Cornerstones

Steven G. Krantz

Geometric Function Theory

Explorations in Complex Analysis

Birkhäuser
Boston • Basel • Berlin

Steven G. Krantz
Washington University
Department of Mathematics
St. Louis, MO 63130
U.S.A.
e-mail to: sk@math.wustl.edu
http://www.math.wustl.edu/~sk/

Cover design by Mary Burgess.

Mathematics Subject Classicification (2000): 30Axx, 30Bxx, 30Cxx, 30Dxx, 30Exx, 30Fxx, 30H05, 32Axx, 32Bxx, 32Dxx, 32Fxx, 31Axx, 35N15

Library of Congress Cataloging-in-Publication Data
Krantz, Steven G. (Steven George), 1951-
Geometric function theory : explorations in complex analysis / Steven G. Krantz.
p. cm. – (Cornerstones)
Includes bibliographical references and index.
ISBN 0-8176-4339-7 (alk. paper)
1. Geometric function theory. 2. Functions of complex variables. I. Title. II. Cornerstones (Birkhäuser)

QA331.7.K733 2005
515'.9–dc22 2005050071

ISBN-10 0-8176-4339-7 eISBN 0-8176-4440-7 Printed on acid-free paper.
ISBN-13 978-0-8176-4339-3

Printed in the United States of America. (TXQ/MP)

9 8 7 6 5 4 3 2 1

www.birkhauser.com

To the memory of Don Spencer

Contents

Preface

Complex analysis is a rich and textured subject. It is quite old, and its history is broad and deep. Yet the basic graduate course in complex variables has become rather cut and dried. The choice of topics, the order of the topics, and the overall flavor of the presentation are strongly influenced by the need to prepare students for the qualifying exams. The qual-level course is designed to serve a limited purpose, and it does that but little more.

Basic complex analysis is startling for its elegance and clarity. One progresses very rapidly from the basics of the Cauchy theory to profound results such as the fundamental theorem of algebra and the Riemann mapping theorem. Many a student is left, at the end of the course, yearning for more—to be advanced to a level where he or she could consider research questions, indeed the possibility of writing a thesis in the subject.

Yet there are few places to turn in such a quest. The book [BEG] of Berenstein and Gay and the book [CON] of Conway give particular takes on some of the more advanced material in the subject. Some of the older books, such as [FUKS], [GOL], [HIL], [MAR] treat topics not usually found in the basic texts. But it seems that there is a need for a book that will open the student's eyes to what this subject has to offer, and give him or her a taste of some of the areas of current research. This is meant to be such a book.

Our prejudice in the subject is geometric, but this does not prevent us from exploring byways that come from analysis, algebra, and other parts of mathematics. Thus, on the one hand, we treat invariant geometry, the Bergman metric, the automorphism groups of domains, and the boundary regularity of conformal mappings. On the other hand, we also explore the Hilbert transform, the Laplacian, the corona problem, harmonic measure, the inhomogenous Cauchy–Riemann equations, and sheaf theory.

The aim of the book is to expose the student to mathematics as it is practiced: as a synthesis of many different areas, exhibiting particular flavors and features that arise from that synthesis. The student who reads this book should be inspired to go further in the subject, to begin to explore the primary literature, and (one hopes) to think about his or her own research problems.

One of the rewards and pleasures that the student will find in reading this book is the rich interactions that are displayed among the various topics. For example, harmonic measure is used to prove a sharp version of the Lindelöf principle. It is also used to establish the three lines (viz. three circles) theorem of Hadamard. This in turn is used to prove (in another chapter) the Riesz–Thorin theorem. In another venue, the Riesz–Thorin theorem is used to prove the L^p-boundedness of the Hilbert transform. The Laplacian is reviewed and used as a device to introduce the Green's function. The Green's function is used, of course, to derive the Poisson kernel. But it is also used to develop the Bergman kernel. The Bergman kernel is used to study boundary regularity of conformal mappings. It is also used in the study of automorphism groups, and to prove various uniqueness theorems for conformal mappings. The Bergman metric interacts with and arises alongside consideration of other conformal metrics, and leads to Ahlfors's version of the Schwarz lemma.

The Poisson kernel is used to study the boundary behavior of harmonic and holomorphic functions. The Dirichlet problem for the Laplacian is used to give a nonstandard proof of the Riemann mapping theorem. The Green's function is used to prove a canonical representation of multiply connected regions on slit-domains. The Ahlfors map gives a new view of uniformization as introduced by these topics.

The Green's function and Stokes's theorem lead to a solution of the inhomogeneous Cauchy–Riemann equations, which are in turn used to make new constructions in function theory. The Green's function and the Hilbert transform, together with our solution of the inhomogeneous Cauchy–Riemann equations and the F. and M. Riesz theorem, are used to derive a proof of the corona theorem. Duality properties of Hardy spaces (covered in an earlier chapter) are exploited along the way.

Our study of the Ahlfors map uses Banach algebra properties of H^∞ that we developed in our study of the corona theorem. It also gives a reprise of the Green's function and harmonic measure. The Green's function and Green's theorem are used extensively in the proof of the corona theorem and also in our development of the uniformization theorem. Nontrivial ideas from functional analysis arise in our study of the Riesz–Thorin theorem, the Hilbert transform, the summation of Fourier series, the corona theorem, and the Ahlfors map.

We provide a discussion of the statement, concept, and proof of Köbe's uniformization theorem, together with various planar variants. Thus we examine uniformization from many different points of view. When we treat automorphism groups, uniformization provides a powerful tool.

Algebra is encountered in various guises throughout the book. Certainly it plays a role in the group-theoretic aspects of automorphisms. It occurs again in our treatment of Banach algebra techniques. And it plays a decisive role in the study of sheaves. Sheaf theory gives the student a new way to view the Weierstrass and Mittag–Leffler theorems, as well as questions of analytic continuation. Thus this text shows the students many different aspects of complex analysis, and how they interact with each other.

With this book, the student and the advanced worker too are introduced to a rich tapestry of function theory as it interacts with other parts of mathematics. There is hardly any other analysis text that offers such a variety and synthesis of mathematical topics.

Part of my own training is to think of the subject of complex analysis as a foil, or perhaps as a gadfly. Many an interesting problem of geometric analysis is very naturally formulated in the language of complex function theory; but then it is best solved by stripping away the complex analysis and applying tools of geometry, or partial differential equations, or harmonic analysis. That is the point of view that we shall, at least in part, take in the present book. Complex analysis will be our touchstone, but it will be the entrée to many another byway of mathematics—from the Cauchy–Riemann equations to interpolation of linear operators to the study of invariant metrics. It is our view that this is a productive and rewarding way to practice mathematics, and we would like to teach it to a new generation.

It is a pleasure to thank my editor, Ann Kostant, for encouraging me to write this book, and for making the process as smooth and carefree as possible. She enlisted strong and insightful reviewers to help me craft this book into a more precise and useful tool. I thank Elizabeth Loew for a marvelous job of editing and TEX typesetting. I look forward to comments and criticisms from the readership, and hope to make future editions more accurate and therefore more useful.

St. Louis, Missouri and Berkeley, California *Steven G. Krantz*

Part I

Classical Function Theory

1

Invariant Geometry

Genesis and Development

The idea of using invariant geometry to study complex function theory has its foundation in the ideas of Poincaré. Certainly he is credited with the creation of a conformally invariant metric on the unit disk D. The uniformization theorem (covered later in this book) may be used to transfer the metric to other planar domains. Later on, Stefan Bergman found a way to define invariant metrics on virtually any domain in any complex manifold. We shall explore his ideas further on in the book.

The geometric approach provides a new way to view the subject of complex variables. It is the source of tantalizing new questions. But it also provides a vast array of powerful new weapons to use on traditional problems. Any number of problems about mappings and conformality—just as an instance—are rendered transparent by way of geometric language.

An appreciation of the concepts in this chapter requires understanding of the idea of a Riemannian metric. Our presentation, however, is self-contained. The reader may learn what such a metric is by *doing*, that is, by studying this chapter. Even the student who is new to geometric language will find that the material in the ensuing pages will introduce him to a new way to approach function theory.

1.1 Conformality and Invariance

Conformal mappings are characterized by the fact that they infinitesimally **(i)** preserve angles, and **(ii)** preserve length (up to a scalar factor). It is worthwhile to picture the matter in the following manner: Let f be holomorphic on the open set $U \subseteq \mathbb{C}$. Fix a point $P \in U$. Write $f = u + iv$ as usual. Thus we may write the mapping f as $(x, y) \mapsto (u, v)$. Then the (real) Jacobian matrix of the mapping is

$$J(P) = \begin{pmatrix} u_x(P) & u_y(P) \\ v_x(P) & v_y(P) \end{pmatrix},$$

where subscripts denote derivatives. We may use the Cauchy–Riemann equations to rewrite this matrix as

$$J(P) = \begin{pmatrix} u_x(P) & u_y(P) \\ -u_y(P) & u_x(P) \end{pmatrix}.$$

Factoring out a numerical coefficient, we finally write this two-dimensional derivative as

$$J(P) = \sqrt{u_x(P)^2 + u_y(P)^2} \cdot \begin{pmatrix} \frac{u_x(P)}{\sqrt{u_x(P)^2+u_y(P)^2}} & \frac{u_y(P)}{\sqrt{u_x(P)^2+u_y(P)^2}} \\ \frac{-u_y(P)}{\sqrt{u_x(P)^2+u_y(P)^2}} & \frac{u_x(P)}{\sqrt{u_x(P)^2+u_y(P)^2}} \end{pmatrix}$$

$$\equiv h(P) \cdot \mathcal{J}(P).$$

The matrix $\mathcal{J}(P)$ is of course a special orthogonal matrix (i.e., its rows form an orthonormal basis of $\mathbb{R}^2$, and it is oriented positively). Thus we see that the derivative of our mapping is a rotation $\mathcal{J}(P)$ (which preserves angles) followed by a positive "stretching factor" $h(P)$ (which also preserves angles).

It would be incorrect to infer from these considerations that a conformal map therefore preserves Euclidean angles, or Euclidean length, in any global sense. Such a mapping would perforce be linear and in fact special orthogonal; and thus complex function theory would be reduced to a triviality. The mapping of the disk $D(0,2)$ given by

$$\phi : z \mapsto (z+4)^4,$$

illustrated in Figure 1.1, exhibits rather dramatically the failure of Euclidean isometry. The mapping is one-to-one, onto its image, yet $\phi(0) = 256$, $\phi(1) = 625$, and hence

$$1 - 0 = 1 \neq 369 = \phi(1) - \phi(0).$$

And yet it is obviously desirable to have a notion of distance that *is preserved* under holomorphic mappings. Part of Klein's *Erlangen program* is to understand geometric objects according to the groups that act on them. Such an understanding is of course facilitated when the group is actually a group of isometries.

It was H. Poincaré who first found a way to carry out this idea when the domain in question is the unit disk D. To understand his thinking, let us recall that the collection of all conformal maps of the unit disk $D = D(0,1)$ can be described explicitly: it consists of

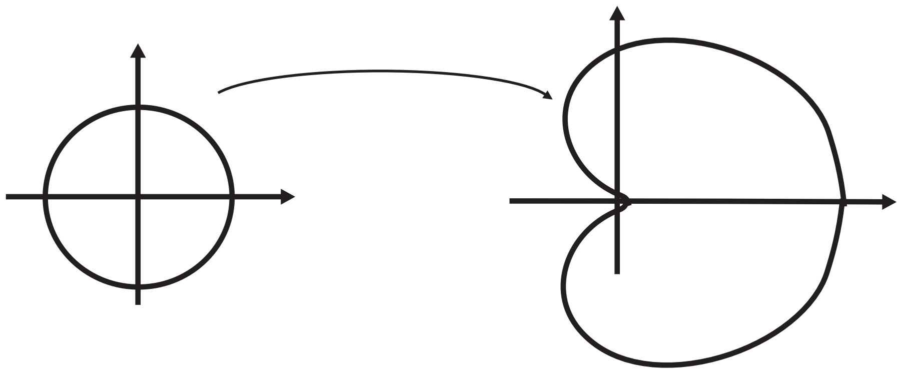

Fig. 1.1. The failure of Euclidean isometry under conformal mappings.

(i) All rotations $\rho_\lambda : z \mapsto e^{i\lambda} \cdot z$, $0 \le \lambda < 2\pi$;
(ii) All Möbius transformations $\varphi_a : z \mapsto [z - a]/[1 - \overline{a}z]$, $a \in \mathbb{C}$, $|a| < 1$;
(iii) All compositions of mappings of type **(i)** and **(ii)**.

We will use Riemann's paradigm for a metric, that is, we shall specify the length of a (tangent) vector at each point of the disk D. If the point is P and the vector is $\mathbf{v}$, then let us denote this length by $|\mathbf{v}|_P$.

Our goal now is to "discover" the Poincaré metric by way of a sequence of calculations. We must begin somewhere, so let us declare that the length of the vector $\mathbf{e} \equiv (1, 0) \equiv 1 + 0i$ at the origin is 1. Thus $|\mathbf{e}|_0 = 1$. Now if ϕ is a conformal self-map of the disk, then invariance means that

$$|\mathbf{v}|_P = |\phi_*(P)\mathbf{v}|_{\phi(P)} , \tag{1.1.1}$$

where $\phi_*(P)\mathbf{v}$ is defined to equal $\phi'(P) \cdot \mathbf{v}$. Here $\phi_*(P)\mathbf{v}$ is called the "push-forward by ϕ" of the vector $\mathbf{v}$.

Now let us apply equation (1.1.1) to the information that we have about the length of the unit vector $\mathbf{e}$ at the origin. Let $\phi(z) = e^{i\lambda} \cdot z$. Then we see that

$$1 = |\mathbf{e}|_0 = |\phi_*(0)\mathbf{e}|_{\phi(0)} = |e^{i\lambda} \cdot \mathbf{e}|_0 = |e^{i\lambda}|_0 . \tag{1.1.2}$$

We conclude that the length of the vector $e^{i\lambda}$ at the origin is 1. So all Euclidean unit vectors based at the origin have length 1 in our new invariant metric.

Now let $\psi(z)$ be the Möbius transformation

$$\psi(z) = \frac{z + a}{1 + \overline{a}z} ,$$

with a a complex number of modulus less than 1. Let $\mathbf{v} = e^{i\lambda}$ be a unit vector at the origin. Then $\psi'(0) = 1 - |a|^2$ and equation (1.1.1) tells us that

$$1 = |\mathbf{v}|_0 = |\psi_*(0)\mathbf{v}|_{\psi(0)} = |(1-|a|^2)\cdot \mathbf{v}|_a \, .$$

We conclude that

$$|\mathbf{v}|_a = \frac{1}{1-|a|^2} \, .$$

Of course our new metric will respect scalar multiplication, so we may apply the preceding calculation to vectors of any length. If we let $\|\mathbf{v}\|$ denote the *Euclidean length* of a vector $\mathbf{v}$, then we may summarize all our calculations as follows:

> Let P be a point of the unit disk D and let $\mathbf{v}$ be any vector based at that point. Then
>
> $$|\mathbf{v}|_P = \frac{\|\mathbf{v}\|}{1-|P|^2} \, .$$
>
> We call this new metric the *Poincaré metric*.

We see in particular that, as P gets ever nearer to the boundary of D, the new metric length of $\mathbf{v}$ becomes greater and greater.

It is well worth spending some time interpreting this new metric and its implications for complex analysis. Let $\gamma : [0,1] \to D$ be a continuously differentiable curve. We define the *length* of γ in the Poincaré metric to be

$$\ell(\gamma) = \int_0^1 |\gamma'(t)|_{\gamma(t)} \, dt \, .$$

Observe that $\gamma'(t)$ is a vector located at $\gamma(t)$, so the definition makes good sense. Of course we can consider curves parametrized over *any* interval, and the definition of length is independent (by the change of variables formula of calculus) of the choice of parametrization. We define the length of a *piecewise continuously differentiable* curve to be the sum of the lengths of its continuously differentiable pieces.

Example 1.1.1. Let $\epsilon > 0$. Consider the curve $\gamma(t) = (1-\epsilon)t$, $0 \le t \le 1$. Then, according to the definition,

$$\begin{aligned}\ell(\gamma) &= \int_0^1 |\gamma'(t)|_{\gamma(t)} \, dt = \int_0^1 \frac{(1-\epsilon)}{1-|\gamma(t)|^2} \, dt \\ &= \int_0^1 \frac{(1-\epsilon)}{1-[(1-\epsilon)t]^2} \, dt = \frac{1}{2} \log\left(\frac{2-\epsilon}{\epsilon}\right) \, .\end{aligned}$$

We see immediately that, as $\epsilon \to 0^+$, the expression $\ell(\gamma)$ tends to $+\infty$. Thus, in some sense, the distance from 0 to the boundary (at least along the given linear path) is infinite.

We define the *Poincaré distance* $d(P,Q)$ between two points $P, Q \in D$ to be the infimum of the Poincaré lengths of all piecewise continuously differentiable curves connecting P to Q.

Proposition 1.1.2. *Let $P \in D$. Then the Poincaré distance of* 0 *to* P *is equal to*

$$d(0, P) = \frac{1}{2} \cdot \log \frac{1 + |P|}{1 - |P|} .$$

Proof. By rotational invariance, we may as well suppose that P is real and positive, so $P = (1 - \epsilon) + i0 \equiv (1 - \epsilon, 0)$. It is an exercise for the reader to see that there is no loss of generality to consider only curves of the form $\gamma(t) = (t, g(t))$, $0 \leq t \leq 1 - \epsilon$. Then

$$\begin{aligned} \ell(g) &= \int_0^{1-\epsilon} \frac{\|\gamma'(t)\|}{1 - |\gamma(t)|^2} \, dt = \int_0^{1-\epsilon} \frac{\sqrt{1^2 + |g'(t)|^2}}{1 - t^2 - |g^2(t)|} \, dt \\ &\geq \int_0^{1-\epsilon} \frac{1}{1 - t^2} \, dt = \frac{1}{2} \cdot \log \left(\frac{2 - \epsilon}{\epsilon} \right) \\ &= \frac{1}{2} \cdot \log \left(\frac{1 + |P|}{1 - |P|} \right) . \end{aligned}$$

Thus we see explicitly that $\mu(t) = (t, 0)$ is the shortest curve from 0 to P, and the distance is as we claimed. □

Combining the result of this proposition with the preceding example, we see that any curve that starts at the origin and runs out to the boundary of the disk will have infinite length. Thus the boundary is infinitely far away.

Exercise for the Reader: Prove that if P is any point of the disk and $r > 0$, then the metric disk

$$\beta(P, r) \equiv \{z \in D : d(z, P) < r\}$$

is actually a Euclidean disk. What are its Euclidean center and radius (expressed in terms of r and P)? Show that the disk $\beta(P, r)$ is relatively compact (i.e., has compact closure) in the disk. It follows that any Cauchy sequence *in the Poincaré metric* has a limit point *in the disk*. Thus the Poincaré metric turns D into a complete metric space. ♢

1.2 Bergman's Construction

Stefan Bergman created a device for equipping virtually *any* planar domain with an invariant metric that is analogous to the Poincaré metric on the disk.[1] Some tracts call this new metric the *Poincaré–Bergman metric*, though it is

[1] Later on, in Section 4.6, we shall treat the uniformization theorem of Köbe. It gives a means for transferring the Poincaré metric from the disk to virtually any planar domain.

more commonly called just the *Bergman metric*. In order to construct the Bergman metric we must first construct the Bergman kernel. For that we need just a little Hilbert space theory (see [RUD2], for example).

A *domain* in $\mathbb{C}$ is a connected open set. Fix a domain $\Omega \subseteq \mathbb{C}$, and define

$$A^2(\Omega) = \left\{ f \text{ holomorphic on } \Omega : \int_\Omega |f(z)|^2 \, dA(z) < \infty \right\} \subseteq L^2(\Omega) .$$

Here dA is an ordinary two-dimensional area measure. Of course $A^2(\Omega)$ is a complex linear space, called the *Bergman space*. The norm on $A^2(\Omega)$ is given by

$$\|f\|_{A^2(\Omega)} = \left[\int_\Omega |f(z)|^2 \, dA(z) \right]^{1/2} .$$

We define an inner product on $A^2(\Omega)$ by

$$\langle f, g \rangle = \int_\Omega f(z)\overline{g(z)} \, dA(z) .$$

The next technical lemma will be the key to our analysis of the space A^2.

Lemma 1.2.1. *Let $K \subseteq \Omega$ be compact. There is a constant $C_K > 0$, depending on K, such that*

$$\sup_{z \in K} |f(z)| \leq C_K \|f\|_{A^2(\Omega)} , \quad \text{all } f \in A^2(\Omega) .$$

Proof. Since K is compact, there is an $r(K) = r > 0$ so that, for any $z \in K, D(z, r) \subseteq \Omega$. Therefore, for each $z \in K$ and $f \in A^2(\Omega)$, we may use the mean value property of harmonic functions to see that

$$\begin{aligned} |f(z)| &= \frac{1}{A(B(z,r))} \left| \int_{B(z,r)} f(t) \, dA(t) \right| \\ &\leq \frac{1}{A(B(z,r))} \cdot \int_{\mathbb{C}} \chi_{B(z,r)}(t) \cdot |f(t)| \, dA(t) , \end{aligned}$$

where

$$\chi_{B(z,r)}(t) = \begin{cases} 1 & \text{if} \quad t \in B(z,r) \\ 0 & \text{if} \quad t \notin B(z,r) . \end{cases}$$

We apply the Cauchy–Schwarz inequality from integration theory (see [RUD2]) to the last expression to find that it is less than or equal to

$$\begin{aligned} &\frac{1}{A(B(z,r))} \cdot \int_{\mathbb{C}} |\chi_{B(z,r)}(t)|^2 \, dA^{1/2} \cdot \int_{\mathbb{C}} |f(t)|^2 \, dA^{1/2} \\ &= (A(B(z,r)))^{-1/2} \|f\|_{A^2(B(z,r))} = \frac{1}{\sqrt{\pi} r} \|f\|_{A^2(\Omega)} \\ &\equiv C_K \|f\|_{A^2(\Omega)} . \qquad \square \end{aligned}$$

Proposition 1.2.2. *The space $A^2(\Omega)$ is complete.*

Proof. Let $\{f_j\}$ be a Cauchy sequence in A^2. Then the sequence is Cauchy in $L^2(\Omega)$, and the completeness of L^2 (see [RUD2]) then tells us that there is a limit function f. So $f_j \to f$ in the L^2 topology. Now the lemma tells us that in fact the convergence is taking place uniformly on compact sets. So $f \in A^2(\Omega)$. That completes the argument. □

Corollary 1.2.3. *The space $A^2(\Omega)$ is a Hilbert space.*

With a little extra effort (see [GRK1], [RUD3]), it can be shown that $A^2(\Omega)$ is in fact a *separable* Hilbert space.

Now our point of view is to find a method for representing certain linear functionals. The key fact is this:

Lemma 1.2.4. *For each fixed $z \in \Omega$, the functional*

$$\Phi_z : f \mapsto f(z), \quad f \in A^2(\Omega)$$

is a continuous linear functional on $A^2(\Omega)$.

Proof. This is immediate from Lemma 1.2.1 if we take K to be the singleton $\{z\}$. □

We may now apply the Riesz representation theorem (see [RUD2]) to see that there is an element $k_z \in A^2(\Omega)$ such that the linear functional Φ_z is represented by inner product with k_z: if $f \in A^2(\Omega)$ then, for all $z \in \Omega$, we have

$$f(z) = \langle f, k_z \rangle \, . \tag{1.2.1}$$

Definition 1.2.5. The *Bergman kernel* is the function $K(z,\zeta) = \overline{k_z(\zeta)}$, $z, \zeta \in \Omega$. It has the reproducing property

$$f(z) = \int K(z,\zeta) f(\zeta) \, dA(\zeta), \quad \forall f \in A^2(\Omega) \, . \tag{1.2.2}$$

Notice that (1.2.2) is just a restatement of (1.2.1).

Proposition 1.2.6. *The Bergman kernel $K(z,\zeta)$ is conjugate symmetric: $K(z,\zeta) = \overline{K(\zeta,z)}$.*

Proof. By its very definition, $\overline{K(\zeta,\cdot)} \in A^2(\Omega)$ for each fixed ζ. Therefore the reproducing property of the Bergman kernel gives

$$\int_\Omega K(z,t)\overline{K(\zeta,t)} \, dA(t) = \overline{K(\zeta,z)} \, .$$

On the other hand,

$$\begin{aligned} \int_\Omega K(z,t)\overline{K(\zeta,t)} \, dA(t) &= \overline{\int K(\zeta,t)\overline{K(z,t)} \, dA(t)} \\ &= \overline{\overline{K(z,\zeta)}} = K(z,\zeta) \, . \end{aligned}$$ □

Proposition 1.2.7. *The Bergman kernel is uniquely determined by the properties that it is an element of $A^2(\Omega)$ in z, is conjugate symmetric, and reproduces $A^2(\Omega)$.*

Proof. Let $\widetilde{K}(z,\zeta)$ be another such kernel. Then

$$\begin{aligned} K(z,\zeta) = \overline{K(\zeta,z)} &= \int \widetilde{K}(z,t)\overline{K(\zeta,t)}\, dA(t) \\ &= \overline{\int K(\zeta,t)\overline{\widetilde{K}(z,t)}\, dA(t)} = \overline{\overline{\widetilde{K}(z,\zeta)}} = \widetilde{K}(z,\zeta)\,. \end{aligned}$$ □

Since $A^2(\Omega)$ is a separable Hilbert space, there is a complete orthonormal basis $\{\phi_j\}_{j=1}^\infty$ for $A^2(\Omega)$.

Proposition 1.2.8. *Let E be a compact subset of Ω. Then the series*

$$\sum_{j=1}^\infty \phi_j(z)\overline{\phi_j(\zeta)}$$

sums uniformly on $E \times E$ to the Bergman kernel $K(z,\zeta)$.

Proof. By the Riesz–Fisher and Riesz representation theorems, we obtain

$$\begin{aligned} \sup_{z\in E}\left(\sum_{j=1}^\infty |\phi_j(z)|^2\right)^{1/2} &= \sup_{z\in E}\left\|\{\phi_j(z)\}_{j=1}^\infty\right\|_{\ell^2} \\ &= \sup_{\substack{\|\{a_j\}\|_{\ell^2}=1 \\ z\in E}}\left|\sum_{j=1}^\infty a_j\phi_j(z)\right| \\ &= \sup_{\substack{\|f\|_{A^2}=1 \\ z\in E}} |f(z)| \le C_E\,. \end{aligned} \tag{1.2.3}$$

In the last inequality we have used Lemma 1.2.1. Therefore

$$\sum_{j=1}^\infty \left|\phi_j(z)\overline{\phi_j(\zeta)}\right| \le \left(\sum_{j=1}^\infty |\phi_j(z)|^2\right)^{1/2}\left(\sum_{j=1}^\infty |\phi_j(\zeta)|^2\right)^{1/2}$$

and the convergence is uniform over $z,\zeta \in E$. For fixed $z \in \Omega$, our preceding calculation shows that $\{\phi_j(z)\}_{j=1}^\infty \in \ell^2$. Hence we have that $\sum \phi_j(z)\overline{\phi_j(\zeta)} \in \overline{A^2(\Omega)}$ as a function of ζ. Let the sum of the series be denoted by $\widetilde{K}(z,\zeta)$. Notice that $\widetilde{K}$ is conjugate symmetric by its very definition. Also, for $f \in A^2(\Omega)$, we have

$$\int \widetilde{K}(\cdot,\zeta)f(\zeta)\, dA(\zeta) = \sum \widehat{f}(j)\phi_j(\cdot) = f(\cdot),$$

where convergence is in the Hilbert space topology. [Here $\widehat{f}(j)$ is the j^{th} Fourier coefficient of f with respect to the basis $\{\phi_j\}$.] But Hilbert space convergence dominates pointwise convergence (Lemma 1.2.1), so

$$f(z) = \int \widetilde{K}(z, \zeta) f(\zeta)\, dA(\zeta), \quad \text{all } f \in A^2(\Omega).$$

Therefore, by Proposition 1.2.7, $\widetilde{K}$ is the Bergman kernel. □

Proposition 1.2.9. *If Ω is a bounded domain in $\mathbb{C}$, then the mapping*

$$P : f \mapsto \int_\Omega K(\cdot, \zeta) f(\zeta)\, dA(\zeta)$$

is the Hilbert space orthogonal projection of $L^2(\Omega, dA)$ onto $A^2(\Omega)$.

Proof. Notice that P is idempotent and self-adjoint and that $A^2(\Omega)$ is precisely the set of elements of L^2 that are fixed by P. □

Lemma 1.2.10. *Let $\phi : \Omega_1 \to \Omega_2$ be a one-to-one, onto, conformal mapping of domains. Then*

$$\int_{\Omega_1} |\phi'(z)|^2\, dx\, dy = \int_{\Omega_2} dx\, dy\,.$$

Proof. Write $\phi = u + iv$. Then the Jacobian determinant of ϕ, thought of as a real mapping from $\mathbb{R}^2$ to $\mathbb{R}^2$, is

$$J = \det \begin{pmatrix} u_x & u_y \\ v_x & v_y \end{pmatrix},$$

which, by the Cauchy–Riemann equations, equals

$$\det \begin{pmatrix} u_x & -v_x \\ v_x & u_x \end{pmatrix} = u_x^2 + v_x^2 = |\phi_x|^2 = |\phi'|^2\,.$$

The last equality is obtained by another application of the Cauchy–Riemann equations.

As a result,

$$\iint\limits_{\Omega_2} 1\, dA = \iint\limits_{\Omega_1} J\, dA = \iint\limits_{\Omega_1} |\phi'(z)|^2\, dx\, dy\,. \qquad \square$$

This last result is called the *Lusin area integral formula.*

Proposition 1.2.11. *Let Ω_1, Ω_2 be domains in $\mathbb{C}$. Let $f : \Omega_1 \to \Omega_2$ be conformal. Then*

$$f'(z) K_{\Omega_2}(f(z), f(\zeta)) \overline{f'(\zeta)} = K_{\Omega_1}(z, \zeta)\,.$$

Proof. Let $\phi \in A^2(\Omega_1)$. Then, by change of variable,

$$\int_{\Omega_1} \left[f'(z) K_{\Omega_2}(f(z), f(\zeta)) \overline{f'(\zeta)} \right] \phi(\zeta)\, dA(\zeta)$$
$$= \int_{\Omega_2} f'(z) K_{\Omega_2}(f(z), \widetilde{\zeta}) \overline{f'(f^{-1}(\widetilde{\zeta}))} \phi(f^{-1}(\widetilde{\zeta})) \cdot \frac{1}{|f'(f^{-1}(\widetilde{\zeta}))|^2}\, dA(\widetilde{\zeta}).$$

In the last equality, we have used Lemma 1.2.10. This simplifies to

$$f'(z) \int_{\Omega_2} K_{\Omega_2}(f(z), \widetilde{\zeta}) \left\{ \frac{1}{f'(f^{-1}(\widetilde{\zeta}))} \phi\left(f^{-1}(\widetilde{\zeta}) \right) \right\} dA(\widetilde{\zeta}).$$

By change of variables, the expression in braces $\{\ \}$ is an element of $A^2(\Omega_2)$. So the reproducing property of K_{Ω_2} applies and the last line equals

$$f'(z) \cdot \frac{1}{f'(z)} \cdot \phi\,(z) = \phi(z)\,.$$

By the uniqueness of the Bergman kernel, the proposition follows. □

Proposition 1.2.12. *For $z \in \Omega \subset\subset \mathbb{C}$ it holds that $K_\Omega(z,z) > 0$.*

Proof. If $\{\phi_j\}$ is a complete orthonormal basis for $A^2(\Omega)$, then

$$K_\Omega(z,z) = \sum_{j=1}^{\infty} |\phi_j(z)|^2 \geq 0\,.$$

If in fact $K(z,z) = 0$ for some z, then $\phi_j(z) = 0$ for all j; hence $f(z) = 0$ for every $f \in A^2(\Omega)$. This is absurd. □

As an exercise, the reader may wish to verify that the last proposition is true with only the simple assumption that the area of Ω is finite.

Definition 1.2.13. For any bounded $\Omega \subseteq \mathbb{C}$ we define a Hermitian metric on Ω by

$$g(z) = \frac{\partial^2}{\partial z \partial \overline{z}} \log K(z,z), \quad z \in \Omega\,.$$

This means that the length of a tangent vector ξ at a point $z \in \Omega$ is given by

$$|\xi|_z = g(z) \|\xi\|.$$

The metric that we have defined is called the *Bergman metric*. Sometimes, for clarity, we denote the metric by g^Ω.

We note that the length of a curve, and the distance $d_\Omega(z,w)$ between two points z, w, is now defined just as for the Poincaré metric on the disk.

Proposition 1.2.14. *Let $\Omega_1, \Omega_2 \subseteq \mathbb{C}$ be domains and let $f : \Omega_1 \to \Omega_2$ be a conformal mapping. Then f induces an isometry of Bergman metrics:*

$$|\xi|_z = |f'(z) \cdot \xi|_{f(z)} \tag{1.2.4}$$

for all $z \in \Omega_1, \xi \in \mathbb{C}$. Equivalently, f induces an isometry of Bergman distances in the sense that

$$d_{\Omega_2}(f(P), f(Q)) = d_{\Omega_1}(P, Q).$$

Proof. This is a formal exercise, but we include it for completeness:

From the definitions, it suffices to check that

$$|g^{\Omega_2}(f(z))f'(z)w| = |g^{\Omega_1}(z)w|$$

for all $z \in \Omega, w \in \mathbb{C}$. But, by Proposition 1.2.11,

$$\begin{aligned} g^{\Omega_1}(z) &= \frac{\partial^2}{\partial z \partial \overline{z}} \log K_{\Omega_1}(z, z) \\ &= \frac{\partial^2}{\partial z \partial \overline{z}} \log \left\{ |f'(z)|^2 K_{\Omega_2}(f(z), f(z)) \right\} \\ &= \frac{\partial^2}{\partial z \partial \overline{z}} \log K_{\Omega_2}(f(z), f(z)) \end{aligned} \tag{1.2.5}$$

since, locally,

$$\log |f'(z)|^2 = \log (f'(z)) + \log \left(\overline{f'(z)} \right) + C \, .$$

But line (1.2.5) is nothing other than

$$\left| g^{\Omega_2}(f(z)) \right| \frac{\partial f(z)}{\partial z} \frac{\partial \overline{f(z)}}{\partial \overline{z}} = \left| g^{\Omega_2}(f(z)) \right| \left| \frac{\partial f(z)}{\partial z} \right|^2$$

and (1.2.4) follows. □

1.3 Calculation of the Bergman Kernel for the Disk

Proposition 1.3.1. *The Bergman kernel for the unit disk D is*

$$K(z, \zeta) = \frac{1}{\pi} \cdot \frac{1}{(1 - z\overline{\zeta})^2} \, .$$

The Bergman metric for the disk is

$$g(z) = \frac{2}{(1 - |z|^2)^2} \, .$$

This is (up to a constant multiple) the well-known Poincaré, or Poincaré–Bergman, metric.

This fact is so important that we now present three proofs. Some interesting function theory will occur along the way.

1.3.1 Construction of the Bergman Kernel for the Disk by Conformal Invariance

Let $D \subseteq \mathbb{C}$ be the unit disk. First we notice that, if either $f \in A^2(D)$ or $\overline{f} \in A^2(D)$, then

$$f(0) = \frac{1}{\pi} \iint_D f(\zeta)\, dA(\zeta)\,. \tag{1.3.1}$$

This is the standard, two-dimensional area form of the mean value property for holomorphic or harmonic functions.

Of course the constant function $u(z) \equiv 1$ is in $A^2(D)$, so it is reproduced by integration against the Bergman kernel. Hence, for any $w \in D$,

$$1 = u(w) = \iint_D K(w,\zeta)u(\zeta)\, dA(\zeta) = \iint_D K(w,\zeta)\, dA(\zeta)\,,$$

or

$$\frac{1}{\pi} = \frac{1}{\pi} \iint_D K(w,\zeta)\, dA(\zeta)\,.$$

By (1.3.1), we may conclude that

$$\frac{1}{\pi} = K(w,0)$$

for any $w \in D$.

Now, for $a \in D$ fixed, consider the Möbius transformation

$$h(z) = \frac{z-a}{1-\overline{a}z}\,.$$

We know that

$$h'(z) = \frac{1-|a|^2}{(1-\overline{a}z)^2}\,.$$

We may thus apply Proposition 1.2.11 with $\phi = h$ to find that

$$\begin{aligned} K(w,a) &= h'(w) \cdot K(h(w),h(a)) \cdot \overline{h'(a)} \\ &= \frac{1-|a|^2}{(1-\overline{a}w)^2} \cdot K(h(w),0) \cdot \frac{1}{1-|a|^2} \\ &= \frac{1}{(1-\overline{a}w)^2} \cdot \frac{1}{\pi} \\ &= \frac{1}{\pi} \cdot \frac{1}{(1-w\overline{a})^2}\,. \end{aligned}$$

This is our formula for the Bergman kernel. The formula for the Bergman metric follows immediately by differentiation.

1.3.2 Construction of the Bergman Kernel by means of an Orthonormal System

Now we will endeavor to write down the Bergman kernel for the disk by means of an orthonormal basis for $A^2(D)$, that is, by applying Proposition 1.2.8. For a general Ω, it can be rather difficult to actually write down a complete orthonormal system for $A^2(\Omega)$. Fortunately, the unit disk $D \subseteq \mathbb{C}$ has enough symmetry that we can actually pull this off.

It is not difficult to see that $\{z^j\}_{j=0}^{\infty}$ is an *orthogonal* system for $A^2(D)$. That is, the elements are pairwise orthogonal; but they are not normalized to have unit length. We may confirm the first of these assertions by noting that, if $j \neq k$, then

$$\begin{aligned}\langle z^j, z^k \rangle &= \iint_D z^j \overline{z^k}\, dA(z) = \int_0^1 \int_0^{2\pi} r^j e^{ij\theta} r^k e^{-ik\theta}\, d\theta\, r dr \\ &= \int_0^1 r^{j+k+1}\, dr \int_0^{2\pi} e^{i(j-k)\theta}\, d\theta = 0\,.\end{aligned}$$

The system $\{z^j\}$ is complete: If $\langle f, z^j \rangle = 0$ for every j, then f will have a null power series expansion and hence be identically zero. It remains to normalize these monomials so that we have a complete ortho*normal* system.

We calculate that

$$\begin{aligned}\iint_D |z^j|^2\, dA(z) &= \int_0^1 \int_0^{2\pi} r^{2j}\, d\theta\, r dr \\ &= 2\pi \int_0^1 r^{2j+1}\, dr \\ &= \pi \cdot \frac{1}{j+1}\,.\end{aligned}$$

We conclude that

$$\|z^j\| = \frac{\sqrt{\pi}}{\sqrt{j+1}}\,.$$

Therefore the elements of our orthonormal system are

$$\phi_j(z) = \frac{\sqrt{j+1} \cdot z^j}{\sqrt{\pi}}\,.$$

Now, according to Proposition 1.2.8, the Bergman kernel is given by

$$\begin{aligned}K(z, \zeta) &= \sum_{j=0}^{\infty} \phi_j(z) \cdot \overline{\phi_j(\zeta)} \\ &= \sum_{j=0}^{\infty} \frac{(j+1) z^j \overline{\zeta^j}}{\pi}\end{aligned}$$

$$= \frac{1}{\pi} \cdot \sum_{j=0}^{\infty} (j+1) \cdot (z\bar{\zeta})^j .$$

Observe that, in this instance, the convergence of the series is manifest (for both $|z| < 1$ and $|\zeta| < 1$).

Of course we easily can sum $\sum_j (j+1)\alpha^j$ by noticing that

$$\begin{aligned}
\sum_{j=0}^{\infty} (j+1)\alpha^j &= \frac{d}{d\alpha} \sum_{j=0}^{\infty} \alpha^{j+1} \\
&= \frac{d}{d\alpha} \left[\alpha \cdot \frac{1}{1-\alpha} \right] \\
&= \frac{1}{(1-\alpha)^2} .
\end{aligned}$$

Applying this result to our expression for $K(z, \zeta)$ yields that

$$K(z, \zeta) = \frac{1}{\pi} \cdot \frac{1}{(1 - z\bar{\zeta})^2} .$$

This is consistent with the formula that we obtained by conformal invariance in Subsection 1.3.1 for the Bergman kernel of the disk. The formula for the Bergman metric follows immediately by differentiation.

1.3.3 Construction of the Bergman Kernel by way of Differential Equations

It is actually possible to obtain the Bergman kernel of a domain in the plane from the Green's function for that domain (see [EPS]). Let us now summarize the key ideas. Chapter 8, especially Section 8.2, contains more detailed information about the Green's function. Unlike the first two Bergman kernel constructions, the present one will work for *any* domain with C^2 boundary.

First, the fundamental solution for the Laplacian in the plane is the function

$$\Gamma(\zeta, z) = \frac{1}{2\pi} \log |\zeta - z|$$

(in Proposition 8.1.4 we shall prove this assertion). This means that $\triangle_\zeta \Gamma(\zeta, z) = \delta_z$. [Observe that δ_z denotes the Dirac "delta mass" at z and $\triangle_\zeta$ is the Laplacian in the ζ variable.] Here the derivatives are interpreted in the sense of distributions. In more prosaic terms, the condition is that

$$\int \Gamma(\zeta, z) \cdot \triangle \varphi(\zeta) \, d\xi d\eta = \varphi(z)$$

for any C^2 function φ with compact support. We write, as usual, $\zeta = \xi + i\eta$.

Given a domain $\Omega \subseteq \mathbb{C}$, the *Green's function* is posited to be a function $G(\zeta, z)$ that satisfies

$$G(\zeta, z) = \Gamma(\zeta, z) - F_z(\zeta)\,,$$

where $F_z(\zeta) = F(\zeta, z)$ is a particular harmonic function in the ζ variable. Moreover, it is mandated that $G(\,\cdot\,, z)$ vanish on the boundary of Ω. One constructs the function $F(\,\cdot\,, z)$, for each fixed z, by solving a suitable Dirichlet problem. Again, the reference [KRA1, p. 40] has all the particulars. It is worth noting, and this point will be discussed later, that the Green's function is a symmetric function of its arguments.

The next proposition establishes a striking connection between the Bergman kernel and the classical Green's function.

Proposition 1.3.2. *Let $\Omega \subseteq \mathbb{C}$ be a bounded domain with C^2 boundary. Let $G(\zeta, z)$ be the Green's function for Ω and let $K(z, \zeta)$ be the Bergman kernel for Ω. Then*

$$K(z, \zeta) = 4 \cdot \overline{\frac{\partial^2}{\partial \zeta \partial \overline{z}} G(\zeta, z)}\,. \tag{1.3.2}$$

Proof. Our proof will use a version of Stokes's theorem written in the notation of complex variables. It says that if $u \in C^1(\overline{\Omega})$, then

$$\oint_{\partial U} u(\zeta)\, d\zeta = 2i \cdot \iint\limits_U \frac{\partial u}{\partial \overline{\zeta}}\, d\xi\, d\eta\,, \tag{1.3.3}$$

where again $\zeta = \xi + i\eta$. The reader is invited to convert this formula to an expression in ξ and η and to confirm that the result coincides with the standard real-variable version of Stokes's theorem that can be found in any calculus book (see, e.g., [THO], [BLK]).

Now we already know that

$$G(\zeta, z) = \frac{1}{4\pi} \log(\zeta - z) + \frac{1}{4\pi} \log \overline{(\zeta - z)} + F(\zeta, z)\,. \tag{1.3.4}$$

Here we think of the logarithm as a multivalued holomorphic function; after we take a derivative, the ambiguity (which comes from an additive multiple of $2\pi i$) goes away.

Differentiating with respect to z (and using subscripts to denote derivatives), we find that

$$G_z(\zeta, z) = \frac{1}{4\pi} \frac{-1}{\zeta - z} + F_z(\zeta, z)\,.$$

We may rearrange this formula to read

$$\frac{1}{\zeta - z} = -4\pi \cdot G_z(\zeta, z) + 4\pi F_z(\zeta, z)\,.$$

We know that G, as a function of ζ, vanishes on $\partial\Omega$. Hence so does G_z. Let $f \in C^2(\overline{\Omega})$ be holomorphic on Ω. It follows that the Cauchy formula

$$f(z) = \frac{1}{2\pi i} \oint_{\partial\Omega} \frac{f(\zeta)}{\zeta - z}\, d\zeta$$

can be rewritten as

$$f(z) = -2i \oint_{\partial\Omega} f(\zeta) F_z(\zeta, z)\, d\zeta\,.$$

Now we apply Stokes's theorem (in the complex form) to rewrite this last as

$$f(z) = 4 \cdot \iint_{\Omega} (f(\zeta)F_z)_{\overline{\zeta}}(\zeta, z)\, d\xi\, d\eta\,,$$

where $\zeta = \xi + i\eta$. Since f is holomorphic and F is real-valued, we may conveniently write this last formula as

$$f(z) = 4 \cdot \iint_{\Omega} f(\zeta)\overline{F_{\zeta\overline{z}}}(\zeta, z)\, d\xi\, d\eta\,.$$

Now formula (1.3.4) tells us that $F_{\zeta\overline{z}} = G_{\zeta\overline{z}}$. Therefore we have

$$f(z) = \iint_{\Omega} f(\zeta) 4\overline{G_{\zeta\overline{z}}}(\zeta, z)\, d\xi\, d\eta\,. \tag{1.3.5}$$

With a suitable limiting argument, we may extend this formula from functions f that are holomorphic and in $C^2(\overline{\Omega})$ to functions in $A^2(\Omega)$.

It is straightforward now to verify that $4\overline{G_{\zeta\overline{z}}}$ satisfies the first three characterizing properties of the Bergman kernel, just by examining our construction. The crucial reproducing property is of course formula (1.3.5). Then it follows that

$$K(z, \zeta) = 4 \cdot \overline{\frac{\partial^2}{\partial\zeta\partial\overline{z}} G(\zeta, z)}\,.$$

That is the desired result. □

It is worth noting that the proposition we have just established gives a practical method for confirming the existence of the Bergman kernel—by relating it to the Green's function, whose existence is elementary.

Now let us calculate. Of course the Green's function of the unit disk D is

$$G(\zeta, z) = \frac{1}{2\pi} \log|\zeta - z| - \frac{1}{2\pi} \log|1 - \zeta\overline{z}|\,,$$

as a glance at any classical complex analysis text will tell us (see, for example, [AHL2] or [HIL]). Verify the defining properties of the Green's function for yourself.

