ 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. If $|r - r'| \le 1/k_n$, then we can choose i/k_n so that both $|r - i/k_n|$ and $|r' - i/k_n|$ are at most $1/k_n$. It follows easily that

$$\operatorname{dist}\Big(x(r),\, x(i/k_n)\Big) \;<\; \epsilon/2^n + \epsilon/2^{n+1} + \cdots,$$

and similarly for $x(r')$. Hence $\operatorname{dist}(x(r),\, x(r')) < 4\epsilon/2^n$. 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 at most 4ϵ from $x(0)$ to $x(1)$. Thus X is locally path-connected, as required. $\square$

Proof of Lemma 17.18. 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 takes the constant value $\alpha(I_j) = f(\partial I_j)$ on each of these closed intervals I_j and which coincides with f outside of these subintervals. Then it is easy to check that α is continuous and that for each $x \in \alpha([0,1])$ the preimage $\alpha^{-1}(x) \subset [0,1]$ is a (possibly degenerate) closed interval of real numbers. Note that the image $A = \alpha([0,1]) \subset X$ can be totally ordered by specifying that $\alpha(s) \ll \alpha(t)$ if and only if $\alpha(s)$ and $\alpha(t)$ are distinct points with $s < t$.

A homeomorphism h between the interval $[0,1]$ and the set A can be constructed as follows. Choose a countable dense subset $\{t_1, t_2, \ldots\}$ in the open interval $(0,1)$ and a countable dense subset $\{a_1, a_2, \ldots\}$ in A, excluding the endpoints $\alpha(0)$ and $\alpha(1)$. Now construct a one-to-one correspondence $i \mapsto j(i)$ by induction: Let $j(1) = 1$, and if $j(1), j(2), \ldots, j(i-1)$ have already been chosen, let $j(i)$ be the smallest positive integer which is distinct from $j(1), \ldots, j(i-1)$ and which satisfies the condition that

$$t_h < t_i \quad \Longleftrightarrow \quad a_{j(h)} \ll a_{j(i)}$$

for $h < i$. The required homeomorphism $h : [0,1] \to A$ is now defined by mapping each Dedekind cut in $\{t_i\}$ to the corresponding Dedekind cut in $\{a_j\}$ so that $h(t_i) = a_{j(i)}$ and so that $t < t_i \iff h(t) \ll h(t_i)$. $\square$

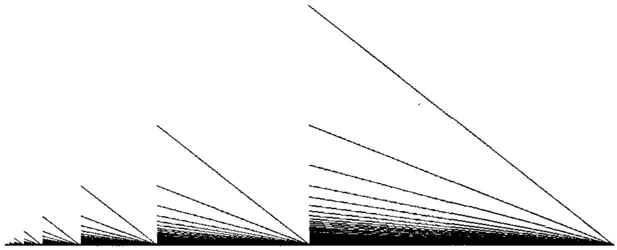


Figure 38. The witch's broom.

Concluding Problems

Problem 17-a. Deformation of paths. With $\psi : \mathbb{D} \xrightarrow{\cong} U$ as in Lemma 17.9, consider a one-parameter family of paths $p_r : [0,1) \to U$, all landing at the same point of ∂U, where $0 \le r \le 1$. Show that the paths $\psi^{-1} \circ p_r$ all land at the same point of $\partial\mathbb{D}$.

Problem 17-b. Local connectivity at a point. Let $X \subset \mathbb{C}$ 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 38). Show that X is locally connected at the origin, but not openly locally connected.

§18. Polynomial Dynamics: External Rays

First recall some definitions from §9. Let $f : \hat{\mathbb{C}} \to \hat{\mathbb{C}}$ be a monic polynomial map

$$f(z) = z^n + a_{n-1}z^{n-1} + \cdots + a_1 z + a_0$$

with $n \geq 2$. (In this section, the degree will be denoted by n.) Then f has a superattracting fixed point at infinity. In particular, it is not difficult to find a constant r_f so that every point z in the neighborhood $|z| > r_f$ of infinity belongs to the basin of attraction $\mathcal{A}(\infty)$. The complement of the basin $\mathcal{A}(\infty)$, that is, the set of all points $z \in \mathbb{C}$ with bounded forward orbit under f, is called the ***filled Julia set*** $K = K(f)$. By Lemma 9.4, this filled Julia set is always a compact subset of the plane, consisting of the Julia set J together with the bounded components (if any) of the complement $\mathbb{C} \smallsetminus J$. These bounded components are all simply connected, and the Julia set J is equal to the topological boundary ∂K. Throughout this section we will assume the following.

Standing Hypothesis. *The Julia set J is connected, or equivalently the filled Julia set K is connected.*

Then by Theorem 9.5 the complement $\mathbb{C} \smallsetminus K$ is conformally isomorphic to $\mathbb{C} \smallsetminus \overline{\mathbb{D}}$ under the Böttcher isomorphism

$$\phi : \mathbb{C} \smallsetminus K \xrightarrow{\cong} \mathbb{C} \smallsetminus \overline{\mathbb{D}}$$

which conjugates the map f outside K to the nth power map $w \mapsto w^n$ outside the closed unit disk, with $\phi(z)$ asymptotic to the identity map at infinity. The continuous function $G : \mathbb{C} \to \mathbb{R}$ defined by

$$G(z) \;=\; \begin{cases} \log|\phi(z)| \;>\; 0 & \text{for} \quad z \in \mathbb{C} \smallsetminus K, \\ 0 & \text{for} \quad z \in K \end{cases}$$

is called the ***Green's function*** for K (Definition 9.6). Note the identity

$$G(f(z)) = nG(z).$$

Each locus $G^{-1}(c) = \{z\ ;\ G(z) = c\}$ with $c > 0$ is called an ***equipotential curve*** around the filled Julia set K. Note that f maps each equipotential $G^{-1}(c)$ to the equipotential $G^{-1}(nc)$ by an n-to-one covering map. The orthogonal trajectories

$$\{z\ ;\ \arg(\phi(z)) \;=\; \text{constant}\}$$

of the family of equipotential curves are called ***external rays*** for K. We

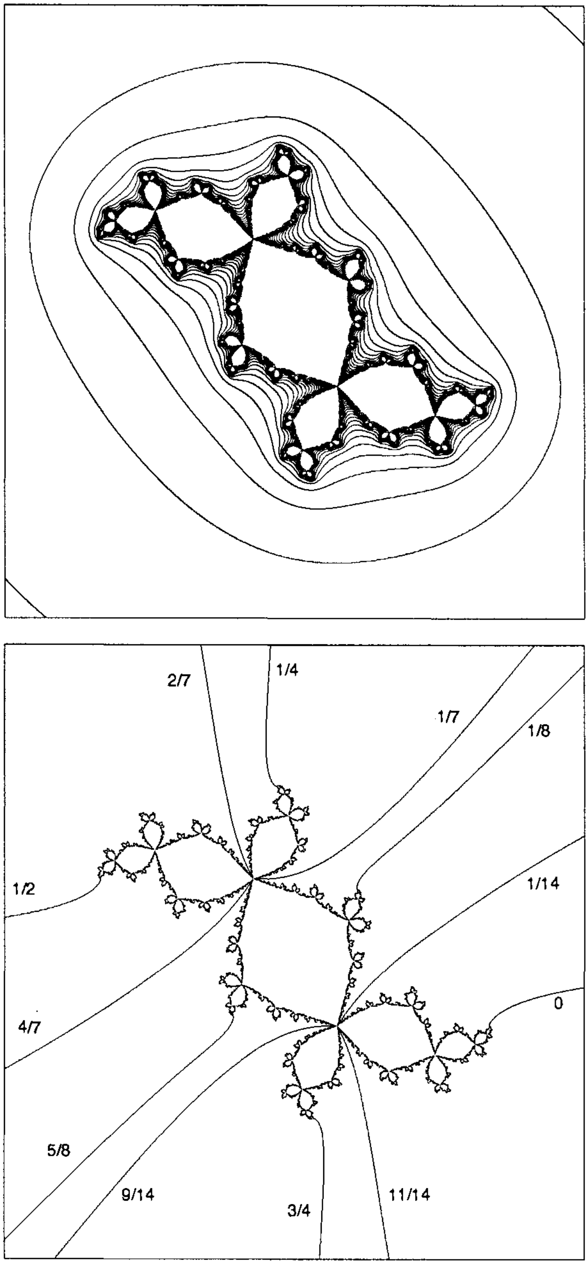


Figure 39. Julia set for the "Douady rabbit"

$$z \mapsto z^2 + c \quad \textit{with} \quad c \approx -.12256 + .74486i.$$

In the top figure, some equipotentials of the form $G = 2^n G_0$ *have been drawn in. The lower figure shows several periodic or pre-periodic external rays.*

will use the notation $R_t \subset \mathbb{C} \smallsetminus K$ for the external ray with angle t, where now we measure angle as a fraction of a full turn, so that $t \in \mathbb{R}/\mathbb{Z}$. By definition, R_t is the image under the inverse Böttcher map ϕ^{-1} of the half-line consisting of all products $re^{2\pi it} \in \mathbb{C} \smallsetminus \overline{\mathbb{D}}$, with $r > 1$. Note the identity

$$f(R_t) = R_{nt}.$$

In particular, if the angle $t \in \mathbb{R}/\mathbb{Z}$ is periodic under multiplication by n, then the ray R_t is periodic. For example, if $n^p t \equiv t \pmod{\mathbb{Z}}$, then it follows that $f^{\circ p}$ maps the ray, R_t onto itself.

Now consider the limit

$$\gamma(t) = \lim_{r \searrow 1} \phi^{-1}(re^{2\pi it}).$$

Whenever this limit exists, we will say that the ray R_t *lands* at the point $\gamma(t)$, which necessarily belongs to the Julia set $J = \partial K$.

Lemma 18.1. *If the ray R_t lands at a single point $\gamma(t)$ of the Julia set, then the ray R_{nt} lands at the point $\gamma(nt) = f(\gamma(t))$. Furthermore each of the n rays of the form $R_{(t+j)/n}$ lands at one of the points in $f^{-1}(\gamma(t))$, and every point in $f^{-1}(\gamma(t))$ is the landing point of at least one such ray.*

Proof. If $z \in J$ is not a critical point, then f maps some neighborhood N of z diffeomorphically onto a neighborhood N' of $f(z)$, carrying any ray $R_s \cap N$ to $R_{ns} \cap N'$. Thus if R_s lands at z then R_{ns} lands at $f(z)$, while if R_t lands at $f(z)$ then for some uniquely determined s of the form $(t+j)/n$ the ray R_s must land at z. If z is a critical point, the situation is similar, except that N maps to N' by a branched covering, so that each ray landing at $f(z)$ is covered by two or more rays landing at z. □

In particular, if the ray R_t is periodic of period $p \geq 1$, and if R_t lands at a point $\gamma(t)$, then it follows that $\gamma(t)$ is a periodic point of f with period dividing p.

Fatou showed that most rays do land, and the Riesz brothers showed that distinct angles usually correspond to distinct landing points. More precisely, applying Theorem 17.4 to the basin of infinity $\mathcal{A}(\infty) \subset \widehat{\mathbb{C}}$, we obtain the following.

Theorem 18.2 (Most Rays Land). *For all $t \in \mathbb{R}/\mathbb{Z}$ outside of a set of measure zero, the ray R_t has a well-defined landing point $\gamma(t) \in J(f)$. Furthermore, for each fixed $z_0 \in J$, the set of t with $\gamma(t) = z_0$ has measure zero.*

However, it is definitely not true that rays land in all cases. Using Carathéodory's work, we can give a precise criterion.

Theorem 18.3 (Landing Criterion). *For any given f with connected Julia set, the following four conditions are equivalent.*

- *Every external ray R_t lands at a point $\gamma(t)$ which depends continuously on the angle t.*
- *The Julia set J is locally connected.*
- *The filled Julia set K is locally connected.*
- *The inverse Böttcher map $\phi^{-1} : \mathbb{C} \smallsetminus \overline{\mathbb{D}} \to \mathbb{C} \smallsetminus K$ extends continuously over the boundary $\partial\mathbb{D}$, and this extended map carries each $e^{2\pi i t} \in \partial\mathbb{D}$ to $\gamma(t) \in J(f)$.*

Furthermore, whenever these conditions are satisfied, the resulting map $\gamma : \mathbb{R}/\mathbb{Z} \to J(f)$ satisfies the semiconjugacy identity

$$\gamma(nt) = f(\gamma(t))$$

and maps the circle $\mathbb{R}/\mathbb{Z}$ onto the Julia set $J(f)$.

Definition. This map γ from $\mathbb{R}/\mathbb{Z}$ onto the locally connected polynomial Julia set $J(f)$ will be called the *Carathéodory semiconjugacy* for $J(f)$.

Proof of Theorem 18.3. First suppose that $\gamma : \mathbb{R}/\mathbb{Z} \to J$ is defined and continuous. Then the image $\gamma(\mathbb{R}/\mathbb{Z})$ is certainly a nonvacuous compact subset of J. Starting with an arbitrary point, say $\gamma(0)$, in this image, we see inductively, using Lemma 18.1, that all iterated preimages also belong to $\gamma(\mathbb{R}/\mathbb{Z})$. Therefore by Corollary 4.13, the image $\gamma(\mathbb{R}/\mathbb{Z})$ is the entire Julia set J, and by Lemma 17.15, J is locally connected. The remaining statements in Theorem 18.3 now follow immediately from Carathéodory's Theorem 17.14, applied to the conformal isomorphism $\hat{\mathbb{C}} \smallsetminus \overline{\mathbb{D}} \to \hat{\mathbb{C}} \smallsetminus K$. □

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 the *symmetric comb*, shown in Figure 40. It consists of the lines $[-1, 1] \times \{\pm c^n\}$ (here $c = .75$), together with the axes $[-1, 1] \times \{0\}$ and $\{0\} \times [-1, 1]$. Evidently, in such an example, the associated landing point function $t \mapsto \gamma(t)$ cannot be continuous.

Here is another result, which follows immediately from Carathéodory's Theorem 17.16.

Corollary 18.4 (Simple Closed Curves). *The Julia set J is a simple closed curve if and only if γ maps $\mathbb{R}/\mathbb{Z}$ homeomor-*

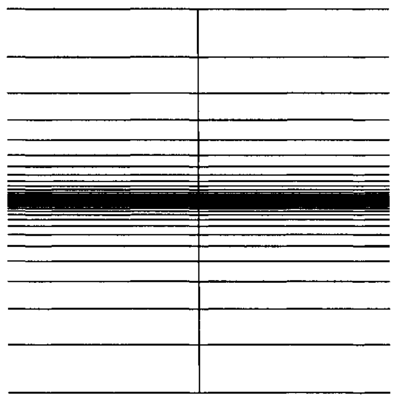


Figure 40. A symmetric comb.

phically onto J.

For examples, see Figures 5a, 11, 14, and 24 (pp. 42, 80, 94, 121).

The following basic result due to Sullivan and Douady will be proved below. (See Sullivan [1983] and also Lyubich [1986, p. 85].)

Theorem 18.5 (Locally Connected Julia Sets). *If the Julia set* J *of a polynomial map* f *is locally connected, then every periodic point in* J *is either repelling or parabolic. Furthermore every cycle of Siegel disks for* f *contains at least one critical point on its boundary.*

Recall from §11 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 statement.

Corollary 18.6 (Some Julia Sets are Not Locally Connected). *If* f *is a polynomial map with a Cremer point, or with a cycle of Siegel disks whose boundary contains no critical point, then the Julia set* $J(f)$ *is not locally connected.*

Remark. It is essential for these results that f be a polynomial. Roesch [1998] has described examples of rational maps with locally connected Julia set which have a Cremer point. In fact the Julia set of a map with Cremer point can be the entire Riemann sphere: If $f(z) = z(z+\alpha)/(\beta z+1)$ where α and β are generic points on the unit circle, then $z=0$ and $z=\infty$ are Cremer points by Corollary 11.3; but using Shishikura [1987] one can show that $J(f)=\widehat{\mathbb{C}}$.

The following supplementary statement holds even for nonpolynomials.

Lemma 18.7 (Siegel Disk Boundaries). *If a rational map f has a Siegel disk $\Delta = f(\Delta)$ such that the boundary $\partial\Delta$ is locally connected or such that the Julia set $J(f)$ is locally connected, then $\partial\Delta$ must be a simple closed curve, and f restricted to $\partial\Delta$ must be topologically conjugate to an irrational rotation. In particular, there can be no periodic points in the boundary.*

Examples of Cremer points were constructed in §11, and examples of Siegel disks with no boundary critical point have been given by Herman [1986] (compare Douady [1987]). However, there is no known example of a Siegel disk which has a boundary periodic point, or a Siegel disk which is not bounded by a simple closed curve.

Proof of Lemma 18.7. Choose a conformal isomorphism $\psi : \mathbb{D} \to \Delta$ satisfying the conjugacy identity

$$\psi(\lambda w) \;=\; f\big(\psi(w)\big),$$

where λ has the form $e^{2\pi i\xi}$ with $\xi \in \mathbb{R} \smallsetminus \mathbb{Q}$. Arguing as in Theorem 17.14, we see that ψ extends to a continuous map $\Psi : \overline{\mathbb{D}} \to \overline{\Delta}$ which must satisfy this same identity. But this implies that $\partial\mathbb{D}$ maps homeomorphically onto $\partial\Delta$, for otherwise, if $\Psi(w_0) = \Psi(uw_0)$ for some $u \neq 1$, $|u| = |w_0| = 1$, then it would follow that $\Psi(\lambda^k w_0) = \Psi(\lambda^k uw_0)$ for $k = 1, 2, \ldots$, and hence that $\Psi(w) = \Psi(uw)$ for all w on the unit circle. If the group of all such u were dense on the circle, then $\Psi(\partial\mathbb{D})$ would be a single point, which is impossible. On the other hand, if this group were generated by some root of unity, then the result of gluing a neighborhood of w_0 in the boundary of $\overline{\mathbb{D}}$ to a neighborhood of uw_0 would be a nonorientable surface embedded in $\widehat{\mathbb{C}}$, which is also impossible. This contradiction proves Lemma 18.7. $\square$

The proof of Theorem 18.5 will be based on the following. Let z_0 be a fixed point in the Julia set J. If J is locally connected, then $\gamma : \mathbb{R}/\mathbb{Z} \to J$ is continuous and onto, hence the set $X = \gamma^{-1}(z_0)$ (consisting of all angles t such that R_t lands at z_0) is a nonvacuous compact subset of the circle. We claim that the n-tupling map $t \mapsto nt$ carries X homeomorphically onto itself. In fact the fixed point z_0 certainly cannot be a critical point of f, since it lies in the Julia set, so f maps a small neighborhood of z_0 diffeomorphically onto a small neighborhood of z_0, carrying external rays landing at z_0 bijectively to external rays landing at z_0.

Lemma 18.8. *Let $n \geq 2$ be an integer and let $X \subset \mathbb{R}/\mathbb{Z}$ be a compact set which is carried homeomorphically into itself by the* *map* $t \mapsto nt \pmod{\mathbb{Z}}$. *Then* X *is finite.*

Proof. In fact we will prove the following more general result. Let X be a compact metric space with distance function $\operatorname{dist}(x,y)$, and let $h : X \to h(X) \subset X$ be a homeomorphism which is *expanding* in the following sense: There should exist numbers $\epsilon > 0$ and $k > 1$ so that

$$\operatorname{dist}(h(x), h(y)) \geq k \operatorname{dist}(x,y) \tag{18 : 1}$$

whenever $\operatorname{dist}(x,y) < \epsilon$. Then we will show that X is finite. Evidently this hypothesis is satisfied in the situation of Lemma 18.8, so this argument will prove the lemma.

Since $h^{-1} : h(X) \to X$ is uniformly continuous, we can choose $\delta > 0$ so that $\operatorname{dist}(x,y) < \epsilon$ whenever $\operatorname{dist}(h(x), h(y)) < \delta$. But this implies that $\operatorname{dist}(x,y) < \delta/k$ by (18 : 1). Since X is compact, we can choose some finite number, say m, of sets of diameter δ which cover X and hence cover $h^{\circ p}(X)$. Applying h^{-p}, we obtain m sets of diameter δ/k^p which cover X. Since p can be arbitrarily large, this proves that X can have at most m points. □

The proof of Theorem 18.5 will also require the following statement, which does not assume local connectivity. We have shown that the set of external rays landing on the fixed point z_0 is finite and maps bijectively to itself under f. Hence the angles of these rays must be periodic under multiplication by the degree n. Replacing f by some iterate if necessary, we may assume that these angles are actually fixed, $nt \equiv t \pmod{\mathbb{Z}}$, so that $f(R_t) = R_t$.

Lemma 18.9. *If a fixed ray* $R_t = f(R_t)$ *lands at* z_0, *then* z_0 *is either a repelling or a parabolic fixed point.*

Proof. This follows easily from Lemma 16.2 (the Snail Lemma). First note that each equipotential $\{z \in \mathbb{C} \; ; \; G(z) = \text{constant} > 0\}$ intersects the ray R_t in a single point. Hence we can parametrize R_t as the image of a topological embedding $p : \mathbb{R} \to \mathbb{C} \smallsetminus K$, which maps each $s \in \mathbb{R}$ to the unique point $z \in R_t$ with $\log G(z) = s$. (In fact, s can be identified with the Poincaré arclength parameter along R_t; compare the proof of Theorem 18.10.) Since $G(f(z)) = n\,G(z)$, we have $f(p(s)) = p(s + \log n)$, and it follows from Corollary 16.3 that the landing point

$$z_0 \;=\; \gamma(t) \;=\; \lim_{s \to -\infty} p(s)$$

is indeed a repelling or parabolic fixed point. □

Proof of Theorem 18.5. If z_0 is a fixed point in a locally connected Julia set, then the preceding two lemmas and the accompanying discussion show that z_0 must be a repelling or parabolic point. The extension to periodic points is straightforward. Now consider an f-invariant Siegel disk $\Delta = f(\Delta)$. Since γ is a continuous map from $\mathbb{R}/\mathbb{Z}$ onto J, it follows that the set $X = \gamma^{-1}(\partial\Delta) \subset \mathbb{R}/\mathbb{Z}$ is compact and infinite, with $nX \subset X$. Hence by Lemma 18.8, the map $t \mapsto nt$ from X to itself cannot be injective. Therefore, there must be two distinct rays R_{t_1} and R_{t_2} landing on $\partial\Delta$ with $f(R_{t_1}) = f(R_{t_2})$. Since $f|_{\partial\Delta}$ is one-to-one, these two rays must land at a common point $\gamma(t_1) = \gamma(t_2)$, and evidently this common landing point must be a critical point of f. This completes the proof of Theorem 18.5 for a Siegel disk of period 1, and the extension to higher periods is straightforward. □

Definition. An external ray R_t is called *rational* if its angle $t \in \mathbb{R}/\mathbb{Z}$ is rational and is called *periodic* if t is periodic under multiplication by the degree n so that $n^p t \equiv t \pmod 1$ for some $p \geq 1$.

Note that R_t is eventually periodic under multiplication by n if and only if t is rational, and is periodic if and only if the number t is rational with denominator relatively prime to n. (If t is rational with denominator d, then the successive images of R_t under f have angles $nt, n^2t, n^3t, \ldots \pmod{\mathbb{Z}}$ with denominators dividing d. Since there are only finitely many such fractions modulo $\mathbb{Z}$, this sequence must eventually repeat. In the special case where d is relatively prime to n, the fractions with denominator d are permuted under multiplication by n modulo $\mathbb{Z}$, hence t is periodic.)

We will prove the following basic results. We continue to assume that K is connected but do not assume local connectivity.

Theorem 18.10 (Rational Rays Land). *Every periodic external ray lands at a periodic point which is either repelling or parabolic. If t is rational but not periodic, then the ray R_t lands at a point which is eventually periodic but not periodic.*

(Compare Douady and Hubbard [1984–85] and Pommerenke [1986].) The converse result, perhaps due to Douady, is more difficult. (Compare Eremenko and Levin [1989], Petersen [1993], and Hubbard [1993].)

Theorem 18.11 (Repelling and Parabolic Points Are Landing Points). *Every repelling or parabolic periodic point is the landing point of at least one periodic ray.*

The following supplementary statement is much easier to prove.

Lemma 18.12. *If one periodic ray lands at z_0, then only finitely many rays land at z_0, and these rays are all periodic of the same* *period (which may be larger than the period of* z_0*).*

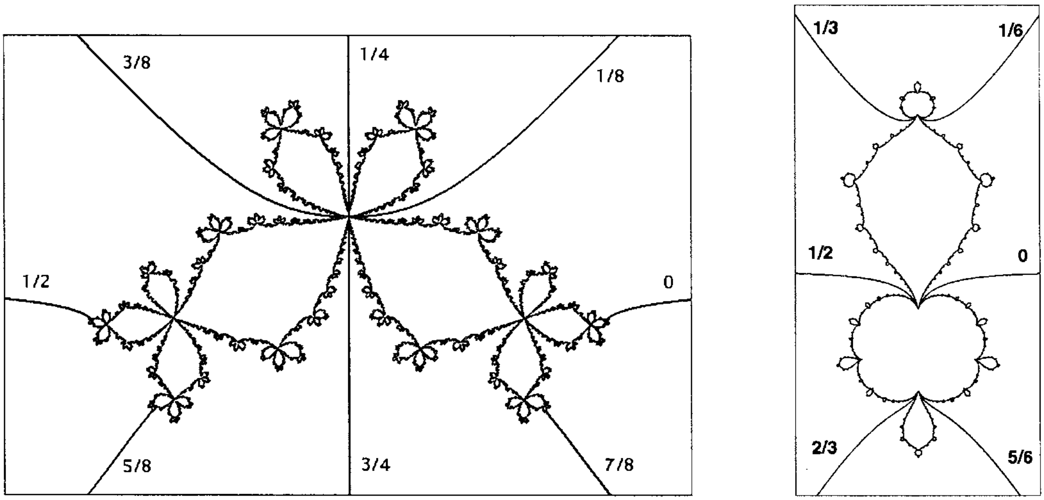


Figure 41. Julia sets for $z \mapsto z^3 - .75z + \sqrt{-7}/4$ *and* $z \mapsto z^3 - iz^2 + z$ *with some external rays indicated.*

As an example, Figure 41 shows the Julia sets for the cubic maps $f(z) = z^3 - 3z/4 + \sqrt{-7}/4$ and $g(z) = z^3 - iz^2 + z$. In the left-hand example, the 0 and 1/2 rays land at distinct fixed points; the 1/8, 1/4, 3/8, and 3/4 rays land at the third fixed point; while the 5/8 and 7/8 rays land on a period 2 orbit. In the right-hand example, the 0 and 1/2 rays must both land at the parabolic fixed point $z = 0$, since the remaining fixed point at $z = i$ is superattracting and hence does not belong to the Julia set. The 1/6, 1/3, 2/3, and 5/6 rays have denominator divisible by 3 and therefore land at pre-periodic points. In fact, these four rays land at the two disjoint preimages of zero. The analogous discussion for Figure 39 (p. 189) will be left to the reader.

In the parabolic case, we can sharpen these statements as follows.

Theorem 18.13 (The Parabolic Case). *Recall that the multiplier at a parabolic fixed point is a primitive* q*th root of unity for some* $q \geq 1$. *Every ray which lands at such a point* z_0 *has period exactly* q. *Furthermore, for every repelling petal* $\mathcal{P}$ *at* z_0, *there is at least one ray landing at* z_0 *through the petal* $\mathcal{P}$.

Example 1. Consider the cubic map $g(z) = z^3 - iz^2 + z$ of Figure 41 (right side). The parabolic fixed point $z = 0$ has multiplier $\lambda = 1$, and there is only one repelling petal, yet two distinct rays R_0 and $R_{1/2}$ land at this point. Figure 19 (p. 106) shows a similar example, with three repelling petals but four fixed landing rays.

Example 2. Now consider the map $f(z) = z^2 + \exp(2\pi i \cdot 3/7)z$ of Figure 21 (p. 109). Here the multiplier is a seventh root of unity, and there are seven repelling petals about the origin. Hence there must be at least seven external rays landing at the origin, and their angles must be fractions with denominator $127 = 2^7 - 1$, so as to be periodic of period 7. 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 [1992].) Thus there are just seven rays which land at zero, one in each repelling petal. The numerators of the corresponding angles are 21, 42, 84, 41, 82, 37, 74.

The proofs begin as follows.

Proof of Lemma 18.12. First consider the special case of a fixed ray $R_{t_0} = f(R_{t_0})$. In other words, suppose that t_0 is a number of the form $j/(n-1)$, so that $t_0 \equiv nt_0 \pmod{\mathbb{Z}}$. If R_{t_0} lands at z_0, then clearly $f(z_0) = z_0$.

Let X be the set of all angles $x \in \mathbb{R}/\mathbb{Z}$ such that the ray R_x lands at z_0. Since f maps a neighborhood of z_0 diffeomorphically onto a neighborhood of z_0, preserving the cyclic order of the rays which land at z_0, it follows that the n-tupling map carries X injectively into itself preserving cyclic order. For every $x \in X$ with $x \not\equiv t_0 \pmod{\mathbb{Z}}$ and hence $n^k x \not\equiv t_0 \pmod{\mathbb{Z}}$, define the sequence $x_0, x_1, \ldots$ of representative points for the orbit of x within the interval (t_0, t_0+1) by the congruence

$$x_k \equiv n^k x \pmod{\mathbb{Z}} \qquad \text{with} \quad t_0 < x_k < t_0 + 1.$$

First suppose that $x_0 < x_1$, and hence $t_0 < x_0 < x_1 < x_2 < \cdots < t_0 + 1$. Then the x_k must converge to some angle $\hat{x}$, which is necessarily a fixed point for the map $t \mapsto nt \pmod{\mathbb{Z}}$. But this is impossible, since this map has only strictly repelling fixed points. Similarly, the case $x_0 > x_1$ is impossible. This proves that $x_0 = x_1$ so that x is one of the $n-1$ fixed points of the map $t \mapsto nt \pmod{\mathbb{Z}}$.

Now suppose that the smallest period of a ray R_t landing at z_0 is $p > 1$. Replacing f by the iterate $g = f^{\circ p}$, the argument above shows that every ray which lands at z_0 is carried into itself by g and hence has period $\le p$ under f. This proves that the period is exactly p. $\square$

Proof of Theorem 18.10. We will make use of the Poincaré metric on $\mathbb{C} \smallsetminus K \cong \mathbb{C} \smallsetminus \overline{\mathbb{D}}$. Note that f is a local isometry for this metric. In fact the universal covering of $\mathbb{C} \smallsetminus \overline{\mathbb{D}}$ is isomorphic to the right half-plane

$\mathbb{H}^+ = \{w = u + iv \ ; \ u > 0\}$ under the exponential map. Here the real part u of w corresponds to the Green's function G on $\mathbb{C} \setminus K$. The map f on $\mathbb{C} \setminus K$ corresponds to the nth power map on $\mathbb{C} \setminus \overline{\mathbb{D}}$, and to the Poincaré isometry $w \mapsto nw$ on $\mathbb{H}^+$. Each external ray corresponds to a horizontal half-line $v =$ constant in $\mathbb{H}^+$, and the Poincaré arclength $\int |dw|/u$ reduces to $\int du/u = \int d \log u$ along each such half-line.

Again, consider first the case of a fixed ray $f(R_t) = R_t$. As in the proof of Lemma 18.9, we can introduce the parameter $s = \log G(z)$ along this ray, so that R_t is the image of a path $p : \mathbb{R} \to \mathbb{C} \setminus K$, where f maps $p(s)$ to $p(s + \log n)$. Thus $R_t = p(\mathbb{R})$ is the union of path segments

$$I_k = p\Big([k \log n, \ (k+1) \log n]\Big)$$

of Poincaré arclength $\log n$, where k ranges over all integers, and where f maps I_k isometrically onto I_{k+1}. On the other hand, $G(p(s)) = e^s$ tends to zero as $s \to -\infty$, so any limit point $\hat{z}$ of $p(s)$ as $s \to -\infty$ must belong to the Julia set $J = \partial K$. Using Theorem 3.4, given any neighborhood N of $\hat{z}$ we can find a smaller neighborhood N' so that any I_k which intersects N' is contained in N. Since f maps one endpoint of I_k to the other, this shows that $N \cap f(N) \neq \emptyset$ for every neighborhood N, so that $\hat{z}$ must be a fixed point of f. But the set of all limit points must be connected. (See Problem 5-b.) Since f has only finitely many fixed points, this proves that the ray R_t must land at a single point, which is necessarily a fixed point of the map f, and is necessarily repelling or parabolic by Lemma 18.9. The corresponding statement for a ray of period p now follows by applying the argument above to the iterate $g = f^{\circ p}$.

