John Milnor

Dynamics in One Complex Variable

Third Edition

单复变动力系统 第3版

PRINCETON UNIVERSITY PRESS

世界图书出版公司
www.wpcbj.com.cn

Dynamics in One Complex Variable

THIRD EDITION

by

John Milnor

PRINCETON UNIVERSITY PRESS

PRINCETON AND OXFORD

2006

Second edition © 2000, published by Vieweg Verlag, Wiesbaden, Germany
First edition © 1999, published by Vieweg Verlag, Wiesbaden, Germany

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.

Annals of Mathematics Studies

Number 160

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).

TABLE OF CONTENTS

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.

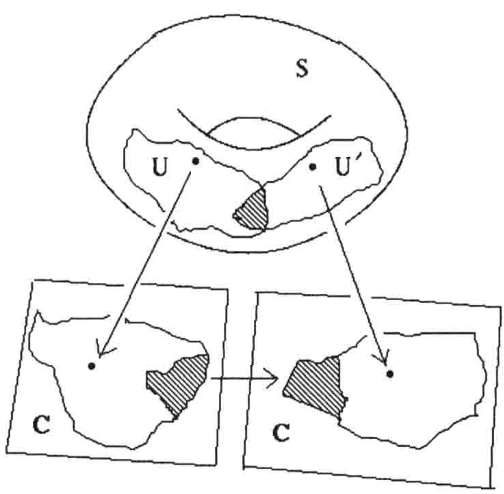

Figure 1. Overlapping coordinate neighborhoods.

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* $\hat{\mathbb{C}}$ *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* $z \ne 0$.

(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 $|q(0)| = |f'(0)| < 1$. □

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. □

(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 $\mathbf{PSL}(2,\mathbb{C})$. □

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 $|f(z)|<1 \iff |z|<1$. 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 $\mathcal{G}(\widehat{\mathbb{C}})$ 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 $\mathcal{G}(\widehat{\mathbb{C}})$). *For every $f \neq I$ in $\mathcal{G}(\widehat{\mathbb{C}})$, the set $\mathrm{Fix}(f) \subset \widehat{\mathbb{C}}$ contains either one point or two points. In general, two nonidentity elements $f, g \in \mathcal{G}(\widehat{\mathbb{C}})$ 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 $\widehat{\mathbb{C}}$. (If an automorphism of $\widehat{\mathbb{C}}$ 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. □

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