With formula (1.3.2) in mind, we can make life a bit easier by writing

$$G(\zeta, z) = \frac{1}{4\pi}\log(\zeta - z) + \frac{1}{4\pi}\log(\overline{\zeta - z})$$
$$-\frac{1}{4\pi}\log\left(1 - \zeta\overline{z}\right) - \frac{1}{4\pi}\log\left(\overline{1 - \zeta\overline{z}}\right) .$$

Here we think of the expression on the right as the concatenation of four multivalued functions, in view of the ambiguity of the logarithm function. This ambiguity is irrelevant for us because the derivative of the Green's function is still well defined (i.e., the derivative annihilates additive constants).

Now we readily calculate that

$$\frac{\partial G}{\partial \overline{z}} = \frac{1}{4\pi} \cdot \frac{-1}{\overline{\zeta} - \overline{z}} + \frac{1}{4\pi} \cdot \frac{\zeta}{1 - \zeta\overline{z}}$$

and

$$\frac{\partial^2 G}{\partial \zeta \partial \overline{z}} = \frac{1}{4\pi} \cdot \frac{1}{(1 - \zeta\overline{z})^2} .$$

In conclusion, we may apply Proposition 1.3.2 to see that

$$K(z, \zeta) = \frac{1}{\pi} \cdot \frac{1}{(1 - z \cdot \overline{\zeta})^2} .$$

This result is consistent with that obtained in the first two calculations (Subsections 1.3.1, 1.3.2). The Bergman metric, as before, is obtained by differentiation.

1.4 A New Application

We now present a rather profound, and relatively new, application of invariant metrics on planar domains. This result was first discovered around 1978 by several different authors. The complete history may be found in [FKKM]. We give here an argument from that paper.

Before we begin the proof, some remarks are in order. First we note that if φ is an automorphism of the disk D that has just two fixed points, then φ is the identity. For we may as well assume that one of the fixed points is the origin. By Schwarz's lemma, we know immediately that φ is a rotation. Now one additional fixed point forces φ to be the identity. The domain $\Omega = \mathbb{C}$ has a similar property. All automorphisms have the form $z \mapsto az+b$, and two fixed points provide the algebraic data to determine a and b. Refer to the cognate discussion in Section 12.3.

Things are a bit more interesting for an annulus, say

$$\mathcal{A} = \left\{ z \in \mathbb{C} : \frac{1}{2} < |z| < 2 \right\} .$$

The inversion mapping $z \mapsto 1/z$ has 1 and -1 as fixed points. And of course it is not the identity. Any third fixed point would (see the theorem below) force an automorphism to be the identity. The Riemann sphere is similar. Any automorphism is a linear fractional transformation, and it is easy to see (on purely algebraic grounds) that three fixed points determine the transformation uniquely.

For a Riemann surface of arbitrary genus, things are more complicated. It is not difficult to see that if k is a positive integer, then there is a Riemann surface with an automorphism φ that is not the identity such that φ has k fixed points. We now provide a couple of quick examples:

Example 1.4.1. Consider the complex one-dimensional torus T generated from the lattice $\{1, i\}$. Let $\pi : \mathbb{C} \to T$ be the standard covering map. Then $z \mapsto -z$ on the complex plane generates an automorphism, say τ, on T. Now τ has *four* fixed points, which are

$$\pi(0)\,, \quad \pi(1/2 + i/2)\,, \quad \pi(1/2)\,, \quad \pi(i/2)\,.$$

Yet τ does not fix $\pi(\frac{1}{4})$, and so it is not the identity map.

Example 1.4.2. We now consider a two-holed torus. This manifold can be generated by a regular octagon centered at the origin of the Poincaré disk together with its reflections. Again, $z \mapsto -z$ generates a nontrivial automorphism of this Riemann surface. The number of fixed points is now *six*, coming from the center (the origin), the vertices, and the corresponding pairs of midpoints of the sides of the octagon.

It is now clear that one can obtain arbitrarily large numbers of fixed points just from among the compact Riemann surfaces. See [FKKM] for the details.

Theorem 1.4.3. *Let $\Omega \subseteq \mathbb{C}$ be any bounded, planar domain and $\phi : \Omega \to \Omega$ a conformal map. If ϕ has three fixed points (i.e., $\phi(p) = p$ for three distinct points p), then ϕ is the identity mapping.*

Proof. We must begin with some terminology. We take it for granted that the reader has at least an intuitive understanding of the concept of "geodesic" (see [STO] or [BOO] for background). Going by the book, a geodesic is defined by a differential equation. For our purposes here, we may think of a geodesic as a locally length-minimizing curve.

Let $x \in M$, where M is a Riemannian manifold. A point $y \in M$ is called a *cut point* of x if there are two or more length-minimizing geodesics from x to y in M. See Figure 1.2. We further use the following basic terminology and facts from Riemannian geometry. Let $\mathrm{dis}(x, y)$ denote the metric distance from x to y. A geodesic $\gamma : [a, b] \mapsto M$ is called a *length-minimizing geodesic* (or alternatively, a *minimal geodesic*, or a *minimal connector*) from x to y if $\gamma(a) = x$, $\gamma(b) = y$, and $\mathrm{dis}\,(x, y) =$ [arc length of γ]. By the Hopf–Rinow theorem, any two points in a complete Riemannian manifold can be connected

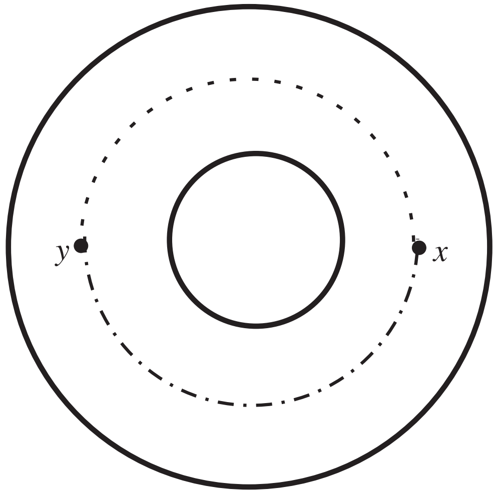

Fig. 1.2. The point y is a cut point for x.

by a minimizing geodesic. If there is a smooth family of minimizing geodesics from x to y, then these two points are said to be *conjugate*. Conjugate points are cut points. The collection of cut points of x in M is called the *cut locus* of x, which we denote by C_x. It is known that C_x is nowhere dense in M ([GKM], [KLI], e.g.), in fact, C_x lies in the singular set for the distance function.

Now equip Ω with the Bergman metric. For convenience, we shall suppose that Ω has C^1 boundary. This will guarantee that the Bergman metric is complete (see [OHS]). We assume that $f \in \mathrm{Aut}\,(\Omega)$ is not the identity map, but has 3 distinct fixed points in Ω. To reach a contradiction, let us start with the fixed point a. If the set of fixed points accumulates at a, we are done. So we may choose as the second fixed point b the closest (with respect to the Hermitian metric) one to a apart from a itself. This choice may not be unique, and hence we simply choose one.

We need to consider only the case that b is a cut point (not conjugate) of a. [For otherwise f would fix the unique geodesic connecting a to b and hence would be the identity.] Then there will be several unit-speed minimal connectors (all of which have the same length, of course), say $\gamma_1, \gamma_2, \ldots$, joining a to b. First notice that no minimal connector can have a self-intersection. Then the automorphism f maps any one of the minimal connectors to another such, since the endpoints a and b are fixed. Note that $f \circ \gamma_1$ cannot intersect γ_1 except at the endpoints. For, if they do intersect at a point other than the endpoints, then they have to intersect at the same time; otherwise one may find an even shorter connector between a and b than the minimal connector, which is a contradiction. Then the intersection point becomes a fixed point of f closer to a than b, which is again not allowed.

Now, γ_1 and $f \circ \gamma_1$ join to form a piecewise smooth Jordan curve in the plane. Thus it bounds a cell, say E, in the plane $\mathbb{C}$. Now consider the third fixed point c that is distinct from a and b. Notice that we may assume that c is not on any of the minimal connectors for a and b. Suppose that c is inside the cell E. Now join c to a by an arc ξ in $E \cap \Omega$ that does not intersect with either γ_1 or $f \circ \gamma_1$, or in fact with any minimal geodesics joining a and b. Notice that the conformality of f at the fixed point a shows that there is a sufficiently small open disk neighborhood U of a on which f must map $U \cap \xi$ to the outside of the cell E. This results in the conclusion that $f \circ \xi$ must cross γ_1 or $f \circ \gamma_1$. But this is impossible, since a point not on any minimal connector from a to b cannot be mapped to a point on a minimal connector from a to b.

If c is outside the cell E, then the arguments are similar. Since there are only finitely many minimal connectors between a and b (since a and b are not conjugate to each other, and the quotient from the universal covering space is formed by a properly discontinuous group action; see [BRE]), some iterate f^m of f will move ξ so that its image has points inside E. Then, $f^m \circ \xi$ again crosses one of these minimal geodesics joining a and b, which leads us to another contradiction. □

Restricting attention back to the unit disk, we recall the difference between fixed points in the interior and fixed points in the boundary (this is a remark of Michael Christ). As we have noted, two fixed points in the interior force an automorphism to be the identity. But it takes three fixed points in the boundary for rigidity. One explanation for this difference is that two fixed points in the interior will give rise, by way of Schwarz reflection, to *four* fixed points of an automorphism of the Riemann sphere. That of course is impossible unless the map is the identity. But of course fixed points in the boundary will not multiply under reflection, so their behavior is different.

1.5 An Application to Mapping Theory

A classical result from complex function theory is this:

Theorem 1.5.1. *Let $\mathcal{A} = \{z \in \mathbb{C} : 1 < |z| < R\}$. If ϕ is a conformal mapping of $\mathcal{A}$ to itself, then ϕ is either a rotation or an inversion of the form $z \mapsto R/z$.*

It is interesting to note that it is quite difficult to compute the Bergman kernel and metric for an annulus. One can certainly see (using Laurent series) that the monomials $\{z^j\}_{j=-\infty}^{\infty}$ form an orthogonal system on an annulus centered at the origin. And one can use a little calculus to normalize these to an orthonormal system. But actually performing the necessary summation is virtually intractable (and involves elliptic functions [BER, pp. 9–10]). Fortunately, the proof that we are about to present requires no detailed knowledge of the Bergman metric of the annulus. Indeed, it uses only the fact that the metric is complete, hence blows up at the boundary. We shall take this last result for granted, although see [APF] for a detailed treatment of this and related matters.

We will need to know that the Bergman metric has geodesics, but that follows from the smoothness and completeness of the metric (see [KON]). We will also utilize the real analyticity of the Bergman kernel and metric in an interesting and surprising way.

Again, we shall indulge the reader's intuition and invoke the heuristic idea of geodesic (see [KON] for all the details). Now fix an annulus $\mathcal{A} = \{z \in \mathbb{C} : 1 < |z| < R\}$. We will be using standard polar coordinates (r, θ). At any point $p \in \mathcal{A}$, a vector in the tangent space decomposes into a component in the $\partial/\partial r$ direction and a component in the $\partial/\partial\theta$ direction. One way to look at the metric is that it assigns a length to $\partial/\partial r$ and to $\partial/\partial\theta$ at each point of the annulus. Consider the set M of points where the Bergman length of $\partial/\partial\theta$ is minimal. Such points exist just because the Bergman metric blows up at the boundary of the domain. What geometric properties will the set M have?

First, the set must be rotationally invariant—because the metric will be rotationally invariant (i.e., the rotations are conformal self-maps of $\mathcal{A}$). Thus M is a union of circles centered at the origin. And M is certainly a closed set by the continuity of the metric. The set has no interior because the Bergman metric ρ is given by a real analytic function (it is the second derivative of the logarithm of the real analytic Bergman kernel), and the zero set of a nontrivial real analytic function can have no interior—see [KRP1]. In fact we may also note that $\|\partial/\partial\theta\|_\rho$ is real analytic in the radial r direction, hence (by the same reasoning) we claim that M can have only finitely many circles in it. More precisely, let $g(r,\theta) = \|\partial/\partial\theta\|_{\rho,(r,\theta)}$. We have already noted that this function is independent of θ, so we may consider $g(r) = \|\partial/\partial\theta\|_{\rho,r}$. This function will assume its minimum value at points r where $g'(r) = 0$. Since the metric is complete, we know that the set of such points forms a compact subset of the interval $(0, R)$. If the set is infinite, then it has an accumulation point, and therefore the real analytic function g' is identically zero. That is impossible, again by the completeness of the metric. Therefore the set M is finite.

Now it is easy to see that any curve (circle) in M will also be one of the curves (the curves that go once around the hole in the middle of the annulus) that minimizes arc length in the Bergman metric, and vice versa. This is so because we have already selected the curve to have $\partial/\partial\theta$ length as small as possible; a curve whose tangents have components in the $\partial/\partial r$ direction will a fortiori be longer. In more detail, the length of any curve $\gamma(t)$, $0 \le t \le 1$, is calculated by

$$\ell_\rho(\gamma) = \int_0^1 \|\gamma'(t)\|_{\rho,\gamma(t)}\, dt = \int_0^1 \|\gamma_r'(t) + \gamma_\theta'(t)\|_{\rho,\gamma(t)}\, dt\,,$$

where γ_r' and γ_θ' are, respectively, the normal and tangential components of γ'. Clearly, if we construct a new curve $\widetilde{\gamma}$ by integrating the vector field γ_θ', then the result is a curve that is shorter than γ.

Thus any circle in M is a length-minimizing geodesic. And the converse is true as well.

Now let φ be a conformal self-map of $\mathcal{A}$. Then, as a result of the considerations in the last paragraph, φ will map circles in M (concentric with the annulus) to circles in M (concentric with the annulus). But more is true: *any* circle c that is centered at 0 has constant Bergman distance from any one given circle C in M. Let τ_c be the Bergman distance from the arbitrary circle c to the fixed circle C. Then c will be mapped by φ to another circle in $\mathcal{A}$ that has Bergman distance τ_c from $\varphi(C)$. In short, φ maps circles to circles. And the same remark applies to φ^{-1}.

Now the orthogonal trajectories to the family of circles centered at P will of course be the radii of the annulus $\mathcal{A}$. Therefore (by conformality) these radii will get mapped to radii. Let $\mathcal{R}$ be the intersection of the positive real axis with the annulus $\mathcal{A}$. After composition with a rotation, we may assume that φ maps $\mathcal{R}$ to $\mathcal{R}$.

If M has just one circle in it, then that circle C must be fixed. Since each circle c in the annulus (centered at P) has a distance τ_c from C, then its image under φ will have the same distance from C. Thus either c is fixed or it is sent to its image under inversion. By continuity, whatever choice is valid for c will be valid for all other circles. Thus, in this case, the map φ is either the identity or inversion.

Now suppose that M contains at least two circles. The only possibilities for the action of φ on $M \cap \mathcal{R}$ are preservation of the order of the points or inversion of the order (since any other permutation of the points is ruled out by elementary topology of a conformal mapping). By composing φ with an inversion ($z \mapsto R/z$), we may assume that φ does *not* act as an inversion on $\mathcal{R}$. But of course φ must preserve $M \cap \mathcal{R}$. We conclude that φ must be the identity.

What we have just proved is that, after we normalize φ so that it maps the positive real axis to itself, then in fact φ must be the identity (or an inversion). In other words, the original map φ must be a rotation (or an inversion). □

Problems for Study and Exploration

1. Prove that if F is an isometry of the unit disk equipped with the Bergman metric, then F is conformal. Can you extend this result to any bounded, simply connected domain?
2. Calculate the collection of all conformal self-maps of the annulus $\mathcal{A} = \{\zeta \in \mathbb{C} : 1/2 < |\zeta| < 2\}$.
3. Let $\phi : \Omega_1 \to \Omega_2$ be a conformal map of bounded domains in $\mathbb{C}$. Derive a formula relating the Bergman projection $P_{\Omega_1} : L^2(\Omega_1) \to A^2(\Omega_1)$ to

$P_{\Omega_2} : L^2(\Omega_2) \to A^2(\Omega_2)$. We shall consider this idea in detail in Section 5.2.

4. What can you say about the Bergman kernel of the annulus $\mathcal{A} = \{\zeta \in \mathbb{C} : 1/2 < |\zeta| < 2\}$? Can you write it as a "principal term" plus a "lower-order term"?

5. Calculate the Bergman kernel on a domain with a corner, such as $\Omega = \{\zeta \in \mathbb{C} : \operatorname{Re}\zeta > 0, \operatorname{Im}\zeta > 0\}$. How does the kernel $K(z, \zeta)$ blow up as z, ζ tend to 0? How does this compare with the boundary behavior of the Bergman kernel on the disk?

6. Write an explicit formula for the Bergman kernel on the upper halfplane $U = \{\zeta \in \mathbb{C} : \operatorname{Im}\zeta > 0\}$.

7. Give an example of a planar domain (not the entire plane) for which the Bergman space consists only of the zero function. Can you characterize all such domains?

8. What is the Bergman projection of the function $\varphi(\zeta) = \overline{\zeta}$ on the unit disk? What about $\varphi_j(\zeta) = \overline{\zeta}^j$, $j = 2, 3, \ldots$?

9. Fix a bounded domain $\Omega \subseteq \mathbb{C}$. Fix also a point $z \in \Omega$. Prove that $\sqrt{K_\Omega(z, z)}$ is the greatest value at z of the modulus of any Bergman space function of norm 1 on Ω.

10. Let Ω be any bounded domain in $\mathbb{C}$. Prove that the Bergman space is separable.

11. Let $\Omega_1 \subseteq \Omega_2 \subseteq \cdots$ be domains and set $\Omega = \cup_j \Omega_j$. How are K_Ω and K_{Ω_j} related?

12. Let $\Omega_1 \subseteq \Omega_2$. How are $K_{\Omega_1}(z, z)$ and $K_{\Omega_2}(z, z)$ related?

13. Let $\Omega \subseteq \mathbb{C}$ be a domain and $z \in \Omega$. Let δ_z be the Dirac delta mass at the point z. Explain why the Bergman kernel function $K(z, \cdot)$ satisfies equation

$$P_\Omega(\delta_z) = K(z, \cdot\,)\,.$$

14. On the domain the unit disk D, imitate the construction of the Bergman kernel using harmonic functions rather than holomorphic functions. What kernel results?

15. Let f be a one-to-one, onto map of the Riemann sphere to itself that maps lines and great circles to lines and great circles. Prove that f is either a linear fractional transformation or the conjugate of a linear fractional transformation.

2

Variations on the Theme of the Schwarz Lemma

Genesis and Development

The Schwarz lemma is one of the simplest results in all of complex function theory. A direct application of the maximum principle, it is merely a statement about the rate of growth of holomorphic functions on the unit disk.

But there is hardly any result that has been quite so influential. Thanks in part to Lars Ahlfors's geometrization of the proof (he showed that the Schwarz lemma can be interpreted in terms of curvature), the Schwarz lemma has assumed a central and powerful role in complex geometry. There are many seminal generalizations of the Schwarz lemma, some of them quite deep. The Carathéodory and Kobayashi metrics—distinctive Finsler metrics that can be used decisively in the study of function-theoretic questions—are constructed with methodologies that are based on the Schwarz lemma. Almost any result in the geometric theory of analytic functions has the Schwarz lemma lurking in the background.

The present chapter explores the Schwarz lemma and its variants. We shall certainly learn about Ahlfors's approach, and some of its consequences. One of the beautiful byproducts of our work will be an elegant new proof of Picard's theorems.

2.1 Introduction

The classical Schwarz lemma is part of the grist of every complex analysis class. A version of it says this:

Lemma 2.1.1. *Let $f : D \to D$ be holomorphic. Assume that $f(0) = 0$. Then*

(a) $|f(z)| \leq |z|$ *for all* $z \in D$*;*
(b) $|f'(0)| \leq 1$*.*

At least as important as these two statements are the cognate uniqueness statements:

(c) *If* $|f(z)| = |z|$ *for some* $z \neq 0$*, then* f *is a rotation:* $f(z) = \lambda z$ *for some unimodular complex constant* λ*;*
(d) *If* $|f'(0)| = 1$*, then* f *is a rotation:* $f(z) = \lambda z$ *for some unimodular complex constant* λ*.*

There are a number of ways to prove this result. The classical argument is to consider $g(z) = f(z)/z$. On a circle $|z| = 1-\epsilon$, we see that $|g(z)| \leq 1/(1-\epsilon)$. Thus $|f(z)| \leq |z|/(1-\epsilon)$. Since this inequality holds for all $\epsilon > 0$, part **(a)** follows. The Cauchy estimates show that $|f'(0)| \leq 1$.

For the uniqueness, if $|f(z)| = |z|$ for some $z \neq 0$, then $|g(z)| = 1$. The maximum modulus principle then forces g to be a unimodular constant, and hence f is a rotation. If instead $|f'(0)| = 1$, then $|g(0)| = 1$ and again the maximum modulus principle yields that f is a rotation.

The Schwarz–Pick lemma observes that there is no need to restrict to $f(0) = 0$. Once one comes up with the right formulation, the proof is straightforward:

Proposition 2.1.2. *Let* $f : D \to D$*. Assume that* $a \neq b$ *are elements of* D *and that* $f(a) = \alpha$*,* $f(b) = \beta$*. Then*

(a) $\left|\dfrac{\beta - \alpha}{1 - \overline{\alpha}\beta}\right| \leq \left|\dfrac{b - a}{1 - \overline{a}b}\right|$;
(b) $|f'(a)| \leq \dfrac{1 - |\alpha|^2}{1 - |a|^2}$.

There is also a pair of uniqueness statements:

(c) *If* $\left|\dfrac{\beta - \alpha}{1 - \overline{\alpha}\beta}\right| = \left|\dfrac{b - a}{1 - \overline{a}b}\right|$*, then* f *is a conformal self-map of the disk* D*;*
(d) *If* $|f'(a)| = \dfrac{1 - |\alpha|^2}{1 - |a|^2}$*, then* f *is a conformal self-map of the disk* D*.*

Proof. We sketch the proof. Recall that for a a complex number in D,

$$\varphi_a(\zeta) = \frac{\zeta - a}{1 - \overline{a}\zeta}$$

defines a *Möbius transformation*. This is a conformal self-map of the disk that takes a to 0. Note that φ_{-a} is the inverse mapping to φ_a.

Now, for the given f, consider

$$g(z) = \varphi_\alpha \circ f \circ \varphi_{-a} \, .$$

Then $g : D \to D$ and $g(0) = 0$. So the standard Schwarz lemma applies to g. By part **(a)** of that lemma,

$$|g(z)| \leq |z| \, .$$

Letting $z = \varphi_a(\zeta)$ yields

$$|\varphi_\alpha \circ f(\zeta)| \le |\varphi_a(\zeta)| \,.$$

Writing this out, and setting $\zeta = b$, gives the conclusion

$$\left|\frac{\beta - \alpha}{1 - \overline{\alpha}\beta}\right| \le \left|\frac{b - a}{1 - \overline{a}b}\right| \,.$$

That is part **(a)**.

For part **(b)**, we certainly have that

$$|(\varphi_\alpha \circ f \circ \varphi_{-a})'(0)| \le 1 \,.$$

Using the chain rule, we may rewrite this as

$$|\varphi'_\alpha(f \circ \varphi_{-a}(0))| \cdot |f'(\varphi_{-a}(0))| \cdot \left|\varphi'_{-a}(0)\right| \le 1 \,. \tag{2.1.1}$$

Now of course

$$\varphi'_a(\zeta) = \frac{1 - |a|^2}{(1 - \overline{a}\zeta)^2} \,.$$

So we may rewrite (2.1.1) as

$$\left(\frac{1 - |\alpha|^2}{(1 - |\alpha|^2)^2}\right) \cdot |f'(a)| \cdot (1 - |a|^2) \le 1 \,.$$

Now part **(b)** follows.

We leave parts **(c)** and **(d)** as exercises for the reader. □

The quantity

$$\rho(a, b) = \frac{|a - b|}{|1 - \overline{a}b|}$$

is called the *pseudohyperbolic metric*. It is actually a metric on D (details left to the reader). It is not identical to the Poincaré–Bergman metric. In fact it is not a Riemannian metric at all. But it is still true that conformal maps of the disk are distance-preserving in the pseudohyperbolic metric. Exercise: Use the Schwarz–Pick lemma to prove this last assertion.

One useful interpretation of the Schwarz–Pick lemma is that a holomorphic function f from the disk to the disk must take each disk $D(0, r)$, $0 < r < 1$, into (but not necessarily onto) the image of that disk under the linear fractional map

$$z \mapsto \frac{z + \alpha}{1 + \overline{\alpha}z} \,,$$

where $f(0) = \alpha$. This image is in fact (in case $-1 < \alpha < 1$) a standard Euclidean disk with center on the real axis at α and diameter (in case $0 < \alpha < 1$) given by the interval

$$\left[\frac{\alpha - r}{1 - \alpha r} \,,\; \frac{\alpha + r}{1 + \alpha r}\right] \,.$$

2.2 Other Versions of Schwarz's Lemma

Here we present some fascinating but less well known versions of the Schwarz lemma concept.

Proposition 2.2.1. *Let f be holomorphic on $D(0,r)$, and assume that $|f(z)| \leq M$ for all z. Then*

$$\left|\frac{f(z)-f(w)}{z-w}\right| \leq \frac{2Mr}{|r^2-\overline{z}w|} .$$

Proof. Define

$$g(z)=\frac{f(rz)}{M} .$$

Then $g: D \to D$ and we may apply Schwarz–Pick to g. The result is

$$\left|\frac{g(z)-g(w)}{1-\overline{g(z)}g(w)}\right| \leq \left|\frac{z-w}{1-\overline{z}w}\right| ,$$

which translates to

$$\left|\frac{f(rz)/M-f(rw)/M}{1-\overline{f(rz)/M}\cdot f(rw)/M}\right| \leq \left|\frac{z-w}{w-\overline{z}w}\right| .$$

Doing some elementary algebra and replacing $z \mapsto z/r$, $w \mapsto w/r$ yields

$$\left|\frac{f(z)-f(w)}{1-[\overline{f(z)}f(w)]/M^2}\right| \leq M \cdot \left|\frac{r(z-w)}{r^2-\overline{z}w}\right| .$$

This may be rearranged as

$$\left|\frac{f(z)-f(w)}{z-w}\right| \leq \frac{rM|1-[\overline{f(z)}f(w)]/M^2|}{|r^2-\overline{z}w|}$$

and the right-hand side here can obviously be majorized by

$$\frac{2Mr}{|r^2-\overline{z}w|} .$$

This completes the argument. □

This last proposition is one way to adapt the ideas of Schwarz–Pick to a disk of arbitrary radius and a function of arbitrary upper bound. The next result, due to Lindelöf, gives a means for majorizing $|f(z)|$, in the classical Schwarz lemma setting, directly *without* assuming that $f(0)=0$.

Proposition 2.2.2. *Let $f: D \to D$ be holomorphic. Then*

$$|f(z)| \leq \frac{|f(0)|+|z|}{1+|f(0)||z|} .$$

Proof. We first note that, for complex numbers $a, b \in D$,

$$\left|\frac{a-b}{1-\overline{a}b}\right|^2 = 1 - \frac{(1-|a|^2)(1-|b|^2)}{|1-\overline{a}b|^2}$$
$$\geq 1 - \frac{(1-|a|^2)(1-|b|^2)}{(1-|a||b|)^2}$$
$$= \frac{(|a|-|b|)^2}{(1-|a||b|)^2}\,.$$

Now we use this inequality to calculate, for $z \in D$, that

$$\frac{|f(z)|-|f(0)|}{1-|f(0)||f(z)|} \leq \frac{|f(z)-f(0)|}{|1-\overline{f(0)}f(z)|}\,.$$

By Schwarz–Pick this last is not greater than

$$\frac{|z-0|}{|1-\overline{0}z|} = |z|\,.$$

Thus

$$|f(z)|-|f(0)| \leq |z| - |z|\cdot|f(0)|\cdot|f(z)|\,.$$

Rearranging the inequality gives

$$|f(z)| \leq \frac{|f(0)|+|z|}{1+|z||f(0)|}\,.$$

That is the desired result. □

2.3 A Geometric View of the Schwarz Lemma

In 1938, Lars Ahlfors [AHL1] caused a sensation by proving that the Schwarz lemma is really an inequality about the curvatures of Riemannian metrics. In the present section we will give an exposition of Ahlfors's ideas. Afterward we can provide some applications.

We shall go into considerable detail in the present discussion, so that the reader has ample motivation and context. Certainly it should be plain that there is definite resonance with the material on the Poincaré and Bergman metrics in Chapter 1.

2.3.1 Geometric Ideas

In classical analysis a *metric* is a device for measuring distance. If X is a set, then a metric λ for X is a function

$$\lambda\colon X \times X \longrightarrow \mathbb{R}$$

satisfying, for all $x, y, z \in X$,

(1) $\lambda(x, y) = \lambda(y, x)$
(2) $\lambda(x, y) \geq 0$ and $\lambda(x, y) = 0$ iff $x = y$;
(3) $\lambda(x, y) \leq \lambda(x, z) + \gamma(z, y)$.

The trouble with a metric defined in this generality is that it does not interact well with calculus. What sort of interaction might we wish to see?

Given two points $P, Q \in X$, one would like to consider the *curve of least length* connecting P to Q. Any reasonable construction of such a curve leads to a differential equation, and thus we require that our metric lend itself to differentiation. Yet another consideration is curvature: *in the classical setting curvature is measured by the rate of change of the normal vector field.* The concepts of normal and rate of change lead inexorably to differentiation. Thus we shall now take a different approach to the concept of "metric." The reader will see definite parallels here with our treatment of "metric" in Chapter 1 on invariant geometry.

Definition 2.3.1. If $\Omega \subseteq \mathbb{C}$ is a domain, then a *metric* on Ω is a continuous function $\rho(z) \geq 0$ in Ω that is twice continuously differentiable on $\{z \in \Omega : \rho(z) > 0\}$. If $z \in \Omega$ and $\xi \in \mathbb{C}$ is a vector, then we define the *length* of ξ *at* z to be

$$|\xi|_{\rho,z} \equiv \rho(z) \cdot \|\xi\| \ ,$$

where $\|\xi\|$ denotes the Euclidean length of the vector ξ. We shall use the notation $\| \ \ \|$ for the remainder of this chapter to denote Euclidean length of a vector. This is a departure from the rest of the book, and we indulge in this conceit in an attempt at local consistency. After that, we shall return to the more familiar $| \ \ |$.

Remark 2.3.2. Usually our metrics ρ will be strictly positive, but it will occasionally be convenient for us to allow a metric to have isolated zeros. These will arise as zeros of holomorphic functions. The zeros of ρ should be thought of as singular points of the metric.

For the record, the metrics we are considering here are a special case of the type of differential metric called *Hermitian*. This terminology need not concern us here. Classical analysts sometimes call these metrics *conformal metrics* and write them in the form $\rho(z)|dz|$.

Technically speaking, our metric lives on the tangent bundle to the domain Ω. That is to say, the metric is a function of the variable (z, v), where v is thought of as a tangent vector at the point z. This is just a mathematically rigorous way of saying that the length of the vector v depends on the point z at which it is positioned.

Definition 2.3.3. Let $\Omega \subseteq \mathbb{C}$ be a domain and ρ a metric on Ω. If

$$\gamma : [a, b] \to \Omega$$

is a continuously differentiable curve, then we define its *length in the metric* ρ to be

$$\ell_\rho(\gamma) = \int_a^b |\dot{\gamma}(t)|_{\rho,\gamma(t)}\, dt.$$

The length of a piecewise continuously differentiable curve is defined to be the sum of the lengths of its continuously differentiable pieces.

Classical sources write the arc length as

$$\ell_\rho(\gamma) = \int_\gamma \rho(z)|dz| \ ,$$

but we shall not use this notation.

If a metric ρ is given on a planar domain Ω, and if P, Q are elements of Ω, then the distance in the metric ρ from P to Q should be defined as follows: Define $\mathcal{C}_\Omega(P,Q)$ to be the collection of all piecewise continuously differentiable curves $\gamma\colon [0,1] \to \Omega$ such that $\gamma(0) = P$ and $\gamma(1) = Q$. Now define the ρ-metric distance from P to Q to be

$$d_\rho(P,Q) = \inf\left\{\ell_\rho(\gamma)\colon \gamma \in \mathcal{C}_\Omega(P,Q)\right\}.$$

Check for yourself that the resulting notion of distance satisfies the classical metric axioms listed at the outset of this section.

There is some subtlety connected with defining distance in this fashion. If $\rho(z) \equiv 1$, the Euclidean metric, and if Ω is the entire plane, then $d_\rho(P,Q)$ is the ordinary Euclidean distance from P to Q. The "shortest curve" from P to Q in this setting is the usual straight line segment. But if Ω and P and Q are as shown in Figure 2.1, then there is no shortest curve in Ω connecting P to Q. The *distance* from P to Q is suggested by the dotted curve, but notice that this curve does not lie in Ω. The crucial issue here is whether the domain is complete in the metric, and we shall have more to say about this point later on.

2.3.2 Calculus in the Complex Domain

In order that we may be able to do calculus computations easily and efficiently in the context of complex analysis, we recast some of the basic ideas in new notation. We define the differential operators

$$\frac{\partial}{\partial z} = \frac{1}{2}\left(\frac{\partial}{\partial x} - i\frac{\partial}{\partial y}\right) \quad \text{and} \quad \frac{\partial}{\partial \overline{z}} = \frac{1}{2}\left(\frac{\partial}{\partial x} + i\frac{\partial}{\partial y}\right).$$

This is, in effect, a new basis for the tangent space to $\mathbb{C}$. In complex analysis it is more convenient to use these operators than to use $\partial/\partial x$ and $\partial/\partial y$.

Proposition 2.3.4. *If f and g are continuously differentiable functions, and if $f \circ g$ is well defined on some open set $U \subseteq \mathbb{C}$, then we have*

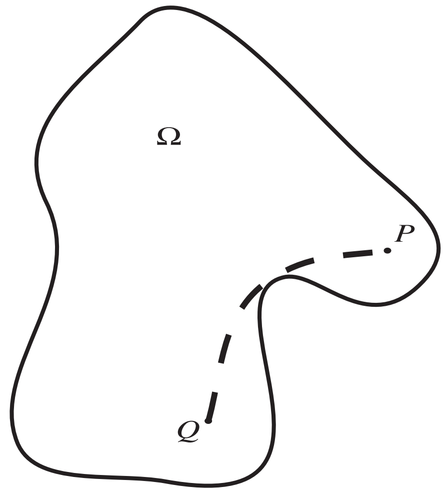

Fig. 2.1. No shortest curve.

$$\frac{\partial}{\partial z}(f \circ g)(z) = \frac{\partial f}{\partial z}(g(z))\frac{\partial g}{\partial z}(z) + \frac{\partial f}{\partial \overline{z}}(g(z))\frac{\partial \overline{g}}{\partial z}(z)$$

and

$$\frac{\partial}{\partial \overline{z}}(f \circ g)(z) = \frac{\partial f}{\partial z}(g(z))\frac{\partial g}{\partial \overline{z}}(z) + \frac{\partial f}{\partial \overline{z}}(g(z))\frac{\partial \overline{g}}{\partial \overline{z}}(z).$$

Proof. We will sketch the proof of the first identity and leave the second as an exercise.

We have

$$\frac{\partial}{\partial z}(f \circ g) = \frac{1}{2}\left(\frac{\partial}{\partial x} - i\frac{\partial}{\partial y}\right)(f \circ g).$$

We write $g(z) = \alpha(z) + i\beta(z)$, with α and β real-valued functions, and apply the usual calculus chain rule for $\partial/\partial x$ and $\partial/\partial y$. We obtain that the last line equals

$$\frac{1}{2}\left(\frac{\partial f}{\partial x}\frac{\partial \alpha}{\partial x} + \frac{\partial f}{\partial y}\frac{\partial \beta}{\partial x} - i\frac{\partial f}{\partial x}\frac{\partial \alpha}{\partial y} - i\frac{\partial f}{\partial y}\frac{\partial \beta}{\partial y}\right). \tag{2.3.1}$$

Now, with the aid of the identities

$$\frac{\partial}{\partial x} = \frac{\partial}{\partial z} + \frac{\partial}{\partial \overline{z}} \qquad \text{and} \qquad \frac{\partial}{\partial y} = i\left(\frac{\partial}{\partial z} - \frac{\partial}{\partial \overline{z}}\right) ,$$

we may reduce the expression (2.3.1) (after some tedious calculations) to the desired formula. □

Corollary 2.3.5. *If either* f *or* g *is holomorphic, then*

$$\frac{\partial}{\partial z}(f \circ g)(z) = \frac{\partial f}{\partial z}(g(z))\frac{\partial g}{\partial z}(z).$$

Here is an example of the utility of our complex calculus notation.

Example 2.3.6. Let f be a nonvanishing holomorphic function on a planar domain Ω. Then

$$\Delta\bigl(\log(\|f\|^2)\bigr) = 0.$$

In other words, $\log(\|f\|^2)$ is harmonic.

To see this, fix $P \in \Omega$ and let $U \subseteq \Omega$ be a neighborhood of P on which f has a holomorphic logarithm. Then on U we have

$$\begin{aligned}\Delta\bigl(\log(|f|^2)\bigr) &= \Delta\bigl(\log f + \log \overline{f}\bigr)\\ &= 4\frac{\partial}{\partial z}\frac{\partial}{\partial \overline{z}}\log f + 4\frac{\partial}{\partial \overline{z}}\frac{\partial}{\partial z}\log \overline{f}\\ &= 0.\end{aligned}$$

Of course the reader may also check that $\log\|f\|$ is harmonic.

We conclude this section with some exercises for the reader.

Exercises

1. Calculate that, for $a > 0$,

$$\frac{\partial^2}{\partial z\partial\overline{z}}\log\biggl(1 + (z\overline{z})^a\biggr) = \frac{a^2(z\overline{z})^{a-1}}{(1+(z\overline{z})^a)^2}\,.$$

2. If g is holomorphic (and f continuously differentiable), then calculate that

$$\triangle(f\circ g) = (\triangle f \circ g)\cdot |g'|^2\,.$$

3. If f is holomorphic (and g continuously differentiable), then calculate that

$$\triangle(f\circ g) = (f'\circ g)\triangle g + (f''\circ g)[(D_x g)^2 + (D_y g)^2]\,.$$

2.3.3 Isometries

In any mathematical subject there are morphisms: functions that preserve the relevant properties being studied. In linear algebra these are linear maps, in Euclidean geometry these are rigid motions, and in Riemannian geometry these are "isometries." We now define the concept of isometry.

Definition 2.3.7. Let Ω_1 and Ω_2 be planar domains and let

$$f : \Omega_1 \to \Omega_2$$

be a continuously differentiable mapping with Jacobian having isolated zeros. Assume that Ω_2 is equipped with a metric ρ. We define the *pullback* of the metric ρ under the map f to be the metric on Ω_1 given by

$$f^*\rho(z) = \rho(f(z))\cdot\left\|\frac{\partial f}{\partial z}\right\|.$$

Remark 2.3.8. The particular form that we use to define the pullback is motivated by the way that f induces mappings on tangent and cotangent vectors, but this motivation is irrelevant for us here.

It should be noted that the pullback of any metric under a conjugate holomorphic f will be the zero metric. Thus we have designed our definition of pullback so that holomorphic pullbacks will be the ones of greatest interest. This assertion will be made substantive in Proposition 2.3.10 below.

Definition 2.3.9. Let Ω_1, Ω_2 be planar domains equipped with metrics ρ_1 and ρ_2, respectively. Let

$$f : \Omega_1 \to \Omega_2$$

be an onto, holomorphic mapping. If

$$f^*\rho_2(z) = \rho_1(z)$$

for all $z \in \Omega_1$, then f is called an *isometry* of the pair (Ω_1, ρ_1) with the pair (Ω_2, ρ_2).

The differential definition of isometry (Definition 2.3.9) is very natural from the point of view of differential geometry, but it is not intuitive. The next proposition relates the notion of isometry to more familiar ideas.

Proposition 2.3.10. *Let Ω_1, Ω_2 be domains and ρ_1, ρ_2 be metrics on these respective domains. If*

$$f : \Omega_1 \to \Omega_2$$

is a holomorphic isometry (in particular, an onto mapping) of (Ω_1, ρ_1) to (Ω_2, ρ_2), then the following three properties hold:

(a) *If $\gamma : [a, b] \to \Omega_1$ is a continuously differentiable curve, then so is the push-forward $f_*\gamma \equiv f \circ \gamma$ and*

$$\ell_{\rho_1}(\gamma) = \ell_{\rho_2}(f_*\gamma).$$

(b) *If P, Q are elements of Ω_1, then*

$$d_{\rho_1}(P, Q) = d_{\rho_2}(f(P), f(Q)).$$

(c) *Part (b) implies that the isometry f is one-to-one. Then f^{-1} is well defined and f^{-1} is also an isometry.*

Proof. Assertion **(b)** is an immediate consequence of **(a)**. Also **(c)** is a formal exercise in definition chasing. Therefore we shall prove **(a)**.

By definition,

$$\ell_{\rho_2}(f_*\gamma) = \int_a^b |(f_*\gamma)'(t)|_{\rho_2, f_*\gamma(t)}\, dt = \int_a^b \left| \frac{\partial f}{\partial z}(\gamma(t)) \cdot \dot{\gamma}(t) \right|_{\rho_2, f_*\gamma(t)} dt\,.$$

With elementary manipulations, we see that the integrand equals

$$\begin{aligned}\left\| \frac{\partial f}{\partial z}(\gamma(t)) \right\| \cdot |\dot{\gamma}(t)|_{\rho_2, f_*\gamma(t)} &= \left\| \frac{\partial f}{\partial z}(\gamma(t)) \right\| \cdot \|\dot{\gamma}(t)\| \cdot \rho_2(f(\gamma(t))) \\ &= |\dot{\gamma}(t)|_{f^*\rho_2, \gamma(t)} \\ &= |\dot{\gamma}(t)|_{\rho_1, \gamma(t)} \,,\end{aligned}$$

since f is an isometry. Substituting this result back into the formula for the length of $f_*\gamma$ gives

$$\ell_{\rho_2}(f_*\gamma) = \int_a^b |\dot{\gamma}(t)|_{\rho_1, \gamma(t)} \, dt = \ell_{\rho_1}(\gamma).$$

This ends the proof. □

In the next section we shall consider the isometries of the Poincaré metric on the disk.

If you have previously studied metric space theory or Banach space theory, then you may have already encountered the term "isometry." The essential notion is that an isometry should preserve distance. In fact you can prove as an exercise that if f is a holomorphic mapping of (Ω_1, ρ_1) onto (Ω_2, ρ_2) that preserves distance, then f is an isometry according to Definition 2.3.9. (Hint: compose f with a curve and differentiate.)

We close this section by noting an important technical fact about isometries:

Proposition 2.3.11. *Let ρ_j be metrics on the domains Ω_j, $j = 1, 2, 3$, respectively. Let $f : \Omega_1 \to \Omega_2$ and $g : \Omega_2 \to \Omega_3$ be isometries. Then $g \circ f$ is an isometry of the metric ρ_1 to the metric ρ_3.*

Proof. We calculate that

$$\rho_3(g([f(z)]) \cdot \|g'(f(z))\| = \rho_2(f(z))$$

hence

$$\rho_3(g([f(z)]) \cdot \|g'(f(z))\| \cdot \|f'(z)\| = \rho_2(f(z)) \cdot \|f'(z)\| = \rho_1(z) \,.$$

In other words,

$$\rho_3(g \circ f(z)) \cdot \|(g \circ f)'(z)\| = \rho_1(z) \,,$$

as required by the definition of an isometry. □

2.3.4 The Poincaré Metric

The Poincaré metric on the disk has occurred frequently in this book. This metric is the paradigm for much of what we want to do in the present chapter, and we shall treat it in some detail here. The Poincaré metric on the disk D is given by

$$\rho(z) = \frac{1}{1 - |z|^2}.$$

(For the record, we note that there is no agreement in the literature as to what constant goes in the numerator; many references use a factor of 2.)

In this and succeeding sections, we shall use the phrase "conformal map" to refer to a holomorphic mapping of one planar region to another that is both one-to-one and onto.

Proposition 2.3.12. *Let ρ be the Poincaré metric on the disk D. Let $h : D \to D$ be a conformal self-map of the disk. Then h is an isometry of the pair (D, ρ) with the pair (D, ρ).*

Proof. We have that

$$h^*\rho(z) = \rho(h(z)) \cdot \|h'(z)\|.$$

We now have two cases:

(i) If h is a rotation, then $h(z) = \mu \cdot z$ for some unimodular constant $\mu \in \mathbb{C}$. So $\|h'(z)\| = 1$ and

$$h^*\rho(z) = \rho(h(z)) = \rho(\mu z) = \frac{1}{1 - \|\mu z\|^2} = \frac{1}{1 - \|z\|^2} = \rho(z)$$

as desired.

(ii) If h is a Möbius transformation, then

$$h(z) = \frac{z - a}{1 - \overline{a}z}, \qquad \text{some constant } a \in D.$$

But then

$$\|h'(z)\| = \frac{1 - \|a\|^2}{\|1 - \overline{a}z\|^2}$$

and

$$\begin{aligned} h^*\rho(z) &= \rho\left(\frac{z - a}{1 - \overline{a}z}\right) \cdot \|h'(z)\| \\ &= \frac{1}{1 - \left\|\frac{z-a}{1-\overline{a}z}\right\|^2} \cdot \frac{1 - \|a\|^2}{\|1 - \overline{a}z\|^2} \\ &= \frac{1 - \|a\|^2}{\|1 - \overline{a}z\|^2 - \|z - a\|^2} \\ &= \frac{1 - \|a\|^2}{1 - \|z\|^2 - \|a\|^2 + \|a\|^2\|z\|^2} \\ &= \frac{1}{1 - \|z\|^2} \\ &= \rho(z). \end{aligned}$$

Since any conformal self-map of D is a composition of maps of the form **(i)** or **(ii)**, the proposition is proved. □

Remark 2.3.13. Certainly the last proposition can be derived from the Schwarz–Pick lemma. Do this as an exercise.

We know from this result and from Chapter 1 that conformal self-maps of the disk preserve Poincaré distance. To understand what this means, consider the Möbius transformation

$$\varphi(z) = \frac{z+a}{1+\overline{a}z} \ .$$

Then φ maps the disk to the disk conformally. It does not preserve Euclidean distance, but it *does* preserve Poincaré distance. See Figure 2.2.