Finally, if t belongs to $\mathbb{Q}/\mathbb{Z}$ then it must certainly be eventually periodic under multiplication by n, so it follows by Lemma 18.1 that the ray R_t lands. This completes the proof of Theorem 18.10. $\square$

The proof that at least one periodic ray lands on a repelling or parabolic point will be based on the following ideas. It clearly suffices to consider the special case of a fixed point at the origin.

Definition. We assume that $0 = f(0)$ is either repelling or parabolic. By a *backward orbit* for the map $f : \mathbb{C} \to \mathbb{C}$ we mean an infinite sequence $\mathbf{z} = (z_0, z_1, \ldots)$ of points $z_k \in \mathbb{C}$ which satisfy $z_k = f(z_{k+1})$, so that $z_0 \hookleftarrow z_1 \hookleftarrow z_2 \hookleftarrow \cdots$. Let E be the space consisting of all such sequences which converge to zero nontrivially in backward time, so that

$$\lim_{k \to \infty} z_k = 0, \qquad \text{but } z_k \neq 0 \text{ for } k \text{ sufficiently large.}$$

To specify a topology for this space E, we can first choose neighborhoods

V_0 and V_1 of 0 so that f maps V_0 diffeomorphically onto V_1. Let $g : V_1 \to V_0$ be the inverse map. The space E can be described as the union of the nested subsets

$$E_0 \subset E_1 \subset E_2 \subset \cdots,$$

where $\mathbf{z} \in E_k$ if and only if $z_j \in V_0 \cap V_1 \smallsetminus \{0\}$ for all $j \geq k$. If $\mathbf{z}$ belongs to E_k, then every coordinate of $\mathbf{z}$ can be expressed as a holomorphic function of z_k. In fact $z_j = f^{\circ(k-j)}(z_k)$ if $j \leq k$, and $z_j = g^{\circ(j-k)}(z_k)$ if $j \geq k$. Now topologize E so that each E_k is an open subset and so that each correspondence $\mathbf{z} \mapsto z_k$ maps E_k homeomorphically into V_0. By this same construction, we can assign a conformal structure to E, so that each connected component of E becomes a Riemann surface. Note that the shift map $(z_0, z_1, z_2, \ldots) \mapsto (z_1, z_2, z_3, \ldots)$ is then a conformal isomorphism, with inverse $\mathbf{f} : E \to E$ defined by the formula

$$\mathbf{f}(z_0, z_1, z_2, \ldots) = (f(z_0), f(z_1), f(z_3), \ldots).$$

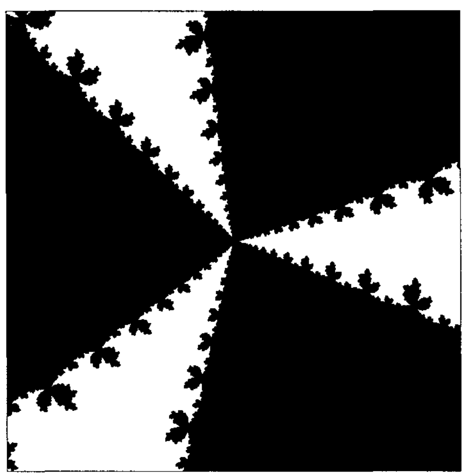


Figure 42. The set $E \cong \mathbb{C}$ for one of the repelling fixed points of Figure 39 (p. 189), with $\tilde{K}$ shaded. Here $\mathbf{f}$ permutes the three components of $E \smallsetminus \tilde{K}$, and $\mathbf{f}^{\circ 3}$ maps each component U^k to itself, expanding by a factor of $\lambda^3 \simeq 1.357 - .004\,i$.

If the origin is a repelling point of multiplier λ, then we can define a canonical diffeomorphism $\kappa : E \xrightarrow{\cong} \mathbb{C} \smallsetminus \{0\}$ by setting

$$\kappa(\mathbf{z}) = \lim_{k \to \infty} \lambda^k z_k.$$

(Compare the proof of Theorem 8.2, Kœnigs Linearization.) Thus E is connected, and hence a Riemann surface, in the repelling case. In the parabolic case, E will be a union of m simply connected Riemann surfaces $E_{\mathcal{P}} \cong \mathbb{C}$, where m is the number of repelling petals in the Leau-Fatou flower around the origin. In fact, replacing f by some iterate if necessary, we may assume that the multiplier of f at the origin is $+1$. For each repelling petal $\mathcal{P}$, let $E_{\mathcal{P}} \subset E$ be the set of $\mathbf{z}$ such that $z_k \in P$ for k sufficiently large, and let $\alpha_{\mathcal{P}} : \mathcal{P} \to \mathbb{C}$ be the Fatou coordinate, satisfying $\alpha_{\mathcal{P}}(f(z)) = \alpha_{\mathcal{P}}(z) + 1$ whenever both z and $f(z)$ belong to $\mathcal{P}$. Thus $\alpha_{\mathcal{P}}(z_k) = \alpha_{\mathcal{P}}(z_{k+1}) + 1$ for large k, and we can define a conformal isomorphism $\Phi_{\mathcal{P}} : E_{\mathcal{P}} \xrightarrow{\cong} \mathbb{C}$ by the formula $\Phi_{\mathcal{P}}(\mathbf{z}) = \alpha_{\mathcal{P}}(z_k) + k$ for any k which is large enough so that the points $z_k, z_{k+1}, \ldots$ all belong to $\mathcal{P}$.

Now let $\tilde{K} \subset E$ be the closed subset consisting of all points $\mathbf{z} = (z_0, z_1, \ldots) \in E$ for which the z_k belong to the filled Julia set K.

Lemma 18.14. *Let U_0 be any connected component of $E \smallsetminus \tilde{K}$. Then U_0 is a universal covering of $\mathbb{C} \smallsetminus K$ under the projection*

$$\pi : (z_0, z_1, \ldots) \mapsto z_0 \quad \textit{from} \quad U_0 \quad \textit{to} \quad \mathbb{C} \smallsetminus K.$$

Proof. To show that $\pi : U_0 \to \mathbb{C} \smallsetminus K$ is actually a covering map, consider any simply connected open set $W \subset \mathbb{C} \smallsetminus K$. Then we will show that W is evenly covered in U_0. Let $\mathbf{z} = (z_0, z_1, \ldots)$ be any point of U_0 with $z_0 \in W$. Since each $f^{\circ k} : \mathbb{C} \smallsetminus K \to \mathbb{C} \smallsetminus K$ is a covering map, there is a unique branch $g_k : W \to \mathbb{C} \smallsetminus K$ of f^{-k} restricted to W such that $g_k(z_0) = z_k$. Since the set $\mathbb{C} \smallsetminus K$ is hyperbolic, the collection of maps $\{g_k\}$ forms a normal family. In fact the sequence $\{g_k\}$ must converge locally uniformly to the zero map, for otherwise we could choose a subsequence converging locally uniformly to a non-zero holomorphic limit. This is impossible since, if k is large enough so that $g_j(z_0) = z_j$ belongs to V_0 for $j \ge k$, then g_k maps a small compact neighborhood N of z_0 into V_0, and it follows easily that the successive images $g_j(N)$ converge uniformly to zero as $j \to \infty$.

Now the required lifting $\mathbf{g} : W \to U_0$ from W to U_0 is defined by

$$\mathbf{g}(z) = \big(z,\, g_1(z),\, g_2(z),\, \ldots\big).$$

Note that $\mathbf{g}(z_0)$ is equal to the specified $\mathbf{z} \in U_0$. Thus W is evenly covered, which proves that $\pi : U_0 \to \mathbb{C} \smallsetminus K$ is a covering map.

On the other hand, each connected component U_0 of $E \smallsetminus \tilde{K}$ is simply connected: Consider any closed loop $\mathbf{h} : \mathbb{R}/\mathbb{Z} \to U_0$. Setting $\mathbf{h}(t) = (h_0(t), h_1(t), \ldots)$ we can choose k so that $h_k(\mathbb{R}/\mathbb{Z})$ is contained in a linearizing neighborhood or a petal. This image loop can be shrunk to a point within $\mathbb{C} \setminus K$, since it cannot enclose any point of the large connected set K. It follows easily that the loop $\mathbf{h}(\mathbb{R}/\mathbb{Z})$ can be contracted within U_0. This completes the proof that the covering map $U_0 \to \mathbb{C} \setminus K$ is a universal covering. $\square$

Lemma 18.15 (Main Lemma). *Each component U_0 of $E \setminus \tilde{K}$ is mapped onto itself by some iterate of $\mathbf{f}$.*

First consider the repelling case, with $\kappa : E \xrightarrow{\cong} \mathbb{C} \setminus \{0\}$. The proof of Lemma 18.15 in this case begins as follows. Consider the sequence of successively larger concentric circles $\mathcal{C}_q \subset E$, defined by the equation $|\kappa(\mathbf{z})| = |\lambda|^q$ for $q \geq 0$, where λ is the multiplier so that $\mathbf{f}(\mathcal{C}_q) = \mathcal{C}_{q+1}$.

Lemma 18.16. *Some image $U_k = \mathbf{f}^{\circ k}(U_0)$, with $k \in \mathbb{Z}$, has the following property for every $q > 0$. Using the Poincaré metric associated with the set U_k, the two subsets $U_k \cap \mathcal{C}_0$ and $U_k \cap \mathcal{C}_q$ have distance less than or equal to $q \log n$ within U_k.*

Proof. Note that any two such images U_k and U_ℓ are either equal or disjoint. Let U be the union of these open sets U_k. The Green's function $G \circ \pi : E \to \mathbb{R}$ is harmonic and strictly positive on U, and identically zero on $\partial U \subset \tilde{K}$. Let $G_0 = (G \circ \pi)(\hat{\mathbf{z}})$ be the maximum value which is attained by $G \circ \pi$ on the set $U \cap \mathcal{C}_0$, and suppose that $\hat{\mathbf{z}} \in U_k$. Since $f^{\circ q}(U \cap \mathcal{C}_0) = U \cap \mathcal{C}_q$, and since $(G \circ \pi \circ \mathbf{f}^{\circ q})(\mathbf{z}) = n^q \, (G \circ \pi)(\mathbf{z})$ where n is the degree, it follows that the maximum value of $G \circ \pi$ on $U \cap \mathcal{C}_q$ is $n^q \, G_0$, attained at the point $\mathbf{f}^{\circ q}(\hat{\mathbf{z}})$. In U_k, just as in $\mathbb{C} \setminus K$, the orthogonal trajectories of the equipotentials $G \circ \pi(\mathbf{z}) = \text{constant}$ can be described as external rays and can be parametrized by their Poincaré arclength $\log G \circ \pi$. Start at the point $\hat{\mathbf{z}} \in U_k \cap \mathcal{C}_0$ and follow the external ray through this point until we first arrive at some $\mathbf{z}' \in U_k \cap \mathcal{C}_q$. The Poincaré length of this ray segment is $\log(G \circ \pi)(\mathbf{z}') - \log(G \circ \pi)(\hat{\mathbf{z}}) \leq \log(n^q G_0) - \log(G_0) = q \log n$, as required. $\square$

To prove Lemma 18.15 in the repelling case, we must show that the sets U_k cannot be pairwise disjoint. Suppose to the contrary that the U_k are pairwise disjoint. Then we will show that the Poincaré distance in U_k between $U_k \cap \mathcal{C}_0$ and $U_k \cap \mathcal{C}_q$ would have to increase more than linearly with q, which is impossible by Lemma 18.16. This contradiction will complete the proof. The argument is based on the following estimate.

Lemma 18.17. *Consider the vertical strip S of width w consisting of all complex numbers $x + iy$ with $0 \leq x \leq w$. Let U be a simply connected region which intersects both boundary lines, which we denote briefly by $\{x = 0\}$ and $\{x = w\}$, and* *let* $\mathbf{a} = \int_{S\cap U} dx\,dy$ *be the Euclidean area of* $S \cap U$. *Then the Poincaré distance between* $U \cap \{x=0\}$ *and* $U \cap \{x=w\}$ *within* U *is greater than* $w^2/4\,\mathbf{a}$.

In particular, as the area tends to zero with w fixed, this distance tends to infinity. The proof will be based on the following inequality. *If* γ *is a smooth path segment in* U *with Poincaré arclength* $L_U(\gamma)$, *then*

$$\tfrac{1}{2} L_U(\gamma) \;\le\; \int_\gamma |dz|/\mathrm{dist}(z,\,\partial U) \;\le\; 2\,L_U(\gamma), \qquad (18:2)$$

using Euclidean distance to the boundary. See Appendix A, Corollary A.8.

Using this inequality, the proof of Lemma 18.17 proceeds as follows. Let $\ell(x_0)$ be the length of the intersection of U with the vertical line $\{x = x_0\}$, so that $\mathbf{a} = \int_0^w \ell(x)\,dx$. Let J be the subset of the interval $I = [0, w]$ consisting of points x such that $\ell(x) \le 2\mathbf{a}/w$. Then the Lebesgue measure of J satisfies $\mathrm{Leb}(J) > w/2$. In fact

$$\mathbf{a} = \int_I \ell(x)\,dx \;\ge\; \int_{I\smallsetminus J} \ell(x)\,dx \;>\; \mathrm{Leb}(I \smallsetminus J)\,2\,\mathbf{a}/w,$$

which implies that $\mathrm{Leb}(I \smallsetminus J) < w/2$, and hence $\mathrm{Leb}(J) > w/2$. Note that the Euclidean distance of any point $z = x + iy \in S \cap U$ from ∂U is at most $\ell(x)/2$. Using $(18:2)$, it follows that the Poincaré length $L_U(\gamma)$ of any path from $\{x = 0\}$ to $\{x = w\}$ within U satisfies

$$L_U(\gamma) \;\ge\; \int_I dx/\ell(x) \;\ge\; \int_J dx/\ell(x) \;\ge\; \mathrm{Leb}(J)\,w/2\mathbf{a} \;>\; w^2/4\mathbf{a}.$$

This proves Lemma 18.17. □

We will use the flat metric $|d\kappa|/\kappa$ on the set $E \cong \mathbb{C} \smallsetminus \{0\}$. (Equivalently, we could identify E with the bi-infinite cylinder $\mathbb{C}/(2\pi i\mathbb{Z})$ with its usual flat Euclidean metric, using the conformal equivalence $\mathbf{z} \mapsto \log\kappa(\mathbf{z})$.) Using this metric, note that $\mathbf{f}$ maps E isometrically onto itself. Let $A_q \subset E$ be the annulus bounded by the circles $\mathcal{C}_q$ and $\mathcal{C}_{q+1}$, and let $\mathbf{a}$ stand for area with respect to this flat metric. Assuming that the U_k are pairwise disjoint, since $\mathbf{a}(A_0)$ is finite it follows that $\mathbf{a}(U_k \cap A_0)$ tends to zero as $|k| \to \infty$. Since $\mathbf{f}^{\circ q}$ maps the intersection $U_{-q} \cap A_0$ isometrically onto $U_0 \cap A_q$, it follows that $\mathbf{a}(U_0 \cap A_q)$ also tends to zero as $q \to \infty$. Using Lemma 18.17, it follows that the Poincaré distance, within U_0, between $\mathcal{C}_q$ and $\mathcal{C}_{q+1}$ tends to infinity as $q \to \infty$. Hence the Poincaré distance within U_0 between $\mathcal{C}_0$ and $\mathcal{C}_q$ increases more than linearly with q. A similar argument applies to the Poincaré distance within each U, but this contradicts Lemma 18.16, and this contradiction completes the proof of Lemma 18.15 in the repelling case. □

Proof of Theorem 18.11 in the Repelling Case. By Lemma 18.15 we can choose $k > 0$ so that $\mathbf{f}^{\circ k}$ maps the simply connected set U_0 biholomorphically onto itself, evidently without fixed points. Hence we can form a quotient Riemann surface $\mathcal{S}$ by identifying each $\mathbf{z} \in U_0$ with $\mathbf{f}^{\circ k}(\mathbf{z})$. We know from §2 that $\mathcal{S}$ can only be an annulus or a punctured disk. However, the punctured disk case is impossible, since Lemma 18.17 easily yields a positive lower bound for the Poincaré arclength of a path joining $\mathbf{z}$ to $\mathbf{f}^{\circ k}(\mathbf{z})$. Thus $\mathcal{S}$ is an annulus. In particular, there is a unique simple closed Poincaré geodesic on $\mathcal{S}$. (Compare Problem 2-f. Intuitively, think of a rubber band placed around a napkin ring which shrinks to the unique simple closed curve of minimal length.) Lifting to the universal covering U_0 of $\mathcal{S}$, we obtain an $\mathbf{f}^{\circ k}$-invariant bi-infinite Poincaré geodesic. Now projecting to $\mathbb{C} \smallsetminus K$ we obtain an $f^{\circ k}$-invariant bi-infinite Poincaré geodesic $p : \mathbb{R} \to \mathbb{C} \smallsetminus K$. Since the Green's function $G(p(s))$ tends to infinity as $s \to +\infty$, this can only be an external ray. (Any Poincaré geodesic in the right half-plane is either horizontal, corresponding to an external ray, or else has bounded real part.) On the other hand, since the Kœnigs coordinate of $p(s)$ tends to zero as $s \to -\infty$, this ray lands at the origin. This proves Theorem 18.11 in the repelling case. $\square$

Proofs in the Parabolic Case. As noted earlier, it suffices to consider the case $\lambda = 1$. Recall that the set $E_{\mathcal{P}}$ consists of backward orbits which converge to zero through a given repelling petal $\mathcal{P}$, with Fatou isomorphism $\Phi_{\mathcal{P}} : E_{\mathcal{P}} \xrightarrow{\cong} \mathbb{C}$ satisfying $\Phi_{\mathcal{P}} \circ \mathbf{f}(\mathbf{z}) = \Phi_{\mathcal{P}}(\mathbf{z}) + 1$. Recall also that every repelling petal $\mathcal{P}$ must intersect the neighboring attracting petals. In fact every $z \in P$ with $\phi(z)$ sufficiently far from the real axis must have orbit converging to zero and hence belong to the filled Julia set K. It follows that any point $z \in E_{\mathcal{P}}$ for which $\Phi_{\mathcal{P}}(\mathbf{z}) = u + iv$ with $|v|$ sufficiently large must belong to the corresponding set $\widetilde{K}$. In other words, the entire image $\Phi_{\mathcal{P}}(E_{\mathcal{P}} \smallsetminus \widetilde{K}) \subset \mathbb{C}$ must be contained in a strip $|v| <$ constant, of finite height. In place of the circles $\mathcal{C}_q$ and annuli A_q of the argument above, we now use the vertical lines

$$L_q \;=\; \{\mathbf{z} \in E_{\mathcal{P}}\ ;\ \mathrm{Re}(\Phi_{\mathcal{P}}(\mathbf{z})) = q\}$$

and the vertical strips bounded by L_q and L_{q+1}. An argument analogous to that of Lemma 18.16 shows that the Poincaré distance between L_0 and L_q within a suitable U_k is less than or equal to $q \log n$. On the other hand, if the images $U_k = \mathbf{f}^{\circ k}(U_0)$ were pairwise disjoint, then an area argument, just like that above, would show that this distance must grow more than linearly with q. This contradiction completes the proof of Lemma 18.15.

In fact, in this parabolic case with $\lambda = 1$, we get the sharper statement that $\mathbf{f}(U_0) = U_0$. For if $U_1 = \mathbf{f}(U_0) \neq U_0$, then the image $\Phi_{\mathcal{P}}(U_1)$ in the strip $\{u + iv\,;\ |v| < \text{constant}\}$ would have to lie either above or below the image $\Phi_{\mathcal{P}}(U_0)$. If, for example, it were above, then $\Phi_{\mathcal{P}}(U_2)$ would have to be above $\Phi_{\mathcal{P}}(U_1)$, and so on, so that $U_k \neq U_0$ for all $k > 0$.

Now, as in the argument above, we can form a quotient annulus $\mathcal{S}$ by identifying each $\mathbf{z} \in U_0$ with $\mathbf{f}(\mathbf{z})$. The unique simple closed geodesic in $\mathcal{S}$ lifts to a bi-infinite geodesic in U_0, which projects to an external ray $R_t = f(R_t)$ which must land at the origin through the given repelling petal $\mathcal{P}$. This completes the proof of Theorems 18.11 and 18.13. $\square$

Further development and application of these ideas may be found, for example, in Goldberg and Milnor [1993], Kiwi [1997, 1999], Milnor [2000a, 2000b], Schleicher [2000], and Schleicher and Zimmer [2003].

§19. Hyperbolic and Subhyperbolic Maps

This section will describe some examples of locally connected Julia sets, using arguments due to Sullivan, Thurston, Douady, and Hubbard.

Definition. A rational map f will be called dynamically *hyperbolic* if f is *expanding* on its Julia set J in the following sense: There exists a conformal metric μ, defined on some neighborhood of J, such that the derivative Df_z at every point $z \in J$ satisfies the inequality

$$\|Df_z(v)\|_\mu > \|v\|_\mu$$

for every nonzero vector v in the tangent space $T\hat{\mathbb{C}}_z$. (Notation as in the proof of Theorem 2.11.) Since J is compact, it follows that there exists an *expansion constant* $k > 1$ with the property that $\|Df_z\|_\mu \geq k$ for all points z in some neighborhood of J. In particular, any smooth path of length L in this neighborhood maps to a smooth path of length $\geq kL$. It follows easily that every $z \in J$ has some open neighborhood N_z in $\hat{\mathbb{C}}$ such that the associated Riemannian distance function satisfies

$$\mathrm{dist}_\mu(f(x), f(y)) \geq k \cdot \mathrm{dist}_\mu(x, y) \qquad (19:1)$$

for all $x, y \in N_z$.

Recall from §11 that the *postcritical set* P of f is the collection of all forward images $f^{\circ j}(c)$ with $j > 0$, where c ranges over the critical points of f.

Theorem 19.1 (Hyperbolic Maps). *A rational map of degree $d \geq 2$ is dynamically hyperbolic if and only if its postcritical closure $\overline{P}$ is disjoint from its Julia set, or if and only if the orbit of every critical point converges to an attracting periodic orbit. In fact if f is hyperbolic, then every orbit in its Fatou set must converge to an attracting periodic orbit.*

Remark. Hyperbolic maps have many other important properties. Every hyperbolic Julia set has area zero (see for example Carleson and Gamelin [1993]). It is not hard to see that every periodic orbit for a hyperbolic map must be either attracting or repelling. If f is hyperbolic, then every nearby map is also hyperbolic. Furthermore, according to Mañé, Sad, and Sullivan [1983] or Lyubich [1983a, 1990], the Julia set $J(f)$ deforms continuously under a deformation of f through hyperbolic maps. In contrast, in the nonhyperbolic case, a small change in f may well lead to a drastic alteration of $J(f)$. For example, if a Siegel point becomes repelling under a small deformation of f, then it will suddenly belong to the Julia set.

The well-known *Generic Hyperbolicity Conjecture* is the conjecture that every rational map can be approximated arbitrarily closely by a hyperbolic map. (See for example McMullen [1994b].) However, this has not been proved, even for quadratic polynomials.

Proof of Theorem 19.1. Let V be the complement $\hat{\mathbb{C}} \smallsetminus \overline{P}$ and let $W = f^{-1}(V) \subset \hat{\mathbb{C}}$. As in the proof of Theorem 11.17, we see that $W \subset V$ and that f maps W onto V by a d-fold covering map. Furthermore, if we exclude the trivial case of a map which is conjugate to $z \mapsto z^{\pm d}$, then every connected component of V or W is conformally hyperbolic.

Now suppose that $\overline{P} \cap J = \emptyset$, or in other words that $J \subset V$. Then W must be strictly smaller than V, for otherwise V would map to itself under f and hence be contained in the Fatou set. In fact any connected component of W which intersects J must be strictly smaller than the corresponding component of V. It follows, arguing as in (11 : 6), that $\|Df_z\|_V > 1$ for every $z \in W$. (Here the subscript V indicates that we use the Poincaré metric associated with the hyperbolic surface V.) Since $J \subset W$, this proves that f is dynamically hyperbolic.

(**Remark:** An alternative version of this argument would be based on the observation that f^{-1} must lift to a single-valued map F from the universal covering surface $\widetilde{V}$ to itself. Then F must be distance decreasing for the Poincaré metric on $\widetilde{V}$, and hence f must be distance increasing for the Poincaré metric on V. Compare the proof of Theorem 19.6.)

Conversely, suppose that f is dynamically hyperbolic. Thus we can choose some conformal metric μ on a neighborhood of J so that f is expanding with expansion factor $\geq k > 1$ throughout some possibly smaller neighborhood V' of J. It certainly follows that there cannot be any critical point in V'. Let $N_\epsilon(J)$ be the open ϵ-neighborhood of J with respect to this metric. Then we can choose $\epsilon > 0$ small enough so that:

(1) every point in the open ϵ-neighborhood $N_\epsilon(J)$ can be joined to J within V' by at least one minimal geodesic for the metric μ; and

$$f^{-1}N_\epsilon(J) \subset V'. \tag{2}$$

For any $z \in f^{-1}N_\epsilon(J)$, it follows that

$$\mathrm{dist}(z, J) \leq \mathrm{dist}(f(z), J)/k .$$

In fact, a minimal geodesic from $f(z)$ to J will necessarily lie within $N_\epsilon(J)$, and one of its d preimages will join the point z to J and will have length at most $\mathrm{dist}(f(z), J)/k$.

It follows that an arbitrary orbit $z_0 \mapsto z_1 \mapsto \cdots$ in the Fatou set can contain at most a finite number of points in $N_\epsilon(J)$. In fact, no point outside of $N_\epsilon(J)$ can map into $N_\epsilon(J)$, while if an orbit starts in $N_\epsilon(J) \smallsetminus J$, then the distance between z_i and J must increase by a factor of k or more with each iteration until the orbit leaves $N_\epsilon(J)$, never to return. Therefore any accumulation point $\hat{z}$ for this orbit lies in the Fatou set. If U is the Fatou component containing $\hat{z}$, then evidently some iterate $f^{\circ p}$ must map U into itself. According to the classification of periodic Fatou components in Theorem 16.1, U must be either an attracting basin, a parabolic basin, or a rotation domain. Since U clearly cannot be a parabolic basin, and by Theorem 11.17 and Lemma 15.7 cannot be a rotation domain, it must be an attracting basin. Therefore the orbit $z_0 \mapsto z_1 \mapsto \cdots$ must converge to the associated attracting periodic orbit. In particular, the orbit of any critical point must converge to an attracting periodic orbit. This clearly implies that $\overline{P} \cap J = \emptyset$ and completes the proof of Theorem 19.1. □

Theorem 19.2 (Local Connectivity). *If the Julia set of a hyperbolic map is connected, then it is locally connected.*

(Note that a Julia set which is not connected can never be locally connected, since it must have infinitely many connected components by Corollary 4.15.)

The proof of Theorem 19.2 will be based on three lemmas.

Lemma 19.3 (Fatou Component Boundaries). *If U is a simply connected Fatou component for a hyperbolic map, then the boundary ∂U is locally connected.*

Proof. First consider the case of an invariant component $U = f(U)$. Choose a conformal isomorphism $\phi : \mathbb{D} \xrightarrow{\cong} U$ so that $\phi(0)$ is the attracting fixed point in U. Then $F = \phi^{-1} \circ f \circ \phi$ is a proper holomorphic map from $\mathbb{D}$ to itself, with $F(0) = 0$, and with at least one critical point by Theorem 8.6. For any $0 < r < 1$, it follows from the Schwarz Lemma that F maps the disk $\mathbb{D}_r$ of radius r into some disk of strictly smaller radius.

If r is sufficiently close to 1, then it is not hard to see that the closure of the annulus $A_0 = F^{-1}(\mathbb{D}_r) \smallsetminus \overline{\mathbb{D}}_r$ is fibered by radial line segments (Figure 43). That is, each radial line $\{te^{i\theta}\ ;\ 0 < t < 1\}$ intersects $\overline{A}_0$ in a closed interval I_θ which varies smoothly with θ. To see this, express F as a Blaschke product

$$F(w) \;=\; e^{i\alpha} \prod (w - a_j)/(1 - \overline{a}_j w).$$

(See Problem 15-c.) If $w\overline{w} = 1$, then a brief computation shows that the

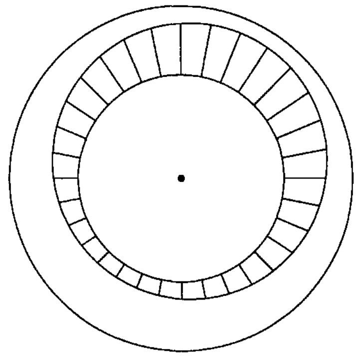


Figure 43. The annulus $A_0 = F^{-1}(\mathbb{D}_r) \setminus \overline{\mathbb{D}}_r$.

radial derivative

$$\frac{d \log F(w)}{d \log(w)} = w \frac{F'(w)}{F(w)}$$

takes the form $\sum (1 - a_j \bar{a}_j)/|w - a_j|^2 > 0$. Hence for $|w|$ close to $+1$ the real part of this radial derivative is still positive, which implies transversality of the required intersection.

Since f is hyperbolic, we can choose a conformal metric in some neighborhood of $J \supset \partial U$ so that f is expanding near J with expansion constant $k > 1$. Choose $r < 1$ large enough so that $\phi(\mathbb{D} \setminus \mathbb{D}_r)$ is contained in this neighborhood. Using the induced metric on $\mathbb{D} \setminus \mathbb{D}_r$, let M be the maximum of the lengths of the radial intervals I_θ. Now consider the sequence of annuli A_0, A_1, A_2, $\ldots$, converging to the boundary of $\mathbb{D}$, where $A_m = F^{-m}(A_0)$. Note that the closure of each A_m is fibered by the connected components of the preimages $F^{-m}(I_\theta)$ and that each such component curve segment has length at most M/k^m. Hence we can inductively construct a sequence of homeomorphisms

$$g_m : \partial \mathbb{D}_r \mapsto F^{-m} \partial \mathbb{D}_r$$

so that g_0 is the identity map and so that $g_m(re^{i\theta})$ and $g_{m+1}(re^{i\theta})$ are the two endpoints of the same fiber in the annulus A_m. Then

$$\operatorname{dist}(g_m(re^{i\theta}),\, g_{m+1}(re^{i\theta})) \le M/k^m,$$

hence the maps

$$\phi \circ g_m : \partial \mathbb{D}_r \to U \subset \widehat{\mathbb{C}}$$

form a Cauchy sequence. Therefore they converge uniformly to a continuous limit map from $\partial \mathbb{D}_r$ onto ∂U. By Theorem 17.14, this proves that ∂U is locally connected.

The case of a periodic Fatou component $U = f^{\circ p}(U)$ now follows by applying this argument to the iterate $f^{\circ p}$. Since any Fatou component U is eventually periodic by Theorem 19.1 (or by Sullivan's Nonwandering Theorem 16.4), and since ∂U is locally homeomorphic to $\partial f^{\circ q}(U)$, the conclusion follows. □

Remark. In the special case of a polynomial map, note that Lemma 19.3 by itself implies Theorem 19.2, since the basin of infinity for a polynomial map is an invariant Fatou component, with boundary equal to the entire Julia set.

For the next lemma, we use the spherical metric on $\widehat{\mathbb{C}}$.

> **Lemma 19.4 (Most Fatou Components Are Small).** *If f is hyperbolic with connected Julia set, then for every $\epsilon > 0$ there are only finitely many Fatou components with diameter greater than ϵ.*

In other words, if there are infinitely many Fatou components, numbered in any order as $U_1, U_2, \ldots$, then the diameter of U_j in the spherical metric must tend to zero as $j \to \infty$. (However, a hyperbolic map with disconnected Julia set may well have infinitely many Fatou components with diameter bounded away from zero. See Figure 6d, p. 43, for McMullen's example. I don't know any such example with connected Julia set, even in the nonhyperbolic case.)