We can use what we have learned so far to calculate the Poincaré metric explicitly.

Proposition 2.3.14. *If P and Q are points of the disk D, then the Poincaré distance of P to Q is*

$$d_\rho(P,Q) = \frac{1}{2}\log\left(\frac{1+\left\|\frac{P-Q}{1-\overline{P}Q}\right\|}{1-\left\|\frac{P-Q}{1-\overline{P}Q}\right\|}\right) \ .$$

Proof. In case $P=0$ and $Q=R+i0$, the result was already noted in Proposition 1.1.2. In the general case, note that we may define

$$\varphi(z) = \frac{z-P}{1-\overline{P}z} \ ,$$

a Möbius transformation of the disk. Then, by Proposition 2.3.12,

$$d_\rho(P,Q) = d_\rho(\varphi(P),\varphi(Q)) = d_\rho(0,\varphi(Q)).$$

Next we have

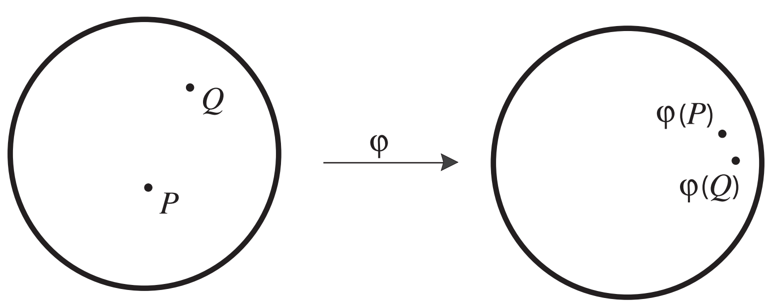

Fig. 2.2. Map preserves Poincaré distance but not Euclidean distance.

$$d_\rho(0, \varphi(Q)) = d_\rho(0, \|\varphi(Q)\|) \tag{2.3.2}$$

since there is a rotation of the disk taking $\varphi(Q)$ to $\|\varphi(Q)\| + i0$. Finally,

$$|\varphi(Q)| = \left\| \frac{P - Q}{1 - \overline{P}Q} \right\| ,$$

so that (2.3.2) together with the special case treated in the the first sentence gives the result. □

One of the reasons that the Poincaré metric is so useful is that it induces the same topology as the usual Euclidean metric topology. That is our next result.

Proposition 2.3.15. *The topology induced on the disk by the Poincaré metric is the usual planar topology.*

Proof. A neighborhood basis for the topology of the Poincaré metric at the origin is given by the balls

$$\mathbf{B}(0, r) = \{z : d_\rho(0, z) < r\}.$$

However, a calculation using Proposition 2.3.14 yields that these balls are the same as the Euclidean disks

$$\left\{ z : \|z\| < \frac{e^{2r} - 1}{e^{2r} + 1} \right\}.$$

These disks form a neighborhood basis for the origin in the Euclidean topology. Thus we find that the two topologies are the same at the origin. Now the origin can be moved to any other point $a \in D$ by the Möbius transformation

$$z \mapsto \frac{z + a}{1 + \overline{a}z}.$$

Since the Poincaré metric is invariant under Möbius transformations, and since Möbius transformations take circles to circles (after all, they are linear fractional) and hence disks to disks, the two topologies are the same at every point. □

One of the most striking facts about the Poincaré metric on the disk is that it turns the disk into a *complete* metric space. How could this be? The boundary is missing! The reason that the disk is complete in the Poincaré metric is the same as the reason that the plane is complete in the Euclidean metric: the boundary is infinitely far away. We now prove this assertion.

Proposition 2.3.16. *The unit disk D, when equipped with the Poincaré metric, is a complete metric space.*

Proof. Let p_j be a sequence in D that is Cauchy in the Poincaré metric. Then the sequence is bounded in that metric. So there is a positive, finite M such that

$$d_\rho(0, p_j) \le M, \qquad \text{all } j.$$

With Proposition 2.3.14 this translates to

$$\frac{1}{2}\log\left(\frac{1+|p_j|}{1-|p_j|}\right) \le M.$$

Solving for $|p_j|$ gives

$$|p_j| \le \frac{e^{2M}-1}{e^{2M}+1} < 1.$$

Thus our sequence is contained in a relatively compact subset of the disk. A similar calculation yields that the sequence must in fact be Cauchy in the Euclidean metric. Therefore it converges to a limit point in the disk, as required for completeness. □

In the proof of Proposition 2.3.14 and in the subsequent remarks we used implicitly the fact, whose verification was sketched earlier, that the curve of least length (in the Poincaré metric) connecting 0 to a point of the form $R + i0$ is in fact a Euclidean segment. We proved this result in Chapter 1. More generally, the shortest path from 0 to any point w is a rotation of the shortest path from 0 to $\|w\| + i0$, which is a segment. Let us now calculate the "curve of least length" connecting any two given points P and Q in the disk.

Proposition 2.3.17. *Let P, Q be elements of the unit disk. The "curve of least Poincaré length" connecting P to Q is*

$$\gamma_{P,Q}(t) = \frac{t\frac{Q-P}{1-Q\overline{P}} + P}{1 + t\overline{P}\cdot\frac{Q-P}{1-Q\overline{P}}}, \qquad 0 \le t \le 1.$$

Proof. Define the Möbius transformation

$$\varphi(z) = \frac{z-P}{1-\overline{P}z}.$$

By what we already know about shortest paths emanating from the origin, the curve $\tau(t) \equiv t\cdot\varphi(Q)$ is the shortest curve from $\varphi(P) = 0$ to $\varphi(Q),\ 0 \le t \le 1$. Applying the isometry

$$\varphi^{-1}(z) = \frac{z+P}{1+\overline{P}z}$$

to τ we obtain that

$$\varphi^{-1}\circ\tau(t) = \frac{(t\varphi(Q)+P)}{(1+\overline{P}t\varphi(Q))}$$

is the shortest path from P to Q. Since $\varphi^{-1}\circ\tau = \gamma_{P,Q}$, we are done. □

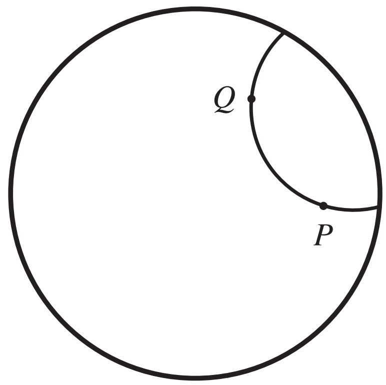

Fig. 2.3. Calculation of geodesic arcs.

Let us analyze the curves discovered in Proposition 2.3.17. First notice that since the curve $\gamma_{P,Q}$ is the image under a linear fractional transformation of a part of a line, the trace of $\gamma_{P,Q}$ is, therefore, either a line segment or an arc of a circle. In fact, if P and Q are collinear with 0, then the formula for $\gamma_{P,Q}$ quickly reduces to that for a segment; otherwise $\gamma_{P,Q}$ traces an arc of a Euclidean circle. Which circle is it?

Matters are simplest if we let t range over the entire real line and look for the whole circle. We find that $t = \infty$ corresponds to the point $1/\overline{P}$. It is now a simple, but tedious, calculation to determine the Euclidean center and radius of the circle determined by the three points $P, Q, 1/\overline{P}$ (note that by symmetry, the circle will also pass through $1/\overline{Q}$). The circle is depicted in Figure 2.3. Notice that since the segment $\{t + i0 : -1 \leq t \leq 1\}$ is orthogonal to ∂D at the endpoints 1 and -1, conformality dictates that the circular arcs of least length provided by Proposition 2.3.17 are orthogonal to ∂D at the points of intersection. Some of these "geodesic arcs" are exhibited in Figure 2.4.

A final note on this matter is that geodesics are particularly easy to calculate in the upper-half-plane realization U of the disk. For the geodesic circles will then have their centers on the boundary (i.e., the real line). If $\widetilde{P}, \widetilde{Q}$ are points of U—not both on the same vertical line—then the perpendicular bisector of the segment connecting these points will intersect the real axis at the center C of the circular arc that forms the geodesic through $\widetilde{P}$ and $\widetilde{Q}$. See Figure 2.5.

We next see that the Poincaré metric is characterized by its property of invariance under conformal maps.

Proposition 2.3.18. *If $\widetilde{\rho}(z)$ is a metric on D that is such that every conformal self-map of the disk is an isometry of the pair $(D, \widetilde{\rho})$ with the pair $(D, \widetilde{\rho})$, then $\widetilde{\rho}$ is a constant multiple of the Poincaré metric ρ.*

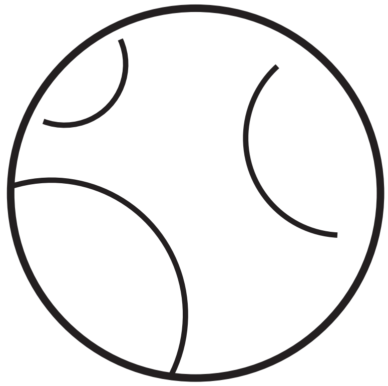

Fig. 2.4. Some geodesic arcs.

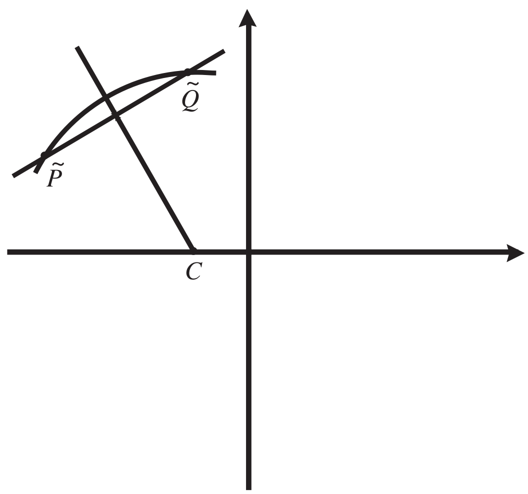

Fig. 2.5. Intersection of geodesics.

Proof. The hypothesis guarantees that if $z_0 \in D$ is fixed and

$$h(z) = \frac{z + z_0}{1 + \overline{z}_0 z}\,,$$

then

$$h^* \widetilde{\rho}(0) = \widetilde{\rho}(0).$$

Writing out the left-hand side gives

$$\|h'(0)\| \widetilde{\rho}(h(0)) = \widetilde{\rho}(0)\,,$$

or

$$\widetilde{\rho}(z_0) = \frac{1}{1 - \|z_0\|^2} \cdot \widetilde{\rho}(0) = \widetilde{\rho}(0) \cdot \rho(z_0).$$

Thus we have exhibited $\widetilde{\rho}$ as the constant $\widetilde{\rho}(0)$ times ρ. □

Now that we know that the Poincaré metric is the right metric for complex analysis on the disk, a natural next question is to determine which other maps preserve the Poincaré metric.

Proposition 2.3.19. *Let $f : D \to D$ be continuously differentiable and let ρ be the Poincaré metric. If f pulls back the pair (D, ρ) to the pair (D, ρ), then f is holomorphic and is one-to-one. We may conclude then that f is the composition of a Möbius transformation and a rotation.*

Proof. First suppose that $f(0) = 0$. For $R > 0$ let C_R be the set of points in D that have Poincaré distance R from 0. Since the Poincaré metric is invariant under rotations (after all, rotations are holomorphic self-maps), it follows that C_R is a Euclidean circle (however, this circle will have Euclidean radius $(e^{2R} - 1)/(e^{2R} + 1)$, not R). Since $f(0) = 0$ and f preserves the metric, it follows that $f(C_R) = C_R$. The fact that f is distance-preserving then shows that f is one-to-one on each C_R. As a result, f is globally one-to-one.

Let $P \in C_R$. Then

$$\frac{\|f(P) - f(0)\|}{\|P - 0\|} = \frac{\|f(P)\|}{\|P\|} = 1.$$

Letting $R \to 0^+$, we conclude that f is conformal at the origin (that is, f preserves length infinitesimally). Since f pulls back the metric ρ, we can be sure that $\partial f/\partial z \neq 0$ at the origin.

Now we drop the special hypothesis that $f(0) = 0$. Pick an arbitrary $z_0 \in D$ and set $w_0 = f(z_0)$. Define

$$\varphi(z) = \frac{z + z_0}{1 + \overline{z}_0 z}, \qquad \psi(z) = \frac{z - w_0}{1 - \overline{w}_0 z}.$$

Also define

$$g = \psi \circ f \circ \varphi.$$

Then $g(0) = 0$ and g is an isometry (since ψ, f, and φ are); therefore the argument in the last paragraph applies and g is conformal at the origin. It follows as before that $\partial g/\partial z \neq 0$ at the origin.

We conclude that f is conformal at every point—since ψ and ϕ are—and $\partial f/\partial z \neq 0$ at every point. Therefore f is holomorphic. □

Isometries are very rigid objects. They are completely determined by their first-order behavior at just one point, as we have discussed in Chapter 1. While a proof of this assertion in general is beyond us at this point, we can certainly prove the result for the Poincaré metric on the disk.

Proposition 2.3.20. *Let ρ be the Poincaré metric on the disk. Let f be an isometry of the pair (D, ρ) with the pair (D, ρ). If $f(0) = 0$ and $\partial f/\partial z(0) = 1$, then $f(z) \equiv z$.*

Proof. By Proposition 2.3.19, f must be holomorphic. Since f preserves the origin and is one-to-one and onto, f must be a rotation. Since $f'(0) = 1$, it follows that f is the identity. □

Corollary 2.3.21. *Let f and g be isometries of the pair (D, ρ) with the pair (D, ρ). Let $z_0 \in D$ and suppose that $f(z_0) = g(z_0)$ and $(\partial f/\partial z)(z_0) = (\partial g/\partial z)(z_0)$. Then $f(z) \equiv g(z)$.*

Proof. We noted earlier in this section that g^{-1} is an isometry. If ψ is a Möbius transformation that takes 0 to z_0, then $\psi^{-1} \circ g^{-1} \circ f \circ \psi$ satisfies the hypothesis of the proposition. As a result, $\psi^{-1} \circ g^{-1} \circ f \circ \psi(z) \equiv z$ or $g(z) \equiv f(z)$. □

2.3.5 The Schwarz Lemma

One of the important facts about the Poincaré metric is that it can be used to study not just conformal maps but all holomorphic maps of the disk. The key to this assertion is the classical Schwarz lemma. We begin with an elegant geometric interpretation of the Schwarz–Pick lemma (see Section 2.1).

Proposition 2.3.22. *Let $f : D \to D$ be holomorphic. Then f is distance-decreasing in the Poincaré metric ρ. That is, for any $z \in D$,*

$$f^*\rho(z) \le \rho(z).$$

The integrated form of this assertion is that if $\gamma : [0, 1] \to D$ is a continuously differentiable curve, then

$$\ell_\rho(f_*\gamma) \le \ell_\rho(\gamma).$$

Therefore, if P and Q are elements of D, we may conclude that

$$d_\rho(f(P), f(Q)) \le d_\rho(P, Q).$$

Proof. Now

$$f^*\rho(z) \equiv \|f'(z)\|\rho(f(z)) = \|f'(z)\| \cdot \frac{1}{1 - \|f(z)\|^2}$$

and

$$\rho(z) = \frac{1}{1 - \|z\|^2},$$

so the asserted inequality is just the Schwarz–Pick lemma. The integrated form of the inequality now follows by the definition of ℓ_ρ. The inequality for the distance d_ρ follows from the definition of distance. □

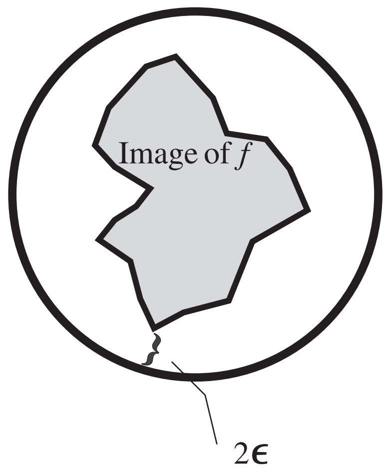

Fig. 2.6. The Farkas–Ritt theorem.

Notice that, in case f is a conformal self-map of the disk, we may apply this corollary to both f and f^{-1} to conclude that f preserves Poincaré distance, giving a second proof of Proposition 2.3.12.

We next give an illustration of the utility of the geometric point of view. We will see that the proposition just proved gives a very elegant proof of Theorem 2.3.23 below (see [EAH] for the source of this proof).

Theorem 2.3.23 (Farkas, Ritt). *Let $f : D \to D$ be holomorphic and assume that the image $M = \{f(z) : z \in D\}$ of f has compact closure in D. Then there is a unique point $P \in D$ such that $f(P) = P$. We call P a fixed point for f.*

Proof. By hypothesis, there is an $\epsilon > 0$ such that if $m \in M$ and $\|z\| \geq 1$, then $\|m - z\| > 2\epsilon$. See Figure 2.6. Fix $z_0 \in D$ and define

$$g(z) = f(z) + \epsilon(f(z) - f(z_0)).$$

Then g is holomorphic and g still maps D into D. Also

$$g'(z_0) = (1 + \epsilon) f'(z_0).$$

By the preceding proposition, g is thus distance-decreasing in the Poincaré metric. Therefore

$$g^* \rho(z_0) \leq \rho(z_0).$$

Writing out the definition of g^* now yields

$$(1 + \epsilon) \cdot f^* \rho(z_0) \leq \rho(z_0).$$

Note that this inequality holds for any $z_0 \in D$. But now if $\gamma : [a,b] \to D$ is any continuously differentiable curve, then we may conclude that

$$\ell_\rho(f_*\gamma) \le (1+\epsilon)^{-1}\ell_\rho(\gamma).$$

If P and Q are elements of D and d is Poincaré distance, then we have that

$$d(f(P), f(Q)) \le (1+\epsilon)^{-1} d(P,Q) \ .$$

We see that f is a contraction in the Poincaré metric. Recall that in Proposition 2.3.16 we proved that the disk D is a complete metric space when equipped with the Poincaré metric. By the contraction mapping fixed-point theorem (see [LOS]), f has a unique fixed point. □

The next result shows how to find the fixed point, and in effect proves the contraction mapping fixed-point theorem in this special case.

Corollary 2.3.24. *If f is as in the theorem and P is the unique fixed point, then the iterates $f, \ f \circ f, \ f \circ f \circ f, \ \ldots$ converge uniformly on compact sets to the constant function P.*

Proof. Let f^n denote the n^{th} iterate of f. Let

$$\overline{\mathbf{B}}(P,R) = \{z \in D : d(z,P) \le R\}$$

be the closed Poincaré metric ball with center P and Poincaré metric radius R. Then the theorem tells us that

$$f(\overline{\mathbf{B}}(P,R)) \subseteq \overline{\mathbf{B}}(P, R/(1+\epsilon))$$

and, more generally,

$$f^n(\overline{\mathbf{B}}(P,R)) \subseteq \overline{\mathbf{B}}(P, R/(1+\epsilon)^n). \tag{2.3.3}$$

Observe that

$$\bigcup_{j=1}^{\infty} \mathbf{B}(P,j) = D.$$

Now Proposition 2.3.15 (or elementary calculation) tells us that these non-Euclidean balls are in fact open disks in the usual Euclidean topology. Therefore every (Euclidean) compact subset K of D lies in some $\mathbf{B}(P,j)$. Then line (2.3.3) implies that

$$f^n(K) \subseteq \overline{\mathbf{B}}\left(P, j/(1+\epsilon)^n\right).$$

The result follows. □

2.4 Ahlfors's Version of the Schwarz Lemma

If $U \subseteq \mathbb{C}$ is a planar domain and ρ is a metric on U, then the *curvature* of the metric ρ at a point $z \in U$ is defined to be

$$\kappa_{U,\rho}(z) = \kappa(z) \equiv \frac{-\Delta \log \rho(z)}{\rho(z)^2}. \tag{2.4.1}$$

(Here zeros of $\rho(z)$ will result in singularities of the curvature function—κ is undefined at such points.)

Since ρ is twice continuously differentiable, this definition makes sense. It assigns to each $z \in U$ a numerical quantity. The most important preliminary fact about κ is its conformal invariance:

Proposition 2.4.1. *Let U_1 and U_2 be planar domains and $h : U_1 \to U_2$ a conformal map (in particular, h' never vanishes). If ρ is a metric on U_2, then*

$$\kappa_{U_1,h^*\rho}(z) = \kappa_{U_2,\rho}(h(z)), \qquad \forall z \in U_1 .$$

Proof. We need to calculate:

$$\begin{aligned}\kappa_{U_1,h^*\rho}(z) &\equiv \frac{-\Delta \log[\rho(h(z)) \cdot \|h'(z)\|]}{[\rho(h(z)) \cdot |h'(z)|]^2} \\ &= \frac{-\Delta \log[\rho(h(z))] - \Delta[\log(\|h'(z)\|)]}{[\rho(h(z)) \cdot \|h'(z)\|]^2}.\end{aligned}$$

Now the second term in the numerator vanishes by Example 2.3.6. We may further simplify the numerator to obtain that the last line equals

$$= \frac{-\left[\Delta \log \rho|_{h(z)}\right] \|h'(z)\|^2}{[\rho(h(z)) \cdot \|h'(z)\|]^2} = \frac{-\Delta \log \rho|_{h(z)}}{\rho(h(z))^2} = \kappa_{U_2,\rho}(h(z)) . \quad \square$$

Remark 2.4.2. In fact, the proof gives the slightly more general fact that if U_1, U_2 are domains and $f : U_1 \to U_2$ is a holomorphic map (not necessarily one-to-one or onto), then the following holds. If ρ is a metric on U_2, then

$$\kappa_{f^*\rho}(z) = \kappa_\rho(f(z))$$

at every point $z \in U_1$ for which $f'(z) \neq 0$ and $\rho(f(z)) \neq 0$. Usually this is all points except for a discrete set.

From the point of view of geometry, any differential quantity that is invariant is automatically of great interest. But why do we call κ "curvature"? The answer is that a standard construction in Riemannian geometry assigns to any Riemannian metric a quantity called Gaussian curvature. In the setting of classical Euclidean geometry this quantity coincides with our intuitive notion of what curvature ought to be. In more abstract settings, the structural equations of Cartan lead to a quantity called curvature, which is invariant but has considerably less intuitive content. The appendix to the book [KRA5] gives a complete derivation—in the complex plane—of the above notion κ of curvature from more classical considerations of Gaussian curvature. One of the pleasant facts about the use of elementary differential geometry in complex function theory is that one does not require all the geometric machinery that leads to the formula for κ in order to make use of κ.

Let us begin our study of curvature by calculating the curvature of the Euclidean metric.

Example 2.4.3. Let U be a planar domain equipped with the Euclidean metric $\rho(z) \equiv 1$. It follows from the definition that $\kappa(z) \equiv 0$. This should be expected, for the Euclidean metric does not change from point to point.

Example 2.4.4. The metric

$$\sigma(z) = \frac{2}{1 + \|z\|^2}$$

on $\mathbb{C}$ is called the *spherical metric*. A straightforward calculation shows that the curvature of σ is identically 1. It is a basic theorem of Riemannian geometry that the sphere is the only manifold of constant positive curvature. In fact, the study of manifolds of constant curvature is an entire subject in itself (see [WOL]).

Next we calculate the curvature of the Poincaré metric on the disk, which calculation will be of great utility for us. Notice that since any point of the disk may be moved to any other by a Möbius transformation, and since curvature is a conformal invariant, we will expect the curvature function to be constant.

Proposition 2.4.5. *Consider the disk D equipped with the Poincaré metric. For any point $z \in D$ it holds that $\kappa(z) = -4$.*

Proof. We notice that

$$-\Delta \log \rho(z) = \Delta \log(1 - \|z\|^2)\,.$$

Now we write $\Delta = 4(\partial/\partial z)(\partial/\partial \bar{z})$ and $\|z\|^2 = z \cdot \bar{z}$ to obtain that this last expression equals

$$-\frac{4}{(1 - \|z\|^2)^2}\,.$$

It follows that $\kappa(z) = -4$. □

The fact that the Poincaré metric on the disk has constant negative curvature turns out to be a special property of the disk. We shall explore this topic later in the book. Now we return to the Schwarz lemma.

It was Ahlfors [AHL1] who first realized that the Schwarz lemma is really an inequality about curvature. In his annotations to his collected works he modestly asserts that, "This is an almost trivial fact and anybody who sees the need could prove it at once." However he goes on to say (most correctly) that this point of view has been a decisive influence in modern function theory. It is a good place to begin our understanding of curvature. Here is the Ahlfors version of Schwarz's lemma:

Theorem 2.4.6. *Let the disk $D = D(0,1)$ be equipped with the Poincaré metric ρ and let U be a planar domain equipped with a metric σ. Assume that, at all points of U, σ has curvature not exceeding -4. If $f : D \to U$ is holomorphic, then we have*

$$f^*\sigma(z) \le \rho(z), \qquad \forall z \in D\,.$$

Proof. We follow the argument in [MIS]. Let $0 < r < 1$. On the disk $D(0,r)$ define the metric

$$\rho_r(z) = \frac{r}{r^2 - z^2}\ .$$

Then straightforward calculations (or change of variable) show that ρ_r is the analogue of the Poincaré metric for $D(0,r)$: it has constant curvature -4 and is invariant under conformal maps. Define

$$v = \frac{f^*\sigma}{\rho_r}\ .$$

Observe that v is continuous and nonnegative on $D(0,r)$ and that $v \to 0$ when $|z| \to r$ (since $f^*\sigma$ is bounded above on $\bar{D}(0,r) \subset\subset D$ while $\rho_r \to \infty$). It follows that v attains a maximum value M at some point $\tau \in D(0,r)$. We will show that $M \le 1$, hence that $v \le 1$ on $D(0,r)$. Letting $r \to 1^-$ then finishes the proof.

If $f^*\sigma(\tau) = 0$, then $v \equiv 0$. So we may suppose that $f^*\sigma(\tau) > 0$. Therefore $\kappa_{f^*\sigma}$ is defined at τ. By hypothesis,

$$\kappa_{f_*\sigma} \le -4\ .$$

Since $\log v$ has a maximum at τ, we have

$$\begin{aligned} 0 &\ge \Delta \log v(\tau) \\ &= \Delta \log f^*\sigma(\tau) - \Delta \log \rho_r(\tau) \\ &= \ -\kappa_{f_*\sigma}(\tau) \cdot (f^*\sigma(\tau))^2 + \kappa_{\rho_r}(\tau) \cdot (\rho_r(\tau))^2 \\ &\ge 4(f^*\sigma(\tau))^2 - 4(\rho_r(\tau))^2\,. \end{aligned}$$

This gives

$$\frac{f^*\sigma(\tau)}{\rho_r(\tau)} \le 1\,,$$

or

$$M \le 1,$$

as desired. □

Observe that the usual Schwarz lemma is a corollary of Ahlfors's version: we take U to be the disk with σ being the Poincaré metric. Let f be a holomorphic mapping of D to $U = D$ such that $f(0) = 0$. Then σ satisfies the hypotheses of the theorem and the conclusion is that

$$f^*\sigma(0) \le \rho(0).$$

Unraveling the definition of $f^*\sigma$ yields

$$\|f'(0)\| \cdot \sigma(f(0)) \le \rho(0).$$

But $\sigma = \rho$ and $f(0) = 0$, so this becomes

$$\|f'(0)\| \le 1.$$

Exercise: The property

$$d(f(P), f(Q)) \le d(P, Q) \ ,$$

where d is the Poincaré distance, is called the *distance-decreasing property* of the Poincaré metric. Use this property to give a geometric proof of the other inequality in the classical Schwarz lemma. ♢

With a little more notation we can obtain a more powerful version of the Ahlfors/Schwarz lemma. Let $D(0, \alpha)$ be the open disk of radius α and center 0. For $A > 0$ define the metric ρ_α^A on $D(0, \alpha)$ by

$$\rho_\alpha^A(z) = \frac{2\alpha}{\sqrt{A}(\alpha^2 - \|z\|^2)}.$$

This metric has constant curvature $-A$. Now we obtain the following more general version of Ahlfors's Schwarz lemma.

Theorem 2.4.7. *Let U be a planar domain that is equipped with a metric σ whose curvature is bounded above by a negative constant $-B$. Then every holomorphic function $f : D(0, \alpha) \to U$ satisfies*

$$f^*\sigma(z) \le \frac{\sqrt{A}}{\sqrt{B}} \rho_\alpha^A(z), \qquad \forall z \in D(0, \alpha).$$

It is a good exercise to construct a proof of this more general result, modeled of course on the proof of Theorem 2.4.6.

In the next section we shall see several elegant applications of Theorems 2.4.6 and 2.4.7.

2.5 Liouville and Picard Theorems

It turns out that curvature gives criteria for when there do or do not exist nonconstant holomorphic functions from a domain U_1 to a domain U_2. The most basic result along these lines is as follows

Theorem 2.5.1. *Let $U \subseteq \mathbb{C}$ be an open set equipped with a metric $\sigma(z)$ having the property that its curvature $\kappa_\sigma(z)$ satisfies*

$$\kappa_\sigma(z) \leq -B < 0$$

for some positive constant B and for all $z \in U$. Then any holomorphic function

$$f : \mathbb{C} \to U$$

must be constant.

Proof. For $\alpha > 0$ we consider the mapping

$$f : D(0, \alpha) \to U.$$

Here $D(0, \alpha)$ is the Euclidean disk with center 0 and radius α, equipped with the metric $\rho_\alpha^A(z)$ as in the last section. Fix $A > 0$. Theorem 2.4.7 yields, for any fixed z and $\alpha > |z|$, that

$$f^*\sigma(z) \leq \frac{\sqrt{A}}{\sqrt{B}} \rho_\alpha^A(z).$$

Letting $\alpha \to +\infty$ yields

$$f^*\sigma(z) \leq 0$$

hence

$$f^*\sigma(z) = 0\,.$$

This can be true only if $f'(z) = 0$. Since z was arbitrary, we conclude that f is constant. □

An immediate consequence of Theorem 2.5.1 is Liouville's theorem:

Theorem 2.5.2. *Any bounded entire function is constant.*

Proof. Let f be such a function. After multiplying f by a constant, we may assume that the range of f lies in the unit disk. However, the Poincaré metric on the unit disk has constant curvature -4. Thus Theorem 2.5.1 applies and f must be constant. □

Remark 2.5.3. This argument does not apply if the range is, for example, $\mathbb{C}$—because the plane $\mathbb{C}$ does not support a metric of strictly negative curvature. We shall refine the proof of Theorem 2.5.2 in the ensuing discussions.

Picard's theorem is a dramatic strengthening of Liouville's theorem. It says that the hypothesis "bounded" may be weakened considerably, yet the same conclusion may be drawn.

Let us begin our discussion with a trivial example.

Example 2.5.4. Let f be an entire function taking values in the set

$$S = \mathbb{C} \setminus \{x + i0 : 0 \leq x \leq 1\}.$$

Following f by the mapping

$$\varphi(z) = \frac{z}{z-1}\,,$$

we obtain an entire function $g = \varphi \circ f$ taking values in $\mathbb{C}$ less the set $\{x + i0 : x \leq 0\}$. If $r(z)$ is the principal branch of the square root function, then $h(z) = r \circ g(z)$ is an entire function taking values in the right halfplane. Now the Cayley map

$$c(z) = \frac{z-1}{z+1}$$

takes the right halfplane to the unit disk. So $u(z) = c \circ h(z)$ is a bounded entire function. We conclude from Theorem 2.5.3 that u is constant. Unraveling our construction, we have that f is constant.

The point of this easy example is that far from being bounded, an entire function need only omit a segment from its values in order that it be forced to be constant. And a small modification of the proof shows that the segment can be arbitrarily short. Picard considered the question of how small a set the image of a nonconstant entire function can omit.

Let us pursue the same line of inquiry rather modestly by asking whether a nonconstant entire function can omit one complex value. The answer is "yes," for $f(z) = e^z$ assumes all complex values except zero. It also turns out that it is impossible to construct a metric on the plane less a point that has negative curvature bounded away from zero.

The next step is to ask whether a nonconstant entire function f can omit two values. The striking answer, discovered by Picard, is "no." Because of Theorem 2.5.1, it suffices for us to prove the following:

Theorem 2.5.5. *Let U be a planar open set such that $\mathbb{C} \setminus U$ contains at least two points. Then U admits a metric μ whose curvature $\kappa_\mu(z)$ satisfies*

$$\kappa_\mu(z) \leq -B < 0$$

for some positive constant B.

Proof. After applying an invertible affine map to U we may take the two omitted points to be 0 and 1 (if there are more than two omitted points, we may ignore the extras). Thus we will construct a metric of strictly negative curvature on $\mathbb{C}_{0,1} \equiv \mathbb{C} \setminus \{0,1\}$.

Define

$$\mu(z) = \left[\frac{(1+\|z\|^{1/3})^{1/2}}{\|z\|^{5/6}}\right] \cdot \left[\frac{(1+\|z-1\|^{1/3})^{1/2}}{\|z-1\|^{5/6}}\right].$$

(After the proof, we will discuss where this nonintuitive definition came from.) Of course μ is positive and smooth on $\mathbb{C}_{0,1}$. We proceed to calculate the curvature of μ.

First notice that, away from the origin,

$$\Delta\left(\log\|z\|^{5/6}\right) = \frac{5}{12}\Delta\left(\log\|z\|^2\right) = 0\,.$$

Thus

$$\begin{aligned}\Delta \log\left[\frac{(1+\|z\|^{1/3})^{1/2}}{\|z\|^{5/6}}\right] &= \frac{1}{2}\Delta\log\left(1+\|z\|^{1/3}\right)\\ &= 2\frac{\partial}{\partial z\partial \bar{z}}\log\left(1+[z\cdot\bar{z}]^{1/6}\right).\end{aligned}$$

Now a straightforward calculation leads to the identity

$$\Delta \log\left[\frac{(1+\|z\|^{1/3})^{1/2}}{\|z\|^{5/6}}\right] = \frac{1}{18}\frac{1}{\|z\|^{5/3}(1+\|z\|^{1/3})^2}\,.$$

The very same calculation shows that

$$\Delta \log\left[\frac{(1+\|z-1\|^{1/3})^{1/2}}{\|z-1\|^{5/6}}\right] = \frac{1}{18}\frac{1}{\|z-1\|^{5/3}(1+\|z-1\|^{1/3})^2}\,.$$

The definition of curvature now yields that

$$\begin{aligned}\kappa_\mu(z) = -\frac{1}{18}\Bigg[&\frac{\|z-1\|^{5/3}}{(1+\|z\|^{1/3})^3(1+\|z-1\|^{1/3})}\\ &+\frac{\|z\|^{5/3}}{(1+\|z\|^{1/3})(1+\|z-1\|^{1/3})^3}\Bigg].\end{aligned}$$

We record the following obvious facts:

a) $\kappa_\mu(z) < 0$ for all $z \in \mathbb{C}_{0,1}$;

b) $\lim_{z\mapsto 0} \kappa_\mu(z) = -\dfrac{1}{36}$;

c) $\lim_{z \mapsto 1} \kappa_\mu(z) = -\frac{1}{36}$;
d) $\lim_{z \mapsto \infty} \kappa_\mu(z) = -\infty$.

It follows immediately that κ_μ is bounded from above by a negative constant. □

Remark 2.5.6. Now we discuss the motivation for the construction of μ. On looking at the definition of μ, one sees that the first factor is singular at 0 and the second is singular at 1. Let us concentrate on the first of these.

Since the expression defining curvature is rotationally invariant, it is plausible that the metric we define would also be rotationally invariant. Thus it should be a function of $\|z\|$. So one would like to choose an α such that $\|z\|^\alpha$ defines a metric of negative curvature. However, a calculation reveals that the α that is suitable for z large is not suitable for z small and vice versa. This explains why the expression has powers both of $\|z\|$ (for behavior near 0) and of $(1 + \|z\|)$ (for behavior near ∞). A similar discussion applies to the factors $\|z - 1\|^\alpha$.

Our calculations thus lead us to design the metric so that it behaves like $\|z\|^{-4/3}$ near ∞ and behaves like $\|z\|^{-5/6}$ (respectively $\|z - 1\|^{-5/6}$) near 0 (respectively 1).

We formulate Picard's little theorem as a corollary of Theorem 2.5.1 and Theorem 2.5.5:

Corollary 2.5.7. *Let f be an entire analytic function taking values in a set U. If $\mathbb{C} \setminus U$ contains at least two points, then f is constant.*

Proof. Since $\mathbb{C} \setminus U$ contains at least two points, Theorem 2.5.5 says that there is a metric of strictly negative curvature on U. But then Theorem 2.5.1 implies that any entire function taking values in U is constant. □

Entire functions are of two types: there are polynomials and non-polynomials *(transcendental* entire functions). Notice that a polynomial has a pole at infinity. Conversely, any entire function with a pole at infinity is a polynomial. So a transcendental entire function cannot have a pole at infinity and, being unbounded (by Liouville), cannot have a removable singularity at infinity. It therefore has an essential singularity at infinity.

Now notice that, by the fundamental theorem of algebra, a polynomial assumes all complex values. For a transcendental function, we analyze its essential singularity at infinity by recalling the Casorati–Weierstrass theorem: if 0 is an essential singularity for a holomorphic function f on a punctured disk $D'(0, \epsilon) \equiv D(0, \epsilon) \setminus \{0\}$, then f assumes values on $D'(0, \epsilon)$ that are *dense* in the complex plane. One might therefore conjecture that the essential feature of Picard's theorem is not that the function being considered is entire, but rather that it has an essential singularity at infinity. This is indeed the case and is the content of the great Picard theorem, which we now state:

Theorem 2.5.8. *Let f be a holomorphic function on a punctured disk $D'(p,r) \equiv D(p,r) \setminus \{p\}$. Assume that f has an essential singularity at p. Then the restriction of f to any punctured disk $D'(p,s)$, $0 < s < r$, omits at most one value.*

We shall provide a proof of the great Picard theorem when we treat Theorem 3.4.5 below.

2.6 The Schwarz Lemma at the Boundary

There has been interest for some time in studying Schwarz lemmas at the boundary of a domain. Löwner conducted early studies of a result weaker than the one presented here; his motivation was the study of distortion theorems. Our methods are quite distinct from Löwner's (see [VEL] for a consideration of classical results with references). The standard reference for this new material is the paper [BUK] of Burns and Krantz.

Theorem 2.6.1. *Let $\varphi : D \to D$ be a holomorphic function such that*

$$\varphi(\zeta) = 1 + (\zeta - 1) + \mathcal{O}(\|\zeta - 1\|^4)$$

as $\zeta \to 1$. Then $\varphi(\zeta) \equiv \zeta$ on the disk.

Compare this result with the uniqueness part of the classical Schwarz lemma. In that context, we assume that $\varphi(0) = 0$ and $\|\varphi'(0)\| = 1$. At the boundary we must work harder. In fact, the following example shows that the size of the error term cannot be reduced from fourth order to third order.

Example 2.6.2. The function

$$\varphi(\zeta) = \zeta - \frac{1}{10}\,(\zeta - 1)^3$$

satisfies the hypotheses of the theorem with the exponent 4 replaced by 3. Yet clearly this φ is not the identity.

To verify this example, we need only check that φ maps the disk to the disk. It is useful to let $\zeta = 1 - \tau$ and consider therefore the function

$$\widetilde{\varphi}(\tau) = 1 - \tau + \left[\frac{1}{10}\,\tau^3\right] .$$

Now the result is clear by inspection. If $\arg \tau$ is near to 0, then of course the expression in brackets cannot push $\widetilde{\varphi}(\tau)$ past the edge of the disk. If $\arg \tau$ is larger, then one checks by hand that the expression in brackets has argument that causes it to push the value of the function *into* the disk. We leave details to the interested reader.

Proof of the Theorem. Consider the holomorphic function

$$g(\zeta) = \frac{1+\varphi(\zeta)}{1-\varphi(\zeta)}\,.$$

Then g maps the unit disk D to the right halfplane. By the Herglotz representation theorem,[1] there must be a positive measure μ on the interval $[0, 2\pi)$ and an imaginary constant $\mathcal{C}$ such that

$$g(\zeta) = \frac{1}{2\pi}\int_0^{2\pi} \frac{e^{i\theta}+\zeta}{e^{i\theta}-\zeta}\,d\mu(\theta) + \mathcal{C}\,. \tag{2.6.1}$$

We use the hypothesis on φ to analyze the structure of g and hence that of μ. To wit,

$$\begin{aligned} g(\zeta) &= \frac{1+\zeta+\mathcal{O}(\zeta-1)^4}{1-\zeta-\mathcal{O}(\zeta-1)^4} \\ &= \frac{1}{1-\zeta}\cdot\frac{(1+\zeta)+(\zeta-1)^4}{1-\mathcal{O}(\zeta-1)^3} \\ &= \frac{1}{1-\zeta}\cdot\left[(1+\zeta)+\mathcal{O}(\zeta-1)^4\right]\cdot\left[1+\mathcal{O}(\zeta-1)^3\right] \\ &= \frac{1+\zeta}{1-\zeta}+\mathcal{O}(\zeta-1)^2\,. \end{aligned}$$

From this and equation (2.6.1) we easily conclude that the measure μ has the form $\delta_0+\nu$, where δ_0 is (2π times) the Dirac mass at the origin and ν is another positive measure on $[0, 2\pi)$. [In fact, a nice way to verify the positivity of ν is to use the equation

$$\frac{1+\zeta}{1-\zeta}+\mathcal{O}(\zeta-1)^2 = \frac{1}{2\pi}\int_0^{2\pi}\frac{e^{i\theta}+\zeta}{e^{e\theta}-\zeta}\,d(\delta_0+\nu)(\theta)+\mathcal{C} \tag{2.6.2}$$

to derive a Fourier–Stieltjes expansion of $\delta_0+\nu$ and then to apply the Herglotz criterion for the expansion of a positive measure (see [KAT, p. 38]).

We may simplify equation (2.6.2) to

$$\mathcal{O}(\zeta-1)^2 = \frac{1}{2\pi}\int_0^{2\pi}\frac{e^{i\theta}+\zeta}{e^{i\theta}-\zeta}\,d\nu(\theta)+\mathcal{C}\,.$$

---

[1] The reference [AHL3] has a thorough treatment of Herglotz's result. The reader may find it convenient to think of the result in this way. For $0 < r < 1$, the functions $G_r(e^{i\theta}) = \operatorname{Re} g(re^{i\theta})$ are all positive and all have mean value $g(0)$. Thus they form a bounded set in $L^1[0, 2\pi) \subseteq \mathcal{M}([0, 2\pi))$. Here $\mathcal{M}$ stands for the space of finite Borel measure, which is certainly the dual of the continuous functions. By the Banach–Alaoglu theorem (see [RUD3]), there is a subsequence G_{r_j} that converges weak-$*$ to a limit measure μ. This measure of course must be positive, and it is the measure that we seek. The Cauchy integral of μ gives a holomorphic function with the same real part as g. That real part determines the imaginary part up to an additive imaginary constant $\mathcal{C}$.

Now pass to the real part of the last equation, thereby eliminating the constant $\mathcal{C}$. Since ν is a positive measure, we see that the real part of the integral on the right represents a positive harmonic function h on the disk that satisfies

$$h(\zeta) = \mathcal{O}(\zeta - 1)^2 . \tag{2.6.3}$$

In particular, h takes a minimum at the point $\zeta = 1$ and is $\mathcal{O}(\zeta - 1)^2$ there as well. This contradicts Hopf's lemma (Lemma 5.3.1 below) unless $h = 0$. More specifically, Hopf's lemma tells us that a positive harmonic function with a minimum at the boundary point 1 must have nonvanishing normal derivative at 1. But line (2.6.3) contradicts that assertion.

The only possible conclusion is that that $\nu \equiv 0$, so that $\mu = \delta_0$. As a result,

$$g(\zeta) = \frac{1+\zeta}{1-\zeta} .$$

In conclusion, $\varphi(\zeta) \equiv \zeta$. The theorem is proved. □

Now we will treat some versions of the boundary Schwarz lemma that are due to Robert Osserman [OSS]. One of the remarkable features of Osserman's work is that he directly relates the interior Schwarz lemma to the boundary Schwarz lemma. The first result is reminiscent of Proposition 2.2.2.

Lemma 2.6.3. *Let $f : D \to D$ be a holomorphic function on the disk such that $f(0) = 0$. Then*

$$\|f(z)\| \le \|z\| \cdot \frac{\|z\| + \|f'(0)\|}{1 + \|f'(0)\|\|z\|} . \tag{2.6.4}$$

Proof. Set $g(z) = f(z)/z$. Then the standard Schwarz lemma tells us that either $\|g(z)\| < 1$ for all $z \in D$ or else f is a rotation. In the latter case, $\|f'(0)\| = 1$ and (2.6.4) holds trivially. Thus we need only consider the case $\|g(z)\| < 1$.

Since inequality (2.6.4) is unaffected by rotations, we may as well suppose that $g(0) = f'(0) = a$, where $0 \le a < 1$. In this case (2.6.4) is equivalent to

$$\|g(z)\| \le \frac{\|z\| + a}{1 + a\|z\|} . \tag{2.6.5}$$

This last inequality is our Proposition 2.2.2. In particular, g must map each disk $D(0, r)$ into the image of that disk under the linear fractional map

$$\varphi_{-a}(\zeta) = \frac{\zeta + a}{1 + a\zeta} .$$

This latter image is the Euclidean disk whose diameter is the interval

$$\left[\frac{a-r}{1-ar}\,,\;\frac{a+r}{1+ar}\right]$$

in the real axis.

In conclusion,

$$\|z\| = r \Rightarrow \|g(z)\| \le \frac{a+r}{1+ar} = \frac{|z|+a}{1+a\|z\|}\,.$$

This estimate establishes (2.6.5), and that in turn proves (2.6.4). □

Now we have a version of the boundary Schwarz lemma:

Proposition 2.6.4. *Let $f : D \to D$ be holomorphic and suppose that $f(0) = 0$. Further assume that, for some $b \in \partial D$, f extends continuously to b, $\|f(b)\| = 1$, and also that $f'(b)$ exists. Then*

$$\|f'(b)\| \ge \frac{2}{1+\|f'(0)\|}\,.$$

Proof. We use Lemma 2.6.3. If $b, c \in \partial D$ with $f(b) = c$, then

$$\left\|\frac{f(z)-c}{\|z\|-\|b\|}\right\| \ge \frac{1-\|f(z)\|}{1-\|z\|} \ge \frac{1+\|z\|}{1+\|f'(0)\|\|z\|}\,.$$

As $\|z\| \to 1$, the right-hand side tends to $2/[1+\|f'(0)\|]$. This reasoning proves that

$$\liminf_{z_j\to b}\left\|\frac{f(z_j)-c}{\|z_j\|-\|b\|}\right\| \ge \liminf_{z_j\to b}\frac{1-\|f(z_j)\|}{1-\|z_j\|} \ge \frac{2}{1+\|f'(0)\|}\,.$$

If we choose $z_j = (1-1/j)b$, then the result follows. □

Osserman's more general boundary Schwarz lemma is as follows:

Theorem 2.6.5. *Let $f : D \to D$ be a holomorphic function such that, for some $b \in \partial D$, f extends continuously to b, $\|f(b)\| = 1$, and $f'(b)$ exists. [Note that we do not assume that $f(0) = 0$.] Define*

$$F(z) = \frac{f(z)-f(0)}{1-\overline{f(0)}f(z)}\,.$$

Then

$$\|f'(b)\| \ge \frac{2}{1+\|F'(0)\|}\cdot\frac{1-\|f(0)\|}{1+\|f(0)\|}\,.$$

Proof. Now F satisfies the hypotheses of Lemma 2.6.3, so that

$$\|F'(b)\| \ge \frac{2}{1+\|F'(0)\|}\,. \tag{2.6.6}$$

But it is easy to calculate that

$$F'(z) = f'(z) \cdot \frac{1 - \|f(0)\|^2}{[1 - \overline{f(0)} f(z)]^2} \,.$$

Observe that $\|f(b)\| = 1$ implies that

$$\|1 - \overline{f(0)} f(b)\| \geq 1 - \|\overline{f(0)} f(b)\| = 1 - \|f(0)\| \,,$$

hence

$$\|F'(b)\| = \|f'(b)\| \cdot \frac{1 - \|f(0)\|^2}{\|1 - \overline{f(0)} f(b)\|^2} \leq \|f'(b)\| \cdot \frac{1 + \|f(0)\|}{1 - \|f(0)\|} \,. \tag{2.6.7}$$

Now (2.6.6) and (2.6.7) yield the desired conclusion. □

An interesting corollary of Proposition 2.6.4 is the following.

Corollary 2.6.6. *Under the hypotheses of Proposition 2.6.4, we have*

$$\|f'(b)\| \geq 1$$

and

$$\|f'(b)\| > 1 \quad \textit{unless } f \textit{ is a rotation.}$$

Proof. These two inequalities follow from Proposition 2.6.4 and the classical Schwarz lemma. □

Another interesting corollary, sort of an integrated form of the boundary Schwarz lemma, is the next result.

Corollary 2.6.7. *Let f satisfy all the conditions of Proposition 2.6.4 except the hypothesis about the boundary point b. Assume that f extends continuously to a boundary arc $C \subseteq \partial D$ with $|f(z)| = 1$ for $z \in C$. Then the length s of C and the length σ of $f(C)$ satisfy*

$$\sigma \geq \frac{2}{1 + \|f'(0)\|} \cdot s \,.$$

Proof. By the Schwarz reflection principle, f extends to be holomorphic on C and therefore automatically satisfies the condition of Proposition 2.6.4 at the boundary point b for every point of C. Thus the corollary is immediate from the proposition. □

Problems for Study and Exploration

1. The Schwarz lemma on the disk implies the Cauchy estimates and vice versa. Explain.

2. Fix a domain $\Omega \subseteq \mathbb{C}$. Let $P \in \Omega$. Define the *Carathéodory metric* at P to be the quantity

$$F_C^{\Omega}(P) = \sup\{\|f'(P)\| : f \text{ is holomorphic on } \Omega, \|f\| < 1, f(P) = 0\}.$$

Define the *Kobayashi metric* at P to be the quantity

$$F_K^{\Omega}(P) = \inf\left\{\frac{1}{\|f'(P)\|} : f \text{ is holomorphic on } D, \right.$$
$$\left. f \text{ takes values in } \Omega, f(0) = P\right\}.$$

Use the Schwarz lemma to prove that $F_C^{\Omega}(P) \le F_K^{\Omega}(P)$ always.

3. Refer to Exercise 2 for terminology. Prove that if Ω_1, Ω_2 are domains and $\varphi : \Omega_1 \to \Omega_2$ is a holomorphic mapping, then

$$F_C^{\Omega_2}(\varphi(P)) \cdot \|\varphi'(P)\| \le F_C^{\Omega_1}(P)$$

for all $P \in \Omega_1$. Likewise show that

$$F_K^{\Omega_2}(\varphi(P)) \cdot \|\varphi'(P)\| \le F_K^{\Omega_1}(P)$$

for all $P \in \Omega_1$.

4. Refer to Exercises 2 and 3. Show that if $\Omega_1 \subseteq \Omega_2$ are domains, then

$$F_C^{\Omega_2}(\varphi(P)) \le F_C^{\Omega_1}(P)$$

for all $P \in \Omega_1$ and

$$F_K^{\Omega_2}(\varphi(P)) \le F_K^{\Omega_1}(P)$$

for all $P \in \Omega_1$.

5. Formulate and prove a version of the Schwarz lemma for holomorphic mappings of the strip $\Omega = \{z \in \mathbb{C} : -1 < \operatorname{Im} z < 1\}$ to itself.

6. Formulate a version of the Burns–Krantz theorem for the strip. Is the exponent still 4? Give an example to show that if the exponent 4 is replaced by 3 then the result is false.

7. Give examples to show that the Schwarz–Pick lemma is sharp at each point of the disk.

8. Give an example to show that Osserman's Schwarz lemma is sharp.

9. Let $f : D \to D$ be a holomorphic function, not necessarily one-to-one or onto. What does the Schwarz lemma say about how many fixed points f can have?

10. Refer to Exercise 9. What can you say about fixed points of f in the boundary?

11. Can you refine the Burns–Krantz theorem if there are two fixed boundary points instead of one?

3

Normal Families

Genesis and Development

The concept of normal family is an outgrowth of the standard technique for proving the Riemann mapping theorem. Recall that the mapping function is produced as the solution of a certain extremal problem, and showing that that extremal problem actually has a solution is a byproduct of a normal families argument.

Normal families have grown into a subject in its own right. There are many interesting geometric conditions that imply that a family of functions is normal. Also, as Hung-Hsi Wu has taught us [WU], the theory of normal families is intimately bound up with ideas about invariant metrics in complex analysis.

Today there are differential conditions, metric conditions, topological conditions, geometric conditions, value-distribution conditions, and many combinations of these that can be used to show that a family of functions is normal. By way of the Ascoli–Arzelà theorem, we can see that "normality" is not only the province of complex function theory. It is actually a very useful property for solution sets of partial differential equations and in many different venues of geometric analysis.

In the present chapter we explore the world of normal families, and their applications to questions in complex analysis. The reader who has progressed this far in the book will have no trouble digesting the ideas in the present chapter.

3.1 Introduction

The concept of "normal family" is one of the most elegant and most powerful in complex function theory. It hinges on several key ideas: **(i)** the topology on the space of holomorphic functions, **(ii)** the Cauchy estimates, **(iii)** the

Ascoli–Arzelà theorem. When properly viewed, Montel's theorem is really just the triangle inequality formulated in certain invariant metrics.

The applications of normal families are manifold. Of course they are used decisively in the modern proof of the Riemann mapping theorem. They are used in the proof of Picard's theorems, especially the great Picard theorem. They are used to study automorphisms of domains. And they can be used in the estimation of invariant metrics.

In the present chapter we shall explore all of these avenues, and imbue the reader with a greater appreciation for normal families.

3.2 Topologies on the Space of Holomorphic Functions

Let $\Omega \subseteq \mathbb{C}$ be a domain. Let $\mathcal{O}(\Omega)$ denote the linear space of holomorphic functions on Ω. The most natural, and most commonly used, topology on $\mathcal{O}(\Omega)$ is the compact-open topology. This is just the same as the topology of uniform convergence on compact sets.[1] Most any complex analysis text will use this topology exclusively. For example, a basic theorem in the subject is that a sequence $\{f_j\}$ of holomorphic functions that converges uniformly on compact sets has a holomorphic function as its limit.

The most elementary topology on functions is that of pointwise convergence. Yet one rarely encounters that topology in the subject of complex function theory. There is, however, at least one interesting result about this topology that ought to be recorded.

Proposition 3.2.1. *Let $\Omega \subseteq \mathbb{C}$ be a domain. Let $\{f_j\}$ be holomorphic functions on Ω. Assume that the f_j converge pointwise to a function f on Ω. Then f is holomorphic on a dense, open subset of Ω.*

Proof. Let V be a subset of Ω that is the closure of an open set, and is bounded. For each $k = 1, 2, \ldots$ we define

$$S_k = \{z \in V : |f_j(z)| \leq k \text{ for all } j = 1, 2, \ldots\} .$$

[Note that we have resumed the habit of using $|\ \ |$ to denote modulus of a complex number.] Then clearly $V = \cup_k S_k$. It follows from the Baire category theorem that some S_k has nonempty interior. Say that $D(p,r) \subseteq S_k$.

For $0 < r' < r$, we consider the Cauchy integral formula on $\overline{D}(p,r')$. Thus, for $z \in \overline{D}(p,r')$,

$$f_j(z) = \frac{1}{2\pi i} \oint_{\partial D(p,r')} \frac{f_j(\zeta)}{\zeta - z}\, d\zeta .$$

[1] The remarkable book [LUR] explores the basic development of complex function theory from the point of view of the holomorphic functions viewed as a topological vector space equipped with this topology.

Now let $\epsilon > 0$. By Egorov's theorem from measure theory, there is a set $E \subseteq \partial D(p, r')$ such that $|{}^c E| < \epsilon \cdot r'/4k$ and $\{f_j\}$ converges uniformly on E.[2] Choose a number $N > 0$ such that $j, j' > N$ implies $|f_j(z) - f_{j'}(z)| < \epsilon \cdot r'/4$ for all $z \in E$. Then, if $z \in D(p, r'/2)$, we have

$$\begin{aligned}
|f_j(z) - f_{j'}(z)| &= \left| \frac{1}{2\pi i} \oint_{\partial D(p,r')} \frac{f_j(\zeta) - f_{j'}(\zeta)}{\zeta - z} \, d\zeta \right| \\
&\le \frac{1}{2\pi} \int_E \left| \frac{f_j(\zeta) - f_{j'}(\zeta)}{\zeta - z} \right| |d\zeta| + \frac{1}{2\pi} \int_{{}^c E} \left| \frac{f_j(\zeta) - f_{j'}(\zeta)}{\zeta - z} \right| |d\zeta| \\
&\le \frac{1}{2\pi} \cdot |E| \cdot \frac{\epsilon \cdot r'}{4} \cdot \frac{1}{r'/2} + \frac{1}{2\pi} |{}^c E| \cdot 2k \cdot \frac{1}{r'/2} \\
&\le \frac{1}{2\pi} \cdot 2\pi \cdot \frac{\epsilon \cdot r'}{4} \cdot \frac{1}{r'/2} + \frac{1}{2\pi} \cdot \frac{\epsilon r'}{4k} \cdot 2k \cdot \frac{1}{r'/2} \\
&< \frac{\epsilon}{2} + \frac{\epsilon}{2} \\
&= \epsilon \, .
\end{aligned}$$

Thus we see that every subset $V \subseteq \Omega$, as specified above, has a disk in it on which the convergence is uniform. We may produce a dense, open subset $\mathcal{V}$ of Ω that is a union of such sets V. By a standard diagonalization argument, we may show that every subsequence of $\{f_j\}$ has a subsequence that converges uniformly on compact subsets of $\mathcal{V}$. It follows then that the full sequence $\{f_j\}$ converges uniformly on compact subsets of $\mathcal{V}$. That is the desired result. $\square$

In fact, Siciak [SIC] has studied the nature of the set $\Omega \setminus \mathcal{V}$ in terms of capacity theory.

3.3 Normal Families in Their Natural Context

Let us begin by giving a definition of normal family of *holomorphic functions*. These considerations will lead us naturally to extend our notion of normal family to a broader class of functions.

Fix a domain $\Omega \subseteq \mathbb{C}$ and let $\{f_j\}$ be a sequence of functions on Ω. We say that the sequence *converges uniformly on compact sets* if there is a function f on Ω such that, for each compact $K \subseteq \Omega$ and each $\epsilon > 0$, there is an $N > 0$ such that, when $j > N$, $|f_j(z) - f(z)| < \epsilon$ for each $z \in K$. We sometimes say that the family *converges normally*.

With Ω a domain and $\{f_j\}$ holomorphic on Ω, we say that $\{f_j\}$ is *compactly divergent* if, for each pair of compact sets $K \subseteq \Omega$ and $L \subseteq \mathbb{C}$, there is an $N > 0$ such that if $j > N$, then $f_j(z) \notin L$ for all $z \in K$.

The correct definition of normal family in view of these definitions is this:

[2] Here we use the notation ${}^c S$ to denote the complement of the set S.

Definition 3.3.1. Let $\Omega \subseteq \mathbb{C}$ be a domain. Let $\mathcal{F} = \{f_\alpha\}_{\alpha \in \mathcal{A}}$ be a family of functions on Ω. We say that $\mathcal{F}$ is a *normal family* if each subsequence $\{f_j\} \subseteq \mathcal{F}$ either has a subsequence $\{f_{j_k}\}$ that converges uniformly on compact sets or else has a subsequence $\{f_{j_\ell}\} \subseteq \mathcal{F}$ that is compactly divergent.

In fact, the last definition makes more sense in a more general context. Instead of holomorphic scalar-valued functions, let us consider *meromorphic functions*. A meromorphic function should be thought of as a holomorphic function taking values in the Riemann sphere $\widehat{\mathbb{C}}$. What does this mean?

The Riemann sphere $\widehat{\mathbb{C}}$ is a complex manifold. If

$$f : \Omega \to \widehat{\mathbb{C}}$$

and $f(p) = \infty$, then we say that f is holomorphic near p if $1/f$ is holomorphic at p in the classical sense. [What is going on here, of course, is that we are applying a coordinate map to the image of f near the point ∞. See [LOS] for more on manifolds.] The important point here is that a sequence of functions

$$f_j : \Omega \to \widehat{\mathbb{C}}$$

that converges uniformly on compact sets might in fact be compactly divergent. Consider as an illustration the next example.

Example 3.3.2. Let $f_j(z) = z^j$. Think of these as functions taking values in $\widehat{\mathbb{C}}$.

On the domain $\Omega_1 = \{z \in \mathbb{C} : |z| < 1\}$, the sequence converges uniformly on compact sets to 0.

On the domain $\Omega_2 = \{z \in \mathbb{C} : |z| > 1\}$, the sequence converges uniformly on compact sets to ∞ (if we think of the functions as taking values in $\widehat{\mathbb{C}}$). We may also say that, on Ω_2, the sequence is compactly divergent.

Now the elegant and all-inclusive formulation of Montel's theorem—in a temporarily heuristic format—is this:

Theorem 3.3.3 (Montel). *Let $\Omega \subseteq \mathbb{C}$ and let $\mathcal{F} = \{f_\alpha\}_{\alpha \in \mathcal{A}}$ be a family of meromorphic functions on Ω. If $\mathcal{F}$ satisfies a suitable growth or value-distribution condition, then $\mathcal{F}$ is a normal family.*

This formulation of Montel's celebrated result is not very satisfying, just because it is too vague. The most naive condition to put on $\mathcal{F}$ is to mandate that $|f(z)| \leq M$ for some universal constant M and for all $f \in \mathcal{F}$, $z \in \Omega$. But that would rule out compact divergence and would also rule out meromorphic functions. We need something more general. Even the more general condition, commonly found in texts (see [RUD2] or [GRK1]), that $|f(z)| \leq M_K$ for each compact set $K \subseteq \Omega$ and some constant M_K depending on K would rule out compact divergence.

In fact, one of the most succinct formulations of Montel's theorem is this:

Theorem 3.3.4 (Montel). *Let $\{\mathcal{F}\}$ be a family of meromorphic functions on a domain $\Omega \subseteq \mathbb{C}$. If there are three values $\alpha, \beta, \gamma \in \widehat{\mathbb{C}}$ such that no $f \in \mathcal{F}$ takes any of the values α, β, γ, then $\mathcal{F}$ is a normal family.*

Note in particular that this new version of Montel includes as a special case the standard hypothesis that the family is uniformly bounded. But it includes a number of other interesting cases as well. It is not, however, a universal or all-inclusive result. For instance, it does not include the result about a family of functions that is uniformly bounded on compact sets.

There is just one unifying theorem in the subject, and that is Marty's theorem. We treat that result now.

As already indicated, it is most convenient to think of normal families in the context of *meromorphic* functions. This is just because we want to allow for convergence to infinity (formerly called "compact divergence"). Thus we will think of our functions as holomorphic maps from a domain $\Omega \subseteq \mathbb{C}$ to the Riemann sphere $\widehat{\mathbb{C}}$. Since we want to look at things geometrically, we shall equip each of Ω and $\widehat{\mathbb{C}}$ with a metric. In fact, since uniform convergence on compact sets is a local property, we may as well take Ω to be the unit disk. We shall equip the disk with the Poincaré or Poincaré–Bergman metric (see Sections 1.1, 1.2). We shall equip the Riemann sphere with the spherical metric.

The latter metric may be unfamiliar, so let us take a few moments to discuss it. Imagine the stereographic projection p from the plane to the unit sphere—see Figure 3.1. For a point $(z, 0) = (x + iy, 0)$ in the plane, the intersection of the sphere with the line

$$t \mapsto (tx, ty, (1 - t))$$

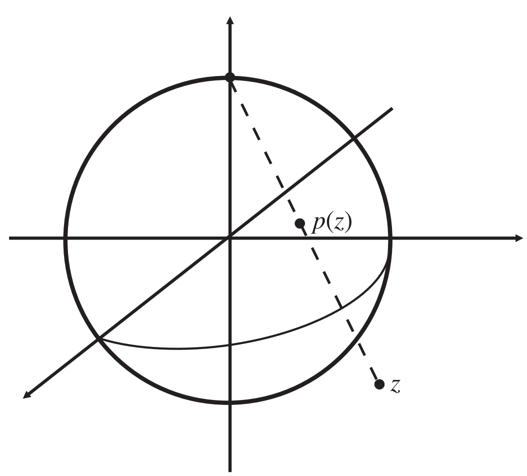

Fig. 3.1. Stereographic projection.

is (along with the north pole $(0,0,1)$) the point

$$p(z) \equiv \left(\frac{2x}{|z|^2+1}, \frac{2y}{|z|^2+1}, \frac{|z|^2-1}{|z|^2+1} \right) .$$

Now if $(z,0) = (x+iy,0)$ and $(w,0) = (u+iv,0)$ are points in the plane, then we measure their distance by taking the 3-dimensional Euclidean distance between $p(z)$ and $p(w)$. The result is

$$\begin{aligned} s(z,w) &= \sqrt{\left(\frac{2x}{|z|^2+1} - \frac{2u}{|w|^2+1} \right)^2 + \left(\frac{2x}{|z|^2+1} - \frac{2v}{|w|^2+1} \right)^2} \\ &\qquad \overline{+ \left(\frac{|z|^2-1}{|z|^2+1} - \frac{|w|^2-1}{|w|^2+1} \right)^2} \\ &= \frac{2|z-w|}{\sqrt{1+|z|^2}\sqrt{1+|w|^2}} . \end{aligned}$$

This is the spherical metric. Some references call this the *chordal metric.*

It is also useful to have the infinitesimal form of the spherical metric. This can be calculated by pulling back the infinitesimal 3-dimensional Euclidean metric under the stereographic projection, or by differentiating the metric that we just calculated. We take the latter approach. In point of fact,

$$\begin{aligned} \lim_{|h|\to 0} \frac{s(z+h,z) - s(z,z)}{|h|} &= \lim_{|h|\to 0} \frac{2|h|/[\sqrt{1+|z|^2}\sqrt{1+|z+h|^2}] - 0}{|h|} \\ &= \lim_{|h|\to 0} \frac{2}{\sqrt{1+|z|^2}\sqrt{1+|z+h|^2}} \\ &= \frac{2}{1+|z|^2} . \end{aligned}$$

Thus if η is a tangent vector to $\widehat{\mathbb{C}}$ at the point w, then

$$|\eta|_{\mathsf{sph},w} = \frac{2\|\eta\|}{1+|w|^2} .$$

Here of course $\| \cdot \|$ denotes the standard Euclidean length.

Remark 3.3.5. As an exercise, the reader may wish to calculate that the *spherical distance* (i.e., the distance along the sphere) between $z,w \in \widehat{\mathbb{C}}$ is

$$2\tan^{-1} \left| \frac{z-w}{1+\overline{w}z} \right| .$$

This quantity in fact exceeds $s(z,w)$, as it should on geometric grounds.

Before we proceed, we offer a few simple remarks about the spherical metric:

(1) It may be calculated directly that the spherical distance is invariant under the inversion $z \mapsto 1/z$, $w \mapsto 1/w$. Details are left as an exercise.

(2) On any compact subset of $\mathbb{C}$, the spherical metric is comparable to the Euclidean metric. This result is immediate, for the numerator of our expression for the spherical metric is already the Euclidean metric. On a compact subset of $\mathbb{C}$, the denominator is bounded above and below.

(3) On any compact subset of $(\mathbb{C} \setminus \{0\}) \times (\mathbb{C} \setminus \{0\})$, we have the estimates

$$c \cdot \left| \frac{1}{z} - \frac{1}{w} \right| \leq s(z,w) \leq C \cdot \left| \frac{1}{z} - \frac{1}{w} \right| .$$

Here c, C are positive constants. We invite the reader to verify this assertion.

(4) Our original definition of "uniform convergence on compact sets" may now be formulated in the language of the spherical metric. To wit, let $\Omega \subseteq \mathbb{C}$ be a domain. A sequence $\{f_j\}_{j=1}^{\infty}$ of functions $f_j : \Omega \to \widehat{\mathbb{C}}$ converges uniformly on compact sets to a limit function f (we sometimes say that the sequence $\{f_j\}$ converges *normally*) if $s(f_j(z), f(z))$ converges uniformly to 0 on compact subsets of Ω. Notice that this definition of normal convergence allows for the possibility of $f_j(z)$ converging to ∞ at certain values of z. In other words, the definition using the spherical metric amalgamates the ordinary notion of uniform convergence on compact sets with the newer notion of compact divergence.

Of course a holomorphic function $g : \Omega \to \widehat{\mathbb{C}}$ is actually meromorphic. But we will sometimes find it convenient to refer to such functions as "holomorphic." No confusion should result if we keep the range of the function clearly in mind.

Lemma 3.3.6. *Suppose that $f_j : \Omega \to \widehat{\mathbb{C}}$ is a normally convergent sequence of holomorphic functions. Then the limit function f is a holomorphic function from Ω to $\widehat{\mathbb{C}}$ (in other words, the limit function is either meromorphic or identically ∞).*

Proof. At any point z where $f(z)$ is a finite complex number, we can be sure that there is a neighborhood U of z such that for j sufficiently large, the f_j are uniformly bounded on U. But then f is necessarily a holomorphic function on U. At any point z where $f(z) = \infty$, we know that $1/f_j$ is uniformly bounded and converges uniformly to $1/f$ on some neighborhood U' of z. Now the same argument shows that $1/f$ is holomorphic in a neighborhood of z. In case $1/f \equiv 0$, then $f \equiv \infty$. Otherwise the zeros of $1/f$ are isolated; so f is meromorphic as claimed. $\square$

We note that, in case the functions f_j in the lemma are actually holomorphic (not meromorphic), then the limit function f is either holomorphic or identically equal to ∞. The proof is essentially the same.

Now here is our philosophical approach to the matter of normal families. We begin with a sketch of the most classical Montel theorem. We have a family $\mathcal{F}$ of holomorphic functions on a domain Ω. The hypothesis is that $|f(z)| \le M$ for all $f \in \mathcal{F}$ and all $z \in \Omega$. Using the Cauchy estimates, this information implies that on a slightly smaller domain $\widetilde{\Omega}$ we have $|f'(z)| \le C \cdot M$ for all $z \in \widetilde{\Omega}$. Now the mean value theorem tells us that, for $z \in \widetilde{\Omega}$ and h small,

$$|f(z) - f(z+h)| \le C \cdot M|h|$$

for all $f \in \mathcal{F}$. But this means that the family $\mathcal{F}$ is equicontinuous on $\widetilde{\Omega}$. Of course $\mathcal{F}$ is equibounded by hypothesis. Now the Ascoli–Arzelà theorem applies to tell us that every sequence in $\mathcal{F}$ has a subsequence that converges uniformly on compact sets. That is the result.

The nub of this proof is the uniform bound on the derivative of any function $f \in \mathcal{F}$. This is the only step in the proof where “holomorphic” is used. Everything else in the proof is elementary topology and calculus. And this will be the inspiration for our formulation of Marty's result. We simply demand, for a family of meromorphic functions $f : D \to \widehat{\mathbb{C}}$, that the derivative be bounded from the Poincaré–Bergman metric on D (the unit disk) to the spherical metric on $\widehat{\mathbb{C}}$. Concretely, this condition is that

$$f^{\#}(z) \equiv \frac{2|f'(z)|}{1 + |f(z)|^2} \le C \cdot \frac{1}{1 - |z|^2} \,.$$

In particular, this inequality means that $f^{\#}$ is bounded—independent of f—uniformly on compact subsets of D.

The expression $f^{\#}(z)$ is called the *spherical derivative* of f at the point z. Before we develop the theory any further, we make a few remarks about the spherical derivative.

1. If $f : \Omega \to \widehat{\mathbb{C}}$ and $\gamma : [0,1] \to \Omega$ is a continuously differentiable curve, then the spherical length of $f \circ \gamma$ is

$$\int_\gamma f^{\#}(z)|dz| \,.$$

2. The spherical derivative is invariant under inversion:

$$[1/f]^{\#}(z) = f^{\#}(z) \,.$$

Now we come to the standard formulation of Marty's theorem as can be found in many standard complex function theory texts (see, for example, [GAM]). First we need a lemma.

Lemma 3.3.7. *If f_j are meromorphic functions on the disk D and $f_j \to f$ uniformly on compact sets, then $f_j^{\#} \to f^{\#}$ uniformly on compact sets in D.*

Proof. This result follows as usual from the Cauchy estimates. □

Theorem 3.3.8 (Marty). *Let $\mathcal{F}$ be a family of meromorphic functions on a domain $\Omega \subseteq \mathbb{C}$. Then $\mathcal{F}$ is normal if and only if, for each compact set $K \subseteq \Omega$ and $z \in K$, we have*

$$f^{\#}(z) \leq C_K ,$$

where C_K is a constant that depends only on K.

Proof. We may take $C_K = 1$. Fix a point $p \in \Omega$ and suppose that the spherical derivatives of elements of $\mathcal{F}$ are uniformly bounded on a neighborhood of p. Say that $f^{\#}(z) \leq C$ for $z \in D(p, r)$. If $q \in D(p, r)$ is any element and if γ is the straight line segment connecting p to q then, for any $f \in \mathcal{F}$, we may estimate the spherical distance from $f(p)$ to $f(q)$ by

$$s(f(p), f(q)) \leq \int_\gamma f^{\#}(z)\, |dz| \leq |p - q| .$$

Thus we see that the family $\mathcal{F}$ is equicontinuous—from the Euclidean metric to the spherical metric. The Ascoli–Arzelà theorem now implies that $\mathcal{F}$ is normal.

Conversely, suppose that the spherical derivatives of the elements of $\mathcal{F}$ are *not* uniformly bounded on compact subsets of Ω. Then there are elements $f_j \in \mathcal{F}$ such that the maximum of $f_j^{\#}$ over some compact $K \subseteq \Omega$ tends to $+\infty$. By the lemma, the sequence $\{f_j\}$ cannot have a normally convergent subsequence. □

Exercise for the Reader: Show that under suitable hypotheses,

$$(g \circ f)^{\#}(z) = g^{\#}(f(z)) \cdot |f'(z)| . \qquad \diamondsuit$$

Exercise for the Reader: Let γ be a compact curve in the complex plane that contains more than one point. Show that a family $\mathcal{F}$ of holomorphic functions on a domain Ω, taking values in $\mathbb{C} \setminus \gamma$, will be normal. ◊

Exercise for the Reader: Suppose that $\mathcal{F}$ is a family of holomorphic functions $f : \Omega \to \mathbb{C}$ and that

$$|f'(z)| \leq e^{|f(z)|}$$

for all $f \in \mathcal{F}$ and all $z \in \Omega$. Then prove that $\mathcal{F}$ is a normal family. ◊

3.4 Advanced Results on Normal Families

One of the unifying themes in the theory of normal families is Lawrence Zalcman's result that follows. It is based on an old idea in function theory called the "Bloch principle." We shall see this technique used to good effect to establish a general version of Montel's theorem and thereby the Picard theorems. We shall also study Abraham Robinson's *Ansatz* about conditions that force a family of holomorphic functions to be normal.

Proposition 3.4.1. *Assume that $\mathcal{F}$ is a family of meromorphic functions $f : \Omega \to \widehat{\mathbb{C}}$ that is not normal. Then there exist*

(a) *Points $z_j \in \Omega$ converging to a point $z \in \Omega$;*
(b) *Numbers $\rho_j > 0$ converging to 0;*
(c) *Functions $f_j \in \mathcal{F}$ such that the dilated functions $g_j(\zeta) \equiv f_j(z_j + \rho_j \zeta)$ converge normally to a nonconstant meromorphic function g on $\mathbb{C}$ satisfying $g^{\#}(0) = 1$ and $g^{\#}(\zeta) \le 1$ for all $\zeta \in \mathbb{C}$.*

Conversely, if these three conditions hold, then the family $\mathcal{F}$ is not normal.

Abraham Robinson's principle (which we shall discuss below) posits that there is a relationship between normal families and entire functions. Zalcman's proposition begins to make this connection palpable. For we see that a family that is not normal can be forced, after dilations and translations, to converge to an entire function with certain extremal properties.

Example 3.4.2. Zalcman's proposition is most useful in theoretical settings. But we shall present now a very simple example of how it works.

Let $f_j(z) = z^j$ and $\mathcal{F} = \{f_j\}_{j=1}^{\infty}$. Let $\Omega = \{z \in \mathbb{C} : 1/2 < |z| < 2\}$. Then $\mathcal{F}$ is not normal.

Now select $z_j = 1$ for all j and $\rho_j = 1/j$. We plainly see that

$$\lim_{j \to \infty} f_j(z_j + \rho_j \zeta) = \lim_{j \to \infty} \left(1 + \frac{\zeta}{j}\right)^j = e^{\zeta} \equiv g(\zeta) .$$

Furthermore,

$$g^{\#}(\zeta) = \frac{2|e^{\zeta}|}{1 + |e^{\zeta}|^2} .$$

Therefore $g^{\#}(0) = 1$ and (with a small calculation) $g^{\#}(\zeta) \le 1$ for all ζ. We conclude that the family $\mathcal{F}$ is not normal.

Proof of Zalcman's Proposition. We may apply Marty's theorem to find a sequence $\{\alpha_j\}$ in some compact subset of Ω and functions $f_j \in \mathcal{F}$ such that $f_j^{\#}(\alpha_j) \to +\infty$. Translating coordinates, we may as well suppose that $\alpha_j \to 0 \in \Omega$. Dilating coordinates, we may also assume that the closed unit disk $\overline{D}(0,1)$ lies in Ω.

Set

$$R_j \equiv \max_{|z| \le 1} [f_j^{\#}(z)] \cdot (1 - |z|) .$$

Since $\alpha_j \to 0$ and $f_j^{\#}(\alpha_j) \to +\infty$, we know immediately that $\lim_{j \to \infty} R_j = +\infty$. Suppose, for each j, that $[f_j^{\#}(z)] \cdot (1 - |z|)$ attains its maximum on the closed disk at the point β_j. Then

$$R_j = [f_j^{\#}(\beta_j)] \cdot (1 - |\beta_j|) .$$

Since $f_j^{\#}(\beta_j) \ge R_j$, we may conclude that $f_j^{\#}(\beta_j) \to +\infty$.

Now we set

$$\rho_j = \frac{1}{f_j^{\#}(\beta_j)} .$$

Certainly $\rho_j > 0$ and $\rho_j \to 0$. Consider the disk parametrized by

$$\zeta \mapsto \beta_j + \rho_j \zeta , \quad |\zeta| < R_j . \tag{3.4.1}$$

This disk has center β_j and radius $1 - |\beta_j| = \rho_j R_j$. The disk defined in (3.4.1) lies in the unit disk and hence certainly is a subset of Ω.

Define

$$g_j(\zeta) = f_j(\beta_j + \rho_j \zeta) , \quad |\zeta| < R_j .$$

Since $R_j \to \infty$, the functions g_j are defined on larger and larger sets that exhaust $\mathbb{C}$. By the chain rule, we calculate that

$$g_j^{\#}(\zeta) = \rho_j \cdot f_j^{\#}(\beta_j + \rho_j \zeta) \ , |\zeta| < R_j .$$

Fix a positive number K. If j is so large that $R_j > K$, then $g_j(\zeta)$ is certainly defined on the disk $D(0, K)$. Since

$$[f_j^{\#}(\beta_j + \rho_j \zeta)] \cdot (1 - |\beta_j + \rho_j \zeta|) \le R_j ,$$

we may conclude that

$$\begin{aligned} g_j^{\#}(\zeta) &\le \rho_j \cdot \frac{R_j}{1 - |\beta_j + \rho_j \zeta|} \\ &\le \frac{\rho_j R_j}{1 - |\beta_j| - \rho_j K} \\ &= \frac{\rho_j R_j}{\rho_j R_j - \rho_j K} \\ &= \frac{1}{1 - K/R_j} , \quad |\zeta| < K . \end{aligned}$$

Marty's theorem now tells us that on the disk $D(0, K)$, the functions $\{g_j\}$ form a normal family (we should restrict attention to j large so that all the functions are defined on $D(0, K)$). Passing to a subsequence, and re-indexing, we may thus assert that $\{g_j\}$ converges normally on $\mathbb{C}$ to a meromorphic function g. We know that, for each fixed K, the quantity $1/[1 - K/R_j]$ tends to 1. Hence the estimate on $g_j^{\#}$ shows that $g^{\#}(\zeta) \le 1$ for all $\zeta \in \mathbb{C}$. Finally, because $g_j^{\#}(0) = \rho_j f_j^{\#}(\beta_j) = 1$, we know that $g^{\#}(0) = 1$. That concludes the proof of the first half of the result.

For the converse direction, suppose that $\mathcal{F}$ is a normal family. Say that $|z_j| < r$ for every j and some $r > 0$. By Marty's theorem, there is a constant $K > 0$ such that

$$\max_{|z| \le [1+r]/2} f^{\#}(z) \le K$$

for all $f \in \mathcal{F}$. Now suppose that

$$f_j(z_j + \rho_j \zeta) \to g(\zeta)$$

spherically uniformly on compact subsets of Ω. Fix a point $\zeta \in \mathbb{C}$. For j large, we know that $|z_j + \rho_j \zeta| \leq [1+r]/2$. Hence

$$\rho_j f^{\#}(z_j + \rho_j \zeta) \leq \rho_j K \, .$$

As a result, for all $\zeta \in \mathbb{C}$, we have

$$g^{\#}(\zeta) = \lim_{j \to \infty} \rho_j f_j^{\#}(z_j + \rho_j \zeta) = 0 \, .$$

We conclude that g is constant (perhaps ∞). That concludes the proof of the converse direction. □

The following terminology will prove useful in our consideration of Picard's theorems.

Definition 3.4.3. Let f be meromorphic on a punctured disk $D'(p,r) \equiv D(p,r) \setminus \{p\}$ (we allow the possibility that f is actually holomorphic on $D(p,r)$). A value $\beta \in \widehat{\mathbb{C}}$ is called an *omitted value* of f at the point p if there is a number $\delta > 0$ such that $f(z) \neq \beta$ for all $0 < |z-p| < \delta$. An omitted value for f at the point $p = \infty$ is defined similarly.

Theorem 3.4.4 (Montel). *Let $\mathcal{F}$ be a family of meromorphic functions on a domain Ω. If there are extended complex numbers $\alpha, \beta, \gamma \in \widehat{\mathbb{C}}$ such that every $f \in \mathcal{F}$ omits each of α, β, γ, then the family $\mathcal{F}$ is normal.*

Proof. Of course normality is a local property, so we may as well suppose that $\Omega = D(0,1)$. By postcomposing each element of $\mathcal{F}$ with a linear fractional transformation, we can assume that the omitted values are $0, 1, \infty$. Thus, in particular, the elements of $\mathcal{F}$ are actually holomorphic. Since these functions are nonvanishing on the simply connected domain D, they have roots of all orders. Define

$$\mathcal{F}_k = \{f^{1/2^k} \in \mathcal{F} : f \in \mathcal{F}\} \, .$$

The functions in $\mathcal{F}_k$ omit the values $0, 1$, and all 2^kth roots of unity. Evidently the family $\mathcal{F}_k$ is normal if and only if the family $\mathcal{F}$ is normal.

Seeking a contradiction, we now suppose that the family $\mathcal{F}$ is *not* normal. Thus, for each k, the family $\mathcal{F}_k$ is not normal. For each k, let g^k be the entire function for the family $\mathcal{F}_k$ that is produced by Zalcman's proposition. Thus $(g^k)^{\#}(0) = 1$ for each k and $(g^k)^{\#}(\zeta) \leq 1$ for all k and all ζ. Also, each g^k is the limit of scaled, translated functions in $\mathcal{F}_k$.

Since the family $\mathcal{F}_k$ omits the 2^kth roots of unity, then so do the translations, dilations, and restrictions of the elements of $\mathcal{F}_k$. By Hurwitz's theorem, so does any limit function. Hence each g^k omits the values that are the 2^kth roots of unity. By Marty's theorem, $\{g^k\}$ is a normal family (because the spherical derivatives are uniformly bounded). Let g^0 be the normal limit of some subsequence. By Hurwitz's theorem, g^0 will omit *all* 2^kth roots of unity for *all* k (it also omits 0 and ∞). Because g^0 is an open mapping, it must be therefore that g^0 omits every value on the unit circle.

Thus the image of g^0, which is certainly connected, lies either in $D(0,1)$ or in $\mathbb{C} \setminus \overline{D(0,1)}$. In the first instance—since g^0 is entire—we see that g^0 is constant. In the second instance, the function $1/g^0$ (remember that g^0 does not vanish!) is constant. This is of course a contradiction because certainly $(g^0)^{\#}(0) = 1$. □

Recall that the Casorati–Weierstrass theorem tells us that the distribution of values of an analytic function near an essential singularity is dense in the complex numbers. The theorems of Picard strengthen this result in a decisive fashion.

Theorem 3.4.5 (Picard's Great Theorem). *Suppose that f is holomorphic on a punctured disk $D'(p,r) \equiv D(p,r) \setminus \{p\}$. If f has an essential singularity at p, then f cannot omit two values at p.*

Proof. We may as well suppose that $p = 0$ and that the function omits the two values 0 and 1. We shall establish then that f cannot have an essential singularity at $p = 0$.

Let $\delta_1 > \delta_2 > \cdots \to 0$. Set

$$f_j(z) = f(\delta_j z) \ , \quad 0 < |z| < r \, .$$

Then $\{f_j\}$ omits three values (the two hypothesized plus ∞). Hence, by Montel, $\{f_j\}$ is a normal family. Let f^0 be a subsequential limit on $D'(p,r)$.

If f^0 is not identically ∞, then f^0 is holomorphic on $D'(p,r)$. Let $0 < s < r$. There is a positive number M such that $|f^0(z)| < M$ when $|z| = s$. For j large, we may then conclude that $|f_j(z)| < M$ when $|z| = s$. In conclusion, $|f(z)| < M$ for $|z| = \delta_j s$. By the maximum principle, $|f(z)| < M$ for $\delta_j s \leq |z| \leq s$, $j = 1, 2, \ldots$. Taking the union over j, we find that $|f(z)| < M$ on the punctured disk $D'(p,s)$. Now the Riemann removable singularities theorem implies that f continues analytically across the puncture at p. That is a contradiction.

If instead f^0 *is* identically ∞, then we simply apply the preceding argument to the function $1/f^0$. The conclusion is that $1/f$ continues analytically across p with value 0 at p. Thus f has a pole at p. That is a contradiction. □

Theorem 3.4.6 (Picard's Little Theorem). *A nonconstant entire function f cannot omit two values.*

Proof. If f is a polynomial, then f takes *all* values by the fundamental theorem of algebra.

If f is transcendental, then f has an essential singularity at ∞. Now apply the result of the preceding theorem (exercise for the reader). □

It is easy to find examples that illustrate the principle of Picard's theorems. First, a polynomial is entire and omits no values. The function $f(z) = e^z$ is entire and omits just one value. The function $f(z) = e^{1/z}$ has an essential singularity at the origin and omits the value 0 near the origin. The function ze^z is entire and omits no values. The function $\cos z$ is entire and omits no values.

3.5 Robinson's Heuristic Principle

In his retiring presidential address to the Association for Symbolic Logic [ROB], Abraham Robinson proposed an intuitive condition for normality of a family of holomorphic or meromorphic functions. It says that any "property" that would cause an entire function to be constant would also cause a family of holomorphic or meromorphic functions to be normal. If we ignore for the moment the fact that we do not have a rigorous definition of "property," then we can certainly appreciate Robinson's idea. For example, "boundedness" will cause an entire function to be constant. It will also, by Montel's theorem, cause a family of holomorphic functions to be normal. Omission of two values will cause an entire function to be constant. It will also cause a family of holomorphic functions to be normal.

In the present section we shall consider Robinson's principle. We shall follow the presentation in [ZAL], which gives the principle a rigorous formulation and a proof. In what follows, we will consider (instead of functions in the ordinary sense) *function elements*. A function element is an ordered pair (f, U), where $U \subseteq \mathbb{C}$ is an open set and f a holomorphic or meromorphic function on U. A *property* P is a collection of function elements.

Definition 3.5.1. Let P be a property that satisfies the following conditions:

(a) If $(f, U) \in P$ and $\widetilde{U} \subseteq U$, then $(f, \widetilde{U}) \in P$.
(b) If $(f, U) \in P$ and $\phi(\zeta) = a\zeta + b$ is an affine holomorphic mapping, then $(f \circ \phi, \phi^{-1}(U)) \in P$.
(c) Let $(f_j, U_j) \in P$, where $U_1 \subseteq U_2 \subseteq \cdots$, and set $U = \cup_j U_j$. If $f_j \to f$ uniformly on compact sets in the spherical metric, then $(f, U) \in P$.

We call such a property *critical*.

Theorem 3.5.2. *Let P be a critical property. Assume that $(f, \mathbb{C}) \in P$ if and only if f is a constant function. Then, for any domain U, the family of functions f satisfying $(f, U) \in P$ is normal.*

This theorem is a rigorous enunciation of Abraham Robinson's idea. According to Zalcman [ZAL], it is the outgrowth of ideas of Pommerenke. We shall now prove the result and then offer some examples and applications. The reference [ALK] offers some generalizations of Zalcman's ideas.

Proof of Theorem 3.5.2. Let $\mathcal{F}$ be the family of all functions on a domain U that has property P. If $\mathcal{F}$ is not normal, then Marty's condition shows that it is not normal on some subdisk, which we may take to be $D(0,1)$.

We apply Proposition 3.4.1 and use that notation. Fix $r < 1$ and define $R_j = (r - |z_j|)/\rho_j$. Since $R_j \to \infty$, we may suppose (passing to a subsequence if necessary) that the R_j form an increasing sequence. Let $g_j(\zeta) = f_j(z_j + \rho_j \zeta)$ and $D_j = D(0, R_j) = \{\zeta \in \mathbb{C} : |\zeta| < R_j\}$. The functions (g_j, D_j) satisfy Property P by conditions **(a)** and **(b)**. Now condition **(c)** implies that the limit element $(g, \mathbb{C})$ also satisfies Property P. Since P has no nonconstant functions that are defined on all of $\mathbb{C}$, we now have a contradiction. It follows that $\mathcal{F}$ must be a normal family. □

It is important now that we illustrate Robinson's heuristic principle with a few applications. We begin with a new proof of Montel's theorem.

Theorem 3.5.3 (Montel). *Let $\mathcal{F}$ be a family of meromorphic functions on the domain Ω. If $\mathcal{F}$ omits the three values α, β, γ, then $\mathcal{F}$ is a normal family.*

Proof. We apply Robinson's principle. Take for Property P the condition "either f is constant or it omits the values α, β, γ on Ω." We see immediately that properties **(a)** and **(b)** of "critical property" hold. Also, condition **(c)** is a consequence of Hurwitz's theorem. Of course we know that any meromorphic function on all of $\mathbb{C}$ that satisfies Property P must be constant—that is Picard's Great Theorem. The proof is complete. □

We confess that our treatment here is not entirely satisfactory. After all, we used Montel's theorem earlier in the chapter to *prove* Picard's Great Theorem. Now we are using Picard to prove Montel. The concerned reader may consult [GRK1], for example, where Picard is proved using the elliptic modular function, and thus in a fashion that is independent of the present discussion. The book [KRA5] presents yet another point of view.

Zalcman [ZAL] presents a rather exotic version of Montel's theorem, one that is not very well known. We recount it here.

Theorem 3.5.4. *Let $\mathcal{F}$ be a family of meromorphic functions on the domain Ω. Assume that each function $f \in \mathcal{F}$ omits three distinct values—but the choice of the three values may depend on f. Call the omitted values $a = a_f$, $b = b_f$, and $c = c_f$. Suppose that s is the spherical distance and there is a constant $\lambda > 0$ (independent of $f \in \mathcal{F}$) such that we always have*

$$s(a,b) \cdot s(b,c) \cdot s(c,a) \geq \lambda > 0\,.$$

Then f is a normal family.

Proof. Let P be the property "f omits three values a, b, c such that $s(a,b) \cdot s(b,c) \cdot s(c,a) \geq \lambda$." By Picard's theorem, no nonconstant meromorphic function can enjoy Property P. Certainly properties **(a)** and **(b)** of "critical property" are satisfied. The proof will be complete if we can show that Property P is preserved under uniform convergence in the spherical metric.