Proof of Lemma 19.4. As in the proof of Theorem 19.1, we can choose a conformal metric μ with expansion constant $k > 1$ in a neighborhood V' of J, and choose $\epsilon > 0$ so that the closure $\overline{N}_\epsilon = \overline{N}_\epsilon(J)$ of the ϵ-neighborhood of J is contained in V', with $f^{-1}(N_\epsilon) \subset N_\epsilon$, Thus, if a Fatou component U is contained in N_ϵ, then every component U' of $f^{-1}(U)$ is contained in N_ϵ, and satisfies

$$\operatorname{diam}(U') \leq \operatorname{diam}(U)/k,$$

using the infimum of μ-lengths of paths within $\overline{N}_\epsilon$ as distance function.

Note that the various Fatou components U_j, together with N_ϵ, form an open cover of the compact space $\widehat{\mathbb{C}}$. Choosing a finite subcover, we see that all but finitely many of the U_j are contained in N_ϵ. For each U_j, we know from Theorem 19.1 that some forward image $f^{\circ \ell}(U_j)$ contains an attracting periodic point, and hence is not contained in N_ϵ. Define the *level* of U_j to be the smallest $\ell \geq 0$ such that $f^{\circ \ell}(U_j) \not\subset N_\epsilon$. Since there are only finitely many U_j of level zero, it follows that there are only finitely many of each fixed level. If M is the maximum $\operatorname{diam}(U_j)$ among the U_j

of level 1, then it follows that

$$\mathrm{diam}(U_j) \le M/k^{\ell-1}$$

for each U_j of level $\ell \ge 1$. Evidently this tends to zero as $\ell \to \infty$. Since all but finitely many of the U_j lie within the compact set $\overline{N}_\epsilon$, it follows easily that the diameters in the spherical metric also tend to zero, as required. □

Lemma 19.5 (Locally Connected Sets in S^2). *If X is a compact subset of the sphere S^2 such that every component of $S^2 \smallsetminus X$ has locally connected boundary and such that there are at most finitely many components with diameter $> \epsilon$ for any $\epsilon > 0$, then X is locally connected.*

(For the converse statement, see Problem 19-f.)

Proof. Given a point $x \in X$ and a neighborhood $N_\epsilon(x)$ of radius ϵ, using the spherical metric, we will find a smaller neighborhood $N_\delta(x)$ so that any point in $X \cap N_\delta(x)$ is joined to x by a connected subset of $X \cap N_\epsilon(x)$. In fact, choose $\delta < \epsilon/2$ so that, for any component U_j of $S^2 \smallsetminus X$, any two points of ∂U_j with distance less than δ are joined by a connected subset of ∂U_j of diameter less than $\epsilon/2$. (See Lemma 17.13. Evidently we need only consider the finitely many U_j which have diameter $\ge \epsilon/2$.) Now for any $y \in X \cap N_\delta(x)$, take the spherical geodesic I from x to y and replace each connected component of $I \smallsetminus X$ by a connected subset of the boundary of the corresponding U_j with diameter less than $\epsilon/2$. The result will be a connected subset of $X \cap N_\epsilon$ containing both x and y. This proves that X is locally connected at the arbitrary point x. □

Proof of Theorem 19.2. This follows immediately from Lemmas 19.3, 19.4, and 19.5. □

Douady and Hubbard, using ideas of Thurston, also consider a wider class of mappings which they call *subhyperbolic*. These may have critical points in the Julia set, but only if their orbits are eventually periodic. The only change in the definition is that the conformal metric in a neighborhood of J is now allowed to have a finite number of relatively mild singularities in the postcritical set. To understand the allowed singularities, consider a smooth conformal metric $\rho(w)|dw|$ which is invariant under rotation of the w-plane through an angle of $2\pi/m$ radians, and consider the identification space in which w is identified with $e^{2\pi i/m}w$. The resulting object is a smooth Riemannian manifold except at the origin, where it has a *cone point*. If we set $z = w^m$, then the induced metric in the z-plane has the form $\gamma(z)|dz|$, where

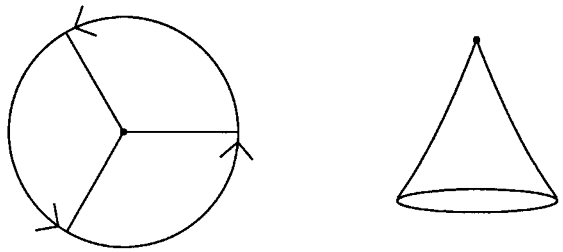


Figure 44. On the left: disk in the w*-plane. Three fundamental domains under* $120°$ *rotation are indicated. On the right: quotient space in the pushed forward metric.*

$$\gamma(z) \;=\; \rho(w)\left|\frac{dz}{dw}\right|^{-1} \;=\; \frac{\rho(\sqrt[m]{z})}{m\,|z|^{1-1/m}}.$$

Thus $\gamma(z) \to \infty$ as $z \to 0$, but the singularity is relatively innocuous since any reasonable path $t \mapsto z(t)$ still has finite length $\int \gamma(z(t))|dz(t)|$.

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 nonzero except at a locally finite collection of "ramified points" $a_1, a_2, \ldots$ where it blows up in the following special way. There should be integers $\nu_j \ge 2$ called the *ramification indices* at the points a_j such that, if we take a local branched covering by setting $z(w) = a_j + w^{\nu_j}$, then the induced metric $\gamma(z(w))|dz/dw| \cdot |dw|$ on the w-plane is smooth and nonsingular throughout some neighborhood of the origin. We will say that f is *expanding* with respect to such a metric if its derivative satisfies $\|Df_p\| \ge k > 1$ whenever p and $f(p)$ are not ramified points. (Note that we cannot expect the sharper condition (19 : 1) near a critical point.)

Definition. The rational map f is *subhyperbolic* if it is expanding with respect to some orbifold metric on a neighborhood of its Julia set.

Following Douady and Hubbard [1984–85], we have the following two results. (Compare Corollary 14.5.)

Theorem 19.6. *A rational map is subhyperbolic if and only if every critical orbit is either finite or converges to an attracting periodic orbit.*

Theorem 19.7. *If* f *is subhyperbolic with* $J(f)$ *connected, then* $J(f)$ *is locally connected.*

The proof of Theorem 19.7 is essentially identical to the proof of Theorem 19.2. In particular, the proofs of Lemmas 19.3 and 19.4 work equally well in the subhyperbolic case. Details will be left to the reader. □

Proof of Theorem 19.6. The argument is just an elaboration of the proof of Theorem 19.1. In one direction, if f is expanding with respect to some orbifold metric defined near the Julia set, then just as in Theorem 19.1 there is a neighborhood $N_\epsilon(J)$ so that every orbit in the Fatou set eventually leaves this neighborhood, never to return. Hence it can only converge to an attractive periodic orbit. On the other hand, if c is a critical point in the Julia set, then every forward image $f^{\circ i}(c)$, $i > 0$, must be one of the ramification points a_j for our orbifold metric, since the map $f^{\circ i}$ has derivative zero at the critical point c and yet must satisfy $\|Df_z^{\circ i}\| \geq k^i$ at points arbitrarily close to c. The collection of ramified points in J is required to be locally finite, so it follows that the orbit of c must be eventually periodic.

Orbifolds. For the proof in the other direction, we must introduce more ideas from Thurston's theory of orbifolds. (See Appendix E for a brief introduction to this theory.)

Definition. 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 (to be called *ramified points*), each of which is assigned a *ramification index* $\nu_j = \nu(a_j) \geq 2$ as above. For any point z which is not one of the a_j we set $\nu(z) = 1$.

Now let f be a rational map such that every critical orbit is either finite or converges to an attracting cycle. We assign to f the associated *canonical orbifold* (S, ν) as follows. As underlying Riemann surface S we take the Riemann sphere $\widehat{\mathbb{C}}$ with all attracting periodic orbits removed. As ramified points a_j we take all *postcritical points*, that is, all points which have the form $a_j = f^{\circ k}(c)$ for some $k > 0$ and for some critical point c of f. Since every critical orbit is either finite or converges to a periodic attractor, we see easily that this collection of points a_j is locally finite in S (although perhaps not in $\widehat{\mathbb{C}}$). 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,

$$f(z) = z_2 + b(z - z_1)^n + (\text{higher terms}),$$

where $b \neq 0$ and $n \geq 1$. 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 $\nu(a_j) \geq 2$ to be the smallest integers which satisfy the following condition.*

> **Condition ($\star$).** For any $z \in S$, the ramification index $\nu(f(z))$ at the image point must be some multiple of the product $n(f, z)\,\nu(z)$.

To construct these integers $\nu(a_j)$, where a_j ranges over all postcritical points in S, we consider all pairs (c, m) where c is a critical point with $f^{\circ m}(c) = a_j$, and choose $\nu(a_j)$ to be the least common multiple of the corresponding branch indices $n(f^{\circ m}, c)$. (Note that a_j itself may be a critical point, since one critical point may eventually map to another.) There are only finitely many such pairs (c, m) since we have removed all superattracting periodic orbits, so this least common multiple is well defined and finite. It is not hard to check that it provides a minimal solution to the required Condition $(\star)$.

(**Remark.** For some purposes it is convenient to extend this definition to all points of $\widehat{\mathbb{C}}$ by setting $\nu(f(z)) = \infty$ whenever z is an attracting periodic point.)

As in Appendix E, we consider the *universal covering surface*

$$\widetilde{S}_\nu \to (S, \nu)$$

for this canonical orbifold, that is, the unique regular branched covering of S which is simply connected and has the given $\nu : S \to \mathbb{Z}$ as ramification function. Such a universal covering could fail to exist only if S were the entire Riemann sphere with at most two ramified points (see Lemma E.1), and it is straightforward to show that this case can never occur for our canonical orbifold.

Since $f^{-1}(S) \subset S$, it is not difficult to see that f^{-1} lifts to a single-valued holomorphic map $F : \widetilde{S}_\nu \to \widetilde{S}_\nu$. In fact Condition $(\star)$ is exactly what is needed in order to guarantee that such a lifting of f^{-1} exists locally, and since $\widetilde{S}_\nu$ is simply connected there is no obstruction to extending to a global lifting. There are now three possible cases.

Conformally Hyperbolic Case. If $\widetilde{S}_\nu$ is hyperbolic, then F must either preserve or decrease the Poincaré metric for this universal covering surface. If F were metric preserving, then f would preserve the orbifold metric for (S, ν), and it would follow that every periodic point for f in S must be indifferent, which is impossible. Hence F must be metric decreasing, and f must be metric increasing. Since J is compact, it follows that $\|DF_w\| \le 1/k < 1$ whenever $w \in \widetilde{S}_\nu$ projects into a suitably chosen neighborhood W of J. Hence $\|Df_z\| \ge k > 1$ for every $z \in W$ such that z and $f(z)$ are not ramified points, which completes the proof in this case.

Conformally Euclidean Case. The easiest way to proceed when $\widetilde{S}_\nu$ is a Euclidean surface is simply to change the ramification function ν. For example, if we choose some periodic orbit in S and replace ν by a ramification function ν' which is equal to 2ν on this orbit and equal to ν

elsewhere, then Condition $(\star)$ will be preserved while $\widetilde{S}_{\nu'}$ as a nontrivial branched covering of $\widetilde{S}_\nu \cong \mathbb{C}$, will certainly be hyperbolic. The proof then goes through as above. (For a more informative argument in the Euclidean case, see Theorem 19.9.)

Spherical Case. If $\widetilde{S}_\nu$ were conformally equivalent to $\widehat{\mathbb{C}}$, then S would have to be the whole Riemann sphere. Furthermore, since F is a lift of f^{-1}, the composition

$$\widetilde{S}_\nu \xrightarrow{F} \widetilde{S}_\nu \xrightarrow{\text{projection}} S \xrightarrow{f} S$$

would have to coincide with the projection map, and yet its degree would have to be strictly larger than the degree of the projection map. Thus this case can never occur. This completes the proof of Theorem 19.6. □

As a corollary, we obtain a sharper version of Corollary 16.5.

Corollary 19.8. *If f is subhyperbolic with no attracting periodic orbits, so that S is the entire Riemann sphere, then f is expanding with respect to its orbifold metric on the entire sphere. Hence the Fatou set is vacuous, and $J(f)$ is the entire sphere.*

In the Euclidean case, a more careful argument yields a much more precise description of the subhyperbolic map f. In order to determine the geometry of $\widetilde{S}_\nu$, we introduce the *orbifold Euler characteristic*

$$\chi(S, \nu) = \chi(S) + \sum \left(\frac{1}{\nu(a_j)} - 1\right),$$

to be summed over all points a_j with $\nu(a_j) \neq 1$. (Here $\chi(S)$ is the ordinary Euler characteristic, equal to $2 - m$, where m is the number of points in the complement $\widehat{\mathbb{C}} \smallsetminus S$.) It follows easily from Lemma E.4 (Appendix E) that the universal covering surface $\widetilde{S}_\nu$ is either spherical, hyperbolic, or Euclidean according to whether $\chi(S, \nu)$ is positive, negative, or zero.

Theorem 19.9. *If $\chi(S, \nu) = 0$, then f induces a linear isomorphism $\tilde{f}(w) = \alpha w + \beta$ from the Euclidean covering space $\widetilde{S}_\nu \cong \mathbb{C}$ onto itself. In this case, the Julia set is either a circle or a line segment, or the entire Riemann sphere. Here the expansion constant $|\alpha|$ is equal to the degree d when J is 1-dimensional and is equal to $\sqrt{d}$ when J is the entire sphere.*

Compare §7. (Caution: The coefficient α itself is not uniquely determined, since 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 $w \mapsto \alpha' w + \beta'$ where α' is some root of unity.)

Proof of Theorem 19.9. Starting with the orbifold (S, ν) associated with a subhyperbolic map f, construct a new orbifold $(S',\, \mu)$ as follows. Let $S' = f^{-1}(S)$ be obtained from the open set S by removing all immediate preimages of attracting periodic points, and define $\mu = f^*(\nu)$ by the formula

$$\mu(z) \;=\; \nu(f(z))/n(f,\, z),$$

where $n(f,\, z)$ is the branch index. It follows from Condition $(\star)$ that $\mu(z)$ is an integer with $\mu(z) \geq \nu(z)$ 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 in Appendix E, since the map $f : (S',\, \mu) \to (S,\, \nu)$ is a "d-fold covering of orbifolds," we conclude that f induces an isomorphism $\widetilde{S}'_\mu \to \widetilde{S}_\nu$ of universal covering surfaces and also that the Riemann-Hurwitz formula takes the form

$$\chi(S',\, \mu) \;=\; \chi(S,\, \nu)\, d.$$

Combining these two statements, we see that

$$\chi(S,\, \nu)\, d \;\leq\; \chi(S,\, \nu),$$

with equality if and only if $S = f^{-1}(S)$ and $\nu = f^*\nu$. Since $d \geq 2$, this provides another proof that $\chi(S,\, \nu) \leq 0$. Furthermore, it shows that we are in the Euclidean case $\chi(S, \nu) = 0$ if and only if $S = f^{-1}(S)$ and $\nu = f^*\nu$, so that f maps (S, ν) to itself by a d-fold covering of orbifolds.

Thus, when $\widetilde{S}_\nu$ is conformally Euclidean, it follows that f lifts to a necessarily linear automorphism $\widetilde{f}(w) = \alpha w + \beta$ of the universal covering surface $\widetilde{S}_\nu \cong \mathbb{C}$. Furthermore, since S is fully invariant under f, it follows from Lemma 4.9 that the complement $\widehat{\mathbb{C}} \smallsetminus S$ has at most two points. We now consider the three possibilities.

Case 0. If $S = \widehat{\mathbb{C}}$, then we can compute the degree by integrating $\|Df_z\|$ over the sphere. In fact, using the (locally Euclidean) orbifold metric, note that f maps a generic small region of area A to a region of area $|\alpha|^2 A$. Integrating over S, we see that the degree d must be equal to $|\alpha|^2$. In fact, it can be shown that such maps f, with Euclidean orbifold and with the entire Riemann sphere as Julia set, are precisely the *Lattès maps*. (Compare Definition 7.4, as well as Milnor [2004b].)

Case 1. If $S \cong \mathbb{C}$, then solving the required equation

$$\chi(S, \nu) = 1 - \sum\Big(\frac{1}{\nu(z)} - 1\Big) = 0,$$

it is not difficult to check that there must be exactly two ramification points, both with index $\nu(z) = 2$. The corresponding universal covering space is isomorphic to $\mathbb{C}$, with the integers as branch points, and with all linear maps of the form $w \mapsto \pm w + m$ as deck transformations. Since $\tilde{f}$ must carry integers to integers, it follows easily that f is a Chebyshev map up to sign, with an interval as Julia set and with degree $|\alpha|$.

Case 2. If $S \cong \mathbb{C} \smallsetminus \{0\}$, then there can be no ramification points, and it follows that f is conjugate to $z \mapsto z^{\pm d}$. Thus the Julia set is a circle, and again $d = |\alpha|$. $\square$

Other Locally Connected Julia Sets. Many other Julia sets are known to be locally connected. For quadratic polynomials, the most important result is the Yoccoz proof that the Julia set is locally connected provided that it is connected, with no Cremer points or Siegel disks, and is not infinitely ***renormalizable***. (See Douady and Hubbard [1985], McMullen [1994a, 1994b, 1996], Milnor [1999, 2000a], and Lyubich [1999] for renormalization; and see Hubbard [1993], Yoccoz [1999/2003], Milnor [2000b], and Lyubich [1997/2000] for the Yoccoz theorem.) Another important example is given by the Petersen-Zakeri proof of local connectivity for a generic quadratic polynomial with Siegel disk. (See Petersen [1996, 1998], Yampolsky [1999], and Petersen and Zakeri [2004].) A rational map is called ***geometrically finite*** if the orbit of every critical point in its Julia set is eventually periodic. Recall from §16 that every critical orbit in the Fatou set must converge to an attracting cycle or to a parabolic cycle. Thus, in the geometrically finite case we have very strict control of all critical orbits. In particular, it follows from Corollary 14.4 and Lemma 15.7 that a geometrically finite map can have no Cremer points, Siegel disks, or Herman rings. Tan Lei and Yin [1996] and Pilgrim and Tan Lei [2000] have proved the following much sharper version of Theorem 19.7: *If f is geometrically finite, then every connected component of its Julia set is locally connected.*

Concluding Problems

Problem 19-a. The nonwandering set. By definition, the ***nonwandering set*** for a continuous map $f : X \to X$ is the closed subset $\Omega \subset X$ consisting of all $x \in X$ such that for every neighborhood U of x there exists an integer $k > 0$ such that $U \cap f^{\circ k}(U) \neq \emptyset$. Using the results of §§4 and 16, show that the nonwandering set for a rational f is the disjoint union of its Julia set, its rotation domains (if any), and its set of attracting periodic points.

Problem 19-b. Axiom A. In the literature on smooth dynamical systems a 1-dimensional map is said to satisfy Smale's Axiom A if and only if the following two conditions are satisfied:

(1) The nonwandering set Ω splits as the union of a closed subset Ω^+ on which f is infinitesimally expanding with respect to a suitable Riemannian metric, and a closed subset Ω^- on which f is contracting.*

(2) Periodic points are everywhere dense in Ω.

(See, for example, Smale [1967], Shub [1987].) Show that a rational map is hyperbolic if and only if it satisfies Axiom A.

Problem 19-c. An orbifold example. **(1)** Show that the Julia set for the rational map $z \mapsto (1-2/z)^{2n}$ is the entire Riemann sphere. **(2)** For $n > 1$, show that the orbifold metric for this example is hyperbolic. **(3)** For $n = 1$, show that it is Euclidean. (In fact f is a Lattès map in this case; compare Definition 7.4. The associated semiconjugacy $\Theta : \mathbb{T} \to \widehat{\mathbb{C}}$ has degree 4 (Milnor [2004b, §8.1]).

Problem 19-d. The Euclidean case. For any subhyperbolic map whose canonical orbifold metric is Euclidean, show that every periodic orbit outside of the finite postcritical set has multiplier λ satisfying $|\lambda| = n^{p/\delta}$ where n is the degree, p is the period, and δ is the dimension (1 or 2) of the Julia set.

Problem 19-e. Expansive maps. A map f from a metric space to itself is said to be *expansive* on a subset X if there exists $\epsilon > 0$ so that, for any two points $x \neq y$ whose orbits remain in X forever, there exists some $k \geq 0$ so that $f^{\circ k}(x)$ and $f^{\circ k}(y)$ have distance greater than ϵ. Using Sullivan's results from §16, show that a rational map is expansive on some neighborhood of its Julia set if and only if it is hyperbolic. (However, a map with a parabolic fixed point may be expansive on the Julia set itself.)

Problem 19-f. Locally connected sets in the 2-sphere. Give a complete characterization of compact locally connected subsets of the 2-sphere as follows. **(1)** Prove the following theorem of Torhorst:

If $X \subset S^2$ is compact and locally connected, then the boundary

*In higher dimensions, (1) is replaced by the assumption that the tangent vector bundle of the underlying manifold, restricted to Ω, splits as the direct sum of a bundle on which the derivative Df is expanding and a bundle on which it is contracting.

of every complementary component must be locally connected.

(See Whyburn [1964] and compare the proof of Theorem 17.14.) **(2)** Furthermore, prove that:

*If there are infinitely many complementary components, then their diameters tend to zero.**

(3) Now using Lemma 19.5, conclude that these two conditions are necessary and sufficient for local connectivity.

* Here is one possible proof outline: Let U be a component of $S^2 \smallsetminus X$ of diameter greater than 4ϵ, with ϵ small. After rotating the sphere, we may assume that U contains a pair of points at distance 2ϵ from the equator, one in the northern hemisphere and one in the southern. Then each parallel at distance less than ϵ from the equator must intersect U in at least one interval whose endpoints cannot be joined within X by any connected set of diameter less than ϵ. Let δ_U be the infimum of the lengths of such intervals. Then U has area $A(U) \geq 2\epsilon\,\delta_U$. Now if there were infinitely many such components U, then their areas would tend to zero, hence the numbers δ_U would tend to zero, and X would not be locally connected.

Appendix A. Theorems from Classical Analysis

This appendix will describe some miscellaneous theorems from classical complex variable theory. We first complete the proofs of Theorem 11.14, Theorem 17.4, and Theorem 18.2 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.

We begin with a discussion of Jensen's† inequality. Let $f : \mathbb{D} \to \mathbb{C}$ be a holomorphic function on the open disk which is not identically zero. Given any radius $0 < r < 1$, we can form the average

$$A(f,r) \;=\; \frac{1}{2\pi}\int_0^{2\pi} \log|f(re^{i\theta})|d\theta \qquad (\mathrm{A}:1)$$

of the quantity $\log|f(z)|$ over the circle $|z| = r$.

Theorem A.1 (Jensen [1899]). *This average $A(f,r)$ is monotone increasing as a function of r. Hence $A(f,r)$ either converges to a finite limit or diverges to $+\infty$ as $r \nearrow 1$.*

In fact the proof will show something much more precise.

Lemma A.2. *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 $\mathbb{D}_r$ 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

$$n \;=\; \frac{1}{2\pi i}\oint_{|z|=r} d\log f(z) \;=\; \frac{1}{2\pi i}\oint_{|z|=r} \frac{f'(z)\,dz}{f(z)}$$

*A function of one complex variable is called *univalent* if it is holomorphic and injective.

†Johan Ludwig Jensen (1859–1925) was president of the Danish telephone company. He was very active in mathematics although he never held an academic position. (His career is reminiscent of that of Jorgen Pedersen Gram (1850–1916), another famous Danish mathematician, who was an insurance executive.)

measures the number of zeros of f inside the disk $\mathbb{D}_r$. In particular, choosing $r_0 < r < r_1$, we see 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 $\Big(\log f(z) - \log z^n\Big)dz/iz$ 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. (Compare the indefinite integral $\int \log|x|dx = x \log|x| - x$, which is continuous as a function of x.) Continuity of $A(f, r)$ as r varies through such a singularity is not difficult and will be left to the reader. □

Theorem A.3 (F. and M. Riesz [1916]). *Suppose that $f : \mathbb{D} \to \mathbb{C}$ is bounded and holomorphic on the open unit disk. If the radial limit*

$$\lim_{r \nearrow 1} f(re^{i\theta})$$

exists and takes some constant value c_0 for θ belonging to a set $E \subset [0, 2\pi]$ of positive Lebesgue measure, then f must be identically equal to c_0.

(Compare Theorem 17.4, which combines this statement with a theorem of Fatou.)

Proof of Theorem A.3. Without loss of generality, we may assume that $c_0 = 0$ and that $f(\mathbb{D}) \subset \mathbb{D}$. 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 sets $E(\epsilon, \delta)$ form a nested family with union equal to E. Therefore the Lebesgue measure $\ell(E(\epsilon, \delta))$ must tend to the limit $\ell(E)$ as $\delta \searrow 0$. 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 (A : 1) for some fixed $r > 1 - \delta$. Since $|f(z)| < 1$ for all $z \in \mathbb{D}$, the expression $\log|f(re^{i\theta})|$ is less than zero everywhere, and less than $\log \epsilon$ throughout a set $E(\epsilon, \delta)$ 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 $\lim_{r \nearrow 1} A(f, r) \;=\; -\infty$, which contradicts Theorem A.1 unless

f is identically zero. □

Now consider the following situation. Let K be a compact connected subset of $\mathbb{C}$ and suppose that the complement $\mathbb{C} \setminus K$ is conformally diffeomorphic to the complement $\mathbb{C} \setminus \overline{\mathbb{D}}$.

Lemma A.4 (Area Formula, Gronwall [1914–15]). *Let*

$$\psi : \mathbb{C} \setminus \overline{\mathbb{D}} \to \mathbb{C} \setminus K$$

be a conformal isomorphism, with Laurent series expansion

$$\psi(w) \;=\; b_1 w + b_0 + b_{-1}/w + b_{-2}/w^2 + \cdots .$$

Then the 2-dimensional Lebesgue measure of K *is given by*

$$\text{area}(K) \;=\; \pi \sum_{n\le 1} n|b_n|^2 \;=\; \pi\Big(|b_1|^2 - |b_{-1}|^2 - 2|b_{-2}|^2 - \cdots\Big).$$

Proof. For any $r > 1$ consider the image under ψ of the circle $|w| = r$. This will be some embedded circle in $\mathbb{C}$ which encloses a region of area say $A(r)$. We can compute this area by Green's Theorem, as follows. Let $\psi(re^{i\theta}) = z = x + iy$. Then

$$A(r) \;=\; \oint x\,dy \;=\; -\oint y\,dx \;=\; \tfrac{1}{2i}\oint \overline{z}\,dz,$$

to be integrated around the image of $|w| = r$. Substituting the Laurent series $z = \sum_{n\le 1} b_n w^n$, with $w = r^n e^{ni\theta}$, this yields

$$A(r) \;=\; \tfrac{1}{2} \sum_{m,n\le 1} n\overline{b}_m b_n r^{m+n} \oint e^{(n-m)i\theta} d\theta.$$

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 $r \searrow 1$, we obtain the required formula. □

Remark. Unfortunately, it is difficult to make any estimate on the rate of convergence of this series. If the set K has a complicated shape, then it seems likely that very high order terms will play an important role.

One trivial consequence of Lemma A.4 is the inequality

$$|b_1|^2 \;\ge\; \sum_1^\infty m|b_{-m}|^2.$$

In particular we have the following.

Corollary A.5 (Gronwall Inequality). *With $K \subset \mathbb{C}$ and $\psi(w) = \sum_{n\le 1} b_n w^n$ as above, we have $|b_1| \ge |b_{-1}|$, with equality if and only if K is a straight line segment.*

Proof. Since $\text{area}(K) \ge 0$, we have $|b_1| \ge |b_{-1}|$. 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 $z = \psi(w)$ coordinate, the Laurent series will reduce to the simple formula $z = w + w^{-1}$. As noted in §7, this transformation carries $\mathbb{C} \smallsetminus \overline{\mathbb{D}}$ diffeomorphically onto the complement of the interval $[-2, 2]$. (Compare Lemma 7.1 as well as Problems 7-c, 7-d.) □

Now consider an open set $U \subset \mathbb{C}$ which contains the origin and is conformally isomorphic to the open disk.

Lemma A.6 (Bieberbach Inequality). *If $\psi : \mathbb{D} \to U$ 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 $\mathbb{C} \smallsetminus U$ is a closed half-line pointing towards the origin.*

Remark. The Bieberbach Conjecture (proposed by Bieberbach [1916] and proved by de Branges [1985]) asserts that $|a_n| \le n|a_1|$ for all n. Again, equality holds if $\mathbb{C} \smallsetminus U$ is a closed half-line pointing towards the origin, for example for $\psi(\eta) = \eta + 2\eta^2 + 3\eta^3 + \cdots = \eta/(1-\eta)^2$.

Proof of Lemma A.6. After a scale change, multiplying $\psi(\eta)$ by a constant, we may assume that $a_1 = 1$. Let us set $\eta = 1/w^2$, so that each point $\eta \ne 0$ in $\mathbb{D}$ corresponds to two points $\pm w \in \mathbb{C} \smallsetminus \overline{\mathbb{D}}$. Similarly, set $\psi(\eta) = \zeta = 1/z^2$, so that each $\zeta \ne 0$ in U corresponds to two points $\pm z$ in some neighborhood $N = -N$ of infinity. A brief computation shows that ψ corresponds to a Laurent series of the form

$$w \mapsto z = 1/\sqrt{\psi(1/w^2)} \;=\; w - \tfrac{1}{2}a_2/w + (\text{terms in } 1/w^3,\ 1/w^5,\ \ldots)$$

which maps $\mathbb{C} \smallsetminus \overline{\mathbb{D}}$ diffeomorphically onto N. Thus $|a_2| \le 2 = 2a_1$ 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. □

Theorem 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 $\mathbb{D}$ diffeomorphically onto an open set

$U \subset \mathbb{C}$. Then the distance r between the origin and the boundary of U satisfies

$$\tfrac{1}{4}|a_1| \;\le\; r \;\le\; |a_1|. \qquad (A:2)$$

Here the first equality holds if and only if $\mathbb{C} \setminus U$ 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 $U = \psi(\mathbb{D})$ necessarily contains the disk $\mathbb{D}_{1/4}$ of radius $1/4$ centered at the origin, so that $\psi^{-1} : \mathbb{D}_{1/4} \to \mathbb{D}$ is well defined and univalent. The left-hand inequality in (A : 2) 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.

Proof of Theorem A.7. Without loss of generality, we may assume that $a_1 = 1$. If $z_0 \in \partial U$ is 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 is also univalent on $\mathbb{D}$, and has the form

$$\eta \;\mapsto\; \psi(\eta)/(1-\psi(\eta)/z_0) \;=\; \eta + (a_2 + 1/z_0)\eta^2 + \cdots.$$

By Bieberbach's Inequality (Lemma A.6), 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 $1/z_0 = -2a_2$. The exact description of U then follows easily.

On the other hand, suppose that $r \ge 1$. Then the inverse mapping ψ^{-1} is defined and holomorphic throughout the unit disk $\mathbb{D}$ and takes values in $\mathbb{D}$. Since its derivative at zero is 1, it follows from the Schwarz Lemma 1.2 that ψ is the identity map, with $r = 1$. $\square$

Here is an interesting restatement of the Quarter Theorem. Let $ds = \rho(z)|dz|$ be the Poincaré metric on the open set U, and let $r = r(z)$ be the distance from z to the boundary of U.

Corollary A.8. *If $U \subset \mathbb{C}$ is simply connected, then the Poincaré metric $ds = \rho(z)|dz|$ on U agrees with the metric $|dz|/r(z)$ up to a factor of 2 in either direction. That is,*

$$\frac{1}{2r(z)} \;\le\; \rho(z) \;\le\; \frac{2}{r(z)}$$

for all $z \in U$. Again, the left equality holds if and only if $\mathbb{C} \setminus U$ 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.

This follows, since we can choose $\psi : \mathbb{D} \to U$ mapping the origin to any given point of U, and since the Poincaré metric at the center of $\mathbb{D}$ is $2|d\eta|$. $\square$

As an example, if U is a half-plane, then the Poincaré metric precisely agrees with the $(1/r)$-metric $|dz|/r$.