So let us say that $f_j \to f$ spherically uniformly on compact subsets of Ω and that each f_j omits the values a_j, b_j, c_j with $s(a_j, b_j) \cdot s(b_j, c_j) \cdot s(c_j, a_j) \geq \lambda > 0$. We may suppose that f is nonconstant, for otherwise it trivially satisfies Property P. Now the sphere is compact, so there exist points $a, b, c \in \widehat{\mathbb{C}}$ and a subsequence—still indexed by j—such that $s(a_j, a) \to 0$, $s(b_j, b) \to 0$, and $s(c_j, c) \to 0$. Continuity implies that $s(a, b) \cdot s(b, c) \cdot s(c, a) \geq \lambda$. We shall show that f omits the values a, b, c.

If $f(w) = a$ and $a \neq \infty$, then select a number $r > 0$ such that $K = \{z \in \mathbb{C} : |z - w| \leq r\} \subseteq \Omega$ and f is holomorphic on K. Of course f is bounded on K, so $f_j(z) - a_j \to f(z) - a$ on K. The function $f(z) - a$ is nonconstant and vanishes at $z = w$. Hurwitz's theorem then implies that $f_j(z) - a_j$ must vanish on K for large j. That is a contradiction.

If instead $a = \infty$ then consider the functions $1/f_j$, $1/f$. Arguing as before, and using the invariance of the spherical metric under inversion, we see that the proof is complete. □

Problems for Study and Exploration

1. Let $\{f_\alpha\}_{\alpha \in \mathcal{A}}$ be a normal family of holomorphic functions. Prove that $\{f'_\alpha\}_{\alpha \in \mathcal{A}}$ is a normal family.
2. If $\{f_\alpha\}_{\alpha \in \mathcal{A}}$ is a normal family and if $\{F_\alpha\}_{\alpha \in \mathcal{A}}$ is a collection of antiderivatives, does it follow that $\{F_\alpha\}_{\alpha \in \mathcal{A}}$ is a normal family?
3. Refer back to Exercises 2, 3, 4 in Chapter 2 for the concepts of Carathéodory and Kobayashi metrics. Prove that if Ω is a bounded domain and $P \in \Omega$, then there is a function f that actually realizes the supremum in the definition of the Carathéodory metric and there is also a function g that realizes the infimum in the definition of the Kobayashi metric.
4. Prove that if $\{f_\alpha\}$ is a family of holomorphic functions on a domain Ω and if $\|f_\alpha\|_{L^2} \leq C < \infty$, then $\{f_\alpha\}$ is a normal family.
5. Let f_j be a sequence of meromorphic functions on a domain $\Omega \subseteq \mathbb{C}$, taking values in $\widehat{\mathbb{C}}$. Assume that $\{f_j\}$ converges pointwise to a limit function f. What can you say about the holomorphicity/meromorphicity of the limit function f? [*Hint:* You will need the Baire category theorem.]
6. Let $\mathcal{S}$ consist of the functions f holomorphic on the unit disk D such that **(i)** $f(0) = 0$, **(ii)** $f'(0) = 1$, and **(iii)** f is univalent. Prove that $\mathcal{S}$ is a normal family.
7. A function f holomorphic on the unit disk is said to be a *normal function* if, whenever $\{\varphi_j\}$ is a collection of automorphisms of the disk, then $\{f \circ \varphi_j\}$ is a normal family. Verify that any bounded holomorphic function on the disk is normal. Verify that any holomorphic function on the disk whose range omits two values is normal.
8. Refer to Exercise 7. Define a concept of *normal meromorphic function.* Formulate properties analogous to those in Exercise 7 for normal meromorphic functions and prove them.

9. Refer to Exercises 7 and 8. Use Marty's characterization of normality to give a differential characterization of normal function—one that makes no reference to automorphisms of the disk.

10. In fact, a successful solution of Exercise 9 will teach you that a normal function is one whose derivative is bounded from the Poincaré metric on the disk to the spherical metric on the Riemann sphere. Formulate this statement precisely and prove it.

11. Prove that if $\{f_j\}$ is a sequence of holomorphic functions on a domain U such that $\lim_{j\to\infty} f_j(z)$ exists for each point z in a set E that has an interior accumulation point in U, then $\{f_j\}$ converges uniformly on compact subsets of U to a holomorphic function on U.

12. Prove that if $\{f_j\}$ is a uniformly bounded sequence of holomorphic functions on the disk D such that $\{f_j\}$ converges at each $a_j \in U$, $j = 1, 2, \ldots$, and if $\sum_j (1 - |a_j|) = \infty$, then $\{f_j\}$ converges uniformly on compact subsets of D to a holomorphic function.

13. Let $\mathcal{F}$ be a normal family of holomorphic functions on a domain U. Suppose that each limit function of this family is *not* equal to a constant a. Then, for each compact set $K \subseteq U$, there is a constant $M = M(K)$ such that, for each $f \in \mathcal{F}$, the number of zeros of $f(z) - a$ in K does not exceed M.

14. Prove Schottky's theorem: Let f be a holomorphic function in the disk $D(0, R)$ that does not assume the values 0 or 1. Let $f(0) = a_0$. Then, for each value θ with $0 < \theta < 1$, there is a positive constant $M(a_0, \theta)$ such that

$$|f(z)| \leq M(a_0, \theta) \qquad \text{for all} \quad |z| \leq \theta R\,.$$

4

The Riemann Mapping Theorem and Its Generalizations

Genesis and Development

The Riemann mapping theorem has been said by some to be the greatest theorem of the nineteenth century. The entire *concept* of the theorem is profoundly original, and its proof introduced many new ideas. Certainly normal families and the use of extremal problems in complex analysis are just two of the important techniques that have grown out of studies of the Riemann mapping theorem.

As has been noted previously, conformal mappings lend themselves well to study by way of invariant geometry. But even the classical Euclidean geometry of the plane—by way of the Jordan curve theorem, the Lusin area integral, the basic idea of conformality, and the study of special mappings like the Schwarz–Christoffel mappings—plays a central role in analytic studies.

The Riemann mapping theorem has many variants and generalizations. Certainly adaptations of the result to nonsimply-connected domains has been an important avenue of exploration. Generalizations to Riemann surfaces (the uniformization theorem) have been profoundly influential. And, even though the Riemann mapping theorem fails dramatically in several complex variables, the pursuit of such a theorem has been a driving force in the subject.

Familiarity with the Riemann mapping theorem, and with the concepts of conformal mapping theory, will be helpful in your study of this chapter.

Certainly the Riemann mapping theorem is a standard part of any graduate course in complex variables. The theorem says this.

Theorem 4.0.1. *Let D be the unit disk. Let $\Omega \subseteq \mathbb{C}$ be a simply connected domain that is not the entire complex plane. Then there is a one-to-one, onto holomorphic mapping*

$$\Phi : D \to \Omega\,.$$

The standard textbook proof of this result considers a fixed point $P \in \Omega$ and the set

$$\mathcal{S} = \{f : D \to \Omega : f(0) = P \text{ and } f \text{ is one-to-one}\}\,.$$

Using normal families and the argument principle, one shows that there is an extremal function f_0 such that $|f_0'(0)|$ is maximal. It is a result of the argument principle that f_0 is univalent. It follows from this extremal property that f_0 is onto, and that completes the proof.

Riemann's original proof of the theorem solved a somewhat different extremal problem related to the Dirichlet problem. His argument was flawed because he in fact *assumed* that the extremal problem had a solution. He did not prove it. The proof was rescued some years later using the Dirichlet principle. Our modern proof—due to L. Fejér and F. Riesz—sidesteps those difficulties, and uses Montel's theorem to show that the extremal problem has a solution.

There are a number of other approaches to proving the Riemann mapping theorem. We shall begin the considerations of this chapter by presenting a proof that relies on the solution of the Dirichlet problem. This is in fact quite similar in spirit to Riemann's original approach to the theorem.

4.1 The Riemann Mapping Theorem by way of the Green's Function

Now fix a domain $\Omega \subseteq \mathbb{C}$. We shall assume that Ω has twice continuously differentiable boundary—just so that we can make good use of some results about the Laplacian and associated ideas (see Chapter 8 for further details).

Fix a point $b \in \Omega$. Let F_b be the solution of the Dirichlet problem with data $(-1/[2\pi]) \log |z - b|$. Define

$$G(b, z) = -\frac{1}{2\pi} \log |z - b| - F_b(z)\,.$$

This is of course the Green's function for Ω with pole at b.

Now let $\widetilde{G}(b, z)$ be the harmonic conjugate of G. Define

$$F(z) = e^{-G(b,z) - i\widetilde{G}(b,z)}\,.$$

Then F is holomorphic and vanishes at b (because of the logarithmic singularity of G at b. Observe that if $z \in \partial\Omega$, then

$$|F(z)| = \left|e^{-G(b,0) - i\widetilde{G}(b,0)}\right| = \left|e^{-G(b,0)}\right| = e^0 = 1\,. \tag{4.1.1}$$

Since the Green's function is positive, $-G$ is negative and hence F maps Ω into the disk. We claim that F is a one-to-one, onto mapping of Ω to the unit disk D.

To verify the claim, first note that $|F(z)| \to 1$ as $z \to \partial\Omega$. The function F has only one zero, a simple zero, at b. If α is any point of the disk, then we may apply the argument principle to conclude that $F(z) - \alpha$ has precisely one zero in Ω. Thus F is both one-to-one and onto.

Just for fun, we now also present an ad hoc argument to see these assertions. By a translation of coordinates, we may take b to be the origin. Now the Riemann mapping theorem provides a conformal mapping $\varphi : \Omega \to D$ such that $\varphi(0) = 0$. We know that the Green's function for D at 0 is $(-1/[2\pi]) \log |z|$. It follows immediately that $(-1/[2\pi]) \log |\varphi(z)|$ is the Green's function for Ω at 0. So

$$G(z,0) = -\frac{1}{2\pi} \log |\varphi(z)| .$$

Exponentiating both sides and adding in a conjugate function of $G(z,0)$, we find that

$$\varphi(z) = e^{-G(z,0) - i\widetilde{G(z,0)}} .$$

Of course this equation is quite similar to (4.1.1). In fact, they differ only by a unimodular multiplicative constant. We conclude that the function F defined above is one-to-one and onto.

4.2 Canonical Representations for Multiply Connected Regions

Any simply connected domain, except the plane, can be conformally represented as the unit disk. What can be said for domains of higher connectivity?

For a doubly connected domain (i.e., a domain whose complement has two connected components, a bounded component and an unbounded component), there is a remarkable and elegant result.

Theorem 4.2.1. *Let $\Omega \subseteq \mathbb{C}$ be a doubly connected domain such that no component of the complement is a point. Then Ω is conformally equivalent to an annulus.*

We will in fact prove something more general.

Definition 4.2.2. Let $\Omega \subseteq \mathbb{C}$ be a domain. We say that Ω is *finitely connected* if its complement $\widehat{\mathbb{C}} \setminus \Omega$ has finitely many connected components. We say that Ω is k-connected if $\widehat{\mathbb{C}} \setminus \Omega$ has k components.

Theorem 4.2.3. *Let $\Omega \subseteq \mathbb{C}$ be a k-connected domain, $k \geq 2$, and assume that no component of the complement is a point. Then there is a one-to-one, onto, holomorphic map $\varphi : \Omega \to \mathcal{B}$, where $\mathcal{B}$ is an annulus $A = \{\zeta \in \mathbb{C} : c < |\zeta| < C\}$ with $k-2$ concentric arcs (lying on circles $|\zeta| = c_1$, $|\zeta| = c_2$, ... , $|\zeta| = c_{k-2}$) removed. See Figure 4.1.*

Our proof will proceed in several steps. First we must review some basic ideas about homology, and about periods of integrals. We collect that material in the next section.

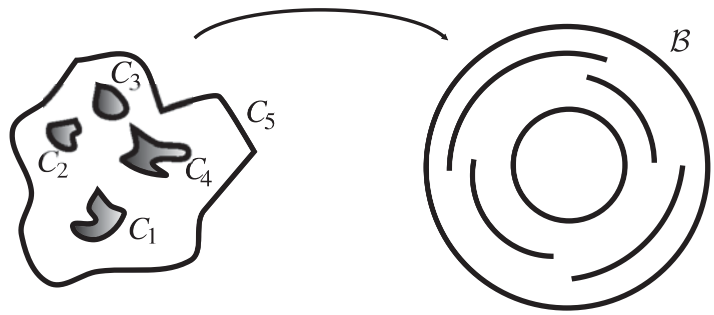

Fig. 4.1. An annulus with $k-2$ concentric arcs removed.

4.3 Review of Some Topological Ideas

We need the concepts of index of a point with respect to a curve, of homology, and of periods of integrals. We now provide a brief treatment.

All curves that we consider in this chapter will be piecewise continuously differentiable. We will not always enunciate this standing hypothesis. A closed curve will be called a *cycle*. In some contexts, it is useful to let a cycle be any formal "sum" of closed curves.

4.3.1 Cycles and Periods

Let γ be a closed curve and a a point not on that curve. Then the *index* of γ with respect to a is

$$n(\gamma, a) = \frac{1}{2\pi i} \oint_\gamma \frac{dz}{z-a} .$$

The index is always an integer (see [AHL2] or [GRK1]). Intuitively, the index measures how many times γ wraps around a, and with what orientation. More precisely, the index is positive when the orientation is counterclockwise and negative when the orientation is clockwise.

In what follows, we shall say that a closed curve γ in a domain Ω is *homologous to zero* if the index of γ with respect to any point $a \in {}^c\Omega$ is zero. In symbols, the condition is $n(\gamma, a) = 0$. Note that a domain Ω is simply connected if all closed curves in Ω are homologous to zero (see the detailed discussion in [GRK1], Chapter 11).

Now let $\Omega \subseteq \mathbb{C}$ be a multiply connected (but *finitely connected*) domain. Let the components of the complement (in the extended complex plane, or Riemann sphere) be $A_1, A_2, \ldots, A_k$ and suppose that A_k is that component that contains ∞. It is easy to find cycles γ_j, $j = 1, \ldots, k-1$, such that

- $n(\gamma_j, a) = 1$ for all $a \in A_j$;
- $n(\gamma_j, a) = 0$ for all $a \in {}^c\Omega \setminus A_j$.

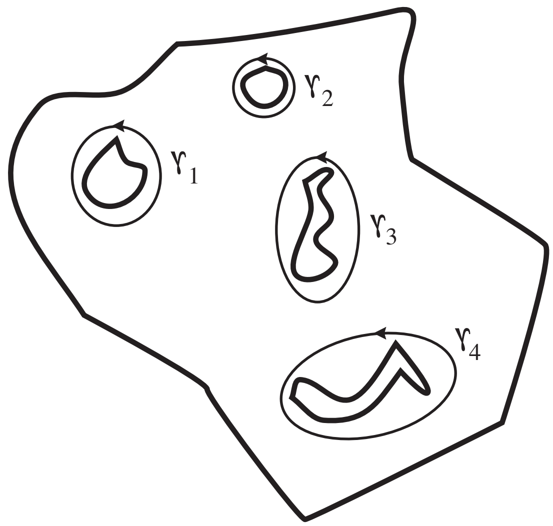

Fig. 4.2. A homology basis.

See Figure 4.2.

Suppose that γ is *any* cycle in Ω. If $a \in A_j$ then the value of $n(\gamma, a)$ is, by connectivity, independent of the choice of $a \in A_j$. Call the constant value c_j. Then the cycle $\mu = \gamma - c_1\gamma_1 - c_2\gamma_2 - \cdots - c_{k-1}\gamma_{k-1}$ has the property that its index with respect to any $a \in {}^c\Omega$ is zero.[1] [Recall that A_k is the component of the complement that contains ∞, thus it is not considered in this list.] Thus μ is homologous to 0. It convenient to phrase this idea as "the curve γ is homologous to a linear combination of $\gamma_1, \ldots, \gamma_{k-1}$." It is easy to see that the coefficients of this linear combination are uniquely determined.

We call $\gamma_1, \ldots, \gamma_{k-1}$ a *homology basis* for Ω. It is not the only homology basis, but certainly the number of elements in any homology basis will be $k - 1$. We have established that every region with finite connectivity has a finite homology basis, and vice versa.

Our homology statement has an interpretation in terms of complex line integrals. Namely, for any function f holomorphic on Ω we have

$$\oint_\gamma f(z)\,dz = c_1 \oint_{\gamma_1} f(z)\,dz + c_2 \oint_{\gamma_2} f(z)\,dz + \cdots + c_{k-1} \oint_{\gamma_{k-1}} f(z)\,dz\,.$$

This is the homology version of Cauchy's theorem. The numbers

$$P_j \equiv \oint_{\gamma_j} f(z)\,dz$$

[1] Since the c_j are integers, this notation makes sense. Here $c_j\gamma_j$ means c_j copies of the cycle γ_j.

are called the *periods* of the differential $f\,dz$.

As we know from our study of the Cauchy theory (see [AHL2] or [GRK1]), it is possible on a suitable domain (a rectangle, for instance) to define an antiderivative of a holomorphic function f by

$$F(z) = \oint_{z_0}^{z} f(\xi)\,d\xi\,. \tag{4.3.1}$$

Here the integral is understood to be along *any* piecewise continuously differentiable curve from the fixed base point z_0 to z. Part of what one proves—just using Cauchy's theorem—is that the value of the integral is independent of the choice of curve. Conversely, if f is a given holomorphic function on the rectangle $\mathcal{R}$ and F is a well-defined, single-valued antiderivative for f, then F and f are related (up to an additive constant) by (4.3.1).

We may put the elementary ideas of the last paragraphs into context by saying that, on a given domain Ω, the vanishing of all the periods is a necessary and sufficient condition for f to have a well-defined, single-valued antiderivative.

4.3.2 Harmonic Functions

If u is a harmonic function on a region Ω, then

$$f(z) = \frac{\partial u}{\partial x}(z) - i\frac{\partial u}{\partial y}(z) \tag{4.3.2}$$

defines a holomorphic function on Ω. This claim may be checked directly using the Cauchy–Riemann equations. Now it is useful to write

$$\begin{aligned} f(z)\,dz &= \left(\frac{\partial u}{\partial x} - i\frac{\partial u}{\partial y}\right)(dx + idy) \\ &= \left(\frac{\partial u}{\partial x}\,dx + \frac{\partial u}{\partial y}\,dy\right) + i\left(-\frac{\partial u}{\partial y}\,dx + \frac{\partial u}{\partial x}\,dy\right)\,. \end{aligned}$$

Observe that the real part in this last expression is simply the ordinary, real-variable differential of u:

$$du = \frac{\partial u}{\partial x}\,dx + \frac{\partial u}{\partial y}\,dy\,.$$

In the case that u has a conjugate harmonic function v on the region Ω, then of course we must have

$$dv = \frac{\partial v}{\partial x}\,dx + \frac{\partial v}{\partial y}\,dy = -\frac{\partial u}{\partial y}\,dx + \frac{\partial u}{\partial x}\,dy\,.$$

We know that in general, even on an annulus, it will not always be the case that u has a well-defined, single-valued harmonic conjugate. So it is best in general not to discuss or write dv. Instead we introduce the notation

$$^*du = -\frac{\partial u}{\partial y}\,dx + \frac{\partial u}{\partial x}\,dy\,.$$

We give $^*du$ the name *conjugate differential of du*. We thus know that

$$f\,dz = du + i\,^*du\,.$$

Now let γ be any curve that is homologous to zero in Ω. Then, by Cauchy's theorem,

$$0 = \oint_\gamma f(z)\,dz = \oint_\gamma du + i\oint_\gamma {}^*du\,.$$

Because du is exact, the first integral on the right vanishes. So we have

$$\oint_\gamma {}^*du = 0\,, \tag{4.3.3}$$

or

$$\oint_\gamma -\frac{\partial u}{\partial y}\,dx + \frac{\partial u}{\partial x}\,dy = 0\,. \tag{4.3.4}$$

We wish to give a geometric interpretation to (4.3.4).

It is convenient to write $\gamma(t) = z(t)$ for our closed curve (cycle). If $\alpha = \arg z'(t)$ is the direction of the tangent to γ, then we may write

$$\begin{aligned} dx &= |dz|\cos\alpha\,,\\ dy &= |dz|\sin\alpha\,. \end{aligned}$$

Let us consider the normal that points to the right of the tangent. Of course it will have direction $\beta = \alpha - \pi/2$. We see that $\cos\alpha = -\sin\beta$ and $\sin\alpha = \cos\beta$. In summary, the directional derivative

$$\frac{\partial u}{\partial n} = \cos\beta\cdot\frac{\partial u}{\partial x} + \sin\beta\cdot\frac{\partial u}{\partial y}$$

is the *normal derivative* of u in the direction of the normal that points to the right of the tangent.

Thus we have

$$\begin{aligned} ^*du &= -\frac{\partial u}{\partial y}\,dx + \frac{\partial u}{\partial x}\,dy\\ &= -\frac{\partial u}{\partial y}|dz|\cos\alpha + \frac{\partial u}{\partial x}|dz|\sin\alpha\\ &= \left(\sin\beta\frac{\partial u}{\partial y} + \cos\beta\frac{\partial u}{\partial x}\right)\,|dz|\\ &= \left(\frac{\partial u}{\partial n}\right)\,|dz|\,. \end{aligned}$$

So now we may rewrite (4.3.3) as

$$\int_\gamma \left(\frac{\partial u}{\partial n}\right) |dz| = 0$$

for all cycles γ that are homologous to 0 in Ω. Of course this last formula is just a restatement of Lemma 8.1.6, which we shall encounter below.

And now let us repeat what we said in the last subsection using our new language. If Ω is a simply connected region, then of course *all* cycles are homologous to zero and hence $\oint_\gamma {}^*du = 0$ for all cycles γ. But then, as we have noted, u will have a well-defined, single-valued conjugate harmonic function on Ω. If instead Ω is multiply connected, then the conjugate function has periods

$$\oint_{\gamma_j} {}^*du = \int_{\gamma_j} \frac{\partial u}{\partial n} |dz|$$

corresponding to the cycles $\gamma_1, \ldots, \gamma_{k-1}$ in a homology basis.

4.4 Uniformization of Multiply Connected Domains

Now we shall use the topological–analytical ideas developed thus far to prove Theorem 4.2.3. It should be noted that there are a number of variants of this result. Instead of an annulus with arcs removed, the target domain can instead be the plane with vertical slits removed. Or it can be the disk with smaller subdisks removed. In all cases, the classical result applies to finitely connected domains. See [AHL2] or [GOL] for detailed discussion of these matters. The reference [HES] treats uniformization on a disk with smaller subdisks removed in the case that ${}^c\Omega$ has *countably many components*.

Now let Ω be a finitely connected domain in $\mathbb{C}$, and assume that the connectivity is $k \geq 2$. Let the connected components of the complement of Ω (in the Riemann sphere) be called $A_1, A_2, \ldots, A_k$. Let us suppose, as usual, that A_k is the unbounded component of the complement (if we think of the complement as taken in $\widehat{\mathbb{C}}$ instead of $\mathbb{C}$, then A_k is the component that contains ∞). We shall assume that none of these components of the complement is a single point. Certainly a single point will be removable for a conformal mapping (by the Riemann removable singularities theorem).

Now we shall make an important geometrical simplification. Notice that $\widetilde{\Omega} \equiv \mathbb{C} \setminus A_k$ is a simply connected region. So of course $\widetilde{\Omega}$ can be conformally mapped to the disk by a mapping φ. Under the mapping φ, Ω itself is mapped to a new k-connected region, and $A_1, \ldots, A_{k-1}$ are mapped to the bounded, connected components of its complement. We shall continue to denote the connected components of the complement of the image of φ by $A_1, \ldots, A_k$. Then A_k must now be $\{\zeta : |\zeta| \geq 1\}$. The unit circle $\{\zeta : |\zeta| = 1\}$, equipped with positive orientation, is now the outer boundary contour of the new (mapped) region, and we call it C_k.

Now let us look at ${}^cA_1$ with respect to the extended plane $\widehat{\mathbb{C}}$. This is a simply connected region, and we map it to the complement in $\widehat{\mathbb{C}}$ of the unit disk, with ∞ mapping to ∞. Under this new mapping, of course C_k will get mapped to a smooth (indeed, real analytic), closed curve, which we still denote by C_k. So now our domain has *two* smooth (real analytic) boundary curves: C_1 and C_k. Now the inner contour C_1 is the unit circle with negative orientation.

Of course we may repeat this process $k-2$ more times to obtain a region bounded by an outer contour C_k and $k-1$ inner contours $C_1, \ldots, C_{k-1}$. It is important to observe that all these contours are smooth (indeed, real analytic) and that the original domain Ω is conformally equivalent to this new region. In our construction, each boundary curve is momentarily a circle. But, in the end, we can say only that each boundary curve (except possibly the last one) is a real analytic curve. We give the components of $\partial\Omega$ the usual positive orientation. This means that the "outer boundary curve" (the curve that bounds the unbounded component of the complement) is oriented *counter*clockwise; and the other boundary curves (that bound the bounded components of the complement) are oriented *clockwise*. See Figure 4.3.

Fig. 4.3. Domain bounded by real analytic curves.

It is important to note that the index of any point in the plane with respect to the various contours present is obvious. Indeed, the index of C_j, $j < k$, with respect to any point in the interior of C_j is -1. The index with respect to any point exterior to C_j is 0. Of course the index of C_k with respect to any point that is in A_k (the unbounded component of the complement of Ω) is 0, and the index with respect to all other points not on C_k is 1. We see that $C_1 + C_2 + \cdots + C_k$ bounds Ω.

We now have the advantage of working on a finitely connected domain whose boundary curves are all real analytic. So certainly every boundary point has a barrier (this term is discussed in detail in Section 8.1, or see [AHL2] or [GRK1]). The Dirichlet problem is therefore solvable (see [AHL] or [GRK1]). Now, for each $1 \leq j \leq k$, we solve the Dirichlet problem with boundary trace equal to 1 on C_j and equal to 0 on the other boundary curves. Call the solution $\omega_j(z)$. Then $\omega_j(z)$ is nothing other than harmonic measure for the curve C_j (see Chapter 9 below). Certainly $0 < \omega_j(z) < 1$ on Ω and

$$\omega_1(z) + \omega_2(z) + \cdots + \omega_k(z) \equiv 1$$

on Ω. Notice that ω_j can be analytically continued across each boundary curve (by Schwarz reflection—just map each boundary curve to the circle). So we may think of ω_j, each j, as harmonic on $\overline{\Omega}$.

Of course the contours $C_1, \ldots, C_k$ form a homology basis for the cycles in Ω. For each j, the conjugate harmonic differential of ω_j has period along C_m given by

$$\alpha_{jm} = \int_{C_m} \frac{\partial \omega_j}{\partial n}\, ds = \int_{C_m} {}^*\, d\omega_j\, .$$

We claim that no linear combination

$$\lambda_1\omega_1(z) + \lambda_2\omega_2(z) + \cdots + \lambda_{k-1}\omega_{k-1}(z) \tag{4.4.1}$$

with constant coefficients λ_j can have a single-valued harmonic conjugate function *unless all the* λ_j *are zero.*

To verify the claim, suppose instead that (4.4.1) were the real part of a holomorphic function f on Ω. By the usual reflection argument, we may suppose that f continues analytically to the closure $\overline{\Omega}$ of Ω. Then $\mathrm{Re}\, f$ would take the value λ_j on the contour C_j, $j = 1, \ldots, k-1$, and it would also take the value 0 on C_k. So each of these contours would be mapped under f to a vertical line segment. If τ lies in the complement of all these line segments, then the harmonic function $\arg(f(z) - \tau)$ has a single-valued branch on each contour $C_1, \ldots, C_{k-1}$. The argument principle then tells us that f certainly does not take the value τ in Ω. But f is a holomorphic function, hence has open image. This is a contradiction. The only possible conclusion is that f is a constant mapping. As a result, the real part of f is identically 0. So the λ_j must all vanish.

Now we give an interpretation of our result in the language of linear algebra. The homogeneous system of linear equations

$$\begin{aligned}
\lambda_1\alpha_{11} + \lambda_2\alpha_{21} + \cdots + \lambda_{k-1}\alpha_{k-1,1} &= 0\,,\\
\lambda_1\alpha_{12} + \lambda_2\alpha_{22} + \cdots + \lambda_{k-1}\alpha_{k-1,2} &= 0\,,\\
&\cdots\\
\lambda_1\alpha_{1,k-1} + \lambda_2\alpha_{2,k-1} + \cdots + \lambda_{k-1}\alpha_{k-1,k-1} &= 0\,,
\end{aligned}$$

has only the trivial solution $\lambda_1 = \lambda_2 = \cdots = \lambda_{k-1} = 0$, for these are precisely the conditions under which $\lambda_1\omega_1 + \lambda_2\omega_2 + \cdots + \lambda_{k-1}\omega_{k-1}$ has a single-valued harmonic conjugate. But now this result tells us that any *inhomogeneous system* of linear equations with these same coefficients must in fact have a unique solution. In particular, it is certainly possible to solve the system

$$\begin{aligned}
\lambda_1\alpha_{11} + \lambda_2\alpha_{21} + \cdots + \lambda_{k-1}\alpha_{k-1,1} &= 2\pi\,,\\
\lambda_1\alpha_{12} + \lambda_2\alpha_{22} + \cdots + \lambda_{k-1}\alpha_{k-1,2} &= 0\,,\\
\lambda_1\alpha_{13} + \lambda_2\alpha_{23} + \cdots + \lambda_{k-1}\alpha_{k-1,3} &= 0\,,\\
&\cdots\\
\lambda_1\alpha_{1,k-1} + \lambda_2\alpha_{2,k-1} + \cdots + \lambda_{k-1}\alpha_{k-1,k-1} &= 0\,.
\end{aligned}$$

We append to this system the additional (redundant) equation

$$\lambda_1\alpha_{1k} + \lambda_2\alpha_{2k} + \cdots + \lambda_{k-1}\alpha_{k-1,k} = -2\pi\,,$$

which is obtained from adding the last $k-1$ equations and using the fact that $\alpha_{j1} + \alpha_{j2} + \cdots + \alpha_{jk} = 0$ for each j.

In the language of function theory, the solution of this last system gives us a multiple-valued holomorphic function f with periods $\pm 2\pi$ along C_1 and C_k and all other periods equal to 0. Also, $\operatorname{Re} f$ is constantly equal to λ_j on C_j, $j = 1, \ldots, k$ (with $\lambda_k = 0$). It follows that the function $F(z) = e^{f(z)}$ is single-valued and holomorphic. We now claim the following result:

Theorem 4.4.1. *The function F is a one-to-one, onto, holomorphic mapping from Ω to the region $\mathcal{B}$ given by the annulus $\mathcal{A} \equiv \{\zeta \in \mathbb{C} : 1 < |\zeta| < e^{\lambda_1}\}$ with $k-2$ concentric arcs in the circles $\{\zeta \in \mathbb{C} : |\zeta| = e^{\lambda_j}\}$, $j = 2, \ldots, k-1$, removed.*

The remainder of this section is devoted to the proof of the theorem. We know that F is a well-defined holomorphic function on Ω and that its image lies in a region as described in the statement of the theorem. What we must establish are the univalence and surjectivity. Note that the concept of the mapping we are discussing is illustrated in Figure 4.4. Observe again that the contours C_1, C_k correspond to the outer and inner circles of the annulus. The contours $C_2, \ldots, C_{k-1}$ correspond to the arcs that are removed from the annulus. It is useful to think of each arc as a closed cut whose two arc-shaped sides have been forced to coincide.

If $\tau \in \mathbb{C}$, τ not on any of the boundary curves of $\mathcal{B}$, then the number of roots of the equation $F(z) = \tau$ is of course given by

$$\frac{1}{2\pi i}\oint_{C_1} \frac{F'(z)\,dz}{F(z)-\tau} + \frac{1}{2\pi i}\oint_{C_2} \frac{F'(z)\,dz}{F(z)-\tau} + \cdots + \frac{1}{2\pi i}\oint_{C_k} \frac{F'(z)\,dz}{F(z)-\tau} \equiv \eta(\tau)\,. \tag{4.4.2}$$

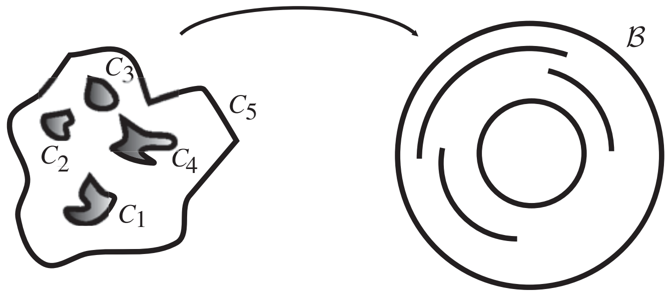

Fig. 4.4. Representation of Ω on an annulus minus arcs.

For $\tau = 0$ the terms in (4.4.2) are known to be $1, 0, 0, \ldots, 0, -1$. Hence the sum in (4.4.2) is 0. The integral over C_1 (which corresponds to the outer circle of the target annulus) is constantly equal to 1 for $|\tau| < e^{\lambda_1}$; it vanishes for $|\tau| > e^{\lambda_1}$. Likewise the integral over C_k (which corresponds to the inner circle of the target annulus) is constantly equal to -1 for $|\tau| < 1$ and is 0 for $|\tau| > 1$. Of course the integrals over C_j, $2 \leq j \leq k-1$, vanish for τ not on the circular arcs.

The aggregate of all this reasoning shows that if $\tau \in \mathcal{B}$, then the expression in (4.4.2) is the sum of $0 + (0 + \cdots + 0) + 1 = 1$, proving that the value τ is taken once and only once by the function F. That is what we wished to show.

We remark that it is possible to calculate the valence of F on the boundary curves. This uses a version of the Plemelj jump formula, and we cannot treat it here. See [AHL2, p.256] for details. Goluzin [GOL] also has a nice treatment of Plemelj.

4.5 The Ahlfors Map

There are many approaches to uniformization of multiply connected domains in the plane. In the present section we will present Ahlfors's ideas that lead to the Ahlfors map. For a detailed reference on this matter, see [FIS].

We fix a bounded domain $\Omega \subseteq \mathbb{C}$ that is finitely connected, that is, its complement has finitely many components. We begin our analysis just as with the proof of the Riemann mapping theorem (RMT). That is, fix a point $p \in \Omega$. Set

$$\gamma = \sup\{|f'(p)| : f \in H^\infty(\Omega), \|f\|_\infty \leq 1\}\,.$$

Any function $f \in H^\infty(\Omega)$ such that $\|f\| = 1$ and $f'(p) = \gamma$ we shall call *extremal*. Just as in the proof of the RMT, we can prove that an extremal exists with a simple normal families argument. The details are omitted.

Proposition 4.5.1. *There is just one extremal function f, and $f(p) = 0$.*

Observe that if we defined extremal by the condition $|f'(p)| = \gamma$ then the extremal would no longer be unique (just postcompose with a rotation). It is the positivity of the derivative at p that gives us uniqueness.

Proof of the Proposition. Let f be an extremal function. Define, for $z \in \Omega$,

$$g(z) = \frac{f(z) - f(p)}{1 - \overline{f(p)}f(z)} .$$

Then

$$g'(z) = \frac{f'(z)(1 - |f(p)|^2)}{[1 - \overline{f(p)}f(z)]^2}$$

and

$$g'(p) = \frac{\gamma}{1 - |f(p)|^2} .$$

Since $\|g\|_\infty \le 1$, we see that the function g is a candidate for the extremal. It must be that $f(p) = 0$ otherwise $g'(p) > \gamma$.

For the uniqueness, suppose that both f_1 and f_2 are extremals. Then certainly $[f_1 + f_2]/2$ is also an extremal. If we can show that any extremal must be an extreme point of the unit ball of $H^\infty(\Omega)$, then it must follow that $f_1 = f_2$ and the proof will be complete.

Thus let $g \in H^\infty$ and suppose that $\|f \pm g\|_\infty \le 1$. It follows that

$$|f|^2 + 2\,\mathrm{Re}\, f\overline{g} + |g|^2 \le 1$$

and

$$|f|^2 - 2\mathrm{Re}\, f\overline{g} + |g|^2 \le 1 .$$

Adding these inequalities together yields

$$|f|^2 + |g|^2 \le 1 .$$

As a result,

$$|g|^2 \le 1 - |f|^2 = (1 + |f|)(1 - |f|) \le 2(1 - |f|) .$$

Now define

$$h = \frac{g^2}{2} .$$

We see that $h \in H^\infty$ and

$$|f| + |h| \le |f| + (1 - |f|) = 1 .$$

We shall prove that this last estimate implies that h vanishes identically. Thus g vanishes identically, and so f is an extreme point of the unit ball of H^∞.

First we claim that $h(p) = 0$. For if not, then we examine the function $f + \lambda f h$ for a constant λ of modulus 1. Notice that

$$|f + \lambda f h| \le |f| + |f||h| \le |f| + |h| \le 1 \,.$$

But then

$$[f + \lambda f h]'(p) = f'(p) + \lambda f'(p)h(p) + \lambda f(p)h'(p) = \gamma + \lambda\gamma h(p) \,.$$

If $h(p) \neq 0$ then an appropriate choice of λ makes this last quantity positive and greater than γ, contradicting the maximal value of γ. That establishes the claim.

If h is not identically 0, then h has some finite order of vanishing r at p, $r > 1$. Let $\epsilon > 0$ and set

$$m(z) = f(z) + \epsilon h(z)(z-p)^{-r+1} \,.$$

Then $|m'(p)| = |\gamma + \epsilon h^{(r)}(p)/r!|$ (where the exponent (r) denotes the r^{th} derivative) and this quantity can be made to exceed γ if we simply choose the argument of ϵ appropriately. Further observe that, if ϵ is small and $|z-p| > \epsilon^{1/[r-1]}$, then $|\epsilon(z-p)^{-r+1}|$ is less than 1. But then

$$|m(z)| \le |f(z)| + |h(z)| \le 1$$

for these values of z. By the maximum principle, we see that $|m| \le 1$, and we have once again contradicted the definition of extremal. In conclusion, $h \equiv 0$, and the proof is complete. □

Definition 4.5.2. The unique extremal function provided by the last proposition will be called the *Ahlfors map* or *Ahlfors function* of the domain Ω.

The Ahlfors function will play the role, for a multiply connected domain, of the Riemann mapping function for a simply connected domain. In honor of Lars Ahlfors, we shall denote this function by $\alpha(z)$. We shall spend the remainder of this section developing its properties.

Definition 4.5.3. A subset $S \subseteq \Omega$ is called *dominating* if

$$\sup_{s\in S} |f(z)| = \sup_{z\in\Omega} |f(z)|$$

for all functions $f \in H^\infty(\Omega)$.

Lemma 4.5.4. *The domain Ω has a countable dominating set that has no limit point in Ω.*

Proof. Let K_j be a sequence of compact subsets of Ω such that

$$K_j \subseteq \overset{\circ}{K}_{j+1}$$

for each j and $\cup_j K_j = \Omega$. Define

$$E_j = K_j \setminus \overset{\circ}{K}_{j-1}\,,\ j = 2, 3, \ldots .$$

Certainly each E_j is compact. Also, for each $f \in H^\infty(\Omega)$,

$$\max_{E_j} |f(z)| = \max_{K_j} |f(z)| \to \|f\|_\infty$$

as $j \to \infty$.

On each E_j, the unit ball $\mathcal{B}$ in $H^\infty(\Omega)$ forms an equicontinuous family. To see this, notice that there is a number $r > 0$ such that $D(e, r) \subseteq \Omega$ for each $e \in E_j$. Now apply the Cauchy estimates to any $f \in \mathcal{B}$ on $D(e, r)$. Now let us select $\delta_j > 0$ such that

$$|f(z) - f(w)| < \left(\frac{1}{2}\right)^j \quad \text{if } |w - z| < \delta_j\,,\ w, z \in E_j\,,\ f \in \mathcal{B}\,.$$

Let T_j be a finite set in E_j such that each point of E_j is distance not greater than $\delta_j/2$ from some element of T_j. [We call T_j a $\delta_j/2$-net in E_j.] It follows that

$$\max_{E_j} |f(z)| \le \max_{T_j} |f(z)| + \left(\frac{1}{2}\right)^j .$$

Now set $S = \cup_j T_j$. This S is the dominating set that we seek. □

Proposition 4.5.5. *Let α be the Ahlfors map for Ω with respect to the point $p \in \Omega$. Then, for each $h \in H^\infty(\Omega)$, we have*

$$\|\alpha h\|_\infty = \|h\|_\infty\,.$$

Proof. Set

$$\Omega' = \{z \in \Omega : \exists h \in H^\infty(\Omega) \text{ such that } |h(z)| > 1 \text{ and } \|\alpha h\| \le 1\}\,.$$

Seeking a contradiction, we suppose that $p \in \Omega'$. Set $\alpha_1 = h{\cdot}\alpha$. Thus $\|\alpha_1\| \le 1$. But

$$|\alpha_1'(p)| = |\alpha'(p)| \cdot |h(p)| > |\alpha'(p)| > 1\,,$$

and that is impossible. So $p \notin \Omega'$.

We shall show that Ω' is open and closed in Ω. Since the last paragraph shows that Ω' is not all of Ω, the only possible conclusion is that $\Omega' = \emptyset$. This will prove the proposition.

It is clear from its very definition that Ω' is open. We concentrate our efforts on showing that the set is closed; in fact we shall show that $\Omega \setminus \Omega'$ is (relatively) open. Fix a point $q \in \Omega \setminus \Omega'$. Then $\|\alpha h\| \le 1$ implies that $|h(q)| \le 1$ for all $h \in H^\infty$. Let $\{s_j\}$ be a countable dominating sequence in Ω with no limit point in Ω (as provided by Lemma 4.5.4). We may of course assume that no s_j is equal to q. Let $\mathbf{M}$ be the maximal ideal space of ℓ^∞ (see

Section 11.1 for a discussion of Banach algebras and maximal ideal spaces). Any function w that is bounded on a neighborhood of $\{s_j\}$ gives rise to a continuous function $\widehat{w}$ on $\mathbf{M}$. More precisely, define

$$\omega_j = w(s_j)$$

for $j = 1, 2, \ldots$. Then $\mathcal{W} = \{\omega_j\} \in \ell^\infty$ and we have the induced function

$$\widehat{w} : \mathbf{M} \to \mathbb{C}$$

given by

$$\widehat{w}(\mathbf{m}) = \mathcal{W}(\mathbf{m}) \qquad \text{for } \mathbf{m} \in \mathbf{M} .$$

In particular, if $w \in H^\infty(\Omega)$, then w induces a continuous function on $\mathbf{M}$.

Now our hypothesis that $q \notin \Omega'$ implies that the linear functional from αH^∞ into $\mathbb{C}$ given by

$$\alpha h \mapsto h(q)$$

has norm 1. The Hahn–Banach theorem now yields that there is a measure μ on $\mathbf{M}$, having norm (i.e., total mass) 1, such that

$$\int \widehat{\alpha}\,\widehat{h}\, d\mu = h(q) \quad \text{for all } h \in H^\infty(\Omega) .$$

Therefore

$$1 = \int \widehat{\alpha}\,\widehat{1}\, d\mu \leq \int |\widehat{\alpha}||d\mu| \leq \|\mu\| = 1 .$$

We conclude that the measure $d\tau = \widehat{\alpha}\, d\mu$ is nonnegative and has mass 1; moreover, τ is supported on the set where $|\widehat{\alpha}| = 1$.

For $\zeta \in \Omega$ a point near q, let

$$s_\zeta(z) = \frac{z - q}{z - \zeta} , \quad z \in \Omega .$$

Set

$$w(\zeta) = \int_\Omega \widehat{s}_\zeta(z)\, d\tau(z) .$$

Clearly the function w is continuous at points near q. Also, since $w(q) = 1$ we may conclude that $w(\zeta) \neq 0$ if ζ is near to q. Let ζ be such a point.

If $h \in H^\infty(\Omega)$, then we define

$$g(z) = \frac{h(z) - h(\zeta)}{z - \zeta} \cdot (z - q) .$$

Then

$$0 = \int \widehat{g}\, d\tau$$

because $g(q) = 0$. Therefore

$$h(\zeta)u(\zeta) = \int \widehat{h}\widehat{s}_\zeta \, d\tau \,,$$

and this implies that

$$|h(\zeta)| \leq C \cdot \sup\{|\widehat{h}(\mathbf{m})| : \mathbf{m} \in \mathbf{A}\} \,,$$

where $\mathbf{A}$ is the support of τ and C is a constant that does not depend on h.

Now we replace h by h^j, take the j^{th} root of both sides, and let $j \to \infty$. We conclude that

$$|h(\zeta)| \leq \|\widehat{h}\|_{\mathbf{A}} = \sup\{|\widehat{h}(\mathbf{m})| : \mathbf{m} \in \mathbf{A}\} \,.$$

The upshot is that if $\|hF\| \leq 1$, then $|\widehat{h}| \leq 1$ on $\mathbf{A}$ and hence $|h(\zeta)| \leq 1$. But then $\zeta \notin \Omega'$. So q has a neighborhood that does not lie in Ω'. Hence Ω' is closed. This completes the proof. □

Definition 4.5.6. Let Ω be a domain as usual. A point $x \in \partial\Omega$ is said to be *essential* if there is a bounded, holomorphic function f on Ω and an $\epsilon > 0$ such that f does not extend analytically to be holomorphic on $D(x, \epsilon)$.

If a point $x \in \partial\Omega$ is not essential, then it is said to be *removable*. If all boundary points of a given domain Ω are essential, then the domain is said to be *maximal*.

In the rest of our exposition about the Ahlfors mapping, we shall assume that $\partial\Omega$ consists of finitely many smooth, closed curves. This extra hypothesis will allow us to focus on the main analytic points and not get bogged down in technical details. Note in particular that every boundary point of such a domain is essential. For if $x \in \partial\Omega$, then there is a conformal mapping φ of Ω to the disk (not, in general, onto the disk) such that the boundary curve containing x goes univalently under φ to the circle and x goes to 1 (see Section 4.4, where we discuss in detail how to construct such a mapping). Since $\lambda(z) = e^{-(1+z)/(1-z)}$ is a bounded, holomorphic function on D that does not analytically continue past 1, we see that $\lambda \circ \varphi$ is a bounded, holomorphic function on Ω that does not analytically continue past x. Likewise, we may pull the peak function $p(z) = (1 + z)/2$ back from the disk to get a peaking function on Ω at x.

Proposition 4.5.7. *Let α be the Ahlfors map for the domain Ω with respect to the point $p \in \Omega$. For each point $x \in \partial\Omega$ we have*

$$\limsup_{\Omega \ni z \to x} |\alpha(z)| = 1 \,. \tag{4.5.1}$$

Proof. As already noted, $x \in \partial\Omega$ is essential. Assume that (4.5.1) fails. In particular, say that

$$\limsup_{\Omega \ni z \to x} |\alpha(z)| = 1 - \delta$$

for some $\delta > 0$. As noted in the discussion immediately preceding the proposition, there exists $h \in H^\infty(\Omega)$ such that

$$\limsup_{\Omega \ni z \to x} |h(z)| = 1$$

while

$$\limsup_{\Omega \ni z \to y} |h(z)| < 1$$

for some $y \in \partial\Omega$, $y \neq x$.