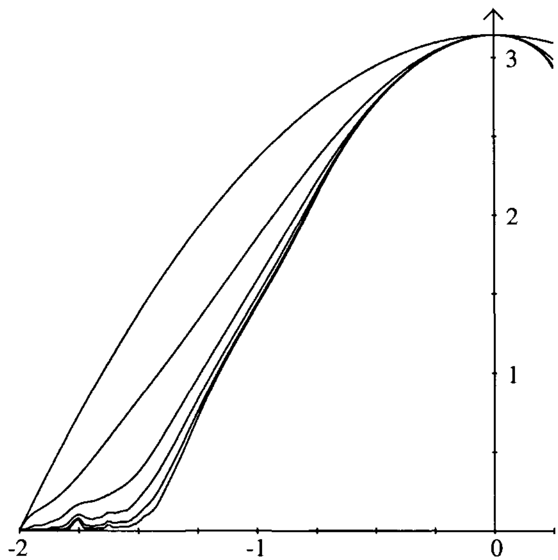


Figure 45. Upper bounds for the area of the filled Julia set for $f_c(z) = z^2 + c$ in the range $-2 \le c \le .25$.

Concluding Problem

Problem A-1. Area of the filled Julia set. Consider the polynomial map $f_c(z) = z^2 + c$. Let $w = \hat{\phi}(z)$ be the associated Böttcher map near infinity, and let $z = \psi(w)$ be the inverse map. **(1)** In analogy with equation (9 : 5), show that ψ satisfies the identity

$$\psi(w^2) \;=\; \psi(w)^2 + c,$$

and conclude that ψ has Laurent series of the form

$$\psi(w) \;=\; w\Big(1 + p_1(c)/w^2 + p_2(c)/w^4 + p_3(c)/w^6 + \cdots\Big),$$

where each $p_k(c)$ is a polynomial of degree k with rational coefficients. **(2)** Let K_c be the filled Julia set for f_c. Show that the area of K_c is upper semicontinuous* as a function of c. **(3)** If K_c is connected, or in other words if c belongs to the Mandelbrot* set, show by Lemma A.4 that

* Compare Lemma 11.15 for semicontinuity and Appendix G for the Mandelbrot set.

its area is given by the formula

$$\text{area}(K_c) = \pi\Big(1 - |p_1(c)|^2 - 3|p_2(c)|^2 - 5|p_3(c)|^2 - \cdots\Big). \qquad (A:3)$$

(Compare Figure 45, which graphs the upper bounds obtained by summing either 1, 10, 100, 1000, 10000, or 100000 terms of this series for real values of c. For example, the topmost graph represents the estimate

$$\text{area}(K_c) \le \pi(1 - |p_1(c)|^2) = \pi(1 - |c|^2/4).$$

Evidently $\text{area}(K_c)$ attains its maximum of $\pi = \text{area}(\mathbb{D})$ for $c = 0$. Unfortunately, the series (A : 3) converges *very* slowly, and I don't know of any useful lower bound for the area.) **(4)** On the other hand, show that equation (A : 3) breaks down when K_c is not connected. In fact the left side is zero but the right side is $-\infty$. Show that the sum

$$|p_1(c)|^2 + 3|p_2(c)|^2 + 5|p_3(c)|^2 + \cdots$$

is infinite. In fact, when K_c is not connected, show that ψ cannot be extended as a holomorphic function over all of $\mathbb{C} \smallsetminus K_c$, and conclude that the sequence of coefficients $p_1(c)$, $p_2(c)$, $\cdots$ must be unbounded. The area is zero in this case, since K_c coincides with the Julia set and since f_c is hyperbolic.

Appendix B. Length-Area-Modulus Inequalities

This appendix will first study the conformal geometry of a rectangle. (Compare Lemma 17.1.) Let $\overline{R} = [0, \Delta x] \times [0, \Delta y]$ be a closed rectangle in the plane of complex numbers $z = x + iy$, and let $R = (0, \Delta x) \times (0, \Delta y)$ be its interior. By a *conformal metric* on R we mean a metric of the form

$$ds = \rho(z)|dz|$$

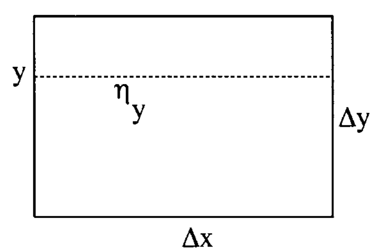
where $z \mapsto \rho(z) > 0$ is any strictly positive real-valued function which is defined and continuous throughout the open rectangle. In terms of such a metric, the *length* of a smooth curve $\gamma : (a, b) \to R$ is defined to be the integral

$$\mathbf{L}_\rho(\gamma) = \int_a^b \rho(\gamma(t))|d\gamma(t)|,$$

and the *area* of a region $U \subset R$ is defined to be

$$\mathrm{area}_\rho(U) = \iint_U \rho(x + iy)^2 dx\, dy.$$

In the special case of the Euclidean metric $ds = |dz|$, with $\rho(z)$ identically equal to 1, the subscript ρ will be omitted.



Lemma B.1 (Main Lemma). *If* $\mathrm{area}_\rho(R)$ *is finite, then for Lebesgue almost every* $y \in (0, \Delta y)$ *the length* $\mathbf{L}_\rho(\eta_y)$ *of the horizontal line* $\eta_y : t \mapsto (t, y)$ *at height* y *is finite. Furthermore, there exists* y *so that*

$$\frac{\mathbf{L}_\rho(\eta_y)^2}{(\Delta x)^2} \le \frac{\mathrm{area}_\rho(R)}{\Delta x \Delta y}. \qquad (\mathrm{B}:1)$$

In fact, the set consisting of all $y \in (0, \Delta y)$ *for which this inequality is satisfied has positive Lebesgue measure.*

Remark. Note that Δx is equal to $\mathbf{L}(\eta_y)$, the Euclidean length, and that the product $\Delta x \Delta y$ is equal to $\mathrm{area}(R)$, the Euclidean area. Evidently the inequality (B : 1) is best possible, for in the special case of the Euclidean metric with $\rho \equiv 1$, both sides of (B : 1) are equal to $+1$.

Proof of Lemma 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. Taking $f(x) \equiv 1$ and $g(x) = \rho(x, y)$ for some fixed y, we obtain

$$\left(\int_0^{\Delta x} \rho(x,y)\,dx\right)^2 \;\le\; \Delta x \int_0^{\Delta x} \rho(x,y)^2\,dx,$$

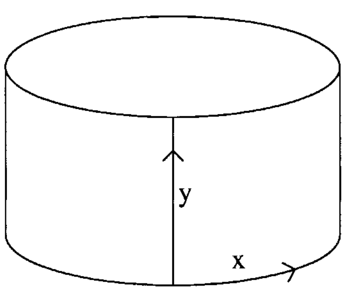
or in other words

$$\mathrm{L}_\rho(\eta_y)^2 \;\le\; \Delta x \int_0^{\Delta x} \rho(x,y)^2 dx,$$

for each constant height y. Integrating this inequality over the interval $0 < y < \Delta y$ and then dividing by Δy, we get

$$\frac{1}{\Delta y}\int_0^{\Delta y} \mathrm{L}_\rho(\eta_y)^2 dy \;\le\; \frac{\Delta x}{\Delta y}\,\mathrm{area}_\rho(A). \tag{B : 2}$$

In other words, the *average* over all y in the interval $(0, \Delta y)$ of $\mathrm{L}_\rho(\eta_y)^2$ is less than or equal to $\frac{\Delta x}{\Delta y}$ $\mathrm{area}_\rho(A)$. Further details of the proof are straightforward. □



Now let us form a cylinder C of circumference Δx and height Δy by gluing the left and right edges of our rectangle together. (Alternatively, C can be described as the Riemann surface 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 $x + \Delta x + iy$. Compare Problem 2-f.)

Definitions. The *modulus* of such a cylinder $C = (\mathbb{R}/(\mathbb{Z}\Delta x)) \times (0, \Delta y)$

is defined to be the ratio of height to circumference,

$$\mathrm{mod}(C) \;=\; \Delta y/\Delta x \;>\; 0.$$

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 $\rho(z)|dz|$ of finite area 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*

$$\mathbf{L}_\rho(\gamma)^2 \;\le\; \mathrm{area}_\rho(C)/\mathrm{mod}(C). \tag{B : 3}$$

Furthermore, this result is best possible: If we use the Euclidean metric $|dz|$ then

$$\mathbf{L}(\gamma)^2 \;\ge\; \mathrm{area}(C)/\mathrm{mod}(C) \tag{B : 4}$$

for every such curve γ.

Proof of Theorem B.2. Just as in the proof of Lemma B.1, we find a horizontal curve η_y with

$$\mathbf{L}_\rho(\eta_y)^2 \;\le\; \frac{\Delta x}{\Delta y}\mathrm{area}_\rho(C) \;=\; \frac{\mathrm{area}_\rho(C)}{\mathrm{mod}(C)}.$$

On the other hand, in the Euclidean case, for any closed curve γ of winding number 1 we have

$$\mathbf{L}(\gamma) \;=\; \oint_\gamma |dz| \;\ge\; \oint_\gamma dx \;=\; \Delta x,$$

hence $\mathbf{L}(\gamma)^2 \;\ge\; (\Delta x)^2 \;=\; \mathrm{area}(C)/\mathrm{mod}(C)$. □

Definitions. A Riemann surface A is said to be an *annulus* if it is conformally isomorphic to some cylinder. (Compare Example 2.4.) An embedded annulus $A \subset C$ is said to be *essentially embedded* if it contains a curve which has winding number 1 around the cylinder C.

Here is an important consequence of Theorem B.2.

Corollary B.3 (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_A. Then*

$$\frac{\mathrm{mod}(C_A)}{\mathrm{mod}(C)} \;\le\; \frac{\mathrm{area}(A)}{\mathrm{area}(C)} \;\le\; 1. \tag{B : 5}$$

In particular, it follows that

$$\mathrm{mod}(C_A) \;\le\; \mathrm{mod}(C). \tag{B : 6}$$

Proof. Let $\zeta \mapsto z$ be the embedding of C_A 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_A, where $\rho(\zeta) = |dz/d\zeta|$. According to Theorem B.2, there exists a curve γ' with winding number 1 about C_A whose length satisfies

$$\mathbf{L}_\rho(\gamma')^2 \le \mathrm{area}_\rho(C_A)/\mathrm{mod}(C_A).$$

This length coincides with the Euclidean length $\mathbf{L}(\gamma)$ of the corresponding curve γ in $A \subset C$, and $\mathrm{area}_\rho(C_A)$ is equal to the Euclidean area $\mathrm{area}(A)$, so we can write this inequality as

$$\mathbf{L}(\gamma)^2 \le \mathrm{area}(A)/\mathrm{mod}(C_A).$$

But according to (B : 4) we have

$$\mathrm{area}(C)/\mathrm{mod}(C) \;\le\; \mathbf{L}(\gamma)^2.$$

Combining these two inequalities, we obtain

$$\mathrm{area}(C)/\mathrm{mod}(C) \;\le\; \mathrm{area}(A)/\mathrm{mod}(C_A),$$

which is equivalent to the required inequality (B : 5). □

Corollary B.4. *If two cylinders are conformally isomorphic, then their moduli are equal.*

Proof. If C' is conformally isomorphic to C then (B : 6) asserts that $\mathrm{mod}(C') \le \mathrm{mod}(C)$, and similarly $\mathrm{mod}(C) \le \mathrm{mod}(C')$. □

It follows that the ***modulus*** of an annulus A can be defined as the modulus of any conformally isomorphic cylinder. If A is essentially embedded in some other annulus A', it then follows from (B : 6) that

$$\mathrm{mod}(A) \;\le\; \mathrm{mod}(A'). \tag{B : 6'}$$

Examples B.5. If $\mathbb{A}_r$ is the annulus consisting of all $w \in \mathbb{C}$ with $1 < |w| < r$, then setting $z = i\log(w) \pmod{2\pi\mathbb{Z}}$ we map $\mathbb{A}_r$ conformally onto a cylinder of height $\log(r)$ and circumference 2π. Hence

$$\mathrm{mod}\,(\mathbb{A}_r) \;=\; \log(r)/2\pi.$$

On the other hand, if we construct an annulus A from the upper half-plane $\mathbb{H}$ by identifying w with kw for some $k > 1$, then setting $z = \log(w) \pmod{\log(k)}$ we map A onto a cylinder of height π and circumference $\log(k)$. Hence

$$\operatorname{mod}\,(\mathbb{H}/(w \equiv kw)) \;=\; \pi/\log(k).$$

Corollary B.6 (Grötzsch Inequality). *Suppose that $A' \subset A$ and $A'' \subset A$ are two disjoint annuli, each essentially embedded in A. Then*

$$\operatorname{mod}(A') + \operatorname{mod}(A'') \;\le\; \operatorname{mod}(A).$$

Proof. We may assume that A is a cylinder C. According to (B : 5) we have

$$\frac{\operatorname{mod}(A')}{\operatorname{mod}(C)} \le \frac{\operatorname{area}(A')}{\operatorname{area}(C)}, \qquad \frac{\operatorname{mod}(A'')}{\operatorname{mod}(C)} \le \frac{\operatorname{area}(A'')}{\operatorname{area}(C)}.$$

where all areas are Euclidean. Using the inequality

$$\operatorname{area}(A') + \operatorname{area}(A'') \le \operatorname{area}(C),$$

the conclusion follows. □

Now consider a *flat torus* $\mathbb{T} = \mathbb{C}/\Lambda$. Here $\Lambda \subset \mathbb{C}$ 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 $\lambda_1/\lambda_2 \notin \mathbb{R}$. Let $A \subset \mathbb{T}$ be an embedded annulus.

By the "*winding number*" of A in $\mathbb{T}$ we will mean the lattice element $w \in \Lambda$ which is constructed as follows. Under the universal covering map $\mathbb{C} \to \mathbb{T}$, the central curve of A lifts to a curve segment which joins some point $z_0 \in \mathbb{C}$ to a translate $z_0 + w$ by the required lattice element. We say that $A \subset \mathbb{T}$ is an *essentially embedded annulus* if $w \ne 0$.

Corollary B.7 (Bers Inequality). *If the annulus A is essentially embedded in the flat torus $\mathbb{T} = \mathbb{C}/\Lambda$ with winding number $w \in \Lambda$, then*

$$\operatorname{mod}(A) \;\le\; \frac{\operatorname{area}(\mathbb{T})}{|w|^2}. \tag{B : 7}$$

Roughly speaking, if A winds many times around the torus, so that $|w|$ is large, then this annulus A must be very thin. A slightly sharper version of this inequality will be given in Problem B-3.

Proof. Choose a cylinder C' which is conformally isomorphic to A. The Euclidean metric $|dz|$ on $A \subset \mathbb{T}$ corresponds to some metric $\rho(\zeta)|d\zeta|$ on C', with

$$\operatorname{area}_\rho(C') \;=\; \operatorname{area}(A).$$

By Theorem B.2 we can choose a curve γ' of winding number 1 on C', or a corresponding curve γ on $A \subset \mathbb{T}$, 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 $\mathbb{C}$, 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 (B : 7). $\square$

Now consider the following situation. Let $U \subset \mathbb{C}$ be a bounded simply connected open set, and let $K \subset U$ be a compact connected subset, so that the difference $A = U \smallsetminus K$ is a topological annulus. As noted in Example 2.4, 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 infinite modulus. (Such an infinite cylinder may be either one-sided infinite and hence conformally isomorphic to a punctured disk, or two-sided infinite and conformally isomorphic to the punctured plane.)

Corollary B.8. *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*

$$4\,\mathrm{diam}(K)^2 \;\le\; \frac{\mathrm{area}(A)}{\mathrm{mod}(A)} \;\le\; \frac{\mathrm{area}(U)}{\mathrm{mod}(A)}. \qquad (\mathrm{B}:8)$$

Proof. According to Theorem B.2, there exists a curve with winding number 1 about A whose length satisfies $\mathbf{L}^2 \le \mathrm{area}(A)/\mathrm{mod}(A)$. Since K is enclosed within this curve, it follows easily that $\mathrm{diam}(K) \le \mathbf{L}/2$, and the inequality (B : 8) follows. Conversely, if K is a single point, then A contains an essentially embedded punctured disk and it follows from (B : 6′) that $\mathrm{mod}(A) = \infty$. $\square$

The following ideas are due to McMullen. (See Branner and Hubbard [1992, §5.4].) 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, Courant and Robbins [1941].) 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

$$\mathrm{area}(K) \;\le\; \frac{\mathrm{area}(U)}{1 + 4\pi\,\mathrm{mod}(A)}. \tag{B : 9}$$

This can be sharpened as follows:

Corollary B.9 (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 (B : 9), 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. Many short lines. In the situation of Lemma B.1 on the unit square $[0,1]\times[0,1]$, show that more than half of the horizontal curves η_y have length $\mathbf{L}_\rho(\eta_y) \le \sqrt{2\,\mathrm{area}_\rho(I^2)}$. (Here "more than half" is to be interpreted in the sense of Lebesgue measure.)

Problem B-2. Defining the modulus by potential theory. Recall that a real-valued function u on a Riemann surface is *harmonic* if and only if it can be described locally as the real part of a complex analytic function $u + iv$. (Compare Problems 9-b, 9-d.) The *harmonic conjugate* of u is the real-valued function v, well defined locally up to an additive constant. **(1)** With the cylinder C as in Theorem B.2, show that there is one and only one harmonic function $u : C \to \mathbb{R}$ such that $u(x+iy) \to 0$ as $x+iy$ tends to the bottom boundary $y = 0$, and $u(x+iy) \to 1$ as $x+iy$ tends to the top boundary. In fact this function is given by the formula $u(x+iy) = y/\Delta y$. **(2)** If γ is a curve with winding number $+1$ around C, show that

$$i\oint_\gamma d(u+iv) \;=\; \oint_\gamma d(x+iy)/\Delta y \;=\; \frac{1}{\mathrm{mod}(C)}.$$

Problem B-3. Bers inequality for multiple annuli. (1) If the flat torus $\mathbb{T} = \mathbb{C}/\Lambda$ contains several disjoint annuli A_i, all with the same winding number $w \in \Lambda$, show that

$$\sum \mathrm{mod}(A_i) \le \mathrm{area}(\mathbb{T})/|w|^2.$$

(2) If two essentially embedded annuli are disjoint, show that they necessarily have the same winding number.

Problem B-4. Branner-Hubbard criterion. Let

$$K_1 \supset K_2 \supset K_3 \supset \cdots$$

be compact connected subsets of $\mathbb{C}$ with each K_{n+1} contained in the interior of K_n. Suppose further that each interior K_n° is simply connected, so that each difference $A_n = K_n^\circ \smallsetminus K_{n+1}$ is an annulus. **(1)** If $\sum_1^\infty \mathrm{mod}(A_n)$ is infinite, show that the intersection $\bigcap K_n$ reduces to a single point. **(2)** Show that the converse statement is false: this intersection may reduce to a single point even though $\sum_1^\infty \mathrm{mod}(A_n) < \infty$. (As a first step, you could consider the open unit disk $\mathbb{D}$ and a closed disk $\overline{\mathbb{D}}'$ of radius $0 < r < 1$ centered at $1 - r - \epsilon$, showing that $\mathrm{mod}(\mathbb{D} \smallsetminus \overline{\mathbb{D}}')$ tends to zero as $\epsilon \searrow 0$.)

Appendix C. Rotations, Continued Fractions, and Rational Approximation

(Compare §§11 and 15.) The study of *recurrence* is a central topic in many parts of dynamics: *How often and how closely does an orbit return to a neighborhood of its initial point?* In the case of irrational rotations of a circle we can give a rather precise description of the answer. This description, which has its roots in classical number theory, turns out to be important not only in holomorphic dynamics, but also in celestial mechanics and other areas where "small divisor" problems occur.

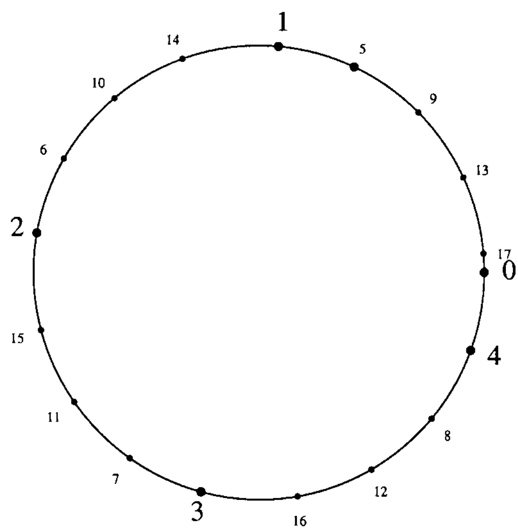


Figure 46. Successive orbit points under a rotation through the angle of $\xi = \sqrt{5} - 2 = .2360\ldots \pmod{\mathbb{Z}}$*, with close returns at times* $q = 1, 4, 17, 72, \ldots$. *The first five orbit points have been emphasized.*

Let $S^1 \subset \mathbb{C}$ be the unit circle, consisting of all complex numbers of absolute value 1. Given some fixed $\lambda = e^{2\pi i\xi} \in S^1$, we are interested in the dynamical system $z \mapsto \lambda z$ for $z \in S^1$. Since any two orbits are isometric under a rotation of the circle, it suffices to study the orbit

$$1 \mapsto \lambda \mapsto \lambda^2 \mapsto \lambda^3 \mapsto \cdots .$$

We are particularly interested in the case where λ is not a root of unity, so that there is no periodic point. Figure 46 illustrates a typical example, showing the first eighteen points on the orbit. (To simplify the picture, each λ^k is labeled simply by k.)

Definition. We will say that the sequence $\lambda^1, \lambda^2, \lambda^3, \ldots$ has a *close return* to $\lambda^0 = 1$ at the *time* q if λ^q is closer to 1 than any of its predecessors:

$$|\lambda^q - 1| < |\lambda^k - 1| \quad \text{for} \quad k = 1, 2, 3, \ldots, q-1.$$

As usual, we can equally well use the additive model $\mathbb{R}/\mathbb{Z}$ for the circle, in place of the multiplicative model $S^1 = \{\lambda \in \mathbb{C} : |\lambda| = 1\}$. These are related by letting each $\xi \in \mathbb{R}/\mathbb{Z}$ correspond to $\lambda = e^{2\pi i\xi} \in S^1$. Thus a completely equivalent problem is to study the dynamical system

$$x \mapsto x + \xi \pmod{\mathbb{Z}},$$

with typical orbit

$$0 \mapsto \xi \mapsto 2\xi \mapsto 3\xi \mapsto \cdots$$

in $\mathbb{R}/\mathbb{Z}$. By definition, the angle $\xi \in \mathbb{R}/\mathbb{Z}$ is called the *rotation number* of this orbit. (Compare §15.)

Any two distinct points $\xi, \eta \in \mathbb{R}/\mathbb{Z}$ cut the circle $\mathbb{R}/\mathbb{Z}$ into two arcs with total length 1. Define the *distance* $\|\xi - \eta\| \le 1/2$ to be the length of the shorter arc, or to be zero if $\xi = \eta$. The distance between the points ξ and 0 will also be described as the *norm* $\|\xi\|$.

With this notation, we can define $q \ge 1$ to be a *close return time* under rotation by ξ if

$$\|q\xi\| < \|k\xi\| \qquad \text{for all} \quad k \quad \text{with} \quad 0 < k < q.$$

This is equivalent to the previous definition, in view of the easily verified identity

$$|\lambda^k - 1| = 2\sin(\pi\|k\xi\|),$$

where the function $x \mapsto 2\sin(\pi x)$ is strictly monotone for $0 \le x \le 1/2$.

If the rotation number is rational, $\xi \equiv p/q$, then there will be only finitely many close return times, with q as the largest. However, to fix our ideas, let us concentrate on the case where ξ is irrational, so that the orbit points $k\xi \in \mathbb{R}/\mathbb{Z}$ are all distinct. Then there are infinitely many close return times, and we can number them as

$$1 = q_1 < q_2 < q_3 < \cdots.$$

The distance $d_n = \|q_n\xi\|$ between 0 and the orbit point $q_n\xi$ will be called the nth *close return distance*. Evidently

$$d_1 > d_2 > d_3 > \cdots > 0,$$

with $d_1 < 1/2$. Here is a preliminary remark. (Compare equation (C : 9) and Problem C-1.)

Lemma C.1. *For each* $n \geq 2$*, the first* q_n *orbit points*

$$0,\ \xi,\ 2\xi,\ \ldots,\ (q_n - 1)\xi$$

divide the circle $\mathbb{R}/\mathbb{Z}$ *up into* q_n *distinct segments, each of which has length at least* d_{n-1}*. Therefore* $d_{n-1} < 1/q_n$*.*

Proof. Given any integers $0 \leq j < k < q_n$, the distance between the orbit points $j\xi$ and $k\xi$ around the circle is equal to the distance $\|(k-j)\xi\|$ between 0 and $(k-j)\xi$. Suppose that one of these distances $\|(k-j)\xi\|$ is strictly less than d_{n-1}. If $k - j \leq q_{n-1}$ this would contradict the defining property of q_{n-1}, while if $q_{n-1} < k - j < q_n$ then it would contradict the defining property of q_n. Thus we must have

$$\|(k-j)\xi\| \ \geq \ d_{n-1},$$

as required. The lengths of these complementary intervals cannot all be equal since ξ is irrational, so the minimum length d_{n-1} must be strictly less than the average length $1/q_n$. □

It will be convenient to set $q_0 = 0$ and $d_0 = 1$.

Theorem C.2. *The sequence of close return times satisfies*

$$q_{n+1} \ \equiv \ q_{n-1} \pmod{q_n} \quad \textit{for all} \quad n \ \geq \ 1.$$

Thus there exist positive integers $a_1, a_2, a_3, \ldots$ (*known as partial quotients*) *which satisfy the equation*

$$q_{n+1} \ = \ q_{n-1} + a_n q_n \qquad \textit{for} \qquad n \geq 1, \tag{C : 1}$$

with $q_0 = 0$*,* $q_1 = 1$*, and* $q_2 = a_1 \geq 2$*. Similarly, the sequence of close return distances satisfies*

$$d_{n+1} \ = \ d_{n-1} - a_n d_n \qquad \textit{for} \qquad n \geq 1, \tag{C : 2}$$

using these same integers a_n*, with* $d_0 = 1$ *and* $d_1 = \|\xi\|$*.*

Thus a_n can be described either as the integer part $\operatorname{int}(q_{n+1}/q_n)$, which governs how rapidly the numbers q_n are growing, or as the integer part $\operatorname{int}(d_{n-1}/d_n)$, which governs how rapidly the d_n, are tending to zero.

Proof of Theorem C.2, by induction on n. For each $k \geq 0$, let x_k be the unique real number in the interval $-1/2 < x_k \leq 1/2$ which represents the residue class $k\xi \pmod{\mathbb{Z}}$, so that $|x_k| = \|k\xi\|$. To start the induction, if a_1 is the largest integer with $a_1 d_1 < 1$, then it is easy to see that the first and second close return times are $q_1 = 1$ and $q_2 = a_1$, with close return distances $d_1 = \|\xi\|$ and $d_2 = 1 - a_1 d_1$, as required. Furthermore, these two close returns occur on opposite sides of zero. For example, if $x_{q1} = +d_1$, then $x_{q2} = -d_2$.

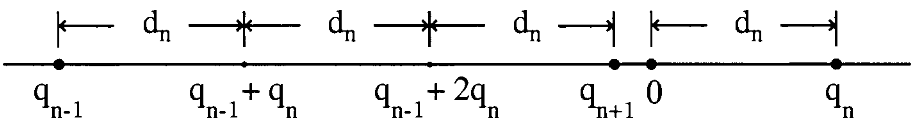


Figure 47. Locations of three successive close returns along the interval $(-1/2, 1/2)$, *illustrated for the case* $a_n = 3$ *so that* $d_{n+1} = d_{n-1} - 3d_n$. *As in the previous figure, each point* x_k *along the orbit is labeled simply by the integer* k. *Depending on the parity of* n, *the orientation of this figure may be reversed.*

Now suppose inductively that we have two consecutive close returns, at $q_{n-1}\xi$ and at $q_n\xi$, lying on opposite sides of zero. Thus

$$d_{n-1} = |x_{q_{n-1}}| > d_n = |x_{q_n}| \qquad \text{with} \qquad x_{q_{n-1}} x_{q_n} < 0.$$

Define $a_n \ge 1$ to be the largest integer for which $a_n d_n < d_{n-1}$. We must show that $q_{n+1} = q_{n-1} + a_n q_n$ and that $d_{n+1} = d_{n-1} - a_n d_n$. The proof will be based on the following auxiliary statement.

Lemma C.3. *For* $0 < k \le q_{n+1}$, *the point* x_k *lies strictly between* $x_{q_{n-1}}$ *and* x_{q_n} *if and only if* k *is a number of the form* $k_j = q_{n-1} + jq_n$ *with* $1 \le j \le a_n$. *In particular, it follows that the next close return time is given by* $q_{n+1} = k_{a_n} = q_{n-1} + a_n q_n$.

Proof of Lemma C.3. Suppose, to fix our ideas, that $x_{q_{n-1}} < 0 < x_{q_n}$, as illustrated in Figure 47, so that $x_{q_{n-1}} = -d_{n-1}$ and $x_{q_n} = d_n$. If x_k lies in the open interval $(x_{q_{n-1}}, x_{q_n})$, then it follows from the definition of q_n that $k > q_n$. Let $\ell = k - q_n > 0$, so that

$$x_\ell = x_k - d_n \in (x_{q_{n-1}} - d_n, 0).$$

There are now two possibilities: If the point x_ℓ has distance less than d_n from $x_{q_{n-1}}$, then $x_\ell = x_{q_{n-1}}$ and hence $\ell = q_{n-1}$ by Lemma C.1. Otherwise, x_ℓ must be strictly to the right of $x_{q_{n-1}}$, in which case we can repeat the argument using ℓ in place of k. After finitely many repetitions of this argument, we must find a difference $m = k - ad_n$ with x_m close to $x_{q_{n-1}}$ and hence $m = q_{n-1}$. This completes the proof of both Lemma C.3 and Theorem C.2. □

Corollary C.4. *The close return times* q_n *increase at least exponentially fast as* $n \to \infty$. *Similarly the close return distances*

d_n decrease to zero at least exponentially fast as $n \to \infty$.

Proof. This follows from the inequalities $q_{n+1} = a_n q_n + q_{n-1} \ge 2q_{n-1}$ and $2d_{n+1} \le a_n d_n + d_{n+1} = d_{n-1}$. □

These constructions are closely related to the classical *continued fraction algorithm*. Given any real number $0 < r < 1$, we can construct a finite or infinite sequence $a_1, a_2, a_3, \ldots$ of positive integers, as well as a finite or infinite sequence $r_1, r_2, r_3, \ldots$ of remainder terms, by induction as follows. We start with $r_0 = r$ and set

$$1/r_{n-1} = a_n + r_n \qquad (\mathrm{C}:3)$$

for $n \ge 1$, with $a_n \in \mathbb{Z}$ and $0 \le r_n < 1$. Thus a_n can be described as the *integer part*, $\mathrm{int}(1/r_{n-1})$, and r_n can be described as the *fractional part* $\mathrm{frac}(1/r_{n-1})$. First suppose that $r_0 = r$ is a rational number p/q. Then r_1 will be a rational number with denominator p strictly less than q. It follows that this procedure must terminate after at most q steps, reaching some r_n which is equal to zero, so that $1/r_{n-1} = a_n \in \mathbb{Z}$, while $1/r_n$ is not defined. In this case we obtain the finite continued fraction equation

$$\frac{p}{q} = \frac{1}{a_1 + r_1} = \frac{1}{a_1 + \dfrac{1}{a_2 + r_2}} = \cdots = \cfrac{1}{a_1 + \cfrac{1}{a_2 + \ddots + \cfrac{1}{a_n}}},$$

or more compactly $p/q = 1/(a_1 + 1/(a_2 + \cdots + 1/(a_{n-1} + 1/a_n)\cdots))$, with $a_n \ge 2$. On the other hand, if r is irrational, then all of the r_n will be irrational, and this inductive procedure continues indefinitely. This construction is often summarized by the infinite continued fraction equation

$$r = \cfrac{1}{a_1 + \cfrac{1}{a_2 + \ddots + \cfrac{1}{a_n + \ddots}}} = \lim_{n\to\infty} \cfrac{1}{a_1 + \cfrac{1}{a_2 + \ddots + \cfrac{1}{a_n}}}.$$

Convergence of this limit will be proved below.