Now consider αh. At the point x we have

$$\limsup_{\Omega \ni z \to x} |\alpha(z) h(z)| \leq 1 - \delta\,.$$

But at $\partial\Omega \ni y \neq x$ we have

$$\limsup_{\Omega \ni z \to y} |\alpha(z) h(z)| < 1$$

since h has boundary limit less than 1 while α has boundary limit not exceeding 1. In conclusion, $\|\alpha h\|_\infty < 1 = \|h\|_\infty$. This contradicts Proposition 4.5.5. In conclusion, for any $x \in \partial\Omega$, (4.5.1) holds.

Our main result about the Ahlfors map is Theorem 4.5.9 below; it gives a representation of a multiply connected domain onto the unit disk. Prior to that result, we have one more general fact about the Ahlfors map.

In what follows, a set E in the Riemann sphere $\widehat{\mathbb{C}}$ is called a (Painlevé) null set if the only holomorphic functions on $\widehat{\mathbb{C}} \setminus E$ are constants. As an example, a singleton is a null set.

Lemma 4.5.8. *Let α be the Ahlfors map for Ω with respect to the point $p \in \Omega$. Let $X \subseteq D$ be those points of the unit disk that are not in the range of α. Then $X_r \equiv X \cap \overline{D}(0,r)$ is a Painlevé null set for $r < 1$.*

The upshot of this last result is that the Ahlfors map is onto. The details of this assertion will follow below. The technical statement of the lemma is just for convenience in proving the result.

Proof of the lemma. If the conclusion fails for some $r < 1$, then let $D' = D \setminus X_r$ and let β be the Ahlfors map for D' with respect to the point 0. Then $\beta'(0) > 1$, otherwise $\beta'(0) = 1$ and therefore uniqueness of the Ahlfors map implies that $\beta(z) \equiv z$. This would contradict Proposition 4.5.7.

Now we see that

$$(\beta \circ \alpha)'(p) = \beta'(\alpha(p)) \cdot \alpha'(p) = \beta'(0) \cdot \alpha'(p) > \alpha'(p)\,.$$

Certainly this estimate contradicts the extremality of the Ahlfors map α. The result is proved. □

For us, the main result about the Ahlfors map is contained in the next theorem.

Theorem 4.5.9. *Let Ω be bounded by $m+1$ disjoint, real analytic Jordan curves $\gamma_0, \gamma_1, \ldots, \gamma_m$ and let α be the Ahlfors function for Ω relative to a point $p \in \Omega$. Then*

(a) *The mapping α maps Ω onto D precisely $m+1$ times;*
(b) *The mapping α continues analytically across each boundary curve γ_j and maps each γ_j diffeomorphically onto the circle;*
(c) *The derivative α' does not vanish on any boundary curve γ_j.*

Proof. Let η denote $\partial\Omega$. We let $A(\Omega)$ denote the continuous functions on $\overline{\Omega}$ that are holomorphic on Ω. Define

$$\sigma = \sup\{|f'(p)| : f \in A(\Omega), \|f\| \leq 1\}.$$

By the Hahn–Banach theorem and the Riesz representation theorem, there is a finite Borel measure μ on $\partial\Omega$ of total variation σ such that

$$\int_\eta f(z)\, d\mu(z) = f'(p) \quad \text{for all } f \in A(\Omega).$$

Let $G(z,p)$ be the Green's function on Ω having pole at p. Let $\widetilde{G}(z,p)$ be the harmonic conjugate of the Green's function, and set $q(z) = G(z) + i\widetilde{G}(z)$. We know from Propositions I and II in the appendix at the end of the section that

$$iq'(z)\, dz = d\omega_p(z) \quad \text{for } z \in \eta.$$

Here ω_p is the harmonic measure (what we usually call $\omega(p, \Omega, \partial\Omega)$) for $\partial\Omega$ with respect to p—see Chapter 9. In addition, we know that

$$q'(z) = -\frac{1}{z-p} + w(z),$$

where w is a function holomorphic in a neighborhood of $\overline{\Omega}$. As a result,

$$\frac{i}{2\pi} \oint_\eta f(z) q'(z)\, dz = f(p) \quad \text{for all } f \in A(\Omega).$$

We conclude that

$$0 = \oint_\eta f(z)\,(z-p) \left[\frac{i}{2\pi} \frac{q'(z)}{z-p}\, dz - d\mu(z)\right] \quad \text{for all } f \in A(\Omega).$$

Now the F. and M. Riesz theorem essentially implies that the expression in brackets, thought of as a measure, is absolutely continuous with respect to dz (see [FIS, p. 85] for a more detailed treatment). We treat this theorem of the Riesz brothers in detail in Section 9.6. Thus, by subtraction, the same is true for $d\mu$. Because $q' \neq 0$ on η, we may write

$$d\mu(z) = r(z) \frac{i}{2\pi} q'(z)\, dz \quad \text{for } z \in \partial\Omega.$$

Here $f \in L^1(\eta, ds)$.

We conclude that

$$0 = \oint_\eta f(z)\,(z-p)\left[\frac{1}{z-p} - r(z)\right]\frac{i}{2\pi}q'(z)\,dz \quad \text{for all } f \in A(\Omega)\,.$$

Now a version of the F. and M. Riesz theorem (see [FIS, p. 91]) tells us that

$$\big[1-(z-p)r(z)\big]q'(z) = h(z)$$

for $z \in \overline{\Omega}$ and some $h \in H^1(\Omega)$. An equivalent formulation of this last identity is

$$r(z) = \frac{1}{z-p} + \sum_{j=1}^{m}\frac{c_j}{z-t_j} \; + \; g(z)$$

for $z \in \overline{\Omega}$. Here $g \in H^1(\Omega)$, the $c_1, \ldots, c_m$ are complex scalars, and $t_1, \ldots, t_m$ are the critical points of q, that is, the zeros of q'. Note that $t_1, \ldots, t_m$ come from the Green's function, and this is where the connectivity of Ω plays a role.

If α is the Ahlfors mapping for Ω relative to p, then

$$\begin{aligned}\eta \equiv \alpha'(p) &= \oint_\eta \alpha(z)r(z)\frac{i}{2\pi}q'(z)\,dz \\ &\le \oint_\eta |\alpha(z)||r(z)|\left|\frac{i}{2\pi}q'(z)\right|\,d|z| \le \|\alpha\|_\infty\,\sigma \\ &= \sigma \le \eta\,.\end{aligned}$$

As a result,

$$\alpha(z)r(z) \ge 0 \quad \text{a.e. } |dz| \text{ on } \eta \tag{4.5.2}$$

and

$$|\alpha(z)| = 1 \quad \text{a.e. on the set where } r \ne 0\,. \tag{4.5.3}$$

Now each point z on η has a neighborhood V such that $V \cap \Omega$ is conformally equivalent to the unit disk D and r is in $H^1(V \cap \Omega)$. Thus, by a result in the Appendix to this section, we may conclude that both α and r may be analytically continued over each point of η. [This argument requires the canonical factorization (Proposition 6.1.21), a concept not covered in full detail in this book. See however [FIS, p. 75] and also [HOF] and [KRA1, Chapter 8]. Also see the general discussion in Section 6.1.] Certainly $|\alpha| = 1$ at every point of η. Thus the meromorphic function $r(z)\alpha(z)$ is real and positive on η hence has as many poles as zeros in Ω. Note that the function has exactly $m+1$ poles and therefore α has at most $m+1$ zeros in Ω.

Next we show that $\arg\alpha$ is locally increasing on η. Fix a point $x \in \partial\Omega$. Let $w = \log|\alpha|$ near x. Then $\widetilde{w} = \arg\alpha$ and the Cauchy–Riemann equations yield that

$$0 \leq \frac{\partial w}{\partial \nu} = \frac{\partial \widetilde{w}}{\partial \tau} \quad \text{on } \eta\,.$$

This is what we wanted to see. Since $\arg \alpha$ must increase on each γ_j by an integer multiple of 2π, we find that α must have at least $m+1$ zeros in Ω. In conclusion, α has precisely $m+1$ zeros in Ω and $\arg \alpha$ increases by 2π on each arc γ_j. Now the argument principle tells us that α is $(m+1)$-to-one.

That completes the proof of the theorem. □

The Ahlfors map has developed into a powerful tool of modern function theory. Just as an instance, Steven Bell [BEL_1] has shown that the Bergman kernel (see Chapter 1) for a finitely connected domain can be constructed from finitely many Ahlfors maps (in analogy with the fact that the Bergman kernel for a simply connected domain can be constructed from the Riemann map).

Appendix to Section 4.5

In this appendix we collect some technical results that were used in Section 4.5.

Proposition I: Let Ω be smoothly bounded. Let ω_p be harmonic measure for $\Gamma = \partial\Omega$ with respect to p—see Chapter 9. As usual, let $G(z,p)$ be the Green's function, $\widetilde{G}(z,p)$ its harmonic conjugate, and $q(z,p) = G(z,p) + i\widetilde{G}(z,p)$. Then

$$d\omega_p(\zeta) = \frac{i}{2\pi} q'(\zeta)\, d\zeta\,.$$

Proof. This is a calculation similar to one that we did in Subsection 4.3.1. Or see [FIS, p. 23] for the full details. □

Proposition II: Let Ω be a smoothly bounded domain each of whose boundary arcs is in fact real analytic. Let $\Gamma = \partial\Omega$ consist of $m+1$ disjoint analytic simple closed curves (in other words, Ω has connectivity $m+1$). Fix a point $p \in \Omega$. Let $G(z,p)$ be the Green's function as usual, and let $\widetilde{G}(z,p)$ be the (multiple-valued) harmonic conjugate of g as usual. Set $q(z) = G(z,p) + i\widetilde{G}(z,p)$. Then

(a) q' does not vanish on Γ.
(b) q' has exactly m zeros in Ω, counting multiplicities.

Proof. Certainly q' has a single pole of order 1 at p; this is by the construction of the Green's function. Further, from Proposition I above, we know that $iq'(z)\, dz$ is a nonnegative measure on Γ. As a result, the total change in the quantity

$$\frac{1}{2\pi} \arg\, q'(z) = \frac{1}{2\pi i}[iq'(z)]$$

as the point z traverses Γ just once is precisely $m-1$ (because each boundary curve that bounds a bounded component of the complement contributes $+1$—there are m of these—and the outside curve that bounds the unbounded component of the complement contributes -1). But, by the argument principle, this number must be precisely the number of zeros of q' minus the number of poles. That quantity is then $m-1$. So q' has precisely m zeros, and part **(b)** is proved.

For part**(a)**, note that the Cauchy–Riemann equations imply that $\widetilde{g}$ is increasing locally on Γ hence q is locally one-to-one in a neighborhood of each point of Γ. So certainly q' cannot vanish. □

4.6 The Uniformization Theorem

The philosophical grandfather of all the theorems discussed in the present chapter is the celebrated uniformization theorem of Poincaré and Köbe. In this section we shall discuss the meaning of the theorem, and then prove a special case of it. The proof hinges on the existence of the Green's function.

If X is any topological space, then it has a simply connected universal covering space $\widehat{X}$. The universal covering space is constructed by fixing a point $x_0 \in X$ and considering the space of all paths in X emanating from x_0. The covering map

$$\pi : \widehat{X} \to X$$

is a local homeomorphism. We refer the reader to [SPA] or [HUS] for details.

In case X is a domain Ω in the complex plane, or more generally a Riemann surface, then the universal covering space $\widehat{\Omega}$ will be a two-dimensional object (because π is a local homeomorphism), and $\widehat{\Omega}$ can be endowed with a complex structure by local pullback under π of the complex structure from Ω. So $\widehat{\Omega}$ is a simply connected analytic object. What is it?

The uniformization theorem answers this question in a dramatic way. Before we present the answer, let us first restate the question—stripped of all the preliminary material that led up to it.

QUESTION: What are all the simply connected Riemann surfaces?

The answer is

ANSWER: The only simply connected Riemann surfaces are **(i)** the disk D, **(ii)** the plane $\mathbb{C}$, and **(iii)** the Riemann sphere $\widehat{\mathbb{C}}$.

And in fact much more can be said. Let us return to the motivational discussion above. If the original analytic object X is a sphere, then it turns out that the universal covering space $\widehat{X}$ will be a sphere, and that is the *only* circumstance under which a sphere arises as the universal covering space.

If the original analytic object is a plane or a punctured plane or a torus or a cylinder, then the universal covering space $\widehat{X}$ is a plane, and these are the only circumstances in which the plane arises as the universal covering space.

In all other circumstances, the universal covering space is the disk D. In other words,

> The universal covering space for any planar domain except $\mathbb{C}$ or $\mathbb{C}\setminus\{0\}$ is the disk D.

This is powerful information, and those who study Riemann surfaces have turned the result into an important tool (see [FAK]). As an instance, our result (Theorem 1.4.3) that any automorphism of a planar domain that fixes three points is the identity can be derived as a corollary of the uniformization theorem (see [FKKM]). Many of our other results about automorphisms can be studied with the aid of uniformization.

Our goal in the remainder of this section is to prove the last displayed statement: that the disk is the universal covering space for virtually all planar domains. For convenience and simplicity, we shall in fact restrict attention to bounded planar domains with finitely many smooth boundary curves.

Theorem 4.6.1. *Let $\Omega \subseteq \mathbb{C}$ be a bounded domain whose boundary consists of finitely many simple, closed, C^2 curves. Then the universal covering space for $\widehat{\Omega}$ for Ω is conformally equivalent to the unit disk D.*

Proof. The important fact for us is that $\widehat{\Omega}$ has a Green's function. This is easily seen by solving a suitable Dirichlet problem. The references [GAM] and [FAK] provide all the details; see also our discussion in Chapter 8. We shall take it for granted that $\widehat{\Omega}$ has a Green's function. The remainder of the proof presented below studies $\widehat{\Omega}$ and its Green's function.

Fix a point $\xi \in \widehat{\Omega}$ and let $G(z, \xi)$ be the corresponding Green's function. Then, by a construction that we have used before (for instance in Section 1 of this chapter), there is a holomorphic function φ defined near ξ with a simple zero at ξ such that

$$|\varphi(z)| = e^{-G(z,\xi)} .$$

The function φ can be analytically continued along any path in $\widehat{\Omega}$ from ξ to any other point z—just by continuing the harmonic conjugate $\widetilde{g}$ of g and then taking an exponential. Since $\widehat{\Omega}$ is simply connected, the monodromy theorem tells us that the analytic continuation does not depend on the choice of path. By this means we define a holomorphic function φ on $\widehat{\Omega}$ such that

$$|\varphi(z)| = e^{-G(z,\xi)} \quad \text{for } z \in \widehat{\Omega} .$$

In particular, since g has boundary values 0, we see immediately that $|\varphi(z)| < 1$ for all $z \in \widehat{\Omega}$, and φ has only one zero: the simple zero at ξ.

Fix a point $\tau \in \widehat{\Omega}$ and define

$$\psi(z) = \frac{\varphi(z) - \varphi(\tau)}{1 - \overline{\varphi(\tau)}\varphi(z)} \quad \text{for } z \in \widehat{\Omega}\,.$$

Then ψ is holomorphic on $\widehat{\Omega}$, $|\psi(z)| < 1$ for $z \in \widehat{\Omega}$, and $\psi(\tau) = 0$. Let u be a subharmonic function on $\widehat{\Omega} \setminus \{\tau\}$ such that **(i)** $u = 0$ off some compact subset of Ω and **(ii)** $u(z) + \log|\psi(z)|$ is subharmonic. Then $u(z) + \log|\psi(z)|$ is subharmonic on $\widehat{\Omega}$ and $u(z) + \log|\psi(z)| < 0$ off a compact subset of $\widehat{\Omega}$. By the maximum principle for subharmonic functions, $u(z) + \log|\psi(z)| < 0$ on all of $\widehat{\Omega}$.

Taking the supremum over all such u, and recalling the construction of the Green's function, we find that

$$G(z,\tau) + \log|\psi(z)| \le 0 \quad \text{for } z \in \widehat{\Omega}\,. \tag{4.6.1}$$

Since $\psi(\xi) = -\varphi(\tau)$, we determine using the symmetry of the Green's function that

$$G(\xi,\tau) + \log|\psi(\xi)| = G(\xi,\tau) - G(\tau,\xi) = 0\,.$$

By the strict maximum principle, we see that equality holds in (4.6.1).

Thus $\log|\psi(z)| = -G(z,\tau)$ for all $z \in \widehat{\Omega}$ and so ψ has no zeros on $\widehat{\Omega} \setminus \{\tau\}$. It follows that φ assumes the value $\varphi(\tau)$ only at $z = \tau$. Since τ was quite arbitrary, we deduce that φ is one-to-one. Hence φ maps $\widehat{\Omega}$ conformally onto a domain $\varphi(\widehat{\Omega})$ in the unit disk, and we see that $\varphi(\widehat{\Omega})$ is simply connected. By the Riemann mapping theorem, $\varphi(\widehat{\Omega})$ is conformally equivalent to the unit disk, and hence so is $\widehat{\Omega}$. □

Problems for Study and Exploration

1. A finitely connected domain in the plane is conformally equivalent to **(i)** a disk with finitely many smaller disks removed or **(ii)** a disk with finitely many concentric circular arcs removed or **(iii)** a plane with finitely many vertical slits removed. Explain why these three statements are equivalent.
2. Prove that the region indicated in Figure 4.5 is simply connected. Then prove that the conformal mapping of the disk to this region will not have nice boundary behavior at the origin.
3. Does there exist a holomorphic mapping of the disk *onto* $\mathbb{C}$? [*Hint:* The holomorphic mapping $z \mapsto (z-i)^2$ takes the upper halfplane onto $\mathbb{C}$.]
4. Prove that if f is entire and one-to-one, then f must be linear. [*Hint:* Use the fact that f is one-to-one to analyze the possibilities for the singularity at ∞.]
5. Let $\Omega \subseteq \mathbb{C}$ be a bounded, simply connected domain. Let $\phi_1 : \Omega \to D$ and $\phi_2 : \Omega \to D$ be conformal maps. How are ϕ_1 and ϕ_2 related to each other? [*Hint:* Look at $\phi_2 \circ \phi_1^{-1}$; it is a conformal self-map of the disk.]

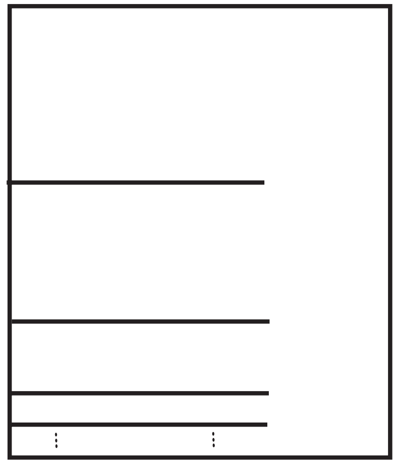

Fig. 4.5. The comb domain.

6. Let Ω be a simply connected domain in $\mathbb{C}$ and let P and Q be distinct points of Ω. Let ϕ_1 and ϕ_2 be conformal self-maps of Ω. If $\phi_1(\mathrm{P}) = \phi_2(\mathrm{P})$ and $\phi_1(Q) = \phi_2(Q)$, then prove that $\phi_1 \equiv \phi_2$.

7. Let Ω be a bounded, simply connected domain in $\mathbb{C}$. Fix a point $P \in \Omega$. Let ϕ_1 and ϕ_2 be conformal mappings of Ω to D. Prove that if $\phi_1(P) = \phi_2(P)$ and $\operatorname{sgn}(\phi_1'(P)) = \operatorname{sgn}(\phi_2'(P))$, then $\phi_1 \equiv \phi_2$. Here $\operatorname{sgn}(\alpha) \equiv \alpha/|\alpha|$ (definition) whenever α is a nonzero complex number.

8. Let Ω be a simply connected planar domain and let ϕ be a conformal mapping of Ω to D. Set $P = \phi^{-1}(0)$. Let $f : \Omega \to D$ be *any* holomorphic function such that $f(P) = 0$. Prove that $|f'(P)| \le |\phi'(P)|$.

9. Complete the following outline to obtain a different proof that the Möbius transformations ϕ_a map the unit disk conformally to the unit disk:
Check that ϕ_{-a} is the inverse of ϕ_a. Hence ϕ_a must be one-to-one.
If $|z| = 1$, then calculate that

$$\left|\frac{z-a}{1-\overline{a}z}\right| = \left|\frac{z-a}{\overline{z}-\overline{a}}\right| = 1.$$

Thus ϕ_a maps the unit circle to the unit circle. Since $\phi_a(a) = 0$, it follows from connectedness of the unit disk that ϕ_a maps the disk D into itself. But, replacing a by $-a$, we see that ϕ_{-a} maps D into D. Hence ϕ_a maps D *onto* D. That is the desired conclusion.

10. Can you explicitly calculate the Ahlfors map of the annulus $\mathcal{A} = \{z \in \mathbb{C} : 1/2 < |z| < 2\}$?

5

Boundary Regularity of Conformal Maps

Genesis and Development

It is a truism that the Riemann mapping theorem allows us to transfer the complex function theory of any simply connected domain (except the plane itself) back to the unit disk, or vice versa. But many of the more delicate questions require something more. If we wish to study behavior of functions at the boundary, or growth or regularity conditions, then we must know something about the boundary behavior of the conformal mapping.

Looking at the matter from another point of view, one of the most important practical applications of conformal mapping is to the study of the Dirichlet problem. In particular, those who design airplane wings make considerable use of this technique. And one needs to know in that context that the relevant conformal mapping extends nicely to the boundary.

It was Painlevé [PAI] who first showed that a conformal mapping of a smoothly bounded domain to the disk extends smoothly and univalently to the boundary (and so does its inverse). In fact this result, part of Painlevé's thesis, preceded by a good many years Carathéodory's better-known result about continuous extension to the boundary. Kellogg, Warschawski, and many others have refined Painlevé's original result so that today we have very precise information about the boundary regularity of conformal mappings.

The subject of boundary regularity of conformal mappings certainly has aesthetic appeal. But, as indicated above, it also offers copious practical applications. The theorems offer a model for many of the standard questions of partial differential equations. The techniques of this chapter are a combination of basic complex function theory and hard analysis.

5.1 Continuity to the Boundary

The Riemann mapping theorem (RMT) states that if $\Omega \subseteq \mathbb{C}$ is a simply connected domain, not all of $\mathbb{C}$, then there is a conformal mapping $\varphi : \Omega \to D$, where D is the unit disk. This result has had a profound influence over the way that we study complex function theory. It has no analogue in the complex analysis of several variables.

The RMT was profoundly generalized by Köbe and Poincaré with their uniformization theorem. A deeper understanding of the Riemann mapping, or of the uniformization mapping, rather naturally raises the question of whether the mapping extends to the boundary in a nice way.

In general, the answer is no. In fact the textbook example to look at is the domain Ω exhibited in Figure 5.1 (see also Exercise 2 in the last chapter).

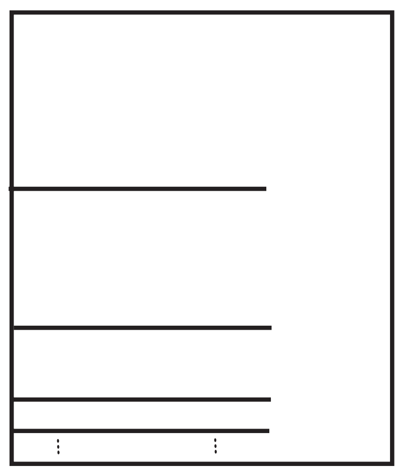

Fig. 5.1. The comb domain.

We call this the "comb domain." The domain is obviously simply connected, and not all of $\mathbb{C}$, so the RMT guarantees the existence of a conformal map $\varphi : \Omega \to D$. This map *cannot* extend continuously to $\partial\Omega$, as **(i)** the boundary is not a Jordan curve and **(ii)** there are portions of the boundary of infinite length in arbitrarily small neighborhoods of the origin. So some condition will be required in order to obtain a positive result. In fact the textbook theorem on the subject is due to Carathéodory:

Theorem 5.1.1 (Carathéodory). *Let Ω_1, Ω_2 be bounded domains in $\mathbb{C}$, each of which is bounded by a single Jordan curve. If $\varphi : \Omega_1 \to \Omega_2$ is a conformal (one-to-one, onto, holomorphic) mapping, then φ extends continuously and one-to-one to $\partial\Omega_1$. That is, there is a continuous, one-to-one function $\widehat{\varphi} : \overline{\Omega}_1 \to \overline{\Omega}_2$ such that $\widehat{\varphi}\big|_{\Omega_1} = \varphi$.*

Remark 5.1.2. Clearly the domains Ω_1, Ω_2 in the theorem are each simply connected since each has complement with just one unbounded, connected component. It will be clear from the proof that in fact Carathéodory's theorem is valid for conformal maps of domains that are bounded by finitely many Jordan curves. There are also ad hoc arguments, presented elsewhere in this book, for establishing the last result.

Proof of Theorem 5.1.1 The proof will be broken up into several lemmas, which will fill out the rest of the section. First, let us remark that the result fails completely for maps $\varphi : \Omega_1 \to \Omega_2$ that are one-to-one and onto but not holomorphic: Even if φ is infinitely differentiable in the real variable sense, it may not extend continuously to even one boundary point. For example, the map $F : D \to D$ given by $F(x) = (1 - \sqrt{1-|z|}) \cdot (z/|z|)$ is a diffeomorphism of the disk to the disk that does not extend smoothly to the boundary (but you can verify for yourself that it *does* extend continuously to the boundary).

The extension-to-the-boundary situation changes if one or both of the domains are unbounded. Consider, for instance, the Cayley transform $\varphi : D \to \{z : \operatorname{Im} z > 0\}$ given by $\varphi(z) = i(1-z)/(1+z)$. Of course φ has no continuous extension to -1 in ∂D.

We shall in fact first prove the result in the special case that Ω_1 is the unit disk D. At the end we shall derive the general result from this special one by the application of a few simple tricks.

Now let us set up the proof. Let $\mu_1 : [0,1] \to \overline{D}$, $\mu_2 : [0,1] \to \overline{D}$ be two curves satisfying $\mu_1(1) = \mu_2(1) = 1$, $\mu_1([0,1)) \subseteq D$, $\mu_2([0,1)) \subseteq D$. Our aim is to see that $\lim_{t\to 1^-} \varphi(\mu_1(t))$ and $\lim_{t\to 1^-} \varphi(\mu_2(t))$ exist and are equal. Once this result is established, we may define $\widehat{\varphi}(1)$ to be this limit (since it is independent of the choice of curve). Then $\widehat{\varphi}$ is automatically a continuous extension of φ to $D \cup \{1\}$. The extension to the other points of ∂D follows in the same way, and the continuity of the whole extension to $D \cup \partial D$ is easy to check.

Thus let μ_1, μ_2 be as above and $V = D(0,1) \cap D(1, 1/2)$. See Figure 5.2. Now V has finite area and $\varphi(V)$ has finite area M since $\varphi(V) \subseteq \Omega_2$.

Notice that if $K \subseteq \Omega_2$ is compact, then (since φ^{-1} is continuous), $\varphi^{-1}(K) \subseteq D$ is compact. Equivalently, if $\{z_j\}$ is a sequence in D that diverges to the boundary of D (i.e., has no subsequence that converges to a point of D), then $\{\varphi(z_j)\}$ diverges to $\partial\Omega_2$. In particular, $\varphi(V)$ does not have compact closure *in* Ω_2. Figure 5.3 suggests what $\varphi(V)$ looks like. The standard mathematical terminology for this general situation is to say that φ is a *proper* map. [A map is said to be *proper* if the inverse image of every compact set is compact.]

Let us introduce polar coordinates on V centered at $1 \in \partial D$. Clearly $0 < r < 1/2$. Given an r small, there is a $\theta_0 = \theta_0(r)$ such that $-\theta_0 < \theta < \theta_0$ describes the corresponding r-constant arc in V. See Figure 5.4. Let

$$\gamma_r : [-\theta_0(r), \theta_0(r)] \to V\,,$$
$$\theta \mapsto 1 - re^{i\theta}.$$

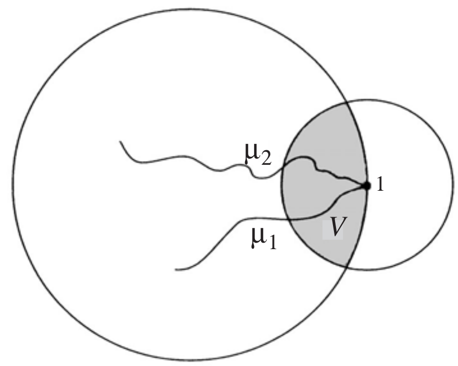

Fig. 5.2. Well-defined boundary limit.

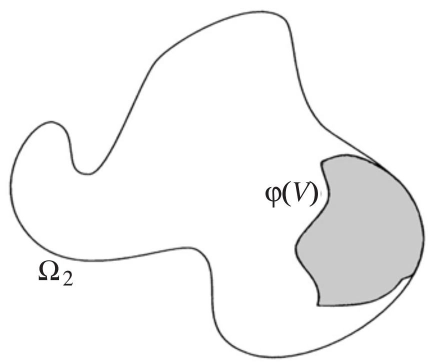

Fig. 5.3. Image of $\varphi(V)$.

Lemma 5.1.3. *Let ℓ_r be the length of $\varphi \circ \gamma_r$. Notice that*

$$\ell_r = \int_{-\theta_0(r)}^{\theta_0(r)} |\varphi'(1 - re^{i\theta})| r \, d\theta.$$

Then

$$\int_0^{1/2} \frac{(\ell_r)^2}{\pi r} \, dr$$

is finite.

Proof. Let M denote the area of $\varphi(V)$. Then

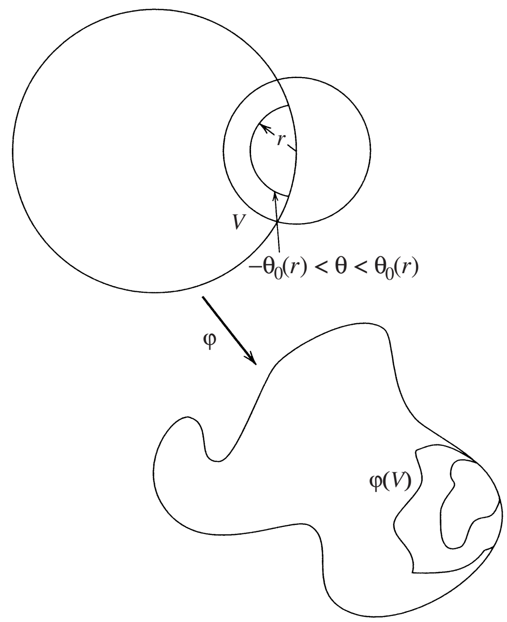

Fig. 5.4. Polar coordinates on V.

$$\begin{aligned}\int_0^{1/2} \frac{(\ell_r)^2}{\pi r}\,dr &= \int_0^{1/2} \left[\int_{-\theta_0}^{\theta_0} \left|\varphi'(1-re^{i\theta})\right| r\,d\theta\right]^2 \frac{1}{\pi r}\,dr \\ &\leq \int_0^{1/2} \left[\int_{-\theta_0}^{\theta_0} \left|\varphi'(1-re^{i\theta})\right|^2 r\,d\theta\right]\left[\int_{-\theta_0}^{\theta_0} r\,d\theta\right] \frac{1}{\pi r}\,dr \\ &\leq M\cdot 1 = M.\end{aligned}$$

As a result,

$$\int_0^{1/2} \frac{(\ell_r)^2}{\pi r}\,dr < \infty.$$

As a consequence of the lemma, because $1/r$ is not integrable at the origin, it must be that there is a sequence $r_j \to 0$ such that $\ell_{r_j} \to 0^+$.

Lemma 5.1.4. *For each such r_j, the limits*

$$\lim_{\theta\to\theta_0(r_j)^-} \varphi(1-r_je^{i\theta})$$

and

$$\lim_{\theta\to-\theta_0(r_j)^+} \varphi(1-r_je^{i\theta})$$

exist.

Proof. Each ℓ_{r_j} is finite. It follows from the definition of limit that the indicated endpoint limits exist. □

The next lemma is a result about metric topology, rather than complex analysis as such.

Lemma 5.1.5. *Let τ be the curve that describes $\partial\Omega_2$. There is a function $\eta(\delta)$, defined for all sufficiently small $\delta > 0$, with $\eta(\delta) \to 0$ as $\delta \to 0^+$, such that if $a, b \in \tau$ with $|a - b| \leq \delta$, then there is one and only one arc of τ having endpoints a, b whose diameter is $\leq \eta(\delta)$.*

Proof. It is convenient to parameterize τ with respect to the unit circle. Let $\psi(e^{it})$ be such a parametrization of τ. Note that ψ is a one-to-one, onto map of compact Hausdorff spaces. Therefore ψ has a continuous inverse. Let $\delta_0 > 0$ be so small that whenever $|\psi(\zeta) - \psi(\zeta')| \leq \delta_0$, then $|\zeta - \zeta'| < 2$.

For ζ, ζ' in the circle with $|\psi(\zeta) - \psi(\zeta')| \leq \delta_0$ we let σ be the unique shorter arc of the circle $S = \{z \in \mathbb{C} : |z| = 1\}$ having endpoints ζ, ζ'. Let $\rho = \psi \circ \sigma$. As usual, we identify ρ with its image. By the continuity of ψ^{-1}, $\operatorname{diam} \rho \to 0$ *uniformly* for $|\psi(\zeta) - \psi(\zeta')| \to 0$.

If $0 < \delta < \delta_0$, then we set

$$\eta(\delta) = \sup\{\operatorname{diam} \rho : |\psi(\zeta) - \psi(\zeta')| \leq \delta\}.$$

Then $\eta(\delta) \to 0$ as $\delta \to 0$. Let $0 < \delta_1 < \delta_0$ be so small that

$$\eta(\delta_1) < \frac{1}{2} \operatorname{diam} \tau.$$

Then the statement of the lemma holds for $\delta \leq \delta_1$.

Definition 5.1.6. Let notation be as in the proof of the last lemma. Let $a, b \in \tau$ with $|a - b|$ sufficiently small. The unique arc of τ with endpoints a, b and having diameter $\leq \eta(|a - b|)$ is called the *smaller arc* of τ that joins a to b. We denote this arc by τ_{ab}.

Lemma 5.1.7. *With $\{r_j\}$ selected as in the paragraph preceding Lemma 5.1.4, there are for each j two possibilities:*

$$\partial\Omega_2 \ni a_j \equiv \lim_{\theta \to \theta_0^-} \varphi(1 - r_j e^{i\theta}) \neq \lim_{\theta \to -\theta_0^+} \varphi(1 - r_j e^{i\theta}) \equiv b_j \in \partial\Omega_2 \qquad (5.1.1)$$

or

$$\partial\Omega_2 \ni \lim_{\theta \to \theta_0^-} \varphi(1 - r_j e^{i\theta}) = \lim_{\theta \to -\theta_0^+} \varphi(1 - r_j e^{i\theta}) \equiv p_j \in \partial\Omega_2 \qquad (5.1.2)$$

See Figure 5.5.

Proof. This is immediate from Lemma 5.1.4 and the properness of φ.

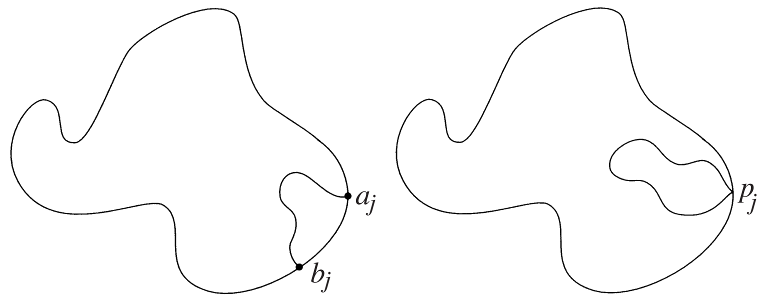

Fig. 5.5. Two possibilities for each j.

Let us write $\Gamma_{r_j} = \varphi \circ \gamma_{r_j}$. In case (5.1.1), let τ_j be the *smaller* of the two boundary arcs of Ω_2 connecting a_j to b_j. Then $\Gamma_{r_j} \cup \tau_j$ forms a simple closed curve.

In case (5.1.2), $\Gamma_{r_j} \cup \{p_j\}$ is a Jordan curve in Ω_2 (remember that φ is one-to-one).

In either case, $\Gamma_{r_j} \cup \tau_j$ or $\Gamma_{r_j} \cup \{p_j\}$ surrounds a region $W_j \subseteq \Omega_2$.

For each j let $V_j = \{1 - re^{i\theta} : 0 < r < r_j, |\theta| < \theta_0(r)\}$. Then for each j there are only two possibilities: Either $\varphi(V_j) = W_j$ or $\varphi(V_j) = \Omega_2 \setminus \overline{W_j}$.

Lemma 5.1.8. *For j sufficiently large, $\varphi(V_j) = W_j$.*

Proof. For each j, let $T_j = D \setminus \overline{V_j}$. Now fix an index j. Select a point $w_0 \in W_j$. Then $w_0 = \varphi(z_0)$ for some $z_0 \in D$. Either $z_0 \in T_j$ or $z_0 \in V_j$. If $z_0 \in V_j$, then we are done by connectivity. If instead $z_0 \in T_j$, then $\varphi(T_j) \subseteq W_j$. We will show that this is impossible for j large.

Now

$$\begin{aligned} \text{area}[\varphi(T_j)] &= \text{area}[\Omega_2] - \text{area}\,[\varphi(V_j)] \\ &= \text{area}[\Omega_2] - \int_{V_j} |\varphi'|^2 \, dx\, dy \\ &\to \text{area}\,[\Omega_2]. \end{aligned}$$

In the penultimate line we have used the Lusin area theorem (see [GRK1] and our discussion in Section 1.2) and in the last line the fact that $\cap_j V_j = \emptyset$.

Let $\ell_j = \ell_{r_j}$. Certainly $|a_j - b_j| \leq \ell_j$. Thus, by Lemma 5.1.5, τ_j has diameter $\leq \eta(\ell_j)$. This last quantity tends to 0 by Lemma 5.1.5.

Let D_j be a disk of radius $\ell_j + \eta(\ell_j)$ and having center a_j. Then the entire Jordan curve $\Gamma_{r_j} \cup \tau_j$ lies in D_j by the above estimates. Hence W_j lies inside D_j.

In conclusion,

$$\text{area}\,[W_j] \leq \pi\big(\ell_j + \eta(\ell_j)\big)^2 \to 0$$

as $j \to \infty$. But we have proved that area$\,(\varphi(T_j)) \to$ area$\,\Omega_2$. Thus it cannot be that $\varphi(T_j) \subseteq W_j$ for j large. We conclude that $\varphi(T_j) \subseteq \Omega_2 \setminus \overline{W_j}$, hence that $\varphi(V_j) \subseteq W_j$.

Lemma 5.1.9. *As $j \to +\infty$,*

$$\operatorname{diam}[W_j] \to 0 \qquad \textit{and} \qquad \operatorname{area}[W_j] \to 0.$$

Proof. Immediate from the proof of the preceding lemma. □

Lemma 5.1.10. *If the curves $\mu_\ell : [0,1] \to \overline{D}$ satisfy $\mu_\ell(1) = 1$ and $\mu_\ell([0,1)) \subseteq D, \ell = 1, 2,$ then*

$$\lim_{t \to 1^-} \varphi(\mu_1(t)) = \lim_{t \to 1^-} \varphi(\mu_2(t)).$$

(Notice that we are asserting that each limit exists and that they are equal.)

Proof. Let $\epsilon > 0$. Choose J so large that $\operatorname{diam}(W_J) < \epsilon$. If S_1 is sufficiently large, then $S_1 < t < 1$ implies that $|\mu_1(t) - 1| < r_J$; hence $\varphi(\mu_1(t)) \in W_J$. Likewise, if S_2 is sufficiently large, then $S_2 < t < 1$ implies that $|\mu_2(t) - 1| < r_J$; hence $\varphi(\mu_2(t)) \in W_J$. Therefore $|\varphi(\mu_1(t)) - \varphi(\mu_2(t))| < \epsilon$.

This proves that

$$\lim_{t \to 1^-} \varphi(\mu_1(t)) = \lim_{t \to 1^-} \varphi(\mu_2(t)) = Q,$$

where $\{Q\}$ is the singleton $\cap \overline{W_j}$.

Lemma 5.1.10 provides the continuous extension of φ to ∂D: If $P \in \partial D$, then choose a curve $\gamma : [0,1] \to \overline{D}$, $\gamma(1) = P, \gamma([0,1)) \subseteq D$. Define $\widehat{\varphi}(P) = \lim_{t \to 1^-} \varphi(\gamma(t))$. The last lemma guarantees that the limit exists and is independent of the choice of γ. The continuity is nearly obvious; think about why $\widehat{\varphi}(P_j) \to \widehat{\varphi}(P)$ when $P_j \in \partial D$ and $P_j \to P \in \partial D$.

Lemma 5.1.11. *Let $F : \overline{D} \to \mathbb{C}$ be continuous on $\overline{D}$ and holomorphic on D. Let $\tilde{\gamma} \subseteq \partial D$ be an open arc. If $F\big|_{\tilde{\gamma}}$ is constantly equal to c, then $F \equiv c$.*

Proof. We can assume that $F\big|_{\tilde{\gamma}} = 0$. If F is not identically zero, then, by composing F with a Möbius transformation, we can suppose that $F(0) \neq 0$. Then Jensen's inequality (Corollaries 6.1.16, 6.1.17) yields

$$-\infty < \log|F(0)| \leq \frac{1}{2\pi} \int_0^{2\pi} \log|F(re^{i\theta})| d\theta$$

for all but countably many $r < 1$. As $r \to 1^-$, this leads to a contradiction. □

Remark 5.1.12. The proof we have just presented of Lemma 5.1.11 is an "analysts's proof." The "geometer's proof" would be to use Schwarz reflection to reflect the function across the arc $\widetilde{\gamma}$. Then we would have a holomorphic function that is constant on an interior set with accumulation point. So it is identically constant.

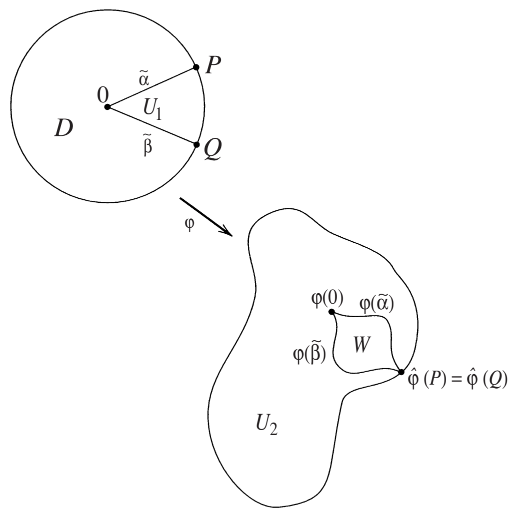

Fig. 5.6. Region bounded by a closed Jordan curve.

Lemma 5.1.13. *The continuous extension $\widehat{\varphi}$ of φ to $\overline{D}$ is one-to-one on $\overline{D}$.*

Proof. Since $\widehat{\varphi}(D) \subseteq \Omega$, $\widehat{\varphi}(\partial D) \subseteq \partial\Omega$, and $\widehat{\varphi}\big|_D$ is one-to-one, it is enough to check that $\widehat{\varphi}$ is one-to-one on ∂D. Suppose that $P, P' \in \partial D$ and $\widehat{\varphi}(P) = \widehat{\varphi}(P')$. Consider the two radial arcs running from 0 out to P and to P'. Call these two arcs $\tilde{\alpha}$ and $\tilde{\beta}$, respectively. Our hypothesis implies that $\varphi(\tilde{\alpha} \cup \tilde{\beta})$ is a closed Jordan curve. Let W be the planar region bounded by this curve. If $D \setminus (\tilde{\alpha} \cup \tilde{\beta}) = U_1 \cup U_2$, then either $\varphi(U_1) = W$ or $\varphi(U_2) = W$. See Figure 5.6.

Say without loss of generality that $\varphi(U_1) = W$. Let $\tilde{\gamma}$ be the circular arc bounding U_1. Then $\varphi(\tilde{\gamma}) \subseteq \overline{W} \cap \partial\Omega_2$. But $\overline{W} \cap \partial\Omega_2$ is the singleton $\varphi(P) = \varphi(P')$. Hence $\widehat{\varphi}$ is *constant* on the entire arc $\tilde{\gamma}$. By Lemma 5.1.11, φ is constant—a clear contradiction.

Finally, Lemma 5.1.10, the subsequent remark, and Lemma 5.1.13 give the desired continuous, one-to-one extension of a conformal map $\varphi : D \to \Omega_2$ to $\overline{D}$.

Consider now the general case. If Ω_1, Ω_2 are bounded, simply connected domains in $\mathbb{C}$, each bounded by a Jordan curve, then let $\varphi_1 : D \to \Omega_1$ be a conformal mapping and let $\varphi_2 : D \to \Omega_2$ be a conformal mapping. If $\varphi : \Omega_1 \to \Omega_2$ is *any* conformal mapping, then consider the diagram in Figure 5.7.

The function φ_1 extends to a continuous map $\widehat{\varphi}_1$ of $\overline{D}$ to $\overline{\Omega}_1$. And $\widehat{\varphi}_1\big|_{\partial D}$ is continuous, one-to-one, and onto $\partial\Omega_1$. It follows that $\left(\widehat{\varphi}_1\big|_{\partial D}\right)^{-1}$ is well defined and continuous (exercise). Thus $\widehat{\varphi}_1^{-1} : \overline{\Omega}_1 \to \overline{D}$ is one-to-one, onto, and continuous. A similar statement holds for $\widehat{\varphi}_2^{-1} : \overline{\Omega}_2 \to \overline{D}$.