To translate Theorem C.2 into the continued fraction terminology, simply set $r = \|\xi\| \le 1/2$ and

$$r_n = d_{n+1}/d_n \qquad \text{or equivalently} \qquad d_n = r_0 r_1 \cdots r_{n-1}.$$

Dividing equation (C : 2) by d_n and rearranging terms, we get

$$d_{n-1}/d_n = a_n + d_{n+1}/d_n,$$

or in other words $1/r_{n-1} = a_n + r_n$, as required. In particular, this shows that the continued fraction for any number $r \in (0, 1/2]$ can be interpreted in terms of the study of close returns. (In fact, continued fractions in the case $1/2 < r < 1$ can easily be reduced to the case $0 < r < 1/2$. See Problem C-2.)

For any $r \in (0,1)$, it is not difficult to check that each product $q_n r$ is congruent to $(-1)^{n+1} d_n$ modulo $\mathbb{Z}$, so that the sum

$$p_n \;=\; q_n r \,+\, (-1)^n d_n \tag{C : 4}$$

is an integer. These numbers can be computed inductively as follows: Multiply equation (C : 1) by the constant r and then add the equation

$$(-1)^{n+1} d_{n+1} \;=\; (-1)^n a_n d_n \,+\, (-1)^{n-1} d_{n-1},$$

which is equivalent to (C : 2), to obtain

$$p_{n+1} \;=\; a_n p_n \,+\, p_{n-1}. \tag{C : 5}$$

The first two values are $p_0 = 1$ and $p_1 = 0$. Applying (C : 5) we find that the subsequent values $p_2 = 1$, $p_3 = a_2$, ... are rapidly increasing. Furthermore, dividing equation (C : 4) by q_n, we see that the initial ratio r is closely approximated by the rational number p_n/q_n. In fact

$$\frac{p_n}{q_n} \;=\; r \,+\, (-1)^n \frac{d_n}{q_n}, \tag{C : 6}$$

where d_n/q_n tends very rapidly to zero by Corollary C.4. These successive approximations, known as *convergents* to r, are ordered as follows:

$$0 = \frac{p_1}{q_1} \,<\, \frac{p_3}{q_3} \,<\, \cdots \,<\, r \,<\, \cdots \,<\, \frac{p_4}{q_4} \,<\, \frac{p_2}{q_2} = \frac{1}{a_1}. \tag{C : 7}$$

Note that each of these convergents can be defined by the finite continued fraction expansion

$$p_{n+1}/q_{n+1} \;=\; 1/(a_1 + 1/(a_2 + 1/(\cdots + 1/(a_{n-1} + 1/a_n)\cdots))).$$

(Caution: Most authors use a different numbering.) Using matrix notation, we can write equations (C : 1) and (C : 5) as

$$\begin{bmatrix} p_n & q_n \\ p_{n+1} & q_{n+1} \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 1 & a_n \end{bmatrix} \begin{bmatrix} p_{n-1} & q_{n-1} \\ p_n & q_n \end{bmatrix},$$

and hence by induction

$$\begin{bmatrix} p_n & q_n \\ p_{n+1} & q_{n+1} \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 1 & a_n \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 1 & a_{n-1} \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 1 & a_{n-2} \end{bmatrix} \cdots \begin{bmatrix} 0 & 1 \\ 1 & a_1 \end{bmatrix}.$$

In particular, it follows that the determinant of this matrix is given by

$$p_n q_{n+1} - q_n p_{n+1} = (-1)^n .$$

In particular, p_n and q_n are relatively prime, so that each p_n/q_n is a fraction in lowest terms. Furthermore, the difference between two successive approximations is given by

$$\left| \frac{p_{n+1}}{q_{n+1}} - \frac{p_n}{q_n} \right| = \frac{1}{q_n q_{n+1}} . \qquad (C:8)$$

Since these numbers converge rapidly to zero, it follows that the successive convergents p_n/q_n (or in other words the successive finite continued fraction expansions) do indeed converge to the limit r.

Combining (C : 6) and (C : 8), we see that

$$\frac{d_n}{q_n} + \frac{d_{n+1}}{q_{n+1}} = \frac{1}{q_n q_{n+1}} ,$$

or in other words $q_n d_{n+1} + q_{n+1} d_n = 1$. (For a geometric interpretation of this equality, see Problem C-1.) Since $q_n d_{n+1} < q_{n+1} d_n$, it follows that

$$1/2 < q_{n+1} d_n < 1, \qquad (C:9)$$

yielding a sharper form of the inequality $d_n < 1/q_{n+1}$ of Lemma C.1.

Now look at the converse problem. Suppose that we are given an infinite sequence $a_1, a_2, \ldots$ of positive integers. Then we can define associated integers p_n and q_n by equations (C : 1) and (C : 5). The fractions p_n/q_n will satisfy (C : 7) and are ordered as in (C : 6). Thus we can define r as the limit of p_n/q_n as $n \to \infty$. It is not difficult to check that we recover the given integers a_n from the continued fraction expansion of this r.

Applying this discussion to the sequence of close return times for a circle rotation, we obtain the following.

Corollary C.5. *The finite or infinite sequence of close return times $q_1, q_2, \ldots$, or equivalently the finite or infinite sequence of integers $a_1, a_2, \ldots$, determines the rotation number ξ up to sign. Here $a_1 \geq 2$ (since $d_1 = \|\xi\| \leq 1/2$) and a given infinite sequence of numbers $a_i \geq 1$ can occur if and only if the initial integer satisfies $a_1 \geq 2$. A finite sequence $(a_1, \ldots, a_n)$ can occur if and only if both $a_1 \geq 2$ and $a_n \geq 2$.*

The proof will be left to the reader. □

The subject of "best" rational approximations to an irrational number $x \in (0, 1)$ is closely related. Setting $r = x$, note the following: *The convergent p_n/q_n is closer to x than any fraction p/q with denominator*

$0 < q < q_n$. In fact we have $|qx - p| \ge |q_n x - p_n| > 0$ by the definition of q_n. Multiplying this by the inequality $1/q > 1/q_n > 0$, we obtain

$$|x - p/q| \;>\; |x - p_n/q_n|,$$

as asserted.

Let $\alpha \ge 2$ be a real number. Recall from §11 that an irrational number x is *Diophantine of order* $\le \alpha$ if there exists $\epsilon > 0$ so that

$$\left| x - \frac{p}{q} \right| \;>\; \frac{\epsilon}{q^\alpha}$$

for all rational numbers p/q, or in other words if the collection of products $|x - p/q| q^\alpha = |qx - p| q^{\alpha - 1}$ is bounded away from zero. We can express this condition as a limitation on the rate of growth of the associated integers q_n as follows.

Lemma C.6. *The irrational number $x \in (0, 1)$ is Diophantine of order $\le \alpha$ if and only if there is a constant C so that $q_{n+1} \;\le\; C\, q_n^{\alpha-1}$ for all n.*

For example, x is Diophantine of order 2 if and only if the ratios q_{n+1}/q_n are bounded, or if and only if the integer parts $a_n = \mathrm{int}(q_{n+1}/q_n)$ are bounded.

Proof of Lemma C.6. If $q_n < q < q_{n+1}$, then we have

$$|qx - p| \;\ge\; d_n \;=\; |q_n x - p_n|$$

for every integer p. Multiplying this inequality by $q^{\alpha-1} > q_n^{\alpha-1}$, it follows that

$$|qx - p| q^{\alpha-1} \;>\; |q_n x - p_n| q_n^{\alpha-1}.$$

But we know from the inequality (C : 9) that the error $d_n = |q_n x - p_n|$ is equal to $1/q_{n+1}$, up to a factor of 2. Thus it suffices to know that the ratios $q_n^{\alpha-1}/q_{n+1}$ are bounded away from zero, or equivalently that the reciprocals $q_{n+1}/q_n^{\alpha-1}$ are bounded. □

Let $\mathcal{D}(\alpha) \subset \mathbb{R} \smallsetminus \mathbb{Q}$ be the set of all irrational numbers which are Diophantine of order $\le \alpha$, and let

$$\mathcal{D}(2+) \;=\; \bigcap_{\alpha > 2} \mathcal{D}(\alpha), \qquad \mathcal{D}(\infty) \;=\; \bigcup_{\alpha < \infty} \mathcal{D}(\alpha).$$

Note that $\mathcal{D}(2) \subset \mathcal{D}(2+) \subset \mathcal{D}(\alpha) \subset \mathcal{D}(\infty)$ whenever $2 < \alpha$.

Lemma C.7. *The complement $\mathbb{R} \smallsetminus \mathcal{D}(\alpha)$ has Hausdorff dimension $\le 2/\alpha$. Hence the set $\mathbb{R} \smallsetminus (\mathcal{D}(\infty) \cup \mathbb{Q})$ of Liouville numbers has Hausdorff dimension zero.*

Proof. It suffices to work in the unit interval $[0,1]$ (or equivalently in $\mathbb{R}/\mathbb{Z}$). We must prove that for any real number $d > 2/\alpha$ and any $\epsilon > 0$ it is possible to cover $[0,1] \cap \mathcal{D}(\alpha)$ by intervals I_j so that $\sum \ell(I_j)^d < \epsilon$, where $\ell(I_j)$ denotes the length of I_j. We proceed as follows. (Compare the proof of Lemma 11.7.) If $\xi \notin \mathcal{D}(\alpha)$, then for every $\epsilon > 0$ there exists a fraction p/q with $|\xi - p/q| \le \epsilon/q^\alpha$. That is, ξ belongs to a union of intervals of length $2\epsilon/q^\alpha$. For each fixed q there are at most $q+1$ different choices for $p/q \in [0,1]$, so the sum of lengths to the dth power is at most $(q+1)(2\epsilon/q^\alpha)^d$ for each q. Summing over q, the total is at most $(2\epsilon)^d \sum_q (q+1)/q^{\alpha d}$, which is finite, and tends to zero as $\epsilon \to 0$, provided that $\alpha d > 2$. Therefore, $\mathbb{R} \smallsetminus \mathcal{D}(\alpha)$ has Hausdorff dimension $\le 2/\alpha$. Taking the intersection as $\alpha \to \infty$, it follows that $\mathbb{R} \smallsetminus \mathcal{D}(\infty)$ has Hausdorff dimension zero. □

Here is a complementary result.

Lemma C.8. *The set $\mathcal{D}(2)$ of numbers of bounded type has measure zero.*

(For a much sharper statement, see Problem C-6.) Combining Lemma C.8 with Lemma 11.7, we see that almost all real numbers belong to the difference set $\mathcal{D}(2+) \smallsetminus \mathcal{D}(2)$.

The proof of Lemma C.8 will be based on some classical elementary number theory. Let us say that two rational numbers $p/q < p'/q'$ are *Farey neighbors* if the determinant $p'q - q'p$ is equal to $+1$, so that

$$\frac{p'}{q'} - \frac{p}{q} = \frac{1}{qq'}. \tag{C : 10}$$

We will need the following fact. *For any fixed $m > 0$, consider the set of all fractions in the unit interval with denominator at most m. Then any two consecutive fractions in this set are Farey neighbors.* To prove this, given two fractions $p/q < p'/q'$ consider the lattice $\Lambda \subset \mathbb{Z}^2$ spanned by the two vectors $\mathbf{v} = (q,p)$ and $\mathbf{w} = (q',p')$. Then the determinant $p'q - q'p$ can be identified with the number of elements in the quotient group $\mathbb{Z}^2/\Lambda$. Hence, if the two are not Farey neighbors, we can find a vector $(s,r) \in \mathbb{Z}^2$ which is not in Λ. Furthermore, we can assume that (s,r) lies in the interior of the fundamental parallelogram consisting of all $\alpha\mathbf{v} + \beta\mathbf{w}$ with $\alpha, \beta \in [0,1]$. Then the slope r/s must lie strictly between p/q and p'/q'. Furthermore, replacing (s,r) by $\mathbf{v} + \mathbf{w} - (s,r)$ if necessary, we can assume that $s \le (q+q')/2 \le \max\{q,q'\}$. *This proves the required assertion: fractions p/q and p'/q' which are not Farey neighbors can never be consecutive elements in the collection of all fractions with denominator $\le m$.*

Proof of Lemma C.8. For any rational $0 < \epsilon < 1$, let X_ϵ be the set of real numbers x which satisfy the inequality $|x - p/q| > \epsilon/q^2$ for every pair of integers $q > 0$ and p. Let $I \subset \mathbb{R}$ be any interval which is "rational" in the sense that it has rational endpoints, and let $\ell(I)$ be its length. Then we will show that it is possible to cover $X_\epsilon \cap I$ by finitely many rational intervals of total length $\le (1-\epsilon)\ell(I)$. Repeating this construction n times, we can cover $X_\epsilon \cap I$ with rational intervals of total length $\le (1-\epsilon)^n \ell(I)$. Since this expression tends to zero as $n \to \infty$, it will follow that X_ϵ has measure zero, and hence that $\mathcal{D}(2) = \bigcup_\epsilon X_\epsilon$ also has measure zero.

First consider the case where the endpoints $p/q < p'/q'$ are Farey neighbors, so that $\ell(I) = 1/qq'$. Let I' be the subinterval $I \smallsetminus (I^- \cup I^+)$ where

$$I^- = [p/q, \;\; p/q + \epsilon/q^2), \qquad I^+ = (p'/q' - \epsilon/(q')^2, \;\; p'/q'].$$

Then $X_\epsilon \cap I \subset I'$. Furthermore, one of these two subintervals must have length $\ell(I^\pm) \ge \epsilon\,\ell(I)$; hence $\ell(I') \le (1-\epsilon)\ell(I)$, as required.

For the more general case, where $\ell(I) > 1/qq'$, we proceed as follows. Let m be the maximum of q and q', and consider the *Farey series* consisting of all fractions p''/q'' in the interval I with denominator $q'' \le m$. These points cut the interval I up into subintervals $I_1, \ldots, I_N$, such that the endpoints of each I_j are Farey neighbors. Applying the construction above to each I_j, the conclusion follows. □

Concluding Problems

Problem C-1. Partition of the circle by orbit points. If the consecutive orbit points $0, \xi, 2\xi, \ldots, (k-1)\xi$ are distinct, then they cut $\mathbb{R}/\mathbb{Z}$ up into k nonoverlapping intervals. **(1)** If a and b are the lengths of the intervals to the left and right of 0, show that each of these k intervals has length either a, b, or $a+b$. In the special case $k = q_n + q_{n+1}$, show that q_n of these intervals have length d_{n+1} while the remaining q_{n+1} have length d_n. (Compare Lemma C.1 and equation (C : 9).)

Problem C-2. Comparing x and $1-x$. If the irrational number $x \in (0, 1/2)$ has continued fraction expansion $x = 1/(a_1 + 1/(a_2 + \cdots))$ with $a_1 \ge 2$, show that $x' = 1 - x$ has expansion $1/(a_1' + 1/(a_2' + \cdots))$ with

$$a_1' = 1, \quad a_2' = a_1 - 1, \quad \text{and} \quad a_n' = a_{n-1} \quad \text{for} \quad n \ge 3.$$

Show similarly that $q_n' = q_{n-1}$ and $d_n' = d_{n-1}$ for $n \ge 3$.

Problem C-3. Fibonacci numbers. (1) In the simplest possible case $a_1 = a_2 = \cdots = 1$, show that

$$\{q_n\} = \{p_{n+1}\} = \{0,\ 1,\ 1,\ 2,\ 3,\ 5,\ 8,\ 13,\ 21,\ \ldots\},$$

yielding the sequence of Fibonacci numbers. **(2)** 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 sequences $\{p_n\}$ and $\{q_n\}$.

Problem C-4. Quadratic irrationals. Show that $x \in \mathbb{R} \smallsetminus \mathbb{Q}$ satisfies a quadratic polynomial equation with integer coefficients if and only if the sequence of partial quotients $a_1, a_2, \ldots$ for its continued fraction expansion is eventually periodic.

Problem C-5. Continued fractions and Euler polynomials. Define the Euler polynomials

$$\begin{aligned} \mathcal{P}(\emptyset) \;=\; 1, \quad \mathcal{P}(x) \;=\; x, \quad \mathcal{P}(x,y) \;&=\; 1+xy, \\ \mathcal{P}(x,y,z) \;&=\; x+z+xyz, \;\ \ldots \end{aligned}$$

by setting $\mathcal{P}(x_1,\, x_2,\, \ldots,\, x_n)$ equal to the sum of all distinct monomials in the variables $x_1, \ldots, x_n$ which can be obtained from the product $x_1x_2\cdots x_n$ by striking out any number of consecutive pairs. (Here $\emptyset$ denotes the empty set of arguments.) Note that

$$\mathcal{P}(x_1,\, x_2,\, \ldots,\, x_n) \;=\; \mathcal{P}(x_n,\, x_{n-1},\, \ldots,\, x_1),$$

and show that

$$\mathcal{P}(x_1,\, \ldots,\, x_{n+1}) \;=\; \mathcal{P}(x_1,\, \ldots,\, x_{n-1}) \,+\, \mathcal{P}(x_1,\, \ldots,\, x_n)\, x_{n+1}$$

for $n \geq 1$. For any continued fraction, show that the numerators p_n and denominators q_n can be expressed as polynomial functions of the partial quotients a_i by the formulas

$$p_{n+1} = \mathcal{P}(a_2,\, \ldots,\, a_n), \qquad q_{n+1} = \mathcal{P}(a_1,\, \ldots,\, a_n),$$

so that

$$\cfrac{1}{a_1 + \cfrac{1}{a_2 + \ddots + \cfrac{1}{a_{n-1} + \cfrac{1}{a_n}}}} \;=\; \frac{\mathcal{P}(a_2,\, \ldots,\, a_n)}{\mathcal{P}(a_1,\, \ldots,\, a_n)}.$$

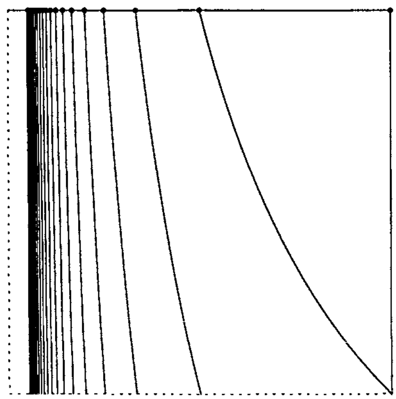


Figure 48. Graph of the Gauss map.

Problem C-6. The Gauss map. Define the *Gauss map* g from the half-open interval $(0,1]$ to itself by the requirement that

$$g(x) \equiv 1/x \pmod{\mathbb{Z}}.$$

Thus g is discontinuous at $1/n$ for every integer $n \ge 1$. (See Figure 48.) **(1)** Show that the probability measure

$$\mu(S) = \frac{1}{\log 2}\int_S \frac{dx}{1+x}$$

on $(0,1]$ is g-invariant, in the sense that $\mu \circ g^{-1} = \mu$. **(2)** It can be shown that this measure is *ergodic*; that is, every measurable g-invariant subset $S = g^{-1}(S)$ must have measure either zero or one. (See, for example, Cornfeld, Fomin, and Sinai [1982].) Assuming this, prove that for all $x \in (0,1)$ outside of a set of Lebesgue measure zero, *every* integer $k \ge 1$ must occur infinitely often among the set of partial quotients $\{a_1, a_2, \ldots\}$ in the continued fraction expansion.

Remark. Using the Birkhoff Ergodic Theorem, one can prove a much more precise statement:

For Lebesgue almost every x, the frequency

$$\lim_{n\to\infty} \frac{1}{n} \sum_{\{i \le n\,:\, a_i = k\}} 1$$

of occurrences of some given integer $k \ge 1$ among the a_i is well defined and equal to

$$\mu\left(\frac{1}{k+1}, \frac{1}{k}\right) = \log_2\left(1+\frac{1}{k}\right) - \log_2\left(1+\frac{1}{k+1}\right).$$

For example, the frequencies of 1, 2, 3 are roughly .4150, .1699, and .0931.

Appendix D. Two or More Complex Variables

Let M be a complex manifold of dimension $n \geq 2$, and let $F : M \to M$ be a holomorphic map. As in the 1-dimensional case, we are interested in studying the behavior of the family of iterates $F^{\circ k}$. Some of the constructions for the 1-dimensional case can be carried over to this higher dimensional case without essential change. However, there are many surprises and new difficulties.

Polynomial Maps. First consider the case of a nonlinear polynomial map

$$F(z_1, \ldots, z_n) = \big(f_1(z_1, \ldots, z_n), \ldots, f_n(z_1, \ldots, z_n)\big)$$

from the coordinate space $\mathbb{C}^n$ to itself. The Jacobian determinant $\det(F')$ of such a mapping is a polynomial function from $\mathbb{C}^n$ to $\mathbb{C}$, with the set of *critical points* of F as its locus of zeros. However, unlike the 1-dimensional case, where the finitely many critical points play a crucial role, there must be either uncountably many critical points or no critical points at all in the higher dimensional case. If there are no critical points, then this Jacobian determinant must be a nonzero constant. In all *known* cases, F will then have a well-defined inverse $F^{-1} : \mathbb{C}^n \to \mathbb{C}^n$ which is also a polynomial map. Such a polynomial map with polynomial inverse is called a *polynomial automorphism* of $\mathbb{C}^n$. (The well-known *Jacobian conjecture* is the assertion that *every* polynomial map of $\mathbb{C}^n$ without critical points must be a polynomial automorphism.)

The *Hénon maps* provide a well-known family of 2-dimensional examples. Suppose that we start with a polynomial function $f : \mathbb{C} \to \mathbb{C}$ of degree $d \geq 2$ and a complex constant $\delta \neq 0$. Consider doubly infinite sequences of complex numbers $\ldots, z_{-1}, z_0, z_1, z_2, \ldots$ satisfying the difference equation

$$z_{k+1} + \delta\, z_{k-1} = f(z_k). \tag{D : 1}$$

Evidently we can solve for (z_k, z_{k+1}) as a holomorphic function of (z_{k-1}, z_k), and the resulting transformation

$$F(z_{k-1}, z_k) = (z_k, f(z_k) - \delta z_{k-1}) \tag{D : 2}$$

will be a polynomial mapping of degree d. Its Jacobian matrix

$$\begin{bmatrix} 0 & 1 \\ -\delta & f'(z_k) \end{bmatrix}$$

has constant determinant δ and trace $f'(z_k)$. Similarly we can solve for $(z_{k-1},\, z_k)$ as a holomorphic function $F^{-1}(z_k,\, z_{k+1})$ with Jacobian determinant δ^{-1}. *This shows that the Hénon map F is a degree d polynomial automorphism of $\mathbb{C}^2$.*

In spite of their lack of critical points, these Hénon maps have highly nontrivial dynamics. For example, the number of fixed points of the iterate $F^{\circ k}$, counted with multiplicity, is equal to d^k, and hence grows exponentially with k. (See Friedland and Milnor [1989].) In contrast with the 1-dimensional case, a polynomial automorphism of $\mathbb{C}^2$ of large degree may well have *infinitely many* attracting periodic orbits. This was proved by Buzzard [1997], making use of ideas introduced by Newhouse [1974] and Fornæss and Gavosto [1992].

There are three different versions of the filled Julia set for a Hénon map—we can look either at the set of points K^+ with bounded forward orbit, the set of points K^- with bounded backward orbit, or the compact set $K = K^+ \cap K^-$ where the interesting dynamics takes place. Just as in 1-dimensional holomorphic dynamics, potential theoretic methods play a very important role, but in the Hénon case there are two different Green's functions: Each point $(z_0, z_1) \in \mathbb{C}^2$ is associated with a bi-infinite sequence $\ldots,\, z_{-1},\, z_0,\, z_1,\, z_2,\, \ldots$ of complex numbers, and we can pass to the limit

$$G^{\pm}(z_0\,,\, z_1) \;=\; \lim_{k\to\pm\infty} (\log^+|z_k|)/d^{|k|}$$

as the index k tends either to $+\infty$ or to $-\infty$. Each of these two functions is continuous, plurisubharmonic (that is, subharmonic on each complex line), and vanishes only on the corresponding set $K^{\pm}$.

For further information on polynomial automorphisms, see, for example, Hubbard [1986], Bedford [1990], Bedford and Smillie [1991–2002], Fornæss and Sibony [1992a], Bedford, Lyubich, and Smillie [1993], and Hubbard and Oberste-Vorth [1995].

Fatou-Bieberbach Domains. A fundamental principle in one-dimensional holomorphic dynamics is that every attracting basin must contain a critical point, but this breaks down for maps of $\mathbb{C}^2$. Given complex constants $\delta \neq 0$ and μ, let F be the Hénon map of determinant δ which is associated with the quadratic function $f(z) = z^2 + \mu z$ via equation (D : 2). Then F has a fixed point at $(0,0)$, and the eigenvalues λ_1 and λ_2 of the associated Jacobian matrix satisfy

$$\lambda_1 + \lambda_2 \;=\; f'(0) \;=\; \mu \qquad \text{and} \qquad \lambda_1\lambda_2 = \delta.$$

Evidently, by the appropriate choice of μ and δ, we can realize any desired nonzero λ_1 and λ_2. In particular, if we choose μ and δ so that both λ_1 and λ_2 lie in the open disk $|\lambda_j| < 1$, then the origin will be an attracting fixed point.

> *Thus we have constructed a nonlinear polynomial map with an attracting fixed point whose basin contains no critical point.*

This attracting basin has another exotic property. To simplify the discussion, let me assume that $\lambda_1 \neq \lambda_2$, so that we can diagonalize the Jacobian matrix by a linear change of coordinates. We will need the following.

Lemma D.1. *Consider any holomorphic transformation in two complex variables with a fixed point at the origin. If the eigenvalues λ_1 and λ_2 of the derivative map at the origin satisfy*

$$1 > |\lambda_1| \geq |\lambda_2| > |\lambda_1^2| \qquad \text{with} \qquad \lambda_1 \neq \lambda_2, \tag{D : 3}$$

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. After a linear change of coordinates, we may assume that $F(x,y)$ is equal to $(\lambda_1 x,\ \lambda_2 y)$ + (higher order terms). We must show that there exists a (nonlinear) change of coordinates $(x,y) \mapsto (u,v) = \phi(x,y)$, defined and holomorphic throughout a neighborhood of the origin, so that $\phi \circ F \circ \phi^{-1} = L$. As in the proof of the Kœnigs Theorem 8.2, 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. □

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. (See Problem D-1. For a much more precise statement as to when linearization is possible, compare Zehnder [1977].)

Now consider a Hénon map $F : \mathbb{C}^2 \xrightarrow{\cong} \mathbb{C}^2$ with a fixed point at the origin with eigenvalues satisfying the inequalities (D : 3). Let $\mathcal{B}$ be the attracting basin of the origin. We claim that ϕ extends to a global diffeomorphism $\Phi : \mathcal{B} \xrightarrow{\cong} \mathbb{C}^2$. For any $(x, y) \in \mathcal{B}$, let

$$\Phi(x,y) \;=\; L^{-n} \circ \phi \circ F^{\circ n}(x,y),$$

taking n to be large. In fact, if n is sufficiently large, then $F^n(x, y)$ is close to the origin, so that this expression is defined. It is independent of the particular choice of n since $\phi \circ F = L \circ \phi$. Similarly,

$$\Phi^{-1}(u,v) \;=\; F^{-n} \circ \phi^{-1} \circ L^{\circ n}(u,v)$$

is well defined for large n. This shows that Φ is a holomorphic diffeomorphism from $\mathcal{B}$ onto $\mathbb{C}^2$.

Note that this basin $\mathcal{B}$ is not the entire space $\mathbb{C}^2$. For example, if $|z_1|$ is sufficiently large compared with $|z_0|$, and if

$$(z_0,\, z_1) \overset{F}{\mapsto} (z_1,\, z_2) \overset{F}{\mapsto} (z_2,\, z_3) \overset{F}{\mapsto} \cdots,$$

then it is not difficult to check that $|z_1| < |z_2| < |z_3| < \cdots$, so that $(z_0,\, z_1)$ is not in $\mathcal{B}$.

> *Thus we have constructed a proper open subset $\mathcal{B} \subset \mathbb{C}^2$ which is biholomorphic (that is, analytically diffeomorphic) to all of $\mathbb{C}^2$.*

Again, such a phenomenon can never occur in one complex variable. Open sets $\mathcal{B}$ with this property are called *Fatou-Bieberbach domains.* Such examples were first constructed by Bieberbach [1933]. (Fatou had much earlier described a many-to-one map from $\mathbb{C}^2$ onto a proper subset of itself.)

Maps of $\mathbb{P}^2$. Now consider a holomorphic map F from the complex projective plane $\mathbb{P}^2 = \mathbb{P}^2(\mathbb{C})$ to itself. Just as in the case of $\mathbb{P}^1(\mathbb{C}) \cong \hat{\mathbb{C}}$, it is not difficult to check that every such F is a *rational map*. That is, using the notation $(x, y, z) \mapsto (x : y : z)$ for the projection from $\mathbb{C}^3 \smallsetminus \{(0,0,0)\}$ to $\mathbb{P}^2$, we can write

$$F(x:y:z) \;=\; \Big(f_0(x,y,z) : f_1(x,y,z) : f_2(x,y,z)\Big), \tag{D : 4}$$

where the f_j are homogeneous polynomial functions, all of the same degree d, with no common factor. In fact the f_j must satisfy the stronger condition of having no common zeros in $\mathbb{C}^3 \smallsetminus \{(0,0,0)\}$. Here d is called the *algebraic degree* of F. Note, however, that the topological degree of such an everywhere defined holomorphic map of $\mathbb{P}^2$ is equal to d^2. For example, a generic point of $\mathbb{P}^2$ has d^2 distinct preimages.

Unlike the case of maps of $\mathbb{C}^2$, where there may be a total lack of critical points, here we have an overabundance. In fact if $d \geq 2$, then there is always an entire algebraic curve of critical points, that is, points at which

F is not locally one-to-one. As in the 1-dimensional case, every attracting basin must intersect this critical locus. (See Ueda [1994].)

Just as in the case of Hénon maps, such holomorphic maps of $\mathbb{P}^2$ may well have infinitely many attracting periodic orbits. This was proved by Gavosto [1998], based on earlier work by Mora and Viana [1993], Fornæss and Gavosto [1992], and Newhouse [1974].

Fatou Components. The Fatou set of a map of $\mathbb{P}^n$ is defined much as in the 1-dimensional case. One can attempt to classify the possible Fatou components. However, unlike the 1-dimensional case (§16 and Appendix F), it is not known whether there may be wandering Fatou components. For the study of invariant Fatou components $U = F(U)$, see, for example, Fornæss and Sibony [1995b] or Fornæss [1996]. As examples of invariant Fatou components in $\mathbb{P}^2$, we have the following construction due to Ueda.

> **Lemma D.2.** *Let U_1 and U_2 be disjoint f-invariant Fatou components for some rational map f of $\mathbb{P}^1 \cong \hat{\mathbb{C}}$. Then there is a holomorphic map $F : \mathbb{P}^2 \to \mathbb{P}^2$ of the same degree with an invariant Fatou component U which is biholomorphic to the product $U_1 \times U_2$, and such that F restricted to U is holomorphically conjugate to $f \times f$ restricted to $U_1 \times U_2$.*

As a typical example, if U_1 is a Herman ring and U_2 is the immediate basin for an attracting fixed point p, then U is the immediate basin for an "attracting Herman ring" which is biholomorphic to $U_1 \times \{p\}$.

Proof. Let ι be the involution $\iota(p, q) = (q, p)$, which interchanges the two factors of $\mathbb{P}^1 \times \mathbb{P}^1$. Then the quotient space $(\mathbb{P}^1 \times \mathbb{P}^1)/\iota$ can be identified with $\mathbb{P}^2$. In fact, to any point $(a : b : c) \in \mathbb{P}^2$ there corresponds a homogeneous polynomial equation $ax^2 + bxy + cy^2$ with two not necessarily distinct roots $(x : y) \in \mathbb{P}^1$. This yields the required one-to-one correspondence between points of $\mathbb{P}^2$ and unordered pairs in $\mathbb{P}^1$. Similarly, the self-map $f \times f$ of $\mathbb{P}^1 \times \mathbb{P}^1$ corresponds to a self-map $F = (f \times f)/\iota$ of $(\mathbb{P}^1 \times \mathbb{P}^1)/\iota \cong \mathbb{P}^2$. Further details will be left to the reader. □

The Green's Function. Potential theoretic methods are important also in the study of holomorphic maps $F : \mathbb{P}^n \to \mathbb{P}^n$. However, the associated *Green's function* is defined, not on the space $\mathbb{P}^n$ itself, but rather on $\mathbb{C}^{n+1}$. By definition, F lifts to a homogeneous polynomial map $\widetilde{F} : \mathbb{C}^{n+1} \to \mathbb{C}^{n+1}$, and we can set

$$G(z_0, \ldots, z_n) \;=\; \lim_{k \to \infty} \Big(\log^+ \|\widetilde{F}^{\circ k}(z_0, \ldots, z_n)\|\Big)/d^k.$$

An open subset $U \subset \mathbb{P}^n$ is contained in the Fatou set of F if and only if the Green's function is *pluriharmonic* (that is, locally the real part of a holomorphic function) throughout the associated open set $\widetilde{U} \subset \mathbb{C}^{n+1} \smallsetminus \{0\}$. See Hubbard and Papadopol [1994], as well as Ueda [1994].

For further information about holomorphic maps of projective space, see, for example, Fornæss and Sibony [1992b, 1994, 1995b, 1998], Fornæss [1996], Smillie [1997], Ueda [1998], Sibony [1999/2003], as well as Morosawa, Nishimura, Taninguchi and Ueda [2000].

Rational Maps and Points of Indeterminacy. To fix our ideas, I will consider only the 2-dimensional case. The preceding discussion is based on the assumption that $F(x:y:z)$ is defined for *every* point $(x:y:z)$ of $\mathbb{P}^2$. But if we start with an arbitrary rational function, defined as in (D : 4), where the f_j are homogeneous polynomials of the same degree with no common factor, then there may well be *points of indeterminacy* $(x:y:z) \in \mathbb{P}^2$ for which the equations

$$f_0(x,y,z) \;=\; f_1(x,y,z) \;=\; f_2(x,y,z) \;=\; 0$$

have a simultaneous nontrivial solution. In this case, the function $F(x:y:z)$ is not everywhere defined, and the situation is quite different. (See, for example, Fornæss and Sibony [1995a].) I will use the notation

$$F : \mathbb{P}^2 \dashrightarrow \mathbb{P}^2,$$

where the dotted arrow indicates that the map may not be everywhere defined.

As an example, every degree d polynomial automorphism F of $\mathbb{C}^2$ extends easily to a rational map $\widehat{F} : \mathbb{P}^2 \dashrightarrow \mathbb{P}^2$ with algebraic degree d, but there are always points of indeterminacy when $d \geq 2$. In fact, a generic point of $\mathbb{C}^2$ has only one preimage, instead of d^2 preimages, as it would have for an everywhere defined holomorphic map of $\mathbb{P}^2$. Note that the inverse automorphism extends similarly to a map $\widehat{F}^{-1} : \mathbb{P}^2 \dashrightarrow \mathbb{P}^2$ and that the composition of $\widehat{F}$ and $\widehat{F}^{-1}$ is the identity wherever it is defined. Such a rational map with rational inverse is said to be *birational*. A birational map of degree $d \geq 2$ necessarily has points of indeterminacy. For further information, see, for example, Fornæss and Sibony [1995a] or Fornæss [1996].

Attractors. Let $f : X \to X$ be a continuous self-map of a locally compact space, and let $N \subset X$ be a nonempty compact proper subset of X with the property that $f(N)$ is contained in the interior of N. It then follows that the intersection $A = \bigcap_k f^{\circ k}(N)$ is a compact f-invariant set, $A = f(A)$. Such an A will be called a *trapped attracting set*, with N as a *trapping neighborhood*. If A contains a dense orbit, then it will be called a *trapped attractor*.

This definition is particularly robust: If we replace f by some nearby map g, then there will be a corresponding attracting set for g which is contained in a small neighborhood of A.

In the case of a holomorphic map of a Riemann surface, it is not hard to see that a trapped attractor is necessarily a finite periodic orbit. The same statement is true for holomorphic maps of $\mathbb{C}^n$. (Compare Problem D-3.) However, for holomorphic maps of $\mathbb{P}^2$, more interesting examples can occur. (Compare Jonsson and Weickert [2000], Fornæss and Weickert [1999], Fornæss and Sibony [2001].)

For some purposes it is useful to broaden the definition so as to include less robust forms of attraction. As one example, the "attracting Herman ring" constructed in Lemma D.2 is certainly attracting in a quite strong sense: It attracts all orbits in the neighborhood U locally uniformly. However, it is not a trapped attractor. In fact, it will disappear under an arbitrarily small perturbation of f. The corresponding statement for an attracting Siegel disk is even easier to check, since an indifferent fixed point becomes repelling under arbitrarily small perturbations.

The following more general definition will allow this example, as well as other much stranger ones.

Definition. Let $f : M \to M$ be a continuous self-map of a smooth manifold, and let $A = f(A)$ be a compact invariant set. Define the *attracting basin* $\mathcal{B}(A)$ to be the set of all points $p \in M$ such that the distance of $f^{\circ k}(p)$ from A tends to zero as $k \to \infty$. The set A will be called a *measure theoretic attracting set* if its basin $\mathcal{B}(A)$ has positive measure. (Compare Milnor [1985].) Again it will be called an *attractor* if it also contains a dense orbit. Here both distance and measure can be defined in terms of some Riemannian metric on M, but the definition does not depend on the particular choice of metric.

With this definition, the possibilities become much more wild. For example, according to Bonifant, Dabija, and Milnor [in preparation], a holomorphic map of $\mathbb{P}^2$ can have *intermingled basins*. (Compare Alexander, Kan, Yorke, and You [1992].) More precisely:

> *There exists a family of degree 4 holomorphic maps of $\mathbb{P}^2$ with one attracting fixed point and two smooth measure theoretic attractors A_1 and A_2, where the three attracting basins are so thoroughly intermingled that any open neighborhood of any point of the Julia set intersects each of the three basins in a set of strictly positive measure.*

Concluding Problems

Problem D-1. A nonlinearizable germ. Let

$$F(x,y) \;=\; (\lambda x,\, \lambda^2 y + x^2) \qquad \text{with} \qquad \lambda \;\neq\; 0,\, 1\,.$$

Show that there is only one smooth F-invariant curve through the origin, namely $x = 0$. (Any smooth curve can be described locally by setting one of the coordinates equal to a power series in the other, for example $y = ax + bx^2 + \cdots$. Assume F-invariance and compare coefficients.) 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.

Problem D-2. Carathéodory distance. Let M be a complex manifold which can be embedded as a bounded subset of some $\mathbb{C}^n$. Define the *Carathéodory distance*, $0 \le d(p,q) = d_M(p,q) \le \infty$, between two points of M to be the supremum, over all holomorphic maps $f : M \to \mathbb{D}$, of the Poincaré distance between $f(p)$ and $f(q)$. **(1)** Show that $d(p,q) < \infty$ if and only p and q belong to the same connected component of M. **(2)** Show that the triangle inequality is satisfied and that

$$d_N(g(p),\, g(q)) \;\le\; d_M(p,q),$$

where g can be any holomorphic map from M to N. **(3)** If M is embedded in N with compact closure, show that there is a constant $c < 1$ so that

$$d_N(p,q) \;\le\; c\, d_M(p,q) \qquad \text{for all} \qquad p,\, q \in M.$$

Problem D-3. No nontrivial trapped attractors in $\mathbb{C}^n$. If A is a trapped attractor for a holomorphic map $F : \mathbb{C}^n \to \mathbb{C}^n$, find bounded neighborhoods U and V with $F(U) \subset V \subset \overline{V} \subset U$. Using the Carathéodory distance, conclude that

$$d_U\big(F(p),\, F(q)\big) \;\le\; c\, d_U(p,q)$$

for some uniform constant $c < 1$, and use this to prove that A must be an attracting periodic orbit.

(**Remark:** Similarly, any trapped attractor which is contained in the Fatou set for a holomorphic map of $\mathbb{P}^n$ must be a periodic orbit. In fact, according to Fornæss and Weickert [1999], any trapped attractor which is infinite must contain a nonconstant holomorphic image of $\mathbb{C}$. But according to Ueda [1994] each component of the Fatou set is Kobayashi hyperbolic, and hence cannot contain such an image.)

Appendix E. Branched Coverings and Orbifolds

This will be an outline of definitions and results due to Thurston. (See Douady and Hubbard [1993].) If

$$f(z) = w_0 + c(z - z_0)^n + (\text{higher terms}),$$

with $n \geq 1$ and $c \neq 0$, recall that the integer $n = n(z_0)$ is called the *local degree* of f at the point z_0. Thus $n(z_0) \geq 2$ if z_0 is a critical point, and $n(z_0) = 1$ otherwise. We will use *ramified point* as a synonym for critical value. Thus if $f(z_0) = w_0$ as above with local degree $n \geq 2$, then w_0 is a ramified point.

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 such 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 local degree $n(z)$ 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 preimage $f^{-1}(w)$. 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 on a discrete closed subset will be called a *Riemann surface orbifold.**

Definition. A branched covering $p : S' \to S$ will be called the *universal covering* for the Riemann surface orbifold (S, ν) if S' is simply connected, and if the local degree at every point $z \in S'$ is equal to $\nu(p(z))$.

Theorem E.1. *With the following exceptions, every Riemann surface orbifold (S, ν) has a universal covering $\tilde{S}_\nu$ which is necessarily a regular branched covering, and which is unique up to conformal isomorphism over S. The only exceptions are given by:*

(1) a surface $S \approx \hat{\mathbb{C}}$ with just one ramified point, or

(2) a surface $S \approx \hat{\mathbb{C}}$ with two ramified points for which $\nu(w_1) \neq \nu(w_2)$.

In these exceptional cases, no such universal covering exists.

Compare Problems E-3 and E-4. We will use the notation $\tilde{S}_\nu \to (S, \nu)$ for this universal branched covering. The associated group Γ of deck transformations is called the *fundamental group* $\pi_1(S, \nu)$ of the orbifold.

By definition, the *Euler characteristic* of an orbifold (S, ν) 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

$$\chi(S) - r \;<\; \chi(S, \nu) \;\leq\; \chi(S) - r/2$$

where r is the number of ramified points and where $\chi(S) \leq 2$. *As an example, if $\chi(S, \nu) \geq 0$, with at least one ramified point, then it follows that $\chi(S) > 0$, so the base surface S can only be $\mathbb{D}, \mathbb{C}$, or $\hat{\mathbb{C}}$, up to isomorphism.* Compare Remark E.6.

---

*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, and this much simpler definition can be used.

If there are infinitely many ramified points, note that we must set $\chi(S, \nu) = -\infty$. Similarly, if S is a connected surface which is not of finite type, then $\chi(S, \nu) = \chi(S) = -\infty$ by definition.

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 an *orbifold covering map* $(S', \mu) \to (S, \nu)$ if the identity

$$n(z)\mu(z) = \nu(f(z))$$

is satisfied for all $z \in S'$, where $n(z)$ is the local degree. Evidently a composition of two orbifold covering maps is again an orbifold covering map. As an example, the universal covering map $\tilde{S}_\nu \to (S, \nu)$, where $\tilde{S}_\nu$ is provided with the trivial ramification function $\tilde{\nu} \equiv 1$, is always an orbifold covering map.

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 (Theorem 7.2) takes the form*

$$\chi(S', \mu) = \chi(S, \nu)d.$$

In particular, if the universal covering of (S, ν) 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. Here we assume that there is at least one ramified point.

Lemma E.3. *Let (S, ν) be any Riemann surface orbifold which possesses a universal covering. Then:*

$\chi(S, \nu) > 0$ if and only if the fundamental group $\pi_1(S, \nu)$ is finite, necessarily of order $2/\chi(S,\nu)$, with $\tilde{S}_\nu \cong \hat{\mathbb{C}}$.

$\chi(S, \nu) = 0$ if and only if the fundamental group contains either $\mathbb{Z}$ or $\mathbb{Z} \oplus \mathbb{Z}$ as a subgroup of finite index.

$\chi(S, \nu) < 0$ if and only if the fundamental group contains a nonabelian free product $\mathbb{Z} * \mathbb{Z}$, and hence does not contain any abelian subgroup of finite index.

The Euler characteristic and the geometry of $\tilde{S}_\nu$ are related as follows.

Lemma E.4. *If S is a compact Riemann surface, then the Euler characteristic $\chi(S, \nu)$ is positive, negative, or zero according to whether the universal covering $\tilde{S}_\nu$ is conformally spherical, hyperbolic, or Euclidean.*

Remark. This lemma is closely related to the Gauss-Bonnet Theorem,

$$\iint K\,dA \;=\; 2\pi\,\chi(S,\,\nu),$$

which holds for any *orbifold metric* (§19) which is complete with finite area, and which is sufficiently well behaved near infinity in the noncompact case. Here K is the Gaussian curvature of $(S,\,\nu)$ and dA is its area element.

Example E.5. If $S=\widehat{\mathbb{C}}$ with four ramified points of index $\nu(w_j)=2$, then the torus T described in §7 (Example 3) provides a regular twofold branched covering. Its universal covering $\tilde{T}\cong\mathbb{C}$ can be identified with the universal covering of $(\widehat{\mathbb{C}},\,\nu)$. In this case, $\chi(\widehat{\mathbb{C}},\,\nu)=0$.

Remark E.6. The relatively few cases in which $\chi(S,\,\nu)\ge 0$ can be listed as follows. Note that a surface with $\chi(S)>0$ can only be a sphere, plane, or disk; while a surface with $\chi(S)=0$ must be a punctured plane, punctured disk, annulus, or torus. The notation $(\nu(w_1),\,\ldots,\,\nu(w_r))$ will be used for the list of ramification indices at all ramified points, for example with $\nu(w_1)\le\cdots\le\nu(w_r)$.

If $\chi(\widehat{\mathbb{C}},\,\nu)>0$ with $r>0$, then the ramification indices must be either (n,n) or $(2,2,n)$ for some $n\ge 2$, 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. (Compare Milnor [1975, p. 179].)

If $\chi(\widehat{\mathbb{C}},\,\nu)=0$, 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 $\mathbb{C}$ by squares, equilateral triangles, alternately colored equilateral triangles, and parallelograms, respectively. (For a more precise description, see Milnor [2004b].) 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 $\chi(\mathbb{C},\,\nu)$ or $\chi(\mathbb{D},\,\nu)$ is strictly positive, then we must have $r\le 1$, while if $\chi(\mathbb{C},\,\nu)$ or $\chi(\mathbb{D},\,\nu)$ is zero, then we must have $r=2$ with ramification indices $(2,2)$. This is the complete list.

Concluding Problems

Problem E-1. The complex plane with 2 ramified points. (1) If $S=\mathbb{C}$ with ramification function satisfying $\nu(1)=\nu(-1)=2$ and with no other ramified points, show that the map $z\mapsto\cos(2\pi z)$ provides a universal covering $\mathbb{C}\to(\mathbb{C},\,\nu)$. **(2)** Show that the Euler characteristic $\chi(\mathbb{C},\,\nu)$ is zero, and the fundamental group $\pi_1(S, \nu)$ consists of all transformations of the form $\gamma : z \mapsto n \pm z$ with $n \in \mathbf{Z}$.

Problem E-2. $\widehat{\mathbb{C}}$ with 3 ramified points. For $S = \widehat{\mathbb{C}}$ with three ramified points $\nu(0) = \nu(1) = \nu(\infty) = 2$, show that the rational map $\pi(z) = -4z^2/(z^2-1)^2$ provides a universal covering $\widehat{\mathbb{C}} \to (\widehat{\mathbb{C}}, \nu)$. Show that $\chi(\widehat{\mathbb{C}}, \nu) = 1/2$, that the degree is equal to $\chi(\widehat{\mathbb{C}})/\chi(\widehat{\mathbb{C}}, \nu) = 4$, and that the fundamental group consists of all transformations $\gamma : z \mapsto \pm z^{\pm 1}$.

Problem E-3. Bad orbifolds. For $\widehat{\mathbb{C}}$ with one ramified point, or with two ramified points with different ramification indices, show that there can be no universal covering surface. (For example, use Lemma E.2.)

Problem E-4. Existence of universal coverings. For an arbitrary Riemann surface orbifold (S, ν) (other than those in Problem E-3), construct a universal covering in several steps, as follows.

Case 1. Suppose that S is the plane $\mathbb{C}$ or the disk $\mathbb{D}$. Let $X \subset S$ be the set of ramified points. Choosing some base point $z_0 \in S \smallsetminus X$, note that the fundamental group $\pi_1(S \smallsetminus X, z_0)$ is a free group with one generator ℓ_x for each $x \in X$, represented by a loop encircling x. Let $N \subset \pi(S \smallsetminus X, z_0)$ be the normal subgroup generated by the powers $\ell(x)^{\nu(x)}$, and let S' be the covering space of $S \smallsetminus X$ with fundamental group $\pi(S') = N$. If $\mathbb{D}_\epsilon(x)$ is a small disk around x, show that the preimage of $\mathbb{D}_\epsilon(x) \smallsetminus \{x\}$ in S' is a union of disjoint punctured disks. Filling in all of these puncture points, show that we obtain the required universal covering.

Case 2. If S is Euclidean or hyperbolic, show that by first passing to the universal covering of S, we are reduced to Case 1.

Case 3. On the other hand, if $S = \widehat{\mathbb{C}}$ with at most three ramification points, show that the universal covering space can be constructed as in Remark E.6 (unless excluded by Problem E-3). The same is true if there are four ramification points, all with ramification index 2.

Case 4. $S = \widehat{\mathbb{C}}$ with four or more ramification points. For any three point subset $X_0 \subset X$, or for any four point set of type $\{2, 2, 2, 2\}$, we see by Case 3 that there exists a simply connected orbifold covering (S', ν') which is ramified only over X_0. Here each of the points of $X \smallsetminus X_0$ will be covered by many ramified points in S'. In most cases, we can choose X_0 so that S' is Euclidean or hyperbolic, and are therefore reduced to Case 2. However, in a few exceptional cases an extra step is needed. For example, if the ramification indices are $\{2, 2, 2, 3\}$ then we can first pass to the fourfold covering, ramifying only over the first three points. Then S' will have ramification indices $\{3, 3, 3, 3, 3\}$, and a further covering will reduce to the Euclidean case. The other exceptional cases are similar.

Appendix F. No Wandering Fatou Components

This appendix will outline a proof of the following. (Compare §16.)

Theorem F.1 (Sullivan Nonwandering Theorem). *Every Fatou component of a rational map is eventually periodic.*

The intuitive idea of the proof is the following. Let U be any Fatou component for f, that is, any connected component of $\widehat{\mathbb{C}} \setminus J(f)$. Suppose that we try to change the conformal structure on U. If f is to preserve this new conformal structure, then we must also change the conformal structure everywhere throughout the grand orbit of U in a compatible manner. If $f(U) = U$, then the condition that f preserves this structure imposes very strong restrictions. Similarly, if U is periodic or even eventually periodic, then there are very strong restrictions. However, if U were a ***wandering component***, that is, if the successive forward images

$$U, \;\; f(U), \;\; f^{\circ 2}(U), \;\; f^{\circ 3}(U), \;\; \cdots$$

were pairwise disjoint, then we could change the conformal structure within U in an arbitrary manner and then propagate this change throughout the entire grand orbit of U. As Sullivan realized, this would be too much of a good thing. He showed that it would yield an infinite-dimensional space of essentially different rational maps of the same degree. But this is patently impossible, since a rational map of given degree is completely determined by a finite number of complex parameters.

There are two key difficulties in carrying out this argument. The first is that conformal structures constructed in this way are usually discontinuous at every limit point of the grand orbit of U, so that it is not easy to make sense of them. However, this problem had been dealt with earlier in the pioneering work of Morrey [1938] and Ahlfors and Bers [1960]. The second difficulty that Sullivan faced was the need for some effective way of showing that he really did get too many distinct rational maps in this way, and not just many different ways of constructing the same rational maps. I will describe a way of dealing with this second problem by means of cross-ratios.

The Beltrami Equation. Before beginning this argument, we must explain the concept of a measurable conformal structure on an open set $U \subset \mathbb{C}$. Intuitively, a conformal structure at a point $z \in U$ can be prescribed by choosing some ellipse centered at the origin in the tangent space $T_zU \cong \mathbb{C}$. We are to think of this ellipse as a "circle" in the new conformal structure. In more technical language, a *conformal structure* at the point $z \in \mathbb{C}$ is determined by a *complex dilatation* $\mu(z) \in \mathbb{D}$. First consider the case where $\mu(z)$ is constant. Then the function $h(z) = z + \bar{z}\,\mu$ satisfies the *Beltrami differential equation*

$$\frac{\partial h}{\partial \bar{z}} = \mu(z)\frac{\partial h}{\partial z} \qquad (\mathrm{F}:1)$$

(named for Eugenio Beltrami, 1835–1900). Here the derivatives $\partial/\partial\bar{z}$ and $\partial/\partial z$ are to be defined by the formula.

$$\frac{\partial}{\partial \bar{z}} = \frac{1}{2}\left(\frac{\partial}{\partial x} + i\frac{\partial}{\partial y}\right), \qquad \frac{\partial}{\partial z} = \frac{1}{2}\left(\frac{\partial}{\partial x} - i\frac{\partial}{\partial y}\right),$$

where $z = x + iy$. (As an illustration, note that $f(z)$ satisfies the *Cauchy-Riemann equation* $\partial f/\partial\bar{z} = 0$ if and only if it is holomorphic, and that $\partial f/\partial z$ is then the usual holomorphic derivative.)

If h satisfies (F : 1) with constant $\mu \in \mathbb{D}$, then a round circle $|h| =$ constant in the h-plane corresponds to an ellipse $|z + \bar{z}\mu| =$ constant in the z-plane, with direction of the major axis controlled by the argument of μ and with eccentricity controlled by $|\mu|$. If $|\mu| = r < 1$, then the ratio of major axis to minor axis is equal to $(1+r)/(1-r)$, which tends to infinity as $r \to 1$.

More generally, if the function $\mu(z)$ is real analytic, then Gauss, in his construction of "isothermal coordinates," showed that an equation equivalent to (F : 1) always has local solutions. Morrey extended this to the case where $\mu(z)$ is measurable, with

$$|\mu(z)| < \text{constant} < 1 \qquad (\mathrm{F}:2)$$

almost everywhere, constructing a solution $z \mapsto h(z)$ which maps a region in the z plane homeomorphically onto a region in the h plane. Furthermore, if h_1 and h_2 are two distinct solutions, he showed that the composition $h_2 \circ h_1^{-1}$ is holomorphic.

Here some explanation is needed, since we are considering a differential equation involving nondifferentiable functions. For any open set $U \subset \mathbb{C}$ let $L^1(U)$ be the vector space consisting of all measurable functions $\phi : U \to \mathbb{C}$ with

$$\int\!\!\int_U |\phi(x+iy)|\,dx\,dy < \infty$$

(where we identify two functions which agree almost everywhere). We will also need the vector space of *test functions* on U, consisting of all C^∞ functions $\tau : U \to \mathbb{C}$ which vanish outside of some compact subset of U.

Definition. A continuous function $h : U \to \mathbb{C}$ has *distributional derivatives in* L^1 if there are complex valued functions h_z and $h_{\overline{z}}$ defined almost everywhere in U and belonging to $L^1(U)$ so that

$$\int\!\!\int_U \Big(h_z(z)\,\tau(z) \;+\; h(z)\,\partial\tau/\partial z\Big)\,dx\,dy \;=\; 0 \qquad \text{(F : 3)}$$

for every such test function τ, with an analogous equation for $h_{\overline{z}}$. (Note that we can change h_z and $h_{\overline{z}}$ on a set of Lebesgue measure zero without affecting (F : 3).) The Beltrami equation for h now requires that

$$h_{\overline{z}}(z) \;=\; \mu(z)\,h_z(z)$$

for almost every $z \in U$. This makes sense, since the pointwise product of an L^1 function and a bounded measurable function is again in L^1. By definition, any continuous one-to-one solution h is called a *quasiconformal mapping* on U, with complex dilatation μ.

More generally, we can consider such a measurable conformal structure on a Riemann surface S. However, it is no longer described by a complex-valued function, but rather by a section of a real analytic $\mathbb{D}$-bundle which is canonically associated with S. Given a local coordinate z on an open set U, we can still describe the conformal structure on U by a dilatation function $\mu : U \to \mathbb{D}$, but on the overlap between two local coordinates z and z' a brief computation shows that the equation

$$\mu'(z') \;=\; \mu(z)\,\frac{\partial z'}{\partial z}\Big/\frac{\partial \overline{z}'}{\partial \overline{z}}$$

must be satisfied in order to make sense of this structure globally.* Note that $|\mu'(z')| = |\mu(z)|$, so that condition (F : 2) is independent of the choice of coordinate system. If this conformal structure is measurable and satisfies (F : 2) everywhere, then the local solutions h form the atlas of local conformal coordinates for a new Riemann surface S_μ which is topologically identical to S, but conformally (and even differentiably) quite different. In the special case where S is the Riemann sphere, it follows from the Uniformization Theorem that S_μ is conformally equivalent to the Riemann sphere. In particular, there is a unique conformal isomorphism $h : S \to S_\mu$ which fixes the points 0, 1, and ∞. If we remember that S_μ is identical to

*In more geometric language, a *Beltrami differential* at a point x of a Riemann surface can be described as an additive homomorphism from the tangent space T_x to itself which is *antilinear*, $\mu_x(\lambda t) = \bar{\lambda}\,\mu_x(t)$, and which multiplies the length of any vector $t \in T_x$ by a constant $|\mu_x| < 1$. In particular, if S is an open subset of $\mathbb{C}$ so that $T_c \cong \mathbb{C}$, then μ_x will have the form $\mu_x(t) = \mu\,\bar{t}$ with $|\mu_x| = |\mu| < 1$.

$S = \widehat{\mathbb{C}}$ as a topological space, then we can also describe $h = h_\mu$ as a quasiconformal homeomorphism from $\widehat{\mathbb{C}}$ to itself (or briefly a *qc-homeomorphism*) with complex dilatation $\mu(z)$.

We can also study the dependence of h_μ on the dilatation μ. For each fixed z_0, Ahlfors and Bers [1960] showed that the correspondence $\mu \mapsto h_\mu(z_0)$ defines a differentiable function from the appropriate space of dilation functions to the Riemann sphere. For further information, see, for example, Ahlfors [1987], Carleson and Gamelin [1993], Lehto [1987], Lehto and Virtanen [1973], or Douady and Buff [2000].

Some Conformal Structures on the Unit Disk. In order to carry out Sullivan's proof, we must construct a large family of essentially distinct conformal structures on the open disk $\mathbb{D}$. This can be done as follows. Fix three base points on the unit circle $\partial\mathbb{D}$, for example ± 1 and i. Let G be the group of all C^∞ diffeomorphisms of $\partial\mathbb{D}$ which fix these three points. I will use the notation $E(t) = e^{2\pi i t}$ for the standard diffeomorphism from $\mathbb{R}/\mathbb{Z}$ onto $\partial\mathbb{D}$. Writing each group element as

$$g \,:\, E(t) \;\mapsto\; E(t + v(t)),$$

we can identify G with the convex set consisting of all C^∞ functions $v : \mathbb{R}/\mathbb{Z} \to \mathbb{R}$ which vanish at three specified points, and such that the correspondence $t \mapsto t + v(t)$ has derivative $1 + v'(t) > 0$ everywhere. *This group G has been constructed so that no g other than the identity map can be extended to a conformal automorphism of the closed disk $\overline{\mathbb{D}}$.* To see this, note that for each nonidentity $g \in G$ we can find four distinct points of $\partial\mathbb{D}$ which map to four points with different cross-ratio, which would be impossible under a conformal automorphism. On the other hand, each $g \in G$ extends to a diffeomorphism $\widehat{g}$ of $\overline{\mathbb{D}}$, as follows. If $\eta : [0,1] \to [0,1]$ is some smooth monotone function with

$$\eta[0, 1/3] \;=\; 0, \qquad \eta[2/3, 1] \;=\; 1,$$

then g extends to the diffeomorphism

$$\widehat{g}\Big(r\,E(t)\Big) \;=\; r\,E\Big(t + \eta(r)v(t)\Big). \tag{F : 4}$$

Evidently this extension depends smoothly on g.

Wandering Components. Suppose that some rational function f has a wandering Fatou component U. Replacing U by some iterated forward image if necessary, we may assume that there are no critical points in the forward images $f^{\circ n}(U)$.

> **Lemma F.2 (Baker).** *If no forward image $f^{\circ n}(U)$, $n \geq 0$, contains a critical point, then U must be simply connected.*

Proof. Since no $U_n = f^{\circ n}(U)$ contains a critical point, it follows that each U_n maps onto U_{n+1} by a covering map. In particular, it follows that each fundamental group $\pi_1(U_n)$ maps injectively into $\pi_1(U_{n+1})$.

After conjugating by a Möbius automorphism of $\widehat{\mathbb{C}}$, we may assume that the point at infinity is in U, so that all other Fatou components lie in the bounded region $\widehat{\mathbb{C}} \smallsetminus U \subset \mathbb{C}$. Let L_0 be an arbitrary simple closed curve contained in U. Since U is contained in the Fatou set, the collection of iterates $f^{\circ n}$ restricted to U forms a normal family. The area of $f^{\circ n}(U)$ must clearly tend to zero as $n \to \infty$, and it follows that any convergent sequence of iterates must converge locally uniformly to a constant map. Therefore the diameter of the compact set $L_n = f^{\circ n}(L_0)$ must converge to zero as $n \to \infty$. Let K_n be the union of L_n and all bounded components of its complement. Then the diameter of K_n also tends to zero as $n \to \infty$. It follows that K_n must map into K_{n+1} for large n. For otherwise, if some point of $K_n \smallsetminus L_n$ mapped outside of K_{n+1}, then the image $f(K_n \smallsetminus L_n)$, being an open set for which the boundary is contained in L_{n+1}, would have to cover the entire complement of K_{n+1}. But this is impossible when K_n and K_{n+1} are sufficiently small. It follows that K_n is contained in the Fatou set when n is large, and hence is contained in U_n. But L_n is clearly contractible within K_n, and hence within U_n. Since $\pi_1(U)$ injects into $\pi_1(U_n)$, this proves that U is simply connected, as required. $\square$

Proof of Theorem F.1. The proof of Sullivan's Theorem begins as follows. Choose some conformal isomorphism ϕ from U to the unit disk $\mathbb{D}$. With $\widehat{g}$ as in (F : 4), we can pull back the conformal structure of $\mathbb{D}$ under the composition $\widehat{g} \circ \phi$ and then use f to transport this conformal structure over the entire grand orbit of U. (There may be isolated points in this grand orbit which are precritical. The induced conformal structure is not defined at such points, but this will not matter.) For points which are not in the grand orbit of U, we simply use the usual conformal structure. Thus we have described a measurable conformal structure almost everywhere on $\widehat{\mathbb{C}}$. The condition $|\mu| \le \text{constant} < 1$ is easily verified. Integrating the resulting Beltrami equation, this yields a family of qc-homeomorphisms h_g, normalized so as to fix three points of the Riemann sphere, and a family of maps f_g so that the following diagram is commutative,

$$\begin{array}{ccccccc}
\mathbb{D} & \xleftarrow{\phi} & U & \subset & \widehat{\mathbb{C}} & \xrightarrow{f} & \widehat{\mathbb{C}} \\
\downarrow \widehat{g} & & \downarrow h_g|_U & & \downarrow h_g & & \downarrow h_g \\
\mathbb{D} & \xleftarrow{\phi_g} & U_g & \subset & \widehat{\mathbb{C}} & \xrightarrow{f_g} & \widehat{\mathbb{C}},
\end{array}$$

where horizontal arrows represent holomorphic maps and vertical arrows represent quasiconformal maps. Here U_g is defined to be the image of U under h_g, and the maps on the bottom row are defined in such a way that the diagram is commutative. Since the conformal structures on $\widehat{\mathbb{C}}$ were constructed so as to be invariant under f, it follows that each f_g is holomorphic, and hence is a rational map of the same degree d.

We must show that the rational map f_g depends smoothly on g. Note that a rational map $p(z)/q(z)$ of degree d is uniquely determined by its values on $2d+1$ distinct points, for a second degree d map $P(z)/Q(z)$ takes the same value as $p(z)/q(z)$ at z_i if and only if the polynomial $p(z)Q(z) - P(z)q(z)$ vanishes at z_i. Furthermore, if such a polynomial equation of degree $2d$ has $2d+1$ distinct solutions, then it must be identically zero. In fact the coefficients of $p(z)$ and $q(z)$, suitably normalized, can be obtained by solving linear equation, and hence depend smoothly on the given data. Now consider the points $h_g(j)$ for $1 \le j \le 2d+1$. Since f_g maps each such point to $h_g(f(j))$, and since both $h_g(j)$ and $h_g(f(j))$ depend smoothly on g, it follows that f_g depends smoothly on g.