Next, $\Psi = \varphi_2^{-1} \circ \varphi \circ \varphi_1$ is a conformal map of D to D; hence Ψ is a rotation composed with a Möbius transformation. It follows that Ψ extends

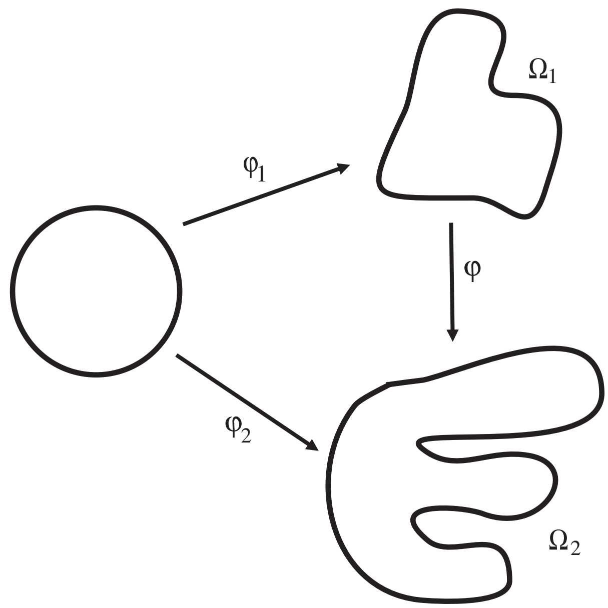

Fig. 5.7. Commutative diagram for conformal mappings.

continuously and one-to-one to a function $\widehat{\Psi} : \overline{D} \to \overline{D}$. Finally, $\widehat{\varphi}_2 \circ \Psi \circ \widehat{\varphi}_1^{-1}$ defines a continuous and one-to-one extension of φ to $\overline{\Omega}_1$.

This completes the proof of Theorem 5.1.1. □

5.2 Preliminary Facts about Boundary Smoothness

For many purposes, continuity to the boundary is insufficient. In the recent work [APF], studies were made of the Bergman kernel on a variety of domains with different boundary geometries. The behavior of the kernel on a domain Ω was studied by considering a conformal mapping $\varphi : \Omega \to D$ and the corresponding transformation formula

$$K_\Omega(z, \zeta) = K_D(\varphi(z), \varphi(\zeta))\varphi'(z)\overline{\varphi'(\zeta)}$$

(see Proposition 1.2.11). In order to pass back and forth, it was essential to know that the expression $|\varphi'(z)|$ is bounded above and below:

$$c_1 \leq |\varphi'(z)| \leq c_2 \,. \tag{5.2.1}$$

Note also that a set of inequalities like (5.2.1) tells us that the amount of stretching that occurs under the map φ is bounded above and below; this fact was also crucial in the results of [APF].

A result that implies the kind of estimates we have been discussing is the following.

Theorem 5.2.1. *Let $\Omega \subseteq \mathbb{C}$ be a bounded domain in $\mathbb{C}$ with sufficiently smooth boundary. Then any conformal mapping $\varphi : \Omega \to D$ will extend univalently and continuously differentiably to $\overline{\Omega}$, and the inverse mapping φ^{-1} will extend univalently and continuously differentiably to $\overline{D}$.*

With this theorem in hand, we know that $|\varphi'|$ extends continuously to the compact set $\overline{\Omega}$. Thus it is bounded. In other words,

$$|\varphi'(z)| \leq C \quad \text{for all } z \in \Omega\,.$$

Likewise, $|(\varphi^{-1})'|$ extends continuously to the compact set $\overline{D}$. So it is bounded. But then

$$\left|\frac{1}{\varphi'(\varphi^{-1}(z)}\right| = \left|(\varphi^{-1})'(z)\right| \leq C\,,$$

and hence[1]

$$\left|\varphi'(\varphi^{-1}(z))\right| \geq \frac{1}{C}\,.$$

In the statement of Theorem 5.2.1, we have intentionally been imprecise about what "sufficiently smooth" means, since that is one of the main points of the rest of this chapter. For the moment, we shall content ourselves with proving the next result (see [BEK]). First some definitions.

We shall require a brief discussion of the concept of boundary smoothness. We will offer two approaches to the question.

Boundary Smoothness by Way of Calculus: We will only consider domains with finitely many boundary components. Each boundary component will be a simple closed curve—see Figure 5.8.

If S^1 is the unit circle in the plane, parametrized by $t \mapsto e^{it}$, then we may think of each boundary curve γ_j as given by

$$\gamma_j : S^1 \to \mathbb{C}\,.$$

Let $k \in \{0, 1, 2, \ldots\}$. We say that $\partial\Omega$ is C^k if γ_j' is never zero, each j, and the first k partial derivatives of the function γ_j exist, each j, and each derivative is continuous.

---

[1] Implicit in this discussion is an important point that ought to be explicitly enunciated. Namely, if Ω_1 and Ω_2 are domains with C^2 boundary and $\varphi : \Omega_1 \to \Omega_2$ is a conformal map, then there are constants $c, C > 0$ such that $c \cdot \text{dist}(\varphi(z), \partial\Omega_2) \leq \text{dist}(z, \partial\Omega_1) \leq C \cdot \text{dist}(\varphi(z), \partial\Omega_2)$. In other words, φ preserves—qualitatively speaking—the distance to the boundary. This assertion follows from the Hopf lemma, a result that we treat in Lemma 5.3.1. See also Lemma 5.3.3.

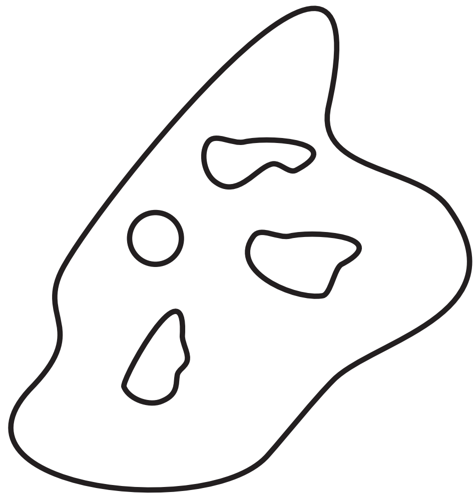

Fig. 5.8. A domain with finitely many boundary components.

Boundary Smoothness by way of a Defining Function: In geometric analysis it is frequently useful to think of a domain Ω in space as the sublevel set of a function $\rho = \rho_\Omega$. For example, the unit disk is given by

$$D = \{z \in \mathbb{C} : \rho(z) \equiv |z|^2 - 1 < 0\} . \qquad (5.2.2)$$

It is a nice exercise with the implicit function theorem to see that any domain whose boundary consists of C^1 curves can be written as the sublevel set of some function ρ, as in equation (5.2.2). We usually demand that $\nabla\rho \neq 0$ on $\partial\Omega$ in order to prevent degeneracies and, more particularly, so that $\nabla\rho(P)$ gives a well-defined outward normal at each boundary point P. We say that $\Omega = \{z \in \mathbb{C} : \rho(z) < 0\}$ has C^k boundary, $k \in \{0, 1, 2, \dots\}$, if there is a defining function ρ that is C^k.

We say that a bounded domain $\Omega \subseteq \mathbb{C}$ has C^∞ (or *smooth*) boundary if there is a function $\rho : \mathbb{C} \to \mathbb{R}$ such that ρ is C^∞,

$$\Omega = \{z \in \mathbb{C} : \rho(z) < 0\},$$

and $\nabla\rho$ is nowhere zero on $\partial\Omega$.

Example 5.2.2. Let

$$\Omega = \left\{ z \in \mathbb{C} : \rho(z) = \left|\frac{z + \overline{z}}{2}\right|^{(2k+1)/2} + \left|\frac{z - \overline{z}}{2}\right|^{(2k+1/2)} - 1 < 0 \right\} ,$$

for $k = 1, 2, \dots$. Then Ω has C^k boundary but not C^{k+1} boundary.

We say that a bounded domain $\Omega \subseteq \mathbb{C}$ has C^∞ (or *smooth*) boundary if there is a function $\rho : \mathbb{C} \to \mathbb{R}$ such that ρ is C^∞,

$$\Omega = \{z \in \mathbb{C} : \rho(z) < 0\},$$

and $\nabla\rho$ is nowhere zero on $\partial\Omega$.

Example 5.2.3. Let

$$\Omega = \left\{ z \in \mathbb{C} : \rho(z) = \left(\frac{z+\overline{z}}{2}\right)^2 + \left(\frac{z-\overline{z}}{8i}\right)^2 - 1 < 0 \right\}.$$

Then Ω is the region bounded by an ellipse. Certainly Ω has C^∞ boundary.

Theorem 5.2.4. *Let $k \in \{0, 1, 2, \ldots\}$. Then there is an integer $N = N(k)$ such that, if $\Omega \subseteq \mathbb{C}$ is a bounded domain with C^N boundary, then any conformal mapping $\varphi : \Omega \to D$ will extend to be univalent and C^k from $\overline{\Omega}$ to $\overline{D}$. Likewise, the inverse mapping φ^{-1} will extend to be univalent and C^k from $\overline{D}$ to $\overline{\Omega}$. It may be shown—although it requires considerable extra effort—that N may be taken to be $k + 1$, or even $k + \epsilon$.*

This theorem is properly attributed to P. Painlevé, who proved it in his thesis [PAI]. Over the years, it has been sharpened through the work of Kellogg, Warschawski, Pommerenke, and others. We shall indicate some of these results in what follows.

Let $\Omega \subseteq \mathbb{C}$ be a domain. We say that Ω satisfies *Bell's Condition R* if, for each $k \in \{0, 1, 2, \ldots\}$, there is an $\ell \in \{0, 1, 2, \ldots\}$ such that, whenever $f : \Omega \to \mathbb{C}$ has bounded derivative up to order ℓ, then $P_\Omega f$ (the Bergman projection of f) has bounded derivatives up to order k. Condition R has proved to be of historical importance in the study of holomorphic mappings. In this section we prove that $D = D(0, 1)$ satisfies a version of Condition R (see Theorem 5.2.13). This will be the key to the proof below of Theorem 5.2.4.

If Ω is a domain and $\Phi : \Omega \to \mathbb{C}$, then we say that Φ is k times boundedly continuously differentiable, and we write $\Phi \in C^k_b(\Omega)$, if all partial derivatives $(\partial/\partial x)^s(\partial/\partial y)^t\Phi$ of order $s + t \le k$ on Ω (the interior of Ω, not the boundary) exist, are continuous, and are bounded. We let $C^\infty_b(\Omega) = \cap_k C^k_b(\Omega)$. If $\Phi \in C^k_b(\Omega)$, then we set

$$\|\Phi\|_{C^k_b(\Omega)} = \sum_{s+t\le k} \sup_{\Omega} \left\| \left(\frac{\partial}{\partial x}\right)^s \left(\frac{\partial}{\partial y}\right)^t \Phi(x+iy) \right\|.$$

We also need an idea of continuity of derivatives at the boundary: A function $\Phi : \overline{\Omega} \to \mathbb{C}$ is called C^k if all partial derivatives of Φ up to order k extend continuously to $\overline{\Omega}$. As before, $C^\infty(\overline{\Omega})$ is defined to be $\cap_k C^k(\overline{\Omega})$.

Definition 5.2.5. If $f, g \in C^k(\overline{\Omega})$, then we say that f and g *agree up to order* k on $\partial\Omega$ if

$$\left(\frac{\partial}{\partial x}\right)^s \left(\frac{\partial}{\partial y}\right)^t (f(z) - g(z))\bigg|_{\partial\Omega} = 0$$

for all $s + t \le k$.

Lemma 5.2.6. *For every $k = 0, 1, 2, \ldots$ there is a function $\lambda_k \in C^k(\mathbb{R})$ such that*

1. $\lambda_k(x) = 0$ *if* $x \le \frac{1}{3}$,
2. $\lambda_k(x) = 1$ *if* $x \ge \frac{2}{3}$,
3. $0 \le \lambda_k(x) \le 1$ *for all* $x \in \mathbb{R}$.

Proof. Let

$$\psi(x) = \begin{cases} x^{k+1} & \text{if } x \ge 0\,, \\ 0 & \text{if } x < 0\,. \end{cases}$$

Obviously $\psi \in C^k(\mathbb{R})$. Define

$$\phi(x) = \psi\left(x - \frac{1}{3}\right) \cdot \psi\left(-x + \frac{2}{3}\right).$$

Then $\phi \in C^k$ and $\phi(x) \ne 0$ only if $1/3 < x < 2/3$. Observe that $\phi(x) \ge 0$ for all x. Let

$$u(x) = \int_{-\infty}^{x} \phi(t)\, dt.$$

Then $u \in C^k(\mathbb{R}), u(x) = 0$ for $x \le 1/3$, and $u \equiv c$ a positive constant for $x \ge 2/3$. Define

$$\lambda_k(x) = \frac{1}{c} u(x).$$

This function has all the required properties. □

It is actually possible to find a C^∞ function with properties **1**, **2**, and **3** of Lemma 5.2.6.

Lemma 5.2.7. *If $\Omega \subseteq \mathbb{C}$ is a domain with C^k boundary, $\Omega = \{z \in \mathbb{C} : \rho(z) < 0\}$, and if U is a neighborhood of $\partial\Omega$ on which $|\nabla\rho| \ge c > 0$, then there is a C^k function α_k on $\overline{\Omega}$ such that*

1. $\alpha_k = 0$ *on* $\Omega \setminus U$,
2. $\alpha_k = 1$ *in a neighborhood of* $\partial\Omega$.

Proof. Choose a number $\epsilon > 0$ such that if $-\epsilon < \rho(z) < 0$, then $z \in U \cap \Omega$. Define

$$\alpha_k(z) = \lambda_k\left(1 + \frac{\rho(z)}{\epsilon}\right).$$

Then α_k has all the desired properties.

Lemma 5.2.8. *Let $\Omega = \{z \in \mathbb{C} : \rho(z) < 0\}$ be a bounded domain with C^∞ boundary. Given $g \in C^\infty(\overline{\Omega})$, choose $f(z) = \rho(z)g(z)$. If $h \in A^2(\Omega)$ (the Bergman space—see Chapter 1)), then*

$$\int_\Omega \overline{h(\zeta)} \frac{\partial}{\partial \zeta} f(\zeta)\, d\xi\, d\eta = 0.$$

Proof. We want to integrate by parts, but h is not defined on $\partial\Omega$ so we need a limiting argument.

Let $\epsilon > 0$ be small and let $\Omega_\epsilon = \{z : \rho(z) < -\epsilon\}$. Then define $f_\epsilon(z) = (\rho(z)+\epsilon)g(z)$. Notice that $f_\epsilon \in C^\infty(\overline{\Omega}_\epsilon)$ and $f_\epsilon\big|_{\partial\Omega_\epsilon} = 0$. Also $h \in C^\infty(\overline{\Omega}_\epsilon)$. So, by Green's theorem (or just the one-variable fundamental theorem of calculus, applied one variable at a time),

$$\int_{\Omega_\epsilon} \overline{h(\zeta)} \frac{\partial}{\partial \zeta} f_\epsilon(\zeta)\, d\xi\, d\eta = -\int_{\Omega_\epsilon} \left(\frac{\partial}{\partial \zeta}\overline{h(\zeta)}\right) \cdot f_\epsilon(\zeta)\, d\xi\, d\eta. \tag{5.2.3}$$

There is no boundary term since $h \cdot f_\epsilon\big|_{\partial\Omega_\epsilon} \equiv 0$. But $h \in A^2(\Omega)$ so $(\partial/\partial\zeta)\overline{h} \equiv 0$ and the last expression is 0.

Finally, let

$$\chi_\epsilon(\zeta) = \begin{cases} 1 \text{ if } \zeta \in \Omega_\epsilon\,, \\ 0 \text{ if } \zeta \notin \Omega_\epsilon. \end{cases}$$

Then, using (5.2.3),

$$\begin{aligned} \left|\int_\Omega \overline{h(\zeta)} \frac{\partial}{\partial \zeta} f(\zeta)\, d\xi\, d\eta\right| &= \left|\int_\Omega \overline{h(\zeta)} \frac{\partial}{\partial \zeta} f(\zeta)\, d\xi\, d\eta - \int_\Omega \overline{h(\zeta)}\chi_\epsilon(\zeta) \frac{\partial}{\partial \zeta} f_\epsilon(\zeta)\, d\xi\, d\eta\right| \\ &\leq \int_\Omega \left|\overline{h(\zeta)}\right| \left|\frac{\partial f}{\partial \zeta}(\zeta)\right| |1 - \chi_\epsilon(\zeta)|\, d\xi\, d\eta \\ &\quad + \int_\Omega |\overline{h(\zeta)}||\chi_\epsilon(\zeta)| \cdot \left|\frac{\partial f}{\partial \zeta}(\zeta) - \frac{\partial f_\epsilon}{\partial \zeta}(\zeta)\right|\, d\xi\, d\eta\,. \end{aligned}$$

By the Cauchy–Schwarz inequality, this is

$$\begin{aligned} &\leq \left(\int_\Omega |\overline{h(\zeta)}|^2 \left|\frac{\partial f}{\partial \zeta}(\zeta)\right|^2\, d\xi\, d\eta\right)^{1/2} \cdot \left(\int_\Omega |1 - \chi_\epsilon(\zeta)|^2\, d\xi\, d\eta\right)^{1/2} \\ &\quad + \left(\int_\Omega |\overline{h(\zeta)}|^2\, d\xi\, d\eta\right)^{1/2} \cdot \left(\int_\Omega \left|\frac{\partial f}{\partial \zeta}(\zeta) - \frac{\partial f_\epsilon}{\partial \zeta}(\zeta)\right|^2\, d\xi\, d\eta\right)^{1/2}. \end{aligned}$$

Now $\int_\Omega |1 - \chi_\epsilon(\zeta)|^2\, d\xi\, d\eta$ clearly tends to zero as $\epsilon \to 0^+$. Also we see that $\partial f_\epsilon/\partial\zeta \to \partial f/\partial\zeta$ uniformly. So, letting $\epsilon \to 0$, we obtain

$$\int_\Omega \overline{h(\zeta)} \frac{\partial f}{\partial \zeta}(\zeta)\, d\xi\, d\eta = 0.$$

That is the desired conclusion. $\square$

If $\Phi : \Omega_1 \to \Omega_2$ is conformal, then we might expect the Bergman kernel for Ω_1 to be related to that for Ω_2. This is indeed the case, as our Proposition 1.2.11 showed. We review it now.

Theorem 5.2.9. *If $\Phi : \Omega_1 \to \Omega_2$ is conformal then*

$$K_{\Omega_1}(z,w) = \Phi'(z) K_{\Omega_2}(\Phi(z),\Phi(w))\overline{\Phi'(w)}.$$

We conclude this section with Bell's projection formula, which is based on Theorem 5.2.9.

If $\Omega \subseteq \mathbb{C}$ is a domain and K its Bergman kernel, then for any square integrable f on Ω we define

$$P_\Omega f(z) = \int_\Omega f(\zeta) K_\Omega(z,\zeta)\, d\xi\, d\eta.$$

This is the *Bergman projection*.

Proposition 5.2.10. *If Ω is a bounded domain in $\mathbb{C}$, then the mapping*

$$P : f \mapsto \int_\Omega K(\cdot,\zeta) f(\zeta)\, dA(\zeta)$$

is the Hilbert space orthogonal projection of $L^2(\Omega,\, dA)$ onto $A^2(\Omega)$.

Proof. Notice that P is idempotent and self-adjoint and that $A^2(\Omega)$ is precisely the set of elements of L^2 that are fixed by P. $\square$

Next is Bell's formula:

Proposition 5.2.11. *Let Ω_1, Ω_2 be domains and let*

$$\Phi : \Omega_1 \to \Omega_2$$

be a conformal map. Let $f \in L^2(\Omega_1)$. Then

$$P_{\Omega_2}((\Phi^{-1})' \cdot (f \circ \Phi^{-1})) = (\Phi^{-1})' \cdot ((P_{\Omega_1} f) \circ \Phi^{-1}).$$

Proof. We sketch the proof. For any $f \in L^2(\Omega_1)$ and $g \in L^2(\Omega_2)$,

$$\begin{aligned}
\langle P_{\Omega_2}((\Phi^{-1})' \cdot (f \circ \Phi^{-1})), g\rangle_2 &= \langle (\Phi^{-1})' \cdot (f \circ \Phi^{-1}), P_{\Omega_2} g\rangle_2 \\
&= \langle f, ((P_{\Omega_2} g) \circ \Phi) \cdot \Phi'\rangle_1 \\
&= \langle P_{\Omega_1} f, ((P_{\Omega_2} g) \circ \Phi) \cdot \Phi'\rangle_1 \\
&= \langle ((P_{\Omega_1} f) \circ \Phi^{-1}) \cdot (\Phi^{-1})', P_{\Omega_2} g\rangle_2 \\
&= \langle ((P_{\Omega_1} f) \circ \Phi^{-1}) \cdot (\Phi^{-1})', g\rangle_2 \,.
\end{aligned}$$

Since $g \in L^2(\Omega_2)$ was arbitrary, we may conclude that Bell's identity is proved. $\square$

The next lemma, due to Bell, is central to the theory.

Proposition 5.2.12. *If $\Omega = \{z \in \mathbb{C} : \rho(z) < 0\}$ is a bounded domain with C^∞ boundary, if $u \in C^{k+1}(\overline{\Omega})$, and if $k \geq 1$, then there is a $g \in C^k(\overline{\Omega})$ that agrees with u up to order k on $\partial\Omega$ and such that $P_\Omega g = 0$.*

Proof. Let α_k be as in Lemma 5.2.7. We define g by induction. For the C^1 case, let

$$v_1(z) = \frac{\partial}{\partial z} w_1(z)\,,$$

where

$$w_1(z) = \frac{\alpha_1(z) \cdot u(z) \cdot \rho(z)}{(\partial\rho/\partial z)(z)}.$$

Then

$$\begin{aligned} v_1(z) &= \alpha_1(z) \cdot u(z) + \rho(z) \cdot \frac{\partial}{\partial z}\left(\frac{\alpha_1(z) \cdot u(z)}{(\partial\rho/\partial z)(z)}\right) \\ &\equiv \alpha_1(z) \cdot u(z) + \rho(z) \cdot \eta_1(z)\,, \end{aligned}$$

where η_1 is defined by this identity and is continuous on $\overline{\Omega}$. Then

$$v_1 - u = \rho(z) \cdot \eta_1(z) \quad \text{near} \ \ \partial\Omega.$$

(So v_1 and u agree to order *zero* on $\partial\Omega$.) In particular, $v_1 - u\big|_{\partial\Omega} = 0$. Also

$$P_\Omega v_1(z) = \int_\Omega K(z,\zeta)\frac{\partial}{\partial\zeta} w_1(\zeta)\, d\xi\, d\eta = 0$$

by Lemma 5.2.8.

Suppose inductively that we have constructed $w_{\ell-1}$ and $v_{\ell-1} = \frac{\partial}{\partial z} w_{\ell-1}$ such that $v_{\ell-1}$ agrees to order $(\ell-1)-1$ with u on $\partial\Omega$ and $P_\Omega v_{\ell-1} = 0$. We shall now construct a function w_ℓ of the form

$$w_\ell = w_{\ell-1} + \theta_\ell \cdot \rho^\ell \tag{5.2.4}$$

such that $v_\ell = \frac{\partial}{\partial z} w_\ell$ agrees with u up to order $\ell - 1$ on $\partial\Omega$ and $P_\Omega v_\ell = 0$.

Let α_ℓ be as in Lemma 5.2.7 and define a differential operator $\mathcal{D}$ on $\overline{\Omega}$ by

$$\mathcal{D}(\phi) = \frac{\alpha_\ell(z)}{|\partial\rho/\partial z|^2}\mathrm{Re}\left(\frac{\partial\rho}{\partial z}\frac{\partial\phi}{\partial\overline{z}}\right).$$

Notice that

$$\mathcal{D}\rho(z) = 1 \tag{5.2.5}$$

when $z \in \partial\Omega$. We define

$$\theta_\ell = \frac{\alpha_\ell \mathcal{D}^{\ell-1}(u - v_{\ell-1})}{\ell!\partial\rho/\partial z}. \tag{5.2.6}$$

Then, with w_ℓ defined as in (5.2.4) and $v_\ell = \frac{\partial}{\partial z} w_\ell$, we have

$$\begin{aligned}\mathcal{D}^{\ell-1}(u - v_\ell) &= \mathcal{D}^{\ell-1}u - \mathcal{D}^{\ell-1}\frac{\partial}{\partial z}(w_{\ell-1} + \theta_\ell \cdot \rho^\ell) \\ &= \mathcal{D}^{\ell-1}(u - v_{\ell-1}) - \theta_\ell \frac{\partial \rho}{\partial z} \cdot (\mathcal{D}\rho)^{\ell-1} \cdot \ell! \\ &\quad + (\text{terms that involve a factor of } \rho).\end{aligned}$$

If $z \in \partial\Omega$, then (using (5.2.5), (5.2.6))

$$\mathcal{D}^{\ell-1}(u - v_\ell) = \left[\mathcal{D}^{\ell-1}(u - v_{\ell-1})\right] \cdot (1 + \varphi) + \psi \cdot \rho\,, \tag{5.2.7}$$

where $\varphi,\ \psi$ are smooth. Since any directional derivative at $P \in \partial\Omega$ is a linear combination of

$$\mathcal{D} = a(P)\frac{\partial}{\partial x} + b(P)\frac{\partial}{\partial y} \quad \text{and} \quad \tau = -b(P)\frac{\partial}{\partial x} + a(P)\frac{\partial}{\partial y},$$

we may re-express our task as follows: We need to see that

$$(\tau)^s \mathcal{D}^t (u - v_\ell)\bigg|_{\partial\Omega} = 0$$

for all $s+t = \ell-1$ (notice that the case $s+t < \ell-1$ follows from the inductive hypothesis and the explicit form of w_ℓ in (5.2.4)). The case $s = 0, t = \ell - 1$ was treated in (5.2.7). If $s \geq 1$, then we write

$$\tau^s \mathcal{D}^t (u - v_\ell) = \tau(\tau^{s-1}\mathcal{D}^t(u - v_\ell)). \tag{5.2.8}$$

Since $(s-1) + t = \ell - 2$, the expression in parentheses is 0 on $\partial\Omega$ by the inductive hypothesis. But τ is a directional derivative *tangent* to $\partial\Omega$ (because $\mathcal{D}$ is normal); hence (5.2.8) is 0.

The induction is now complete, and v_ℓ has been constructed. We set $v_\ell = g$. Then

$$P_\Omega g(z) = \int_\Omega K_\Omega(z,\zeta) g(\zeta) d\xi\, d\eta = \int_\Omega K_\Omega(z,\zeta)\frac{\partial}{\partial \zeta} w_k(\zeta)\, d\xi\, d\eta = 0$$

by Lemma 5.2.8. □

We shall use Bell's lemma (Proposition 5.2.12) twice. Our first use right now is on the disk. First note the following two simple facts:

(a) If K_D is the Bergman kernel for the disk, then

$$\left|\left(\frac{\partial}{\partial z}\right)^k K(z,w)\right| = \left|\frac{(k+1)!\overline{w}^k}{\pi(1 - z \cdot \overline{w})^{k+2}}\right| \leq \frac{(k+1)!}{(1 - |w|)^{k+2}}.$$

(b) If $u \in C^k(\overline{D})$ and if u *vanishes to* order k at ∂D (i.e., u agrees with the zero function to order k at ∂D), then there is a $C > 0$ such that $|u(z)| \leq C \cdot (1 - |z|)^k$.

Theorem 5.2.13 (Condition R for the disk). *If $k \geq 1$ and $u \in C^{k+2}(\overline{D})$, then $\|P_D u\|_{C_b^{k-1}(D)} < \infty$.*

Proof. Use Proposition 5.2.12 to find a function $v \in C^{k+1}(\overline{D})$ that agrees with u to order $k+1$ on ∂D and such that $Pv = 0$. Then $P_D(u - v) = P_D u$, and $u - v$ vanishes to order $k+1$ on ∂D. In particular, by observation **(b)** above, $|u(\zeta) - v(\zeta)| \leq C \cdot (1 - |\zeta|)^{k+1}$. Then, for $j \leq k-1$, we have

$$\begin{aligned}
\left| \left(\frac{\partial}{\partial z} \right)^j P_D u(z) \right| &= \left| \left(\frac{\partial}{\partial z} \right)^j P_D(u - v)(z) \right| \\
&= \left| \int_D \left(\frac{\partial}{\partial z} \right)^j K_D(z, \zeta)(u - v)(\zeta)\, d\xi\, d\eta \right| \\
&\leq \int_D (j+1)!(1 - |\zeta|)^{-j-1} C \cdot (1 - |\zeta|)^{k+1}\, d\xi\, d\eta\,,
\end{aligned}$$

where we have used observation **(a)** above. This last integral is clearly bounded, independent of z. □

Remark 5.2.14. Item **(b)** above actually holds on any bounded domain Ω with C^k boundary: If $u \in C^k(\overline{\Omega})$ vanishes to order k on $\partial\Omega$, then there is a $C > 0$ such that

$$|u(z)| \leq C \cdot \delta_\Omega(z)^k.$$

Here, for $z \in \Omega$,

$$\delta_\Omega(z) = \inf_{w \notin \Omega} |z - w|.$$

Of course δ_Ω could also be replaced here by the modulus of any C^k-smooth defining function. This is immediate from the definitions and the Taylor expansion in the normal direction (in terms of powers of the defining function ρ).

5.3 Smoothness to the Boundary of Conformal Mappings

Let $\Omega = \{z \in \mathbb{C} : \rho(z) < 0\}$ be a bounded and simply connected domain with C^∞ boundary. Let $F : D \to \Omega$ be a conformal mapping. We wish to show that the one-to-one continuation of F to $\overline{D}$ (provided by Theorem 5.1.1) is actually in $C^\infty(\overline{D})$. For this we need a few lemmas. The first is a classical result from the theory of partial differential equations due to Hopf.

Lemma 5.3.1. *Let $U \subseteq \mathbb{C}$ be smoothly bounded. Let u be a harmonic function on U, continuous on the closure $\overline{U}$. Suppose that u assumes a local maximum value at $P \in \partial U$. Let ν be the unit outward normal vector to ∂U at P. Then the one-sided lower derivative $\partial u/\partial \nu$, defined to be*

$$\frac{\partial u}{\partial \nu} = \liminf_{t \to 0^+} \frac{u(P) - u(P - t\nu)}{t} ,$$

is positive.

Proof. It is convenient to make the following normalizations: Assume that u assumes the value 0 at P and is negative nearby and inside U; finally take the negative of our function so that u has a local *minimum* at P.

Now, since U has smooth boundary, there is an internally tangent disk at P. After scaling, we may as well suppose that it is the unit disk and that $P = 1 + i0$. Thus we may restrict our positive function u, with the minimum value 0 at $P = 1$, to the closed unit disk. Note in particular that $u(0) > 0$. Set $C = u(0) > 0$.

The Harnack inequality shows that $u(r) \geq [(1 - r)/(1 + r)]u(0)$, hence

$$\frac{u(1) - u(r)}{1 - r} = \frac{-u(r)}{1 - r} \leq -\frac{u(0)}{1 + r} \equiv \frac{-C}{1 + r} \leq -\frac{C}{2} .$$

The desired inequality for the normal derivative of u now follows. □

Remark 5.3.2. It is worth noting that the definition of the derivative and the fact that P is a local maximum guarantees—just from first principles—that the indicated one-sided lower normal derivative will be nonnegative. The Hopf lemma asserts that this derivative is actually positive.

Lemma 5.3.3. *If Ω is a bounded, simply connected domain with C^∞ boundary and if $F : D \to \Omega$ is a biholomorphic mapping, then there is a constant $C > 0$ such that*

$$\delta_\Omega(F(z)) \leq C(1 - |z|), \quad \text{all} \quad z \in \Omega.$$

Here $\delta_\Omega(z) = \inf_{w \notin \Omega} |z - w|$.

Proof. The issue has to do only with points z near the boundary of D, and thus with points where $F(z)$ is near the boundary of Ω. Consider the function $w \mapsto \log |F^{-1}(w)|$. This function is defined for all w sufficiently near $\partial \Omega$ and indeed on $\Omega \setminus \{F(0)\}$. And it is harmonic there. Moreover, it is continuous on $(\Omega \setminus \{F(0)\}) \cup \partial \Omega$, with value 0 on $\partial \Omega$. In particular, it attains a (global) maximum at every point of $\partial \Omega$, since $|F^{-1}(w)| < 1$ if $w \in \Omega \setminus \{F(0)\}$. So the Hopf lemma applies. The logarithm function has nonzero derivative at all points. The conclusion of the lemma follows from combining this fact with the "normal derivative" conclusion of the Hopf lemma. That is, we have that

$$\frac{\partial}{\partial \nu}|F^{-1}|\bigg|_P \geq c > 0$$

at each $P \in \partial D$. [Note that here, as in the Hopf lemma, no differentiability at boundary points is assumed: The derivative estimates are on the "lower derivative" only, which, as a lim inf, always exists and has value in the extended reals.]

Lemma 5.3.4. *With F, Ω as above and $k \in \{0, 1, 2, \ldots\}$ it holds that*

$$\left|\left(\frac{\partial}{\partial z}\right)^k F(z)\right| \leq C_k(1 - |z|)^{-k}.$$

Proof. Since F takes values in Ω, it follows that F is bounded. Now apply the Cauchy estimates on $D(z, (1 - |z|))$. □

Lemma 5.3.5. *If $\psi \in C^{2k+2}(\overline{\Omega})$ vanishes to order $2k + 1$ on $\partial\Omega$, then $F' \cdot (\psi \circ F) \in C_b^k(D)$. That is, $F' \cdot (\psi \circ F)$ has bounded derivatives up to and including order k.*

Proof. For $j \leq k$ we have

$$\left|\left(\frac{\partial}{\partial z}\right)^j (F' \cdot (\psi \circ F))\right| = \left|\sum_{\ell=0}^{j} \binom{j}{\ell} \left(\frac{\partial}{\partial z}\right)^\ell (F') \cdot \left(\frac{\partial}{\partial z}\right)^{j-\ell} (\psi \circ F)\right|.$$

But

$$\left(\frac{\partial}{\partial z}\right)^{j-\ell} (\psi \circ F) \tag{5.3.1}$$

is a linear combination, with complex coefficients, of terms of the form

$$\left[\left(\left(\frac{\partial}{\partial z}\right)^m \psi\right)(F(z))\right] \cdot \left(\frac{\partial}{\partial z}\right)^{n_1} F(z) \cdots \left(\frac{\partial}{\partial z}\right)^{n_k} F(z)$$

where $m \leq j - \ell$ and $n_1 + \cdots + n_k \leq j - \ell$. So (5.3.1) is dominated by

$$\begin{aligned} &C \cdot \delta_\Omega(F(z))^{2k+1-(j-\ell)}(1 - |z|)^{-n_1} \cdots (1 - |z|)^{-n_k} \\ &\leq C \cdot \delta_\Omega(F(z))^{k+1} \cdot (1 - |z|)^{-k-1}. \end{aligned}$$

By Lemma 5.3.3, this is

$$\leq C \cdot (1 - |z|)^{k+1} \cdot (1 - |z|)^{-k-1} \leq C. \ \square$$

Remark 5.3.6. There is a remarkable formula of Faa di Bruno [KRP1] that formalizes the expansion for the higher derivatives of a composition. While the identity was first discovered in the eighteenth century, it is still being studied today (in the higher-dimensional version).

Lemma 5.3.7. *Let $G : D \to \mathbb{C}$ be holomorphic and have the property that*

$$\left| \left(\frac{\partial}{\partial z} \right)^j G(z) \right| \leq C_j < \infty$$

for $j = 0, \ldots, k+1$. Then each $(\partial/\partial z)^j G$ extends continuously to $\overline{D}$ for $j = 1, \ldots, k$.

Proof. It is enough to treat the case $k = 0$. The general case follows inductively.

If $P \in \partial D$, then we define

$$G(P) = \int_0^1 G'(tP) \cdot P dt + G(0).$$

It is clear that this defines a continuous extension of G to ∂D. □

Theorem 5.3.8 (Painlevé). *If $\Omega = \{z \in \mathbb{C} : \rho(z) < 0\}$ is a bounded, simply connected domain with C^∞ boundary and $F : D(0,1) \to \Omega$ is a conformal mapping, then $F \in C^\infty(\overline{D})$ and $F^{-1} \in C^\infty(\overline{\Omega})$.*

Proof. Let $k \in \{1, 2, \ldots\}$. By Proposition 5.2.12 applied to the function $u = 1$ on Ω, there is a function $v \in C^{2k+8}(\overline{\Omega})$ such that v agrees with u up to order $2k+8$ on $\partial\Omega$ and $P_\Omega v = 0$. Then $\phi \equiv 1 - v$ satisfies $P_\Omega(\phi) = P_\Omega 1 - P_\Omega v = 1$ and ϕ vanishes to order $2k+8$ on $\partial\Omega$. By Lemma 5.3.5, $F' \cdot (\phi \circ F) \in C_b^{k+3}(D)$. By Lemma 5.3.7, $F' \cdot (\phi \circ F) \in C^{k+2}(\overline{D})$. By Theorem 5.2.13, $P_D(F' \cdot (\phi \circ F)) \in C^{k-1}(D)$ and has $k - 1$ bounded derivatives. But the transformation law (Proposition 5.2.11) tells us that

$$P_D(F' \cdot (\phi \circ F)) = F' \cdot ((P_\Omega \phi) \circ F) = F' \cdot 1 = F'.$$

Thus F' has bounded derivatives up to order $k - 1$. By Lemma 5.3.7, all derivatives of F up to order $k - 2$ extend continuously to $\overline{D}$. Since k was arbitrary, we may conclude that $F \in C^\infty(\overline{D})$.

To show that $F^{-1} \in C^\infty(\overline{\Omega})$, it is enough to show that the Jacobian determinant of F as a real mapping on $\overline{D}$ does not vanish at any boundary point of D (we already know it is everywhere nonzero on D). Since F is holomorphic, F continues to satisfy the Cauchy–Riemann equations on $\overline{D}$. So it is enough to check that, at each point of $\overline{D} \setminus D$, some first derivative of F is nonzero. This assertion follows from a Hopf lemma argument analogous to the proof of Lemma 5.3.1. Details are left as an exercise. □

Classically, Theorem 5.3.8 was proved by studying Green's potentials. The result dates back to P. Painlevé's thesis. All the ideas in the proof we have presented here are due to S. Bell and E. Ligocka [BELL]. An account of Bell's approach, in a more general context, can be found in [BEK].

Problems for Study and Exploration

1. Carathéodory's theorem guarantees that if Ω is a planar domain whose boundary consists of a single Jordan curve, then any conformal mapping of the disk D to Ω extends continuously and univalently to the boundary (and so does its inverse mapping). Can you give an example of such a domain for which the mapping does not extend with any Lipschitz smoothness?

2. Assume that we know that the Dirichlet problem on a smoothly bounded domain with smooth data will have a solution that is smooth on the closure. We will use this information to study the boundary regularity of conformal mappings. Complete the following outline to obtain a proof. Let $\phi : U \to D$ be conformal, U smoothly bounded.

(a) Let W be a collared neighborhood of ∂U (see [HIR]). Set $\partial U' = \partial W \cap U$ and let $\partial D' = \phi(\partial U')$. Define B to be the region bounded by ∂D and $\partial D'$. We solve the Dirichlet problem on B with boundary data

$$f(\zeta) = \begin{cases} 1 \text{ if } \zeta \in \partial D\,, \\ 0 \text{ if } \zeta \in \partial D'\,. \end{cases}$$

Call the solution u. Draw a picture to illustrate these ideas.

(b) Consider $v \equiv u \circ \phi : U \cap W \to \mathbb{R}$. Then of course v is still harmonic. By Carathéodory's theorem, v extends to $\partial U, \partial U'$, and

$$v = \begin{cases} 1 \text{ if } \zeta \in \partial U\,, \\ 0 \text{ if } \zeta \in \partial U'\,. \end{cases}$$

If we consider a first-order derivative $\mathcal{D}$ of v, we obtain

$$|\mathcal{D}v| = |\mathcal{D}(u \circ \phi)| = |\nabla u|\, |\nabla \phi| \le C.$$

It follows that

$$|\nabla \phi| \le \frac{C}{|\nabla u|}. \tag{5.e.1}$$

(c) Now let us return to the u from the Dirichlet problem that we considered in part **(a)**. Hopf's lemma (Lemma 5.13) tells us that $|\nabla u| \ge c' > 0$ near ∂D. Thus, from (5.e.1), we conclude that

$$|\nabla \phi| \le C. \tag{5.e.2}$$

Thus we have bounds on the first derivatives of ϕ.

(d) To control the second derivatives, we calculate that

$$\begin{aligned} C \ge |\nabla^2 v| &= |\nabla(\nabla v)| = |\nabla(\nabla(u \circ \phi))| \\ &= |\nabla(\nabla u(\phi) \cdot \nabla \phi)| = |\big(\nabla^2 u \cdot [\nabla \phi]^2\big) + \big(\nabla u \cdot \nabla^2 \phi\big)|. \end{aligned}$$

Here the reader should think of ∇ as representing a generic first derivative and ∇^2 a generic second derivative. We conclude that

$$|\nabla u|\,|\nabla^2 \phi| \le C + |\nabla^2 u|\,|(\nabla\phi)^2| \le C'.$$

Hence (again using Hopf's lemma)

$$|\nabla^2\phi| \le \frac{C'}{|\nabla u|} \le C''.$$

(e) In the same fashion, we may prove that $|\nabla^k \phi| \le C_k$, any $k \in \{1, 2, \dots\}$. This means (use the fundamental theorem of calculus) that $\phi \in C^\infty(\bar{U})$.

3. Let D be the unit disk as usual. For $P \in \partial D$ and $\alpha > 1$, let

$$\Gamma_\alpha(P) = \{z \in D : |z - P| < \alpha(1 - |z|)\}\,.$$

We call Γ_α a *nontangential approach region* at the point P. A function f on the disk has *nontangential limit* λ at P if, for some $\alpha > 0$,

$$\lim_{\Gamma_\alpha(P)\ni z \to P} f(z) = \lambda\,.$$

Carathéodory's theorem says that if f is any holomorphic function on D and if f has nontangential limit zero on a boundary set of positive measure, then $f \equiv 0$.

Complete this sketch for a proof. Let $E \subseteq \partial D$ be a compact set of positive measure such that f has nontangential boundary limit 0 at each point of E. Associate to each point $e \in E$ a nontangential cone $\Gamma_{\alpha(e)}(e)$ and set

$$\Omega = \bigcup_{e\in E} \Gamma_{\alpha(e)}(e)\,.$$

Then Ω is a domain with Lipschitz boundary. Conformally map Ω to the disk D with a conformal map φ, and consider $g \equiv f \circ \varphi^{-1}$. Argue that g is a *bounded* holomorphic function on D with boundary limit 0 on a set of positive measure. Now use Jensen's inequality, for example, to conclude that $g \equiv 0$.

4. Carathéodory's theorem says that, under fairly general hypotheses, a conformal map of a domain Ω to the disk extends continuously and univalently to the boundary. Can you use this theorem to derive an analogous result for the Ahlfors map?

5. We know that a finitely connected domain has a canonical presentation on a slit domain, or on a disk with finitely many smaller disks removed. Formulate and prove a boundary regularity result for this mapping.

6. Statements about boundary regularity of conformal mapping must be suitably modified when unbounded domains are involved. For example, the Cayley map sends the disk to the upper halfplane, and the boundary point -1 in the disk is sent to the point at ∞ in the boundary of the half-plane. Suggest how to modify the statement of Carathéodory's theorem, for instance, in order to handle unbounded domains.

7. Refer to Exercise 3 for terminology. Let $\Omega \subseteq \mathbb{C}$ be a simply connected domain with C^2 boundary, and let $\gamma : [0,1) \to \Omega$ be a curve that approaches $P \in \partial\Omega$ nontangentially (i.e., through some nontangential approach region $\Gamma_\alpha(P)$). Let $\varphi : \Omega \to D$ be a conformal mapping of Ω onto the unit disk. What can you say about $\varphi \circ \gamma$? What can you say about $\varphi(\Gamma_\alpha)$?

8. What happens in Exercise 7 if we do *not* assume that Ω has C^2 boundary?

9. Explain how the boundary regularity for the Dirichlet problem of a smoothly bounded, simply connected domain may be reduced to the study of the analogous question on the disk (using the Poisson kernel). Of course conformal mapping, and its boundary behavior, will play a role in your answer.

10. Let Ω be a bounded, simply connected domain with C^2 boundary. Let φ be a C^2 function on $\partial\Omega$ and let $\mathcal{C}f$ be the Cauchy integral of f to the interior of Ω. Explain why $\mathcal{C}f$ extends to be at least C^1 on the closure of Ω. [*Hint:* Conformal mapping, and its boundary regularity, will play a role in your answer.] Do not expect, however, that $\mathcal{D}f$ will agree with f at the boundary.

6

The Boundary Behavior of Holomorphic Functions

Genesis and Development

P. Fatou, G. H. Hardy, and F. Riesz were the pioneers in the study of the boundary behavior of holomorphic functions. In 1906, quick on the heels of Lebesgue's first publications on measure theory, Fatou proved a seminal result about the almost-everywhere boundary limits of bounded, holomorphic functions on the disk. Interestingly, be was able to render the problem as one about convergence of Fourier series, and he solved it in that language.

Today we see the study of boundary behavior as properly part of complex function theory. Reproducing kernels, such as the Poisson and Cauchy (Chapters 1 and 8), are basic tools in the subject. So are harmonic measure (Chapter 9) and conformal mapping (Chapters 4 and 5). Conversely, results about the boundary behavior of holomorphic functions shed light on mapping theory and the construction of kernels.

Modern developments by Fefferman, Stein, and Weiss give us a way to think of this topic as properly part of harmonic analysis. Thus there is a new wedding of real variable techniques with complex variable techniques. This has indeed proved to be a fecund meeting ground, and we shall indicate some of its features in the present chapter.

The reader of this chapter will want to be comfortable with elementary hard analysis and estimates.

6.0 Introductory Remarks

Let f be a holomorphic function on the unit disk D. It is a matter of considerable interest to determine whether f has some realization on the boundary ∂D of the disk. What does this mean? In the ideal situation, there is some boundary function f^*, that can be recovered by some limiting process from

f, such that f is the Cauchy integral (or perhaps the Poisson integral) of f^*. In fact results of this sort are true for fairly large classes of holomorphic functions f.

First, it is important to understand that without additional hypotheses on f, there is no hope. Let us begin by just discussing smooth f—even real analytic f. The function

$$\phi(z) = \sin\left(\frac{1}{1-|z|^2}\right) \tag{6.0.1}$$

is real analytic on the open unit disk and is bounded, but it has no limit at any boundary point in any reasonable sense. The Bagemihl–Seidel theorem, discussed below (Section 6.2), gives even more dramatic examples.