Let Rat_d be the $(2d+1)$dimensional manifold consisting of all rational maps of degree d. Thus we have a smooth mapping $g \mapsto f_g$ from the infinite dimensional space G to the finite dimensional manifold Rat_d. It is then not hard to construct a smooth nonconstant path in G which maps to a point in Rat_d. To avoid working with infinitely many dimensions, we can first choose some $(2d+2)$-dimensional submanifold $M_0 \subset G$. Choose some point $g_0 \in M_0$ where the rank of the first derivative of the correspondence $g \mapsto f_g$ from M_0 to the space Rat_d of rational maps takes its maximal value $r \le 2d+1$. Then a neighborhood N of g_0 maps smoothly onto an r-dimensional submanifold $M_1 \subset \mathrm{Rat}_d$. Taking the preimage in N of a regular value in M_1, we obtain a submanifold $M_2 \subset N$ of dimension $2d+2-r \ge 1$, with the property that the corresponding maps f_g are all the same. Any nonconstant path $t \mapsto g(t)$ in M_2 will then have the required property of mapping to a point in Rat_d.

Now consider the one-parameter family of qc-homeomorphisms $h'_t = h_{g(t)} \circ h^{-1}_{g(0)}$ which conjugate $f_{g(0)}$ to $f_{g(t)}$. Since $f_{g(t)} = f_{g(0)}$, this means that each h'_t must commute with $f_{g(0)}$. It follows that each h'_t must restrict to the identity map on the Julia set $J(f_{g(0)})$, since the periodic points of $f_{g(0)}$ cannot move under a deformation which commutes with it.

This leads to a contradiction as follows. Suppose first, to simplify the discussion, that U is bounded by a Jordan curve. For every four points of ∂U we can define the *cross-ratio relative to $\overline{U}$* by choosing a conformal isomorphism $U \to \mathbb{D}$, extending continuously to the boundary (see Theorem 17.16), and then taking the usual cross-ratio in $\mathbb{D}$. But if $g(t) \ne g(0)$

then we can choose four points of the unit circle whose cross-ratio definitely changes under the composition $g(t) \circ g(0)^{-1}$. Hence h'_t cannot fix the corresponding four points of $\partial U_{g(0)}$, contradicting our previous statement. This proves Theorem F.1 under the hypothesis that ∂U is a Jordan curve.

If ∂U is not a Jordan curve, then we must elaborate this argument using ideas from §17. Recall that the *Carathéodory compactification* $\widehat{U}$ consists of U together with a circle of ideal points which are called *prime ends*. These are constructed using only the topology of the pair $(\overline{U}, \partial U)$. Recall also that any Riemann map $U \to \mathbb{D}$ extends to a homeomorphism $\widehat{U} \to \overline{\mathbb{D}}$. Thus, if we are given four distinct prime ends in $\partial\widehat{U}$, it follows that the cross-ratio of the corresponding points in $\partial\mathbb{D}$ is well defined, independent of the particular choice of Riemann map. To complete the proof, we will need the following supplementary statement.

Lemma F.3. *If an orientation-preserving homeomorphism h from the pair $(\overline{U}, \partial U)$ to itself restricts to the identity map on ∂U, then h maps each prime end of $(\overline{U}, \partial U)$ to itself.*

(The example of the complex conjugation map on the pair $(\widehat{\mathbb{C}}, [0,1])$ shows that the orientation condition is necessary.) To prove Lemma F.3, recall from §17 that a prime end is determined by a fundamental chain $\{A_j\}$ of transverse arcs, with associated neighborhoods $N(A_1) \supset N(A_2) \supset \cdots$. If the corresponding neighborhoods $h(N(A_j))$ were disjoint from the $N(A_j)$, then each $\overline{N(A_j)} \cup h(\overline{N(A_j)})$ would be a region bounded by a Jordan curve. The homeomorphism h must preserve orientation on this region, yet reverse orientation on its boundary, which is impossible. □

Using this result, the proof of Theorem F.1 goes through just as in the Jordan curve case. □

Appendix G. Parameter Spaces

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

$$f(z) = z^2 + c \qquad (G:1)$$

by an affine change of coordinates. (A closely related† normal form is

$$w \mapsto \lambda w(1-w) \qquad (G:2)$$

with a preferred fixed point of multiplier λ at the origin.) 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. (See Branner [1989] for further information.)

The first crude pictures of this type were made by Brooks and Matelski [1981], as part of a study of Kleinian groups. They used the normal form (G : 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 $\mathcal{H}$ for this set, since its points represent hyperbolic maps. At about the same time, Hubbard (unpublished) made much better pictures of a quite different parameter space arising from Newton's method for cubic equations. Two years later Mandelbrot [1980], perhaps inspired by Hubbard, made corresponding pictures for quadratic polynomials, using the normal form (G : 2), which is essentially the same as that of Figure 29 (p. 145), and also using a variant of (G : 1). In order to avoid confusion, let me translate all of Mandelbrot's definitions to the normal form (G : 1). He introduced two different sets, which I will call M and M'. (Mandelbrot did not give these sets different names, since he believed that they were identical.) By definition, a parameter value c belongs to M' if the corresponding filled Julia set contains an interior point, and belongs to M if its filled Julia set contains the critical point $z = 0$.

* Compare Figures 29, 30, 49 on pages 145, 151, 267.

† These are conjugate under the change of coordinates $z = -\lambda w + \lambda/2$ provided that $\lambda \neq 0$, with $4c = \lambda(2-\lambda)$.

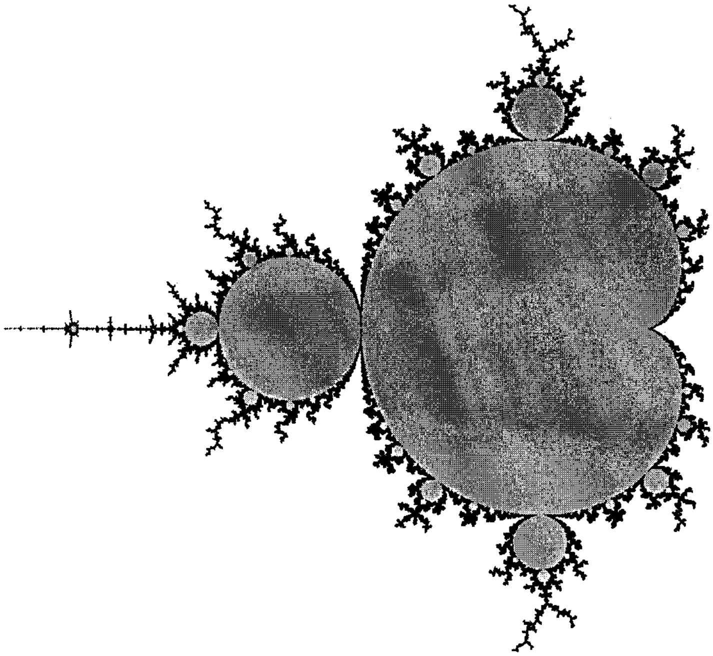


Figure 49. The Mandelbrot set M. The boundary (or bifurcation locus) is shown in black, and the interior, which is conjecturally the same as the locus of hyperbolic maps in M, is shown in grey.

(According to Theorem 9.5, this is equivalent to requiring that the Julia set be connected.) The Brooks-Matelski set $\mathcal{H}$ satisfies $\mathcal{H} \subset M' \subset M$. Mandelbrot made somewhat better computer pictures, which seemed to show a number of isolated "islands." Therefore, he conjectured that M' [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 a smaller set $M'' \subset M'$ which he believed to be the largest connected component of M'. This set M'' consists of a central cardioid with some boundary points included, together with countably many smaller nearly-round (satellite) disks which are attached 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 [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, Douady and Hubbard 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 $\mathcal{H} \subset M' \subset M$ in parameter space. They are certainly different since $\mathcal{H}$ is open, M is compact, and M' is neither. In fact, M' 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 hyperbolic set $\mathcal{H}$ is equal to the interior of M and that M is equal to the closure of $\mathcal{H}$. (Douady and Hubbard have shown that these conjectures are true if the set M is locally connected. The work of Yoccoz lends support to the belief that M may indeed be locally connected. Compare Hubbard [1993].)

As far as practical computing is concerned, it should be noted that we can test whether a given point c belongs to M by following the orbit of zero under $z \mapsto z^2 + c$ to see whether it remains bounded or diverges rapidly to infinity. Similarly, we can test whether $c \in \mathcal{H}$ by seeing whether the critical orbit converges to a finite periodic attractor (compare Theorem 8.6). However, such tests can never prove conclusively that a point lies in $M \smallsetminus \mathcal{H}$ since there is no effective bound for the number of iterations which may be needed. For example, as noted in Remark 8.8, the point $c = -1.5$ certainly belongs to M and conjecturally belongs to $M \smallsetminus \mathcal{H}$; but there is no known way to verify such a statement. Furthermore, it is particularly difficult to decide whether a given point of M corresponds to a polynomial with a Siegel disk or Cremer point. For a more detailed discussion of computational difficulties, see Appendix H.

Another important development came with the work of Mañé, Sad, and Sullivan [1983], and independently Lyubich [1983a], showing that for any holomorphically parametrized family of rational maps there is a dense open subset for which the topological structure of the Julia set $J(f)$ remains stable under deformation of f. (Compare the discussion of Theorem 19.1.)

The study of parameter space for higher degree polynomials began some five years later with the work of Branner and Hubbard [1988, 1992]. 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 Problem 9-e.) In particular, this set is compact and connected. The corresponding result for polynomials of higher degree has been obtained by Lavaurs (unpublished). Further information about polynomial parameter spaces is given in Milnor [1992a, 1992b].

Parameter spaces for rational maps have been studied, for example, by Rees [1990,1992, 1995], Milnor [1993, 2000c], Epstein [2000], and DeMarco [2003, 2004]. The situation is more awkward than for polynomials, since there is no obvious preferred normal form for rational maps. The space Rat_d consisting of all rational maps of degree d is a well behaved complex manifold. In the degree 2 case, the *moduli space*, consisting of quadratic rational maps up to holomorphic conjugation, is also a smooth manifold, canonically diffeomorphic to $\mathbb{C}^2$. (Milnor [1993].) However, for degrees $d > 2$ the corresponding moduli space has singularities (Problem G-3).

Concluding Problems

Problem G-1. Polynomial moduli space. **(1)** Show that every polynomial map of degree $d \ge 2$ is conjugate, under an affine change of coordinates, to one in the "Fatou normal form"

$$f(z) = z^d + a_{d-2} z^{d-2} + \cdots + a_1 z + a_0 .$$

(2) Let $P(d) \cong \mathbb{C}^{d-1}$ be the space of all such maps. Show that the cyclic group $\mathcal{C}_{d-1}$ 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 $P(d)/\mathcal{C}_{d-1}$ can be identified with the *moduli space* of degree d polynomials up to affine conjugation. If $d \ge 4$, this moduli space is not a manifold. As an example, for $d = 4$, let U be a small neighborhood of $f_0(z) = z^4$ in moduli space. **(3)** Show that $U \setminus \{f_0\}$ does not have the mod 3 homology of a 5-dimensional sphere, and conclude that this moduli space is not a manifold.

Problem G-2. Quadratic rational maps. Let λ, μ, ν be the multipliers at the three fixed points of a quadratic rational map f, taking $\lambda = \mu = 1$ in the case of a double fixed point or $\lambda = \mu = \nu = 1$ in the case of a triple fixed point. **(1)** Show that the holomorphic conjugacy class of f is uniquely determined by the unordered triple $\{\lambda, \mu, \nu\}$, or equivalently by the three elementary symmetric functions

$$\sigma_1 = \lambda + \mu + \nu\,, \quad \sigma_2 = \lambda\mu + \lambda\nu + \mu\nu\,, \quad \sigma_3 = \lambda\mu\nu\,.$$

(Compare Problem 12-b for the case where there are at least two distinct fixed points.) **(2)** Show that these invariants are subject only to the relation given by the rational fixed point formula (Theorem 12.4), or equivalently to the relation

$$\sigma_3 = \sigma_1 - 2\,.$$

Conclude that the moduli space of holomorphic conjugacy classes can be identified with the coordinate space $\mathbb{C}^2$, using σ_1 and σ_2 as coordinates. (Compare Milnor [1993].) **(3)** Show however that it is not always possible to choose a smooth two-parameter family of maps which map bijectively onto a given region in moduli space. (In other words, this moduli space has an essential orbifold structure.) For example, consider a map of the form $f_\mu(z) = (z + z^{-1})/\mu$. Since f_μ is an odd function, it has a symmetry $z \mapsto -z$ which interchanges the two finite fixed points. We can embed f_μ into a two-parameter family of maps $f_\mu(z) + c$. However, the map which assigns the conjugacy class of $f_\mu + c$ to each pair (μ, c) in parameter space has local degree 2, since $f_\mu + c$ is linearly conjugate to $f_\mu - c$.

Problem G-3. Cubic rational maps. Consider a cubic rational map of the form

$$f_0(z) = \frac{z^3 + \mu_0 z}{\nu_0 z^2 + 1},$$

with fixed points of multiplier μ_0 and ν_0 at zero and infinity. Show that the moduli space of conjugacy classes of cubic rational maps is singular near the class of f_0. In fact, for generic choice of μ_0 and ν_0, show that any nearby cubic map has a normal form

$$f(z) = \frac{z^3 + \epsilon z^2 + \mu}{\nu z^2 + \delta z + 1}$$

with $\mu \approx \mu_0$ and $\nu \approx \nu_0$ which is unique up to the involution

$$(\epsilon, \delta) \mapsto (-\epsilon, -\delta).$$

Show that the quotient under this involution is not a manifold.

Appendix H. Computer Graphics and Effective Computation

In order to make a computer picture of some complicated compact subset of $\mathbb{C}$, for example a Julia set or filled 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" on the computer screen. Each pixel represents a small square in the complex plane, and the color which is assigned must tell us something about the relationship between this square and the specified subset.

One method which works for arbitrary rational maps involves iterating the inverse map f^{-1} many times, starting at a repelling point, taking all possible branches, and plotting all of the resulting points; that is, coloring the pixel which contains each such point. (Compare Corollary 4.13.) This method is especially convenient in the degree 2 case since there are fewer inverse branches to follow, and since it is easy to solve quadratic equations. This method yields a good picture of what we might call the "outer" parts of the Julia set, but shows very little detail in the "inner" parts.* In the polynomial case, if we think of the electrostatic field produced by an electric charge on $J(f)$, this method will emphasize only the highly charged parts of the Julia set, or equivalently the points where most external rays land.

In the polynomial case, a slower but much better procedure for plotting the filled Julia set involves iterating the map f for some large number of times (perhaps 50 to 50,000), 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. This method can be refined by computing not only the value of the nth 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. (See Appendix G.) In this case one takes the quadratic map corresponding to the midpoint of the square and follows the orbit of its critical point.

Limitations. In order to understand some of the limitations of this method, consider the situation near a fixed point $z_0 = f(z_0)$ in the Julia set. First suppose that z_0 is repelling, for example, with multiplier satisfying $|\lambda| \approx 2$. If we start at a point z at distance $1/1000$ from z_0, then the

*In the limit, the points which are plotted are evenly distributed with respect to the measure of maximal entropy which was described by Lyubich [1983b], or by Freire, Lopes, and Mañé [1983] together with Mañé [1983].

distance from z_0 will roughly double with each iteration. Hence, after perhaps 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. (See for example Figure 12, p. 87).

Now suppose that we try to construct a picture for $z \mapsto z + z^4$ by the same method. Suppose that we start with a point $z = \epsilon > 0$, with orbit escaping monotonely to infinity. Taking $\epsilon = 1/1000$ and examining the proof of Lemma 10.1, we see that the associated coordinate $w = -1/3z^3$ is equal to $-1/3\epsilon^3 \approx -3.3 \times 10^8$. *Since w increases by approximately $+1$ under each iteration, we would have to follow such an orbit for more than 300,000,000 iterations in order to escape from a small neighborhood of $z = 0$.* The result, if we used fewer iterations, would be a false picture which shows the ϵ-neighborhood of the origin to be in the filled Julia set.

This difficulty was eliminated in Figures 19 and 21 (pp. 106, 109) by a special computer program which extrapolated iterates of f near the parabolic point in order to make a more accurate picture. Similarly, Figures 26, 28, and 32 (pp. 127, 132, 164) were made with a special-purpose program. But in general, no such convenient trick is known. For the fixed points of Cremer type, the situation seems particularly bad. As far as I know, no useful picture of the Julia set near such a point has ever been produced, either by computer or by theory.

Distance estimates. 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. In particular, if the filled Julia set has measure zero, then all of the center points of our pixels will quite likely 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. Distance estimates are also important when plotting the Mandelbrot set M, which contains not only large regions but also very fine filaments. Indeed, it was precisely the difficulty of seeing such filaments which led to Mandelbrot's initial belief that M has many components.

Here is an example of how first derivatives can be used to make distance estimates. (Compare Fisher [1988], Milnor [1989, Lemma 5.6], or Peitgen [1988].) Let $f : \hat{\mathbb{C}} \to \hat{\mathbb{C}}$ be a rational map with a superattractive fixed point of local degree n at the origin. Assume that the basin of attraction U for this fixed point is connected, simply connected, and contains no other critical point. Then the Böttcher coordinate of §9 can be defined throughout U and yields a conformal isomorphism $\phi : U \to \mathbb{D}$ with $\phi(f(z)) = \phi(z)^n$.

Define the *Green's function* $G : U \smallsetminus \{0\} \to \mathbb{R}$ by

$$G(z) \;=\; -\log|\phi(z)| \;>\; 0.$$

(Compare §§9 and 18.) Let G' denote the gradient vector of G. Then:

(1) *the function* G *and the norm* $\|G'\| = |\phi'(z)/\phi(z)|$ *can be effectively computed for* $z \in U$, *provided that the orbit of* z *comes close to zero within a reasonable number of iterations.*

(2) *The distance of* z *from the boundary of* U *can be computed, up to a factor of* 2, *from a knowledge of* G *and* $\|G'\|$.

In fact, for any orbit $z_0 \mapsto z_1 \mapsto \cdots$ in U, 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

$$\|G'(z_0)\| = \lim_{k\to\infty} |dz_k/dz_0|/(n^k|z_k|).$$

In both cases, the successive terms can easily be computed inductively, and we obtain good approximations by iterating until $|z_k|$ is small. However, if many iterations do not yield any small z_k, then we do not obtain any definite information (though we can set $G = 0$ in the *hope* that $z_0 \notin U$).

The distance between z and $\widehat{\mathbb{C}} \smallsetminus U$ can now be estimated quite precisely as follows. Setting $\phi(z) = w$, a brief computation shows that the Poincaré metric on U can be written as

$$\frac{2|dw|}{1-|w|^2} \;=\; \frac{2|\phi'(z)dz|}{1-|\phi(z)|^2} \;=\; \frac{\|G'(z)dz\|}{\sinh G(z)}.$$

The following is an immediate consequence of Corollary A.8 to the Quarter Theorem.

Corollary H.1. *With* U *and* G *as above, if* $U \subset \mathbb{C}$, *then the distance between a point* $z \in U \smallsetminus \{0\}$ *and the complement of* U *is equal to* $\sinh(G)/\|G'\|$ *up to a factor of* 2.

If z is very close to ∂U, then G is small and this distance estimate is very close to the ratio $G/\|G'\|$ (equal to the step size which would be prescribed if we tried to solve the equation $G(z) = 0$ by Newton's method).

Now consider a polynomial map with connected Julia set. Conjugating by the inversion map $z \mapsto 1/z$, we reduce to the case above and conclude easily that the distance from z to the filled Julia set $K(f)$ is asymptotically equal to a corresponding ratio $G/\|G'\|$, up to a factor of 2, as $G \to 0$, with G as in Definition 9.6. If the orbit of z escapes to infinity expeditiously, this yields good distance estimates. However, suppose we have iterated

1000 or 100,000 times and the orbit stays persistently bounded. Perhaps this means that z is in the filled Julia set or extremely close to it. But it could be that z is quite far from the filled Julia set, but is separated from infinity within $\mathbb{C} \smallsetminus K(f)$ by an extremely narrow neck. *If the orbit does not escape within a reasonable number of iterates, then we do not have enough information to make any sharp estimate.*

Effective Computability. Recent work of Braverman and Yampolsky shows that there exist polynomial Julia sets which in principle can never be effectively computed. The precise statements are somewhat subtle and need explanation.

Following Turing [1936–37], a real or complex number z is called *computable* if there is a finite algorithm (formally implemented as a *Turing Machine*) which, given some integer $n > 0$ as input, will compute some 2^{-n}-approximation to z by a finite number of discrete steps. But what should we mean by "computability" for a *set* of real or complex numbers, for example, a Julia set? We want a formal theory which supports our empirical experience that some Julia sets are easy to compute and some are very hard.

It would not be reasonable to ask for a finite Turing Machine which can plot the set with arbitrary specified accuracy, since there are uncountably many well-behaved Julia sets but only countably many Turing Machines. What we need is a relative formulation of the problem, in terms of a finite machine or algorithm which can plot the set to any specified accuracy, *provided* that it is given the coefficients to whatever accuracy may be required. By definition, a theoretical device which can input the coefficients to any required finite accuracy is called an *oracle*; and a modified Turing Machine which can accept such oracle input is called an *Oracle Turing Machine*. The Turing Machine part of such an Oracle Turing Machine is physically realizable, except for the requirement of unlimited memory; but, except in countably many cases, the oracle part is not physically realizable. (However, such an "impossible" oracle could consist of something as simple as an infinite sequence of digits such that our Oracle Turing Machine is allowed to read any finite initial segment.)

We also need to ask what it means to plot a set to some specified accuracy. Here the concept of Hausdorff distance is the key. If X and Y are nonvacuous compact subsets of $\mathbb{C}$, then the *Hausdorff distance* $\mathbf{d}(X, Y)$ is defined to be the smallest number $r \geq 0$ such that each of these two sets is contained in the closed neighborhood consisting of all points at distance $\leq r$ from the other.

Combining these two ideas, we define a compact set, depending on finitely many parameters, to be ***computable**** if there exists an Oracle Turing Machine, which given a positive integer n as input and given an oracle which can compute the parameters to any specified accuracy, produces a finite description of an explicit compact set which has Hausdorff distance at most 2^{-n} from the required set. For example, this new set might be a finite union of pixels in some sufficiently fine square grid.

Theorem H.2 (Braverman [2004] and Rettinger [2004]). *If f is any hyperbolic rational function, then there exists an Oracle Turing Machine which computes $J(f)$ with accuracy 2^{-n} in a number of steps bounded by a polynomial function of n.*

This is extended to maps with parabolic fixed points in Braverman [2005]. Another result covers much harder cases, but with no time estimate:

Theorem H.3 (Binder, Braverman, and Yampolsky [2005]). *If a polynomial f has filled Julia set with no interior, then there exists an Oracle Turing Machine which computes $J(f) = K(f)$.*

In particular, this proves the existence of polynomial Julia sets with Cremer point which are computable (although the time needed to produce a useful picture might well be measured in geological ages). Here is an intuitive proof of Theorem H.3: For any rational function f with repelling fixed point p, the finite sets $A_n = f^{-n}\{p\}$ form a family $A_0 \subset A_1 \subset A_2 \subset \cdots$ of subsets, which converge to J in the Hausdorff metric by Corollary 4.13. In the polynomial case, if B_0 is a closed disk large enough so that $f^{-1}(B_0) \subset B_0$, then the compact sets $B_n = f^{-n}(B_0)$ form a decreasing sequence $B_0 \supset B_1 \supset B_2 \cdots$ with intersection K. Therefore, $\{B_n\}$ converges to K in the Hausdorff metric. Now if $J = K$, it follows that $\mathbf{d}(A_n, B_n)$ tends to zero as $n \to \infty$. Choosing n with $\mathbf{d}(A_n, B_n) < \epsilon$, it follows that both A_n and B_n are ϵ-aproximations to the required set $J = K$. (To make this discussion into an actual proof, we would have to provide algorithms for

*In the case of bounded *open* sets, a very different (measure theoretic) definition of computability has been given by Chou and Ko [1995]. The best known concept of computability for closed sets, due to Blum, Shub, and Smale [1989], is not suitable for our purpose, since even Julia sets which are easy to plot in practice are usually uncomputable in their setup. They posit a (physically impossible) machine which can carry out arithmetic operations and comparisons with precise, infinite precision real numbers, but require this machine to decide in finitely many steps whether or not some specified point belongs to the set. As an illustrative example, the graph of the function $y = \sin(x)$ is uncomputable according to their definition.

computing these sets A_n and B_n to any specified accuracy by an Oracle Turing Machine.) There is a similar intuitive proof of Theorem H.2.

We might hope to construct a single Oracle Turing Machine which could compute *any* Julia set to any degree of accuracy, but this is too much to ask for. It is not hard to see that such a machine could never work for maps f_0 at which the correspondence $f \mapsto J(f)$ is discontinuous. As one example, this correspondence is never continuous at a polynomial f_0 which has a Siegel disk centered at the origin. For the origin is well separated from $J(f_0)$, but polynomials arbitrarily close to f_0 have the origin as a repelling point, with $0 \in J(f)$. We could try to get around this difficulty, by allowing different algorithms (or different Oracle Turing Machines) for different rational maps, but in some cases no such machine exists.

Theorem H.4 (Braverman and Yampolsky [2004]). *There exist polynomials of the form* $f_t(z) = z^2 + e^{2\pi i t}z$ *whose Julia set cannot be computed by any Oracle Turing Machine.*

The proof uses delicate estimates due to Buff and Cheritat [2003] to show that $\mathbb{R}/\mathbb{Z}$ is not a countable union of sets on which the correspondence $t \mapsto J(f_t)$ is continuous.

REFERENCES

L. Ahlfors [1966], "Complex Analysis," McGraw-Hill, New York.

L. Ahlfors [1973], "Conformal Invariants," McGraw-Hill, New York.

L. Ahlfors [1987], "Lectures on Quasiconformal Mappings," Wadsworth, Belmont, CA.

L. Ahlfors and L. Bers [1960], *Riemann's mapping theorem for variable metrics,* Annals of Math. **72**, 385–404.

L. Ahlfors and L. Sario [1960], "Riemann Surfaces," Princeton U. Press, Princeton, NJ.

D. S. Alexander [1994], "A History of Complex Dynamics from Schröder to Fatou and Julia," Vieweg, Braunschweig.

J. C. Alexander, I. Kan, J. Yorke, and Z. You [1992], *Riddled basins*, Int. J. Bifurcation and Chaos **2**, 795–813.

V. Arnold [1965], *Small denominators I: On the mappings of the circumference into itself*, Amer. Math. Soc. Transl. (2) **46**, 213–284.

M. Atiyah and R. Bott [1966], *A Lefschetz fixed point formula for elliptic differential operators*, Bull. Amer. Math. Soc. **72**, 245–250.

C. Babbage [1815], *An essay on the calculus of functions*, Phil. Trans. Royal Soc. London **105**, 389–423. (Available through www.jstor.org.)

I. N. Baker [1968], *Repulsive fixedpoints of entire functions*, Math. Zeit. **104**, 252–256.

I. N. Baker [1976], *An entire function which has wandering domains*, J. Austral. Math. Soc. **22**, 173–176.

A. Beardon [1984], "A Primer on Riemann Surfaces," Cambridge U. Press, Cambridge, UK.

A. Beardon [1991], "Iteration of Rational Functions," Grad. Texts Math. **132**, Springer-Verlag, New York.

E. Bedford [1990], *Iteration of polynomial automorphisms of* $\mathbb{C}^2$, Proc. Int. Cong. Math. Kyoto, 847–858.

E. Bedford, M. Lyubich, and J. Smillie [1993], *Polynomial diffeomorphisms of* $\mathbb{C}^2$*, IV: The measure of maximal entropy and laminar currents*, Inv. Math. **112**, 77–125.

E. Bedford and J. Smillie [1991–2002], *Polynomial diffeomorphisms of* $\mathbb{C}^2$, I, Invent. Math. [1991] **103**, 69–99; II, J. Amer. Math. Soc. [1991] **4**, 657–679; III, Math. Ann. [1992] **294**, 395–420; V, Ann. of Math. [1998]

148, 695–735; VI, J. Geom. Anal. [1998] **8**, 349–383; VII, Ann. Sci. École Norm. Sup. Paris [1999] **32**, 455–497; VIII, Amer. J. Math. [2002] **124**, 221–271.

W. Bergweiler [1993], *Iteration of meromorphic functions*, Bull. Amer. Math. Soc. **29**, 151–188.

F. Berteloot and V. Mayer [2001], "Rudiments de Dynamique Holomorphe," Soc. Math. France, Paris.

L. Bieberbach [1916], *Über die Koeffizienten derjenigen Potenzreihen welche eine schlichte Abbildung des Einheitskreises vermitteln*, S.-B. Preuss. Akad. Wiss. 940–955.

L. Bieberbach [1933], *Beispiel zweier ganzer Funktionen zweier komplexer Variablen, welche eine schlichte volumtreue Abbildung des* $\mathbf{R}_4$ *auf einen Teil seiner selbst vermitteln*, S.-B. Preuss. Akad. Wiss. **14/15**, 476–479.

I. Binder, M. Braverman, and M. Yampolsky [2005], *Filled Julia sets with empty interior are computable*, e-print math.DS/0410580 at Arxiv.org.

P. Blanchard [1984], *Complex analytic dynamics on the Riemann sphere*, Bull. Amer. Math. Soc. **11**, 85–141.

P. Blanchard [1986], *Disconnected Julia sets*, in "Chaotic Dynamics and Fractals," Edit. Barnsley and Demko, Academic Press, New York, 181–201.

P. Blanchard and A. Chiu [1991], *Conformal dynamics: An informal discussion*, in "Fractal Geometry and Analysis" (Montreal 1989), NATO Adv. Sci. Inst. Ser. C Math. Phys. Sci., **346**, Kluwer Acad. Publ., Dordrecht, 45–98.

L. Blum, M. Shub, and S. Smale [1989], *On a theory of computation and complexity over the real numbers: NP-completeness, recursive functions, and universal machines*, Bull. Amer. Math. Soc. **21**, 1–46.

A. Bonifant, M. Dabija, and J. Milnor, *Elliptic curves as attractors in* $\mathbb{P}^2$ *Part I: Dynamics*, in preparation.

L. E. Böttcher [1904], *The principal laws of convergence of iterates and their application to analysis* (Russian), Izv. Kazan. Fiz.-Mat. Obshch. **14**, 155–234.

L. de Branges [1985], *A proof of the Bieberbach conjecture,* Acta Math. **154**, 137–152.

B. Branner [1989], *The Mandelbrot set*, 75–105 of "Chaos and Fractals," Edit. Devaney and Keen, Proc. Symp. Applied Math. **39**, Amer. Math. Soc.

B. Branner and J. H. Hubbard [1988], *The iteration of cubic polynomials,* *Part I: The global topology of parameter space,* Acta Math. **160**, 143–206.

B. Branner and J. H. Hubbard [1992], *The iteration of cubic polynomials, Part II: Patterns and parapatterns,* Acta Math. **169**, 229–325.

M. Braverman [2004], *Computational complexity of Euclidean sets: hyperbolic Julia sets are poly-time computable*, Master's thesis, U. Toronto. Available at www.cs.toronto.edu/~mbraverm.

M. Braverman [2005], *Parabolic Julia sets are polynomial time computable*, e-print math.DS/0505036 at Arxiv.org.

M. Braverman and M. Yampolsky [2004], *Non-computable Julia sets*, e-print math.DS/0501448 at Arxiv.org.

H. Brolin [1965], *Invariant sets under iteration of rational functions*, Arkiv för Mat. **6**, 103–144.

R. Brooks and P. Matelski [1981], *The dynamics of* 2*-generator subgroups of* PSL(2, $\mathbb{C}$), in "Riemann Surfaces and Related Topics", Proceedings of the 1978 Stony Brook Conference, Ann. of Math. Stud., **97**, Princeton Univ. Press, Princeton, NJ, 65–71.

M. Brown [1960], *A proof of the generalized Schoenflies theorem,* Bull. Amer. Math. Soc. **66**, 74–76. (See also: *The monotone union of open n-cells is an open n-cell*, Proc. Amer. Math. Soc. [1961] **12**, 812–814.)

H. Bruin, G. Keller, T. Nowicki, and S. van Strien [1996], *Wild Cantor attractors exist,* Annals of Math. **143**, 97–130.

A. D. Bryuno (=Brjuno) [1965], *Convergence of transformations of differential equations to normal forms*, Dokl. Akad. Nauk USSR **165**, 987–989 (Soviet Math. Dokl., 1536–1538).

X. Buff [2003], *Virtually repelling fixed points*, Publ. Mat. **47**, 195–209.

X. Buff and A. Cheritat [2003], *Quadratic Siegel disks with rough boundaries*, e-print math.DS/0309067 at Arxiv.org. (See also 0305080 and 0401044.)

X. Buff and A. Epstein [2002], *A parabolic Pommerenke-Levin-Yoccoz inequality*, Fund. Math. **172**, 249–289.

G. T. Buzzard [1997], *Infinitely many periodic attractors for holomorphic maps of* 2 *variables*, Ann. of Math. **145**, 389–417.

C. Camacho [1978], *On the local structure of conformal mappings and holomorphic vector fields*, Journées Singulières de Dijon, Astérisque **59–60**, Soc. Math. France, Paris, 3, 83–94.

C. Carathéodory [1913], *Über die Begrenzung einfach zusammenhängender Gebiete,* Math. Ann. **73**, 323–370. (Gesam. Math. Schr., v. 4.)

L. Carleson and T. Gamelin [1993], "Complex Dynamics," Springer-Verlag, New York.

A. Cayley [1879], *Application of the Newton-Fourier method to an imaginary root of an equation*, Quart. J. Pure Appl. Math. **16**, 179–185.

T. M. Cherry [1964], *A singular case of iteration of analytic functions: A contribution to the small-divisor problem*, "Nonlinear Problems of Engineering," Academic Press, New York. 29–50.

A. Chou and K. Ko [1995], *Computational complexity of two dimensional regions*, SIAM J. Computation **24**, 923–947.

E. Coddington and N. Levinson [1955], "Theory of Ordinary Differential Equations," McGraw-Hill, New York.

I. Cornfeld, S. Fomin, and Y. Sinai [1982], "Ergodic Theory," Springer-Verlag, New York.

R. Courant and H. Robbins [1941], "What Is Mathematics?," Oxford U. Press, Oxford, UK.

H. Cremer [1927], *Zum Zentrumproblem*, Math. Ann. **98**, 151–163.

H. Cremer [1938], *Über die Häufigkeit der Nichtzentren*, Math. Ann. **115**, 573–580.

L. DeMarco [2003], *Dynamics of rational maps: Lyapunov exponents, bifurcations, and capacity*, Math. Ann. **326**, 43–73.

L. DeMarco [2004], *Iteration at the boundary of the space of rational maps*, e-print math.DS/0403078 at Arxiv.org.

A. Denjoy [1926], *Sur l'itération des fonctions analytiques*, C. R. Acad. Sci. Paris **182**, 255–257.

A. Denjoy [1932], *Sur les courbes définies par les équations différentielles à la surface du tore,* Journ. de Math. **11**, 333–375.

R. Devaney [1986], *Exploding Julia sets*, in "Chaotic Dynamics and Fractals," Edit. Barnsley and Demko, Academic Press, New York, 141–154.

R. Devaney [1989], "An Introduction to Chaotic Dynamical Systems," 2^{nd} Ed., Addison-Wesley, Reading, MA.

R. Devaney [2004], *Cantor and Sierpinski, Julia and Fatou: Complex topology meets complex dynamics*, Notices Amer. Math. Soc., January, 9–15.

A. Douady [1982–83], *Systèmes dynamiques holomorphes*, Séminar Bourbaki, 35^{e} année, n^{o} 599; Astérisque **105/106**, 39–63.

A. Douady [1986], *Julia sets and the Mandelbrot set,* in "The Beauty of Fractals," Edit. Peitgen and Richter, Springer-Verlag, New York, 161–173.

A. Douady [1987], *Disques de Siegel et anneaux de Herman*, Sém. Bourbaki 39^{e} année (1986–87) n^{o} 677; Astérisque **152-153**, 151–172.

A. Douady and X. Buff [2000], *Le théorème d'intégrabilité des structures* *presque complexes*, in "The Mandelbrot Set, Theme and Variations," London Math. Soc. Lecture Note Ser. **274**, 307–324.

A. Douady and J. H. Hubbard [1982], *Itération des polynômes quadratiques complexes*, C. R. Acad. Sci. Paris **294**, 123–126.

A. Douady and J. H. Hubbard [1984–85], "Étude Dynamique des Polynômes Complexes I and II," Publ. Math. Orsay.

A. Douady and J. H. Hubbard [1985], *On the dynamics of polynomial-like mappings*, Ann. Sci. École Norm. Sup. Paris **18**, 287–343.

A. Douady and J. H. Hubbard [1993], *A proof of Thurston's topological characterization of rational functions*, Acta Math. **171**, 263–297.

J. Écalle [1975], *Théorie itérative: Introduction a la théorie des invariants holomorphes*, J. Math. Pure Appl. **54**, 183–258.

A. L. Epstein [1999], *Infinitesimal Thurston rigidity and the Fatou–Shishikura inequality*, Stony Brook I.M.S. Preprint 1999#1 or e-print math.DS/9902158 at Arxiv.org.

A. L. Epstein [2000], *Bounded hyperbolic components of quadratic rational maps*, Ergodic Theory Dynam. Systems **20**, 727–748.

D. B. A. Epstein [1981], *Prime Ends*, Proc. London Math. Soc. **42**, 385–414.

A. Eremenko and G. Levin [1989], *Periodic points of polynomials*, (Russian) Ukrain. Mat. Zh. [1989] **41**, 1467–1471, 1581; English translation in Ukrainian Math. J. [1989/1990] **41**, 1258–1262.

A. Eremenko and M. Lyubich [1990], *The dynamics of analytic transformations*, Leningr. Math. J. **1**, 563–634.

A. Eremenko and M. Lyubich [1992], *Dynamical properties of some classes of entire functions*, Ann. Inst. Fourier, Grenoble **42**, 989–1020. (See also Sov. Math. Dokl. [1984] **30**, 592–594; Funct. Anal. Appl. [1985] **19**, 323–324; and J. Lond. Math. Soc. [1987] **36**, 458–468.)

K. Falconer [1990], "Fractal Geometry: Mathematical Foundations and Applications," Wiley.

H. Farkas and I. Kra [1980], "Riemann Surfaces," Springer-Verlag, New York.

P. Fatou [1906], *Sur les solutions uniformes de certaines équations fonctionnelles*, C. R. Acad. Sci. Paris **143**, 546–548.

P. Fatou [1919–20], *Sur les équations fonctionnelles*, Bull. Soc. Math. France, Paris **47**, 161–271, and **48**, 33–94, 208–314.

P. Fatou [1926], *Sur l'itération des fonctions transcendantes entières*, Acta Math. **47**, 337–370.

Y. Fisher [1988], *Exploring the Mandelbrot set*, in "The Science of Fractal Images," Edit. Peitgen and Saupe, Springer-Verlag, New York, 287–296.

Y. Fisher, J. Hubbard, and B. Wittner [1988], *A proof of the uniformization theorem for arbitrary plane domains*, Proceedings Amer. Math. Soc. **104**, 413–418.

J. E. Fornæss [1996], "Dynamics in Several Complex Variables," CBMS Regional Conference Series in Mathematics **87**, Amer. Math. Soc., Providence, RI.

J. E. Fornæss and E. Gavosto [1992], *Existence of generic homoclinic tangencies for Hénon mappings*, J. Geom. Anal. **2**, 429–444.

J. E. Fornæss and N. Sibony [1992a], *Complex Hénon mappings in $\mathbb{C}^2$ and Fatou-Bieberbach domains*, Duke Math. J. **65**, 345–380.

J. E. Fornæss and N. Sibony [1992b], *Critically finite rational maps on $\mathbb{P}^2$*, in "The Madison Symposium on Complex Analysis" (Madison, 1991), Contemp. Math. **137**, Amer. Math. Soc., Providence, RI, 245–260.

J. E. Fornæss and N. Sibony [1994], *Complex dynamics in higher dimension, I*, "Complex Analytic Methods in Dynamical Systems" (Rio de Janeiro, 1992], Astérisque **222**, 201–231.

J. E. Fornæss and N. Sibony [1995a], *Complex dynamics in higher dimension, II*, in "Modern Methods in Complex Analysis" (Princeton, 1992), Ann. of Math. Stud., **137**, Princeton U. Press, Princeton, NJ, 135–182.

J. E. Fornæss and N. Sibony [1995b], *Classification of recurrent domains for some holomorphic maps*, Math. Ann. **301**, 813–820.

J. E. Fornæss and N. Sibony [1998], *Hyperbolic Maps on $\mathbb{P}^2$*, Math. Ann. **311**, 305–333.

J. E. Fornæss and N. Sibony [2001], *Dynamics of $\mathbb{P}^2$ (examples)*, in "Laminations and Foliations in Dynamics, Geometry and Topology" (Stony Brook, NY, 1998), Contemp. Math. **269**, Amer. Math. Soc., Providence, RI, 47–85.

J. E. Fornæss and B. Weickert [1999], *Attractors in $\mathbb{P}^2$*, "Several Complex Variables" (Berkeley, CA, 1995–1996), Math. Sci. Res. Inst. Publ. **37**, Cambridge Univ. Press, Cambridge, UK, 297–307.

J. Franks [1982], "Homology and Dynamical Systems," Conference Board Math. Sci., Regional Conference **49**, Amer. Math. Soc., Providence, RI.

A. Freire, A. Lopes, and R. Mañé [1983], *An invariant measure for rational maps*, Bol. Soc. Brasil. Mat. **14**, 45–62.

S. Friedland and J. Milnor [1989], *Dynamical properties of plane polynomial automorphisms*, Ergodic Theory Dynam. Systems, **9**, 67–99.

E. Gavosto [1998], *Attracting Basins in* $\mathbb{P}^2$, J. Geom. Anal. **8**, 433–440.

E. Ghys [1999/2003], *Holomorphic dynamical systems* in "Complex Dynamics and Geometry," SMF/AMS Texts and Monographs **10**, Soc. Math. France, Paris; Amer. Math. Soc., Providence, RI., 2003, 1–9. (Translated from "Dynamique et Géometrie Complexes", Soc. Math. Fr. 1999.)

L. Goldberg [1992], *Fixed points of polynomial maps, Part I: Rotation subsets of the circle*, Ann. Sci. École Norm. Sup. Paris **25**, 679–685.

L. Goldberg and L. Keen [1986], *A finiteness theorem for a dynamical class of entire functions*, Ergodic Theory Dynam. Systems, **6**, 183–192.

L. Goldberg and J. Milnor [1993], *Fixed points of polynomial maps, Part II: Fixed point portraits*, Ann. Sci. École Norm. Sup. Paris **26**, 51–98.

T. H. Gronwall [1914–15], *Some remarks on conformal representation*, Ann. of Math. **16**, 72–76.

D. H. Hamilton [1995], *Length of Julia curves*, Pac. J. Math. **169**, 75–93.

G. H. Hardy and E. M. Wright [1938], "An Introduction to the Theory of Numbers," Clarendon Press, Oxford, England.

M. Herman [1979], *Sur la conjugation différentiables des difféomorphismes du cercle à les rotations*, Pub. I.H.E.S. **49**, 5–233.

M. Herman [1984], *Exemples de fractions rationnelles ayant une orbite dense sur la sphère de Riemann*, Bull. Soc. Math. France, Paris **112**, 93–142.

M. Herman [1986], *Recent results and some open questions on Siegel's linearization theorem of germs of complex analytic diffeomorphisms of* $\mathbb{C}^n$ *near a fixed point*, Proc. 8th Int. Cong. Math. Phys., World Scientific, River Edge, NJ, 138–198.

J. Hocking and G. Young [1961], "Topology," Addison-Wesley, Reading, MA.

K. Hoffman [1962], "Banach Spaces of Analytic Functions," Prentice-Hall, Englewood Cliffs, NJ.

J. H. Hubbard [1986], *The Hénon mapping in the complex domain*, in "Chaotic Dynamics and Fractals," Edit. M. Barnsley and S. Demko, Academic Press, New York, 101–111.

J. H. Hubbard [1993], *Local connectivity of Julia sets and bifurcation loci: Three theorems of J.-C. Yoccoz*, in "Topological Methods in Modern Mathematics," Edit. Goldberg and Phillips, Publish or Perish, Houston, TX, 467–511.

J. H. Hubbard and R. Oberste-Vorth [1995], *Hénon mappings in the complex domain, II: Projective and inductive limits of polynomials*, in "Real and Complex Dynamical Systems (Hillerød, 1993)," NATO Adv. Sci. Inst.

Ser. C. Math. Phys. Sci. **464**, Kluwer Acad. Publ., Dordrecht, 89–132.

J. H. Hubbard and P. Papadopol [1994], *Superattractive fixed points in* $\mathbb{C}^n$, Indiana Univ. Math. J. **43**, 321–365.

J. L. W. V. Jensen [1899], *Sur un nouvel et important théoreme de la théorie des fonctions*, Acta Math. **22**, 219–251.

M. Jonsson and B. Weickert [2000], *A nonalgebraic attractor in* $\mathbb{P}^2$. Proc. Amer. Math. Soc. **128**, 2999–3002.

G. Julia [1918], *Memoire sur l'itération des fonctions rationnelles*, J. Math. Pure Appl. **8**, 47–245.

G. Julia [1919], *Sur quelques problèmes relatifs à l'itération des fractions rationnelles*, C. R. Acad. Sci. **168**, 147–149.

E. Kasner [1912], *Conformal geometry*, Proc. Fifth Intern. Congr. Math. Cambridge, **2**, 81–87.

L. Keen [1988], *The dynamics of holomorphic self-maps of* $\mathbb{C}^*$, in "Holomorphic Functions and Moduli," Edit. Drasin et al., Springer-Verlag, New York, 9-30.

L. Keen [1989], *Julia sets* in "Chaos and Fractals, the Mathematics behind the Computer Graphics," Edit. Devaney and Keen, Proc. Symp. Appl. Math. **39**, Amer. Math. Soc., Providence, RI, 57–74.

A. Khintchine [1963], "Continued Fractions," Noordhoff, Groningen, The Netherlands.

J. Kiwi [1997], *Rational rays and critical portraits of complex polynomials,* Thesis SUNY at Stony Brook (SUNY IMS Preprint #1997/15).

J. Kiwi [1999], *From the shift loci to the connectedness loci of complex polynomials*, in "Complex Geometry of Groups" (Olmué, 1998), Contemp. Math., **240**, Amer. Math. Soc., Providence, RI, 231–245.

P. Koebe [1907], *Über die Uniformizierung beliebiger analytischer Kurven*, Nachr. Akad. Wiss. Göttingen, Math.-Phys. Kl., 191–210.

G. Kœnigs [1884], *Recherches sur les intégrales de certaines équations fonctionelles*, Ann. Sci. École Norm. Sup. Paris (3e ser.) **1**, supplém., 1–41.

K. Kuratowski [1968], "Topology II," Acad. Press, New York.

S. Lattès [1918], *Sur l'itération des substitutions rationelles et les fonctions de Poincaré*, C. R. Acad. Sci. Paris **16**, 26–28.

P. Lavaurs [1989], "Systèmes dynamiques holomorphes: Explosion de points périodiques paraboliques," Thèse, Univ. Paris-Sud., Orsay.

L. Leau [1897], *Étude sur les equations fonctionelles à une ou plusièrs variables*, Ann. Fac. Sci. Toulouse **11**, E1–E110.

O. Lehto [1987], "Univalent functions and Teichmüller spaces," Springer-Verlag, New York.

O. Lehto [1998] "Mathematics Without Borders: A History of the International Mathematical Union," Springer-Verlag, New York.

O. Lehto and K. J. Virtanen [1973], "Quasiconformal Mappings in the Plane," Springer-Verlag, New York.

E. R. Love [1969], *Thomas Macfarland Cherry*, Bull. London Math. Soc. **1**, 224–245.

M. Lyubich [1983a], *Some typical properties of the dynamics of rational maps*, Russian Math. Surveys **38**, 154–155. (See also Sov. Math. Dokl. **27**, 22–25.)

M. Lyubich [1983b], *Entropy properties of rational endomorphisms of the Riemann sphere*, Ergodic Theory Dynam. Systems, **3**, 351–385.

M. Lyubich [1986], *The dynamics of rational transforms: The topological picture*, Russian Math. Surveys **41:4**, 43–117.

M. Lyubich [1987], *The measurable dynamics of the exponential map*, Siber. J. Math. **28**, 111–127. (See also Sov. Math. Dokl. **35**, 223–226.)

M. Lyubich [1990], *An analysis of the stability of the dynamics of rational functions*, Selecta Math. Sovietica **9**, 69–90. (Russian original published in 1984.)

M. Lyubich [1997/2000], *Dynamics of quadratic polynomials I, II, III*, Acta Math. **178**, 185–247, 247–297, and Astérisque **261**, 173–200.

M. Lyubich [1999], *Feigenbaum-Coullet-Tresser universality and Milnor's hairiness conjecture*, Ann. of Math. **149**, 319–420.

B. Malgrange [1981/82], *Travaux d'Écalle et de Martinet-Ramis sur les systèmes dynamiques*, Sém. Bourbaki, 34e ann., n° 582.

B. Mandelbrot [1980], *Fractal aspects of the iteration of* $z \mapsto \lambda z(1-z)$ *for complex* λ, z, Annals NY. Acad. Sci. **357**, 249–259.

R. Mañé [1983], *On the uniqueness of the maximizing measure for rational maps*, Bol. Soc. Brasil. Mat. **14**, 27–43.

R. Mañé, P. Sad, and D. Sullivan [1983], *On the dynamics of rational maps*, Ann. Sci. École Norm. Sup. Paris (4) **16**, 193–217.

J. Martinet and J. P. Ramis [1983], *Classification analytique des équations différentielles non linéaires résonnantes du premier ordre*, Ann. Sci. Éc. Norm. Sup. **16**, 571-621.

J. Mather [1982], *Topological proofs of some purely topological consequences of Carathéodory's theory of prime ends*, in "Selected Studies," Edit. T. and G. Rassias, North-Holland, New York, 225–255.

C. McMullen [1987], *Area and Hausdorff dimension of Julia sets of entire functions*, Trans. Amer. Math. Soc. **300**, 329–342.

C. McMullen [1988], *Automorphisms of rational maps*, in "Holomorphic Functions and Moduli I," Edit. Drasin, Earle, Gehring, Kra, and Marden, Springer-Verlag, New York, 31–60.

C. McMullen [1994a], "Complex Dynamics and Renormalization," Ann. Math. Studies **135**, Princeton U. Press, Princeton, NJ.

C. McMullen [1994b], *Frontiers in complex dynamics*, Bull. Amer. Math. Soc. **31**, 155–172.

C. McMullen [1996], "Renormalization and 3-Manifolds which Fiber over the Circle," Ann. Math. Studies **142**, Princeton U. Press, Princeton, NJ.

W. de Melo and S. van Strien [1993], "One Dimensional Dynamics," Springer-Verlag, New York. (See also de Melo, "Lectures on One Dimensional Dynamics," 17° Col. Brasil. Mat., IMPA, 1990.)

J. Milnor [1975], *On the 3-dimensional Brieskorn manifolds* $M(p,q,r)$, in "Knots, Groups, and 3-Manifolds," Edit. Neuwirth, Ann. Math. Studies **84**, Princeton U. Press, Princeton, NJ, 175–225.

J. Milnor [1985], *On the concept of attractor*, Comm. Math. Phys. **99**, 177–195; Correction and remarks: ibid. **102**, 517–519.

J. Milnor [1989], *Self-similarity and hairiness in the Mandelbrot set*, in "Computers in Geometry and Topology," Edit. Tangora, Lect. Notes Pure Appl. Math. **114**, Dekker, New York, 211–257.

J. Milnor [1992a], *Remarks on iterated cubic maps*, Experimental Math. **1**, 5–24.

J. Milnor [1992b], *Hyperbolic components in spaces of polynomial maps*, Stony Brook I.M.S. Preprint 1992#3. (available at www.math.sunysb.edu/preprints.html).

J. Milnor [1993], *Geometry and dynamics of quadratic rational maps*, Experimental Math. **2**, 37–83.

J. Milnor [1999], *The work of Curtis T. McMullen*, Notices Amer. Math. Soc., January, 23–26.

J. Milnor [2000a], *Periodic orbits, externals rays and the Mandelbrot set: An expository account*, in "Géométrie Complexe et Systèmes Dynamiques, Colloque en l'honneur d'Adrien Douady," Astérisque, **261**, 277–333.

J. Milnor [2000b], *Local connectivity of Julia sets: Expository lectures*, in "The Mandelbrot Set, Theme and Variations," Edit. Tan Lei, Cambridge U. Press, Cambridge, UK, 67–116.

J. Milnor [2000c], *Rational maps with two critical points*, Experimental

Math. **9**, 481–522.

J. Milnor [2004a], *Pasting together Julia sets: A worked out example of mating*, Experimental Math. **13**, 55–92.

J. Milnor [2004b], *On Lattès maps*, in "Dynamics on the Riemann Sphere. A Bodil Branner Festschrift", edited by P. Hjorth and C. L. Petersen, European Math. Soc., 2006 (or Stony Brook IMS preprint ims04-01; arxiv math.DS/0402147).

J. Milnor and W. Thurston [1988], *Iterated maps of the interval*, in "Dynamical Systems (Maryland 1986–87)," Edit. J. C. Alexander, Lect. Notes Math. **1342**, Springer-Verlang, New York, 465–563.

M. Misiurewicz [1981], *On iterates of* e^z, Ergodic Theory Dynam. Systems, **1**, 103–106.

P. Montel [1927], "Leçons sur les Familles Normales," Gauthier-Villars, Paris.

L. Mora and M. Viana [1993], *Abundance of strange attractors*, Acta Math. **171**, 1–71.

S. Morosawa, Y. Nishimura, M. Taninguchi, and T. Ueda [2000], "Holomorphic Dynamics," Cambridge Studies in Advanced Mathematics **66**, Cambridge U. Press, Cambridge, UK.

C. B. Morrey, Jr [1938], *On the solutions of quasi-linear elliptic partial differential equations*, Trans. Amer. Math. Soc. **43**, 126–166.

J. Munkres [1975], "Topology: A First Course," Prentice-Hall, Englewood Cliffs, NJ.

V. I. Naĭshul′ [1983], *Topological invariants of analytic and area preserving mappings ...*, Trans. Moscow Math. Soc. **42**, 239–250.

R. Nevanlinna [1967], "Uniformisierung," Springer-Verlag, New York.

S. Newhouse [1974], *Diffeomorphisms with infinitely many sinks*, Topology **16**, 9–18.

M. Ohtsuka [1970], "Dirichlet Problem, Extremal Length and Prime Ends," Van Nostrand, New York.

H.-O. Peitgen [1988], *Fantastic deterministic fractals*, in "The Science of Fractal Images," Edit. Barnsley et al., Springer-Verlag, New York, 169–218.

R. Perez-Marco [1990], "Sur la dynamique des germes de difféomorphismes holomorphes de $(\mathbb{C}, 0)$ et des difféomorphismes analytiques du cercle," Thèse, Paris-Sud.

R. Perez-Marco [1992], *Solution complete au probleme de Siegel de linearisation d'une application holomorphe au voisinage d'un point fixe (d'apres*

J.-C. Yoccoz), Sem. Bourbaki, n° 753: Astérisque **206**, 273–310.

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

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

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

C. L. Petersen [1998], *Puzzles and Siegel disks*, in "Progress in Holomorphic Dynamics," Pitman Res. Notes Math. Ser., **387**, Longman, Harlow, 50–85.

C. L. Petersen and S. Zakeri [2004], *On the Julia set of a typical quadratic polynomial with a Siegel disk*, Ann. of Math. **159**, 1–52.

G. A. Pfeiffer [1917], *On the conformal mapping of curvilinear angles: The functional equation* $\phi[f(x)] = a_1\phi(x)$, Trans. Amer. Math. Soc., **18**, 185–198. (Available from www.jstor.org.)

K. Pilgrim and Tan Lei [2000], *Rational maps with disconnected Julia set*, in "Géométrie Complexe et Systèmes Dynamiques" (Orsay, 1995). Astérisque **261**, xiv, 349–384.

C. Pommerenke [1986], *On conformal mapping and iteration of rational functions*, Complex Variables Theory Appl. **5**, 117–126.

M. Rees [1984], *Ergodic rational maps with dense critical point forward orbit*, Ergodic Theory Dynam. Systems **4**, 311–322.

M. Rees [1986a], *Positive measure sets of ergodic rational maps*, Ann. Sci. École Norm. Sup. Paris (4) **19**, 383–407.

M. Rees [1986b], *The exponential map is not recurrent*, Math. Zeit. **191**, 593–598.

M. Rees [1990], *Components of degree two hyperbolic rational maps*, Invent. Math. **100**, 357–382.

M. Rees [1992], *A partial description of parameter space of rational maps of degree two: Part* 1, Acta Math. **168**, 11–87.

M. Rees [1995], *A Partial description of the parameter space of rational maps of degree two: Part* 2, Proc. London Math. Soc. **70**, 644–690.

R. Rettinger [2004], *A fast algorithm for the Julia sets of hyperbolic rational functions*, Proceedings Computability and Complexity in Analysis Workshop. (See also R. Rettinger and K. Weihrauch, *The computational complexity of some Julia sets*, Proceedings of 35th ACM Symposium on Theory of Computing (2003), 177–185.)

F. and M. Riesz [1916], *Über die Randwerte einer analytischen Funktion*, Quatr. Congr. Math. Scand. Stockholm, 27–44.

J. F. Ritt [1920], *On the iteration of rational functions*, Trans. Amer. Math. Soc. **21**, 348–356.

P. Roesch [1998], *Topologie locale des méthodes de Newton cubiques: Plan dynamique*, C. R. Acad. Sci. Paris, I Math. **326**, 1221–1226.

J. T. Rogers Jr. [1998], *Diophantine conditions imply critical points on the boundaries of Siegel disks of polynomials*, Comm. Math. Phys. **195**, 175–193. (See also: *Critical points on the boundaries of Siegel disks*, Bull. Amer. Math. Soc. **32** [1995], 317–321.)

K. F. Roth [1955], *Rational approximations to algebraic numbers*, Mathematika **2**, 1–20, 168.

H. Rüssmann [1967], *Über die Iteration analytischer Funktionen*, J. Math. and Mech. **17**, 523–532.

D. Schleicher [2000], *Rational parameter rays of the Mandelbrot set*, in "Géométrie Complexe et Systèmes Dynamiques" (Orsay, 1995). Astérisque **261**, xiv–xv, 405–443.

D. Schleicher and J. Zimmer [2003], *Periodic points and dynamic rays of exponential maps*, Ann. Acad. Sci. Fenn. Math. **28**, 327–354.

E. Schröder [1871], *Ueber iterirte Functionen*, Math. Ann. **3**, 296-322. (Available through http://gdz.sub.uni-goettingen.de/en.)

H. Schubert [to appear], *Area of Fatou sets of trigonometric functions*.

G. Segal [1979], *The topology of spaces of rational functions*, Acta Math. **143**, 39–72.

M. Shishikura [1987], *On the quasiconformal surgery of rational functions*, Ann. Sci. École Norm. Sup. Paris **20**, 1–29.

M. Shishikura [1998], *The Hausdorff dimension of the boundary of the Mandelbrot set and Julia sets*, Ann. of Math. **147**, 225–267.

M. Shub [1987], "Global Stability of Dynamical Systems," Springer-Verlag, New York.

N. Sibony [1999/2003], *Dynamics of rational maps on* $\mathbb{P}^k$, in "Complex Dynamics and Geometry," SMF/AMS Texts and Monographs **10**, Soc. Math. France, Paris; Amer. Math. Soc. 2003, 85–166. (Translated from "Dynamique et Géometrie Complexes", Soc. Math. Fr. 1999.)

C. L. Siegel [1942], *Iteration of analytic functions*, Ann. of Math. **43**, 607–612.

C. L. Siegel and J. Moser [1971], "Lectures on Celestial Mechanics," Springer-Verlag, New York.

S. Smale [1967], *Differentiable dynamical systems*, Bull. Amer. Math. Soc.

73, 747–817.

J. Smillie [1997], *Complex dynamics in several variables; with notes by G. T. Buzzard*, in "Flavors of Geometry," Math. Sci. Res. Inst. Publ. **31**, Cambridge U. Press, Cambridge, UK, 117-150.

G. Springer [1957], "Introduction to Riemann Surfaces," Addison-Wesley, Reading, MA.

N. Steinmetz [1993], "Rational Iteration: Complex Analytic Dynamical Systems," de Gruyter, Berlin.

D. Sullivan [1983], *Conformal dynamical systems*, in "Geometric Dynamics," Edit. Palis, Lecture Notes Math. **1007**, Springer-Verlag, New York, 725–752.

D. Sullivan [1985], *Quasiconformal homeomorphisms and dynamics I: Solution of the Fatou-Julia problem on wandering domains*, Ann. Math. **122**, 401–418.

Tan Lei [1997], *Branched coverings and cubic Newton maps*, Fund. Math. **154**, 207–226.

Tan Lei and Y. Yin [1996], *Local connectivity of the Julia set for geometrically finite rational maps*, Science in China A **39**, 39–47.

W. Thurston [1997], "Three-dimensional Geometry and Topology, Vol. 1," Edit. S. Levy, Princeton Mathematical Series **35**, Princeton U. Press, Princeton, NJ.

A. Turing [1936–37], *On computable numbers, with an application to the Entscheidungsproblem*, Proc. London Math. Soc. **42**, 230-265; **43**, 544-546. (Also in "The Undecidable," Edit. M. Davis, Raven Press, 1995, and in Turing's "Collected Works," North-Holland, New York, 1992.)

T. Ueda [1994], *Fatou sets in complex dynamics on projective spaces*, J. Math. Soc. Japan **46**, 545–555.

T. Ueda [1995], *Complex dynamics on projective spaces: Index formula for fixed points*, in "Dynamical Systems and Chaos, Vol. 1" (Hachioji, 1994), World Scientific, River Edge, NJ, 252–259.

T. Ueda [1998], *Critical orbits of holomorphic maps on projective spaces*, J. Geom. Anal. **8**, 319–334.

S. Ulam and J. von Neumann [1947], *On combinations of stochastic and deterministic processes*, Bull. Amer. Math. Soc. **53**, 1120.

S. M. Voronin [1981], *Analytic classification of germs of conformal maps* $(\mathbb{C},0) \to (\mathbb{C},0)$ *with identity linear part*, Funct. Anal. Appl. **15**, 1–17.

G. T. Whyburn [1964], "Topological Analysis," Princeton U. Press, Princeton, NJ.

T. Willmore [1959], "Introduction to Differential Geometry," Clarendon, Oxford, UK.

M. Yampolsky [1999], *Complex bounds for renormalization of critical circle maps*, Ergodic Theory Dynam. Systems **19**, 227–257.

M. Yampolsky and S. Zakeri [2001], *Mating Siegel quadratic polynomials*, J. Amer. Math. Soc. **14**, 25–78.

J.-C. Yoccoz [1984], *Conjugation différentiable des difféomorphismes du cercle dont le nombre de rotation vérifie une condition diophantienne*, Ann. Sci. École Norm. Sup. Paris (4) **17**, 333–359.

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

J.-C. Yoccoz [1995], *Petits Diviseurs en Dimension 1*, Astérisque **231**, Soc. Math. France, Paris.

J.-C. Yoccoz [1999/2003], *Dynamics of quadratic polynomials*, in "Complex Dynamics and Geometry," SMF/AMS Texts and Monographs **10**, Soc. Math. France, Paris; Amer. Math. Soc., 2003, 167–197. (Translated from "Dynamique et Géometrie Complexes," Soc. Math. Fr. 1999.)

J.-C. Yoccoz [2002], *Analytic linearization of circle diffeomorphisms*, in "Dynamical Systems and Small Divisors (Cetraro, 1998). Lecture Notes in Math. **1784**, 125–173, Springer-Verlag, Berlin.

S. Zakeri [1999], *Dynamics of cubic Siegel polynomials*, Comm. Math. Phys. **206**, 185–233.

E. Zehnder [1977], *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-Verlag, New York, 855-866.

This page intentionally left blank

INDEX