By contrast, in the year 1906—shortly after Lebesgue formulated his new theory of measure and the integral—Pierre Fatou proved the following remarkable theorem:

> **Theorem:** Let f be a bounded holomorphic function on the unit disk. Then, for almost every $\theta \in [0, 2\pi)$, the limit
>
> $$\lim_{r\to 1^-} f(re^{i\theta})$$
>
> exists.

Fatou's proof reflected the values and sensibilities of his time. He viewed the putative limit as Abel summation of the Fourier series for f. It was known at the time that Césaro summation implies Abel summation. So Fatou proved almost everywhere Césaro summability of the Fourier series of a bounded, holomorphic f.

In common parlance, we say that Fatou proved that a bounded analytic function has *radial* boundary limits almost everywhere. See Figure 6.1. A number of natural questions now come to mind:

- Is there a more general class of functions than "bounded" for which the result is true?
- Is there a more general sense than "radial" in which one can compute the boundary limit?
- Is the result true on a more general class of domains?

In the present section we will provide detailed answers to the first two of these queries. The third question can be answered with the uniformization theorem, or with various localization techniques. We cannot treat them here.

6.1 Review of the Classical Theory of H^p Spaces on the disk

Throughout this section we let $D \subseteq \mathbb{C}$ denote the unit disk. Let $0 < p < \infty$. We define

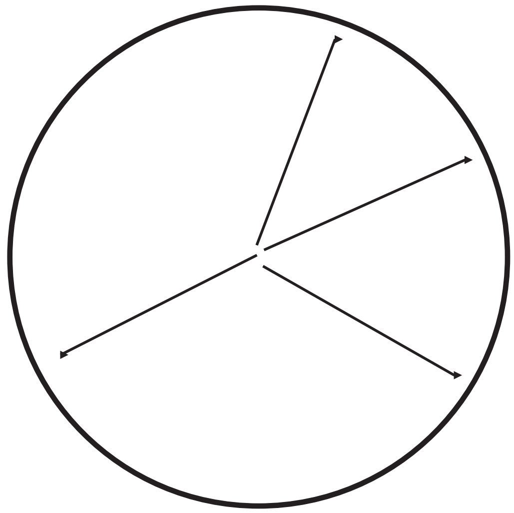

Fig. 6.1. Radial boundary limits.

$$H^p(D) = \left\{ f \text{ holomorphic on } D : \sup_{0<r<1} \left[\int_0^{2\pi} |f(re^{i\theta})|^p d\theta \right]^{1/p} \right.$$
$$\left. \equiv \|f\|_{H^p} < \infty \right\}.$$

Also define

$$H^\infty(D) = \left\{ f \text{ holomorphic on } \ D : \sup_D |f| \equiv \|f\|_{H^\infty} < \infty \right\}.$$

The fundamental result in the subject of H^p, or *Hardy*, spaces (and also the fundamental result of this section) is that, if $f \in H^p(D)$, then the limit

$$\lim_{r \to 1^-} f(re^{i\theta}) \equiv f^*(e^{i\theta})$$

exists for almost every $\theta \in [0, 2\pi)$. For $1 \le p \le \infty$, the function f can be recovered from f^* by way of the Cauchy or Poisson integral formulas; for $p < 1$ this "recovery" process is more subtle and must proceed by way of distributions. Once this pointwise boundary limit result is established, then an enormous and rich mathematical structure unfolds (see Y. Katznelson [KAT], K. Hoffman [HOF], J. Garnett [GAR], and S. Krantz [KRA1]).

Recall from Chapter 4 that the Poisson kernel for the disk is

$$P_r(e^{i\theta}) = \frac{1}{2\pi} \frac{1 - r^2}{1 - 2r\cos\theta + r^2}.$$

Let

$$\mathbf{h}^p(D) = \left\{ f \text{ harmonic on } D : \sup_{0<r<1} \left[\frac{1}{2\pi} \int_0^{2\pi} |f(re^{i\theta})|^p d\theta \right]^{1/p} \equiv \|f\|_{\mathbf{h}^p} < \infty \right\}$$

and

$$\mathbf{h}^\infty(D) = \left\{ f \text{ harmonic on } D : \sup_D |f| \equiv \|f\|_{\mathbf{h}^\infty} < \infty \right\}.$$

Throughout this section, arithmetic and measure theory on $[0, 2\pi)$ (equivalently on ∂D) is done by identifying $[0, 2\pi)$ with $\mathbb{R}/2\pi\mathbb{Z}$. See Y. Katznelson [KAT] for more on this identification procedure.

Proposition 6.1.1. *Let $1 < p \leq \infty$ and $f \in \mathbf{h}^p(D)$. Then there is an $f^* \in L^p(\partial D)$ such that*

$$f(re^{i\theta}) = \int_0^{2\pi} f^*(e^{i\psi}) P_r(e^{i(\theta-\psi)}) d\psi.$$

Proof. Define $f_r(e^{i\theta}) = f(re^{i\theta}), 0 < r < 1$. Then $\{f_r\}_{0<r<1}$ is a bounded subset of $L^p(\partial D) = (L^{p'}(\partial D))^*, p' = p/(p-1)$. By the Banach–Alaoglu theorem (see W. Rudin [RUD3]), there is a subsequence f_{r_j} with $r_j \to 1$ that converges weak-$*$ to some f^* in $L^p(\partial D)$. For any $0 < r < 1$, let $r < r_j < 1$. Then

$$f(re^{i\theta}) = f_{r_j}\left((r/r_j)e^{i\theta})\right) = \int_0^{2\pi} f_{r_j}(e^{i\psi}) P_{r/r_j}\left(e^{i(\theta-\psi)}\right) d\psi$$

because $f_{r_j} \in C(\overline{D})$. Now $P_{r/r_j} \in C(\partial D) \subseteq L^{p'}(\partial D)$. Thus the right-hand side of the last equation is

$$\int_0^{2\pi} f_{r_j}(e^{i\psi}) P_r(e^{i(\theta-\psi)}) d\psi + \int_0^{2\pi} f_{r_j}(e^{i\psi}) \left[P_{r/r_j}(e^{i(\theta-\psi)}) - P_r(e^{i(\theta-\psi)}) \right] d\psi.$$

As $j \to \infty$, the second integral vanishes (because the expression in brackets converges uniformly to 0) and the first integral tends (by the definition of weak-$*$ limit) to

$$\int_0^{2\pi} f^*(e^{i\psi}) P_r(e^{i(\theta-\psi)}) d\psi.$$

This is the desired result.

Remark 6.1.2. It is easy to see that the proof breaks down for $p = 1$ since L^1 is not the dual of any Banach space. This breakdown is not merely ostensible: the harmonic function

$$f(re^{i\theta}) = P_r(e^{i\theta})$$

satisfies

$$\sup_{0<r<1} \int_0^{2\pi} |f(re^{i\theta})| d\theta < \infty,$$

but the Dirac δ mass is the only measure of which f is the Poisson integral.

Exercise for the Reader

If $f \in \mathbf{h}^1$, then there is a Borel measure μ_f on ∂D such that $f(re^{i\theta}) = P_r(\mu_f)(e^{i\theta})$. ♢

Proposition 6.1.3. *Let $f \in L^p(\partial D), 1 \le p < \infty$. Then $\lim_{r\to 1^-} P_r f = f$ in the L^p norm.*

Remark 6.1.4. The result is false for $p = \infty$ if f is discontinuous. The correct analogue in the uniform case is that if $f \in C(\partial D)$, then $P_r f \to f$ uniformly.

As an exercise, consider a Borel measure μ on ∂D. Show that its Poisson integral converges in the weak-$*$ topology to μ.

Proof of Proposition 6.1.3. If $f \in C(\partial D)$, then the result is clear by the solution of the Dirichlet problem. If $1 \le p < \infty$ and $f \in L^p(\partial D)$ is arbitrary, let $\epsilon > 0$ and choose $\varphi \in C(\partial D)$ such that $\|f - \varphi\|_{L^p} < \epsilon$. Then

$$\begin{aligned}\|P_r f - f\|_{L^p} &\le \|P_r(f-\varphi)\|_{L^p} + \|P_r\varphi - \varphi\|_{L^p} + \|\varphi - f\|_{L^p} \\ &\le \|P_r\|_{L^1}\|f-\varphi\|_{L^p} + \|P_r\varphi - \varphi\|_{L^p} + \epsilon \\ &\le \epsilon + o(1) + \epsilon\end{aligned}$$

as $r \to 1^-$. [Here we have used the generalized Minkowski inequality, for which see [KAT] or [STW] or [KRA4]]. Letting $\epsilon \to 0$, the result follows. □

Proposition 6.1.5. *Let $K \subseteq \partial D$ be compact, and let $\{I_\alpha\}_{\alpha\in A}$ be a covering of K by open intervals. Then there is a subcovering $\{I_{\alpha_j}\}_{j=1}^M$ such that every point of K is contained in at least one but not more than two of the I_{α_j}'s. [We call such a subcover a cover of "valence two."]*

Proof. Since K is compact, $\{I_\alpha\}_{\alpha\in A}$ has a finite subcollection $\{I_{\alpha_j}\}_{j=1}^J$ that still covers K. Now, in sequence, discard any I_{α_j} that is contained in the union of the other intervals. □

Definition 6.1.6. If $f \in L^1(\partial D)$, let

$$Mf(\theta) = \sup_{R>0} \frac{1}{2R}\int_{-R}^{R} |f(e^{i(\theta-\psi)})|\,d\psi.$$

The function Mf is called the *Hardy–Littlewood maximal function* of f.

Definition 6.1.7. Let (X,μ) be a measure space and $f : X \to \mathbb{C}$ be measurable. We say that f is of *weak type* p, with $0 < p < \infty$, if $\mu\{x : |f(x)| > \lambda\} \le C/\lambda^p$, all $0 < \lambda < \infty$. The space weak type ∞ is defined to be L^∞.

The next estimate of Chebyshev is elementary but essential.

Lemma 6.1.8. *If $f \in L^p(X, d\mu)$, then f is weak type p, $1 \leq p < \infty$.*

Proof. Let $\lambda > 0$. Then

$$\mu\{x : |f(x)| > \lambda\} \leq \int_{\{x:|f(x)|>\lambda\}} |f(x)|^p/\lambda^p d\mu(x) \leq \lambda^{-p}\|f\|_{L^p}^p. \qquad \square$$

Exercise for the Reader: There exist functions that are of weak type p but not in $L^p, 1 \leq p < \infty$. $\diamondsuit$

Definition 6.1.9. An operator $T : L^p(X, d\mu) \to$ {measurable functions} is said to be of *weak type* (p, p), $1 \leq p < \infty$, if

$$\mu\{x : |Tf(x)| > \lambda\} \leq C\|f\|_{L^p}^p/\lambda^p \ , \qquad \text{all} \quad f \in L^p, \quad \lambda > 0.$$

Proposition 6.1.10. *The operator M is of weak type* $(1, 1)$.

Proof. Let $\lambda > 0$. Set $S_\lambda = \{\theta : |Mf(e^{i\theta})| > \lambda\}$. Let $K \subseteq S_\lambda$ be a compact subset with $2m(K) \geq m(S_\lambda)$. For each $k \in K$, there is an interval I_k centered at k with $|I_k|^{-1} \int_{I_k} |f(e^{i\psi})|d\psi > \lambda$. Then $\{I_k\}_{k \in K}$ is an open cover of K. By Proposition 6.1.5, there is a subcover $\{I_{k_j}\}_{j=1}^M$ of K of valence not exceeding 2. Then

$$\begin{aligned} m(S_\lambda) \leq 2m(K) \leq 2m\left(\bigcup_{j=1}^M I_{k_j}\right) &\leq 2\sum_{j=1}^M m(I_{k_j}) \\ &\leq \sum_{j=1}^M \frac{2}{\lambda}\int_{I_{k_j}} |f(e^{i\psi})|d\psi \\ &\leq \frac{4}{\lambda}\|f\|_{L^1}. \qquad \square \end{aligned}$$

Definition 6.1.11. If $e^{i\theta} \in \partial D$, $1 < \alpha < \infty$, then define the *Stolz region* (or *nontangential approach region* or *cone*) with vertex $e^{i\theta}$ and aperture α to be

$$\Gamma_\alpha(e^{i\theta}) = \{z \in D : |z - e^{i\theta}| < \alpha(1 - |z|)\}.$$

See Figure 6.2.

Proposition 6.1.12. *If $e^{i\theta} \in \partial D, 1 < \alpha < \infty$, then there is a constant $C_\alpha > 0$ such that if $f \in L^1(\partial D)$, then*

$$\sup_{re^{i\phi} \in \Gamma_\alpha(e^{i\theta})} |P_r f(e^{i\phi})| \leq C_\alpha M f(e^{i\theta}).$$

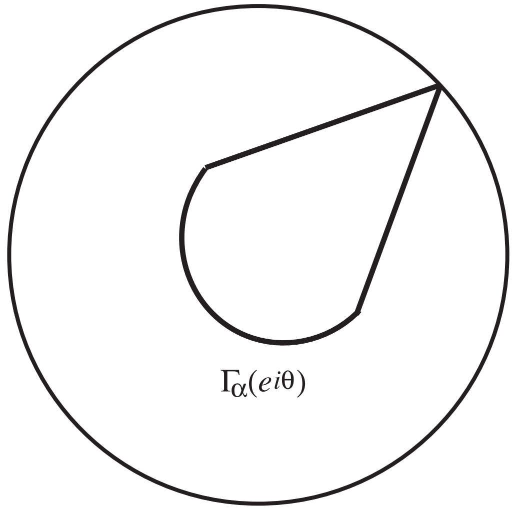

Fig. 6.2. A Stolz region.

Proof. For $re^{i\phi} \in \Gamma_\alpha(e^{i\theta})$, we have

$$|\theta - \phi| \leq 2\alpha(1-r).$$

Therefore, for $1/\alpha \leq r < 1$, we obtain

$$\begin{aligned}
|P_r f(e^{i\phi})| &= \left| \frac{1}{2\pi} \int_0^{2\pi} f(e^{i(\phi-\psi)}) \frac{1-r^2}{1-2r\cos\psi + r^2} d\psi \right| \\
&= \left| \frac{1}{2\pi} \int_0^{2\pi} f(e^{i(\phi-\psi)}) \frac{1-r^2}{(1-r)^2 + 2r(1-\cos\psi)} d\psi \right| \\
&\leq \frac{4}{2\pi} \sum_{j=0}^{\log_2(\pi/\alpha(1-r))} \int_{S_j} |f(e^{i(\phi-\psi)})| \frac{1-r^2}{(1-r)^2 + 2r(2^{j-1}\alpha(1-r))^2} d\psi \\
&\quad + \frac{1}{2\pi} \int_{|\psi|<\alpha(1-r)} |f(e^{i(\phi-\psi)})| \frac{1-r^2}{(1-r)^2} d\psi,
\end{aligned}$$

where $S_j = \{\psi : 2^j\alpha(1-r) \leq |\psi| < 2^{j+1}\alpha(1-r)\}$. Now this is

$$\begin{aligned}
&\leq \frac{4\alpha}{4\pi\alpha^2} \sum_{j=0}^{\infty} \frac{1}{2^{2j-2}(1-r)} \int_{|\psi|<(2+2^{j+1})\alpha(1-r)} |f(e^{i(\theta-\psi)})| d\psi \\
&\quad + \frac{2}{2\pi} \frac{1}{1-r} \int_{|\psi|<3\alpha(1-r)} |f(e^{i(\theta-\psi)})| d\psi \\
&\leq \frac{32}{\pi} \sum_{j=0}^{\infty} 2^{-j} \left[\frac{1}{2\alpha(2+2^{j+1})(1-r)} \int_{|\psi|<(2+2^{j+1})\alpha(1-r)} |f(e^{i(\theta-\psi)})| d\psi \right]
\end{aligned}$$

$$
\begin{aligned}
&+\frac{6\alpha}{\pi}\frac{1}{2\cdot 3\alpha(1-r)}\int_{|\psi|<3\alpha(1-r)}|f(e^{i(\theta-\psi)})|d\psi \\
&\le \frac{32}{\pi}\cdot\sum_{j=0}^{\infty}2^{-j}Mf(e^{i\theta})+\frac{6\alpha}{\pi}Mf(e^{i\theta}) \\
&\le \frac{64}{\pi}Mf(e^{i\theta})+\frac{6\alpha}{\pi}Mf(e^{i\theta}) \\
&\equiv C_\alpha Mf(e^{i\theta}) .
\end{aligned}
$$

If $0 < r \le 1/\alpha$, then

$$
|P_r f(\phi)| \le \frac{1}{2\pi}\int_0^{2\pi}|f(e^{i(\phi-\psi)})|(2\alpha/(\alpha-1))d\psi \le \frac{2\alpha}{\alpha-1}Mf(\theta). \quad \square
$$

Theorem 6.1.13. *Let $f \in \mathbf{h}^p(D)$ and $1 < p \le \infty$. Let f^* be as in Proposition 6.1.1 and $1 < \alpha < \infty$. Then*

$$
\lim_{\Gamma_\alpha(e^{i\theta})\ni z\to e^{i\theta}} f(z) = f^*(e^{i\theta}) \ , \quad a.e. \quad e^{i\theta}\in\partial D.
$$

Proof. It suffices to handle the case $p < \infty$ and f real-valued. If $\epsilon > 0$, then choose $g \in C(\partial D)$ real-valued such that $\|f^* - g\|_{L^p(\partial D)} < \epsilon^2$. We know by the theory of the Dirichlet problem that

$$
\lim_{\Gamma_\alpha(e^{i\theta})\ni z\to e^{i\theta}} g(z) = g(e^{i\theta}) \ , \qquad \text{all} \quad e^{i\theta}\in\partial D, \tag{6.1.1}
$$

where $g(re^{i\theta}) \equiv P_r g(e^{i\theta})$. Therefore

$$
\begin{aligned}
&m\{e^{i\theta} : \limsup_{\Gamma_\alpha(e^{i\theta})\ni z\to e^{i\theta}} |f(z)-f^*(e^{i\theta})| > \epsilon\} \\
&\le m\{e^{i\theta} : \limsup_{\Gamma_\alpha(e^{i\theta})\ni z\to e^{i\theta}} |f(z)-g(z)| > \epsilon/3\} \\
&\qquad + m\{e^{i\theta} : \limsup_{\Gamma_\alpha(e^{i\theta})\ni z\to e^{i\theta}} |g(z)-g(e^{i\theta})| > \epsilon/3\} \\
&\qquad + m\{e^{i\theta} : |g(e^{i\theta})-f^*(e^{i\theta})| > \epsilon/3\} \\
&\le \{e^{i\theta} : C_\alpha M(f^*-g) > \epsilon/3\} + 0 + (3\|g-f^*\|_{L^p}/\epsilon)^p .
\end{aligned}
$$

In the last estimate we used Lemma 6.1.8 and (6.1.1). Now the last line is majorized by

$$
C'_\alpha\|f^*-g\|_{L^1}/(\epsilon/3) + 3^p\epsilon^p \le C''_\alpha\|f^*-g\|_{L^p}/(\epsilon/3) + 3^p\epsilon^p \le C'''_\alpha\epsilon .
$$

It follows that

$$
\lim_{\Gamma_\alpha(e^{i\theta})\ni z\to e^{i\theta}} f(z) = f^*(e^{i\theta}) \ , \qquad a.e. \quad e^{i\theta}\in\partial D.
$$

The informal statement of Theorem 6.1.13 is that f has nontangential boundary limits almost everywhere.

Theorem 6.1.13 contains essentially all that can be said about the boundary behavior of harmonic functions. Why are holomorphic functions better? The classical way of answering this question is to use Blaschke factorization:

Definition 6.1.14. If $a \in \mathbb{C}$, $|a| < 1$, then the *Blaschke factor* at a is

$$B_a(z) = \frac{z - a}{1 - \bar{a}z}.$$

It is elementary to verify that B_a is holomorphic on a neighborhood of $\overline{D}$ and that $|B_a(e^{i\theta})| = 1$ for all θ.

Lemma 6.1.15. *If $0 < r < 1$ and f is holomorphic on a neighborhood of $\overline{D}(0,r)$, let $p_1, \ldots, p_k$ be the zeros of f (listed with multiplicity) in $D(0,r)$. Assume that $f(0) \neq 0$ and that $f(re^{it}) \neq 0$, all t. Then*

$$\log|f(0)| + \log \prod_{j=1}^{k} r|p_j|^{-1} = \frac{1}{2\pi} \int_0^{2\pi} \log|f(re^{it})| dt.$$

Proof. The function

$$F(z) = \frac{f(z)}{\prod_{j=1}^{k} B_{p_j/r}(z/r)}$$

is holomorphic on a neighborhood of $\overline{D}(0,r)$, hence $\log|F|$ is harmonic on a neighborhood of $\overline{D}(0,r)$. Thus

$$\log|F(0)| = \frac{1}{2\pi} \int_0^{2\pi} \log|F(re^{it})| dt = \frac{1}{2\pi} \int_0^{2\pi} \log|f(re^{it})| dt$$

or

$$\log|f(0)| + \log \prod_{j=1}^{k} \frac{r}{|p_j|} = \frac{1}{2\pi} \int_0^{2\pi} \log|f(re^{it})| dt.$$

Notice that, by the continuity of the integral, Lemma 6.1.15 holds even if f has zeros on $\{re^{it}\}$. The next result of Jensen is crucial.

Corollary 6.1.16. *If f is holomorphic in a neighborhood of $\overline{D}(0,r)$, then*

$$\log|f(0)| \leq \frac{1}{2\pi} \int_0^{2\pi} \log|f(re^{it})| dt.$$

Proof. The term $\log \prod_{j=1}^{k} |r/p_j|$ in Lemma 6.1.15 is positive. □

Corollary 6.1.17. *If f is holomorphic on $D, f(0) \neq 0$, and $\{p_1, p_2, \ldots\}$ are the zeros of f counting multiplicities, then*

$$\log|f(0)| + \log \prod_{j=1}^{\infty} \frac{1}{|p_j|} \leq \sup_{0<r<1} \frac{1}{2\pi} \int_0^{2\pi} \log^+ |f(re^{it})| dt.$$

Proof. Apply Lemma 6.1.15, letting $r \to 1^-$. □

Corollary 6.1.18. *If $f \in H^p(D), 0 < p \leq \infty$, and $\{p_1, p_2, \ldots\}$ are the zeros of f counting multiplicities, then $\sum_{j=1}^{\infty}(1 - |p_j|) < \infty$.*

Proof. Since f vanishes to finite order k at 0, we may replace f by $f(z)/z^k$ and assume that $f(0) \neq 0$. It follows from Corollary 6.1.17 that

$$\log \prod_{j=1}^{\infty} \left\{ \frac{1}{|p_j|} \right\} < \infty$$

or $\prod(1/|p_j|)$ converges; hence $\prod |p_j|$ converges. So $\sum_j (1 - |p_j) < \infty$. □

Proposition 6.1.19. *If $\{p_1, p_2, \ldots\} \subseteq D$ satisfy $\sum_j (1 - |p_j|) < \infty, p_j \neq 0$, for all j, then*

$$\prod_{j=1}^{\infty} \frac{-\overline{p}_j}{|p_j|} B_{p_j}(z)$$

converges normally on D.

Proof. Restrict attention to $|z| \leq r < 1$. Then the assertion that the infinite product converges uniformly on this disk is equivalent to the assertion that

$$\sum_j \left| 1 + \frac{\overline{p}_j}{|p_j|} B_{p_j}(z) \right|$$

converges uniformly. But

$$\begin{aligned} \left| 1 + \frac{\overline{p}_j}{|p_j|} B_{p_j}(z) \right| &= \left| \frac{|p_j| - |p_j|\overline{p}_j z + \overline{p}_j z - |p_j|^2}{|p_j|(1 - z\overline{p}_j)} \right| \\ &= \left| \frac{(|p_j| + z\overline{p}_j)(1 - |p_j|)}{|p_j|(1 - z\overline{p}_j)} \right| \\ &\leq \frac{(1 + r)(1 - |p_j|)}{1 - r}, \end{aligned}$$

so the convergence is uniform. □

Definition 6.1.20. Let $0 < p \leq \infty$ and $f \in H^p(D)$. Let $\{p_1, p_2, \dots\}$ be the zeros of f counted according to multiplicity. Let

$$B(z) = \prod_{j=1}^{\infty} \frac{-\overline{p}_j}{|p_j|} B_{p_j}(z)$$

(where each $p_j = 0$ is understood to give rise to a factor of z). Then B is a well-defined holomorphic function on D by Proposition 6.1.19. Let $F(z) = f(z)/B(z)$. By the Riemann removable singularities theorem, F is a well-defined, nonvanishing holomorphic function on D. The representation $f = F \cdot B$ is called the *canonical factorization* of f.

Proposition 6.1.21. *Let $f \in H^p(D), 0 < p \leq \infty$, and let $f = F \cdot B$ be its canonical factorization. Then $F \in H^p(D)$ and $\|F\|_{H^p(D)} = \|f\|_{H^p(D)}$.*

Proof. Trivially, $|F| = |f/B| \geq |f|$, so $\|F\|_{H^p} \geq \|f\|_{H^p}$. If $N = 1, 2, \dots$, let

$$B_N(z) = \prod_{j=1}^{N} \frac{-\overline{p}_j}{|p_j|} B_{p_j}(z)$$

(where the factors corresponding to $p_j = 0$ are just z).

Let $F_N = f/B_N$. Since $|B_N(e^{it})| = 1$, all t, it follows that $\|F_N\|_{H^p} = \|f\|_{H^p}$ (use Lemma 2.1.17 in [KRA4] and the fact that $B_N(re^{it}) \to B_N(e^{it})$ uniformly in t as $r \to 1^-$.) If $0 < r < 1$, then

$$\int_0^{2\pi} |F(re^{it})|^p dt^{1/p} = \lim_{N\to\infty} \int_0^{2\pi} |F_N(re^{it})|^p dt^{1/p} \leq \lim_{N\to\infty} \|F_N\|_{H^p} = \|f\|_{H^p} .$$

Therefore $\|F\|_{H^p} \leq \|f\|_{H^p}$. □

Corollary 6.1.22. *If $\{p_1, p_2, \dots\}$ is a sequence of points in D satisfying $\sum_j (1 - |p_j|) < \infty$ and if $B(z) = \prod_j (-\overline{p}_j/|p_j|) B_{p_j}(z)$ is the corresponding Blaschke product, then*

$$\lim_{r\to 1^-} B(re^{it})$$

exists and has modulus 1 almost everywhere.

Proof. The conclusion that the limit exists follows from Theorem 6.1.13 and the fact that $B \in H^\infty$. For the other assertion, note that the canonical factorization for B is $B = 1 \cdot B$. Therefore, by Proposition 6.1.21,

$$\left[\int |B^*(e^{it})|^2 dt\right]^{1/2} = \|B\|_{H^2} = \|1\|_{H^2} = 1\,;$$

hence $|B^*(e^{it})| = 1$ almost everywhere. □

Theorem 6.1.23. *If $f \in H^p(D), 0 < p \leq \infty,$ and $1 < \alpha < \infty,$ then*

$$\lim_{\Gamma_\alpha(D) \ni z \to e^{i\theta}} f(z)$$

exists for almost every $e^{i\theta} \in \partial D$ and equals $f^*(e^{i\theta})$. Also, $f^* \in L^p(\partial D)$ and

$$\|f^*\|_{L^p} = \|f\|_{H^p} \equiv \sup_{0<r<1} \int_0^{2\pi} |f(re^{i\theta})|^p d\theta^{1/p}.$$

Proof. By Definition 6.1.20, write $f = B \cdot F$, where F has no zeros and B is a Blaschke product. Then $F^{p/2}$ is a well-defined H^2 function and thus has the appropriate boundary values almost everywhere. A fortiori, F has nontangential boundary limits almost everywhere. Since $B \in H^\infty, B$ has nontangential boundary limits almost everywhere. It follows that f does as well. The final assertion follows from the corresponding fact for H^2 functions (exercise). □

6.2 The Lindelöf Principle

We begin these considerations by recalling a little-known classical result of Bagemihl and Seidel [BAS].

Theorem 6.2.1. *Let $E \subseteq \partial D$ be the countable union of closed, nowhere dense sets (note that E could be of full measure). Let ϕ be a continuous function on the open disk D. Then there is a holomorphic function f on D such that*

$$\lim_{r \to 1^-} |f(re^{i\theta}) - \phi(re^{i\theta})| = 0$$

for each point $e^{i\theta} \in E$.

If we apply this theorem with ϕ the continuous function given in line (6.0.1), then we obtain a holomorphic function f that fails to have a radial boundary limit on a set of full measure in ∂D. We stress that f *will not be bounded*—even though ϕ is. For, by Theorems 6.1.13 and 6.1.23, a bounded holomorphic function *will* have radial boundary limits almost everywhere.

In any event, we see that some extra care must be taken if we expect that the holomorphic functions we are studying will have boundary limits.

Much of modern harmonic analysis, as well as the function theory of the disk, is predicated on the interplay of Fourier analysis on the boundary and complex analysis on the interior that is facilitated by Theorem 6.1.23. It is a fundamental result.

Lindelöf's idea was to consider boundary limits point by point. If we take Theorem 6.1.13 as optimal—and it is in many respects—then there is no reason to suppose that there will be a boundary limit at any particular point; after all, the result is "almost everywhere." Thus a very elementary form of the Lindelöf principle takes the following form:

Proposition 6.2.2. *Let f be a bounded holomorphic function on the unit disk. Suppose that $e^{i\theta} \in \partial D$ and that*

$$\lim_{r \to 1^-} f(re^{i\theta}) = \ell\,.$$

Then, for any $\alpha > 1$, we have

$$\lim_{\Gamma_\alpha(e^{i\theta}) \ni z \to e^{i\theta}} f(z) = \ell\,.$$

Proof. Fix $\alpha > 1$. We may assume that $\ell = 0$ and $\theta = 0$. Define, for $j \in \mathbb{N}$,

$$\Omega_j = \{z \in \mathbb{C} : 1 - 2^{-j+1}/(2\alpha) \le \operatorname{Re} z \le 1 - 2^{-j}/(2\alpha), |\operatorname{Im}\, z| < 2^{-j+1}\}.$$

See Figure 6.3.

For $j_0(\alpha)$ sufficiently large,

$$D \supseteq \bigcup_{j=j_0}^{\infty} \Omega_j \supseteq \Gamma_\alpha(\mathbf{1}) \cap \{z \in D : \operatorname{Re} z \ge 1 - 2^{-j_0+1}/(2\alpha)\}.$$

Also, the map $\phi_j : \Omega_j \to \Omega_{j_0}$ given by $\phi_j(z) = 2^{j-j_0}(z-1)+1$ is biholomorphic. There is a K compact in Ω_{j_0} such that

$$\bigcup_{j \ge j_0} \phi_j^{-1}(K) \supseteq \Gamma_\alpha(\mathbf{1}) \cap \{z \in D : \operatorname{Re} z \ge 1 - 2^{-j_0-1}/(2\alpha)\}.$$

Now the functions $\{f \circ \phi_j^{-1}\}$ form a normal family on Ω_{j_0}. Choose a subsequence $f \circ \phi_{j_k}^{-1}$ that converges uniformly on K. Call the limit function f_0. By

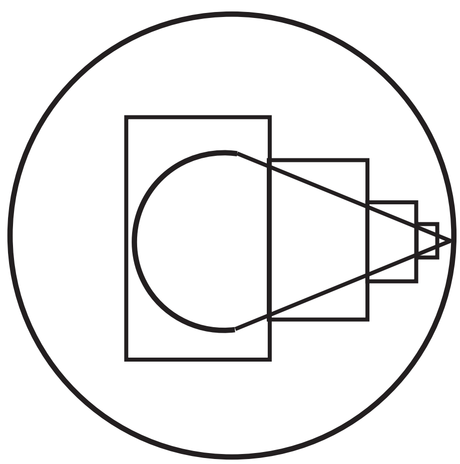

Fig. 6.3. The regions Ω_j.

hypothesis, $f_0(x+i0)=0$, all $x+i0 \in K$. Thus $f_0 \equiv 0$. Therefore $f \circ \phi_{j_k}^{-1} \to 0$ uniformly on K. In fact every subsequence of $f \circ \phi_j^{-1}$ has a subsequence with this property. Hence

$$\lim_{\Gamma_\alpha(1)\ni z \to 1} f(z) = 0. \qquad \square$$

In their seminal paper [LEV], Lehto and Virtanen extracted one of the key ideas from this last proof. To wit, they define a holomorphic (or meromorphic) function f on the unit disk D to be *normal* if, for any sequence φ_j of conformal self-maps of D, the family $\{f \circ \varphi_j\}$ is a normal family. For example, any bounded holomorphic function is certainly normal. Any holomorphic function that omits two values is certainly normal. Any univalent function is normal. The proof of the Lindelöf principle that we have already presented now admits of this generalization:

Theorem 6.2.3. *Let f be a normal holomorphic (or meromorphic) function on the unit disk. Suppose that $e^{i\theta} \in \partial D$ and that*

$$\lim_{r\to 1^-} f(re^{i\theta}) = \ell\,.$$

Then, for any $\alpha > 1$, we have

$$\lim_{\Gamma_\alpha(e^{i\theta})\ni z \to e^{i\theta}} f(z) = \ell\,.$$

It is a deep insight of Lehto and Virtanen [LEV] that the hypotheses of these two results may be considerably weakened. In fact we may assume that f has a limit along *any* curve—even a curve in the boundary. And the same conclusion follows. The key to getting a more general version of the Lindelöf principle is to use harmonic measure (see Chapter 9). We shall present some of these ideas now. A good reference for further reading is [GOL].

For convenience we shall formulate our new Lindelöf results on the upper halfplane. We leave it to the reader to translate the ideas over to the disk (by way of the Cayley transform). We begin with an estimate for harmonic measure.

Lemma 6.2.4. *Let f be holomorphic and bounded in a domain Ω with boundary curve(s) S. Write $S = S_1 \cup S_2$. Let $\omega(z, \Omega, S_1)$ be the solution of the Dirichlet problem on Ω, with boundary data identically 1 on S_1 and 0 elsewhere, evaluated at the point z. Let $\omega(z, \Omega, S_2)$ be defined similarly.*[1] *Assume that the limiting values of $|f|$ as z approaches S_j do not exceed M_j, $j = 1, 2$. Then, in Ω,*

$$\log|f(z)| \le \omega(z, \Omega, S_1)\log M_1 + \omega(z, \Omega, S_2)\log M_2\,. \tag{6.2.1}$$

If equality holds in (6.2.1) at a point $z \in \Omega$, then it holds at all points of Ω.

[1] We shall discuss this construction, in the more general context of harmonic measure, in Chapter 9.

The perceptive reader will notice that this is a version of the three lines theorem (see Section 9.4), rendered in the language of harmonic measure.

Proof of the Lemma. The function

$$u(z) = \log|f(z)| - \omega(z, \Omega, S_1)\log M_1 - \omega(z, \Omega, S_2)\log M_2$$

is harmonic on all of Ω *except* at the zeros of f. If $z \in \Omega$ approaches any boundary point of Ω except at the endpoints of the S_j, then the limiting values of u are nonpositive. But also note that $u(z)$ approaches $-\infty$ as z approaches any of the zeros of f. We conclude that $u \leq 0$ at *all* points of Ω.

This establishes inequality (6.2.1). The statement about equality is just the maximum principle. □

Remark 6.2.5. Of course the traditional statement of Hadamard's three lines theorem (Section 9.4) may be derived as a corollary of this lemma. We leave the details to the reader, or see [GOL, pp. 343–44].

Proposition 6.2.6. *Let f be holomorphic and bounded on the upper halfplane U and continuous up to the boundary halfline $\{z = x + i0 : x \in \mathbb{R}, x > 0\}$. Assume that*

$$\lim_{x \to 0^+} f(x + i0) = \alpha \in \mathbb{C}\,.$$

For any $\epsilon > 0$, let

$$\Gamma = \{z \in U : 0 \leq \arg z \leq \pi - \epsilon\}\,.$$

Then

$$\lim_{\Gamma \ni z \to 0} f(z) = \alpha\,.$$

Proof. We may certainly suppose that $\alpha = 0$ and that $|f| < 1$. Fix a small $\epsilon > 0$ and choose $r > 0$ such that $|f(z)| < \epsilon$ on the interval $\lambda_r \equiv \{x + i0 : 0 \leq x \leq r\}$. We apply the lemma to the half-disk $\mathcal{D}_r \equiv \{z \in U : |z| < r\}$. The result is

$$\log|f(z)| \leq \omega(z, \mathcal{D}_r, \lambda_r)\log\epsilon \tag{6.2.2}$$

for all $z \in \mathcal{D}_r$. We require a lower bound for $\omega(z, \mathcal{D}_r, \lambda_r)$, and we find it by calculating the quantity explicitly.

Observe that the function

$$h(z) = \frac{rz}{(r - z)^2}$$

maps the half-disk $\mathcal{D}_r$ into the halfplane U so that $0 \mapsto 0$ and the interval λ_r is mapped to the positive real axis $\mathbb{R}^+$. Notice that, for any $z \in \mathcal{D}_r$,

$$\omega(z, \mathcal{D}_r, \lambda_r) = \omega(h(z), U, \mathbb{R}^+)\,.$$

By the calculation in Example 9.3.1 below and the remark following, we know that

$$\omega(h(\zeta), U, \mathbb{R}^+) = \frac{\pi - \arg\zeta}{\pi}$$

for $0 < \arg\zeta < \pi$. As a result,

$$\omega(z, \lambda_r, \mathcal{D}_r) = 1 - \frac{1}{\pi}\arg\frac{rz}{(r-z)^2} = 1 - \frac{1}{\pi}\arg\left[\frac{z}{r}\cdot\frac{1}{(1-\frac{z}{r})^2}\right].$$

We conclude that all limiting values of the quantity $\omega(rz, \lambda_r, B_r)$ are not less than $1 - (\pi - \epsilon)/\pi = \epsilon/\pi$ as $z \to 0$ in the angle Γ. As a result, there is a number $r' \in (0, r)$ such that $\omega(z, \lambda_r, \mathcal{D}_r) \geq \epsilon/2\pi \equiv \delta$ in the domain $\widetilde{\mathcal{D}} = \{z \in \mathbb{C} : |z| < r', 0 \leq \arg z \leq \pi - \epsilon\}$. Thus, by (6.2.2), $|f(z)| \leq \epsilon^\delta$. Observe that δ does not depend on either r or ϵ. So $f(z) \to 0$ uniformly, as $z \to 0$, in the angle Γ. That ends the proof. □

The last proposition is very much in the spirit of the Lindelöf principle; what is new is that the hypothesized limit along a curve is for a curve that *lies in the boundary*. Our goal in generalizing the Lindelöf principle is to obtain a result where the hypothesized limit is along a fairly arbitrary curve in the interior. Our next result is of that type.

We first formulate a general concept about harmonic measure, called *the extension principle*, that is useful in many contexts.

Lemma 6.2.7. *Let $\Omega \subseteq \mathbb{C}$ be a domain whose boundary consists of finitely many Jordan arcs. Write $\partial\Omega = A \cup B$. If Ω is "extended" to a larger domain Ω' in such a way that $\partial\Omega'$ differs from Ω only in that the arc B has been moved to a new arc B', then $\omega(z, \Omega', A) > \omega(z, \Omega, A)$. If instead the domain Ω is "extended" to a larger domain Ω' in such a way that $\partial\Omega'$ differs from $\partial\Omega$ only in that the arc A has been moved to a new arc A', then $\omega(z, \Omega', A') < \omega(z, \Omega, A)$. See Figure 6.4 for an illustration of these ideas.*

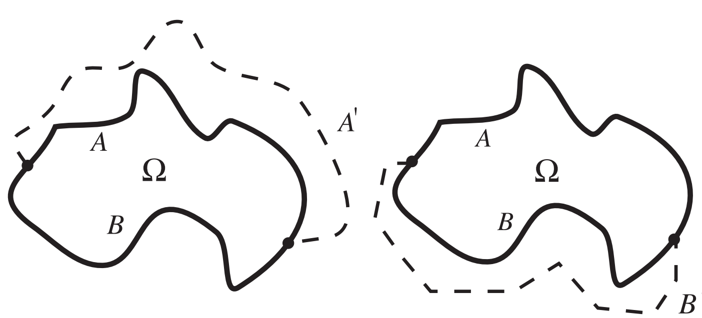

Fig. 6.4. Extension of the domain Ω.

Proof. Of course this is just the maximum principle. Specifically, in the first circumstance, the function

$$u(z) = \omega(z, \Omega', B) - \omega(z, \Omega, B)$$

is harmonic and bounded in Ω. Furthermore, u is equal to 0 on A and is nonnegative on B. It follows that $u \geq 0$ on all of Ω, and that establishes our first assertion.

The second assertion follows from the simple identity $\omega(z, \Omega, A) = 1 - \omega(z, \Omega, B)$. □

Theorem 6.2.8. *Let f be holomorphic and bounded on the upper halfplane U. Let $\gamma : [0, 1] \to U$ be a Jordan curve in U that terminates at $z = 0$ (i.e., $\gamma(1) = 0$). Assume that*

$$\lim_{t \to 1} f(\gamma(t)) = \alpha \in \mathbb{C}\,.$$

Let $\epsilon > 0$ and define $\Lambda_\epsilon = \{z \in U : \epsilon \leq \arg z \leq \pi - \epsilon\}$. Then

$$\lim_{\Lambda_\epsilon \ni z \to 0} f(z) = \alpha\,.$$

Proof. We shall continue to use notation and normalizations from the previous proposition.

So we assume that $\alpha = 0$ and $|f| < 1$. Fix $0 < \epsilon < \pi/2$ small. Choose $r > 0$ so that $|f(z)| \leq \epsilon$ on $\gamma \cap \{z \in \mathbb{C} : |z| < r, \operatorname{Im} z > 0\}$. Set $\mathcal{D}_r = \{z \in U : |z| < r\}$. Let γ_r denote that portion of the curve that extends from the origin to the first point at which the curve meets the circle $|z| = r$. Of course the curve γ_r partitions $\mathcal{D}_r$ into two pieces, $\mathcal{D}_r'$ and $\mathcal{D}_r''$. One of these two pieces is adjacent to the negative real axis; say it is the first piece. See Figure 6.5.

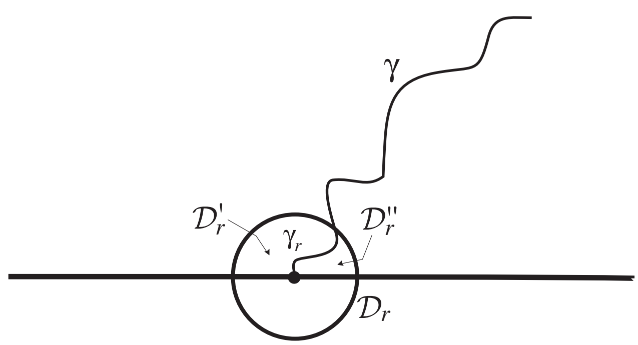

Fig. 6.5. The curve γ_r partitions $\mathcal{D}_r$.

Applying inequality (6.2.1) from our first lemma to the domain $\mathcal{D}_r'$, we see that

$$\log |f(z)| \le \omega(z, \mathcal{D}'_r, \gamma_r) \cdot \log \epsilon \,.$$

Now we must find a lower bound for $\omega(z, \mathcal{D}'_r, \gamma_r)$.

Let $\beta_r = \partial \mathcal{D}'_r \setminus \gamma_r$. And set $\gamma'_r = \partial \mathcal{D}_r \setminus \beta_r$. Finally, let $\lambda_r = \{x + i0 : 0 < x < r\}$. Now the extension principle tells us that

$$\omega(z, \mathcal{D}'_r, \gamma_r) \ge \omega(z, \mathcal{D}_r, \gamma'_r) \ge \omega(z, \mathcal{D}_r, \lambda_r) \,.$$

The proof of the preceding proposition tells us that the quantity $\omega(z, \mathcal{D}_r, \lambda_r)$ has a positive lower bound η_1 in a region of the form $T_r \equiv \{z \in \mathbb{C} : |z| < r', 0 \le \arg z \le \pi - \epsilon\}$ for sufficiently small $r' \in (0, r)$.

As before, $|f(z)| \le \epsilon^{\eta_1}$ in the part of $\mathcal{D}'_r$ lying in the indicated sector. Applying the same argument again, we can show that there are numbers $\eta_2 > 0$ and $r'' \in (0, r)$ such that $|f(z)| < \epsilon^{\eta_2}$ in that part of $\mathcal{D}''_r$ lying in the sector $S_r = \{z \in \mathbb{C} : |z| < r'', \epsilon \le \arg z \le \pi\}$. Let $\eta = \max(\eta_1, \eta_2)$ and let $r_0 = \min(r', r'')$. Then $|f(z)| \le \epsilon^{\eta}$ for $|z| < r_0$ and $z \in \Lambda_\epsilon$.

This proves the uniform convergence of $f(z)$ to 0 as $z \to 0$ in the sector Λ_ϵ. □

Problems for Study and Exploration

1. Review the Bagemihl/Seidel theorem from Section 6.2. Use it to produce a holomorphic function f on the disk such that $\lim_{r \to 1^-} |f(re^{i\theta})| = \infty$ for almost every $\theta \in [0, 2\pi)$. Now produce a nonzero holomorphic function h such that $\lim_{r \to 1^-} |h(re^{i\theta})| = 0$ for almost every $\theta \in [0, 2\pi)$.

2. The following statement is *not* true: If ϕ is a continuous function on ∂D then there is a bounded, holomorphic function on D whose boundary limits agree with ϕ almost everywhere. Prove the failure of this assertion. However, this statement *is* true: If ϕ is a *positive*, continuous function on ∂D, then there is a bounded, holomorphic function f on D such that $|f|$ has boundary limits that agree with ϕ almost everywhere. Prove it. [*Hint:* Your argument will involve a logarithm.]

3. Define what it means for a curve $\gamma : [0, 1) \to D$ to approach the boundary of D *tangentially*. Show that there is a bounded holomorphic function f on the disk D such that $\lim_{t \to 1^-} f(\gamma(t) \cdot e^{i\theta})$ will fail to agree with the radial boundary function f^* for θ in a set of positive measure. [*Hint:* A Blaschke product may help.]

4. Let f be a holomorphic function on the disk. We say that f is in the *Nevanlinna class* if

$$\sup_{0<r<1} \int_0^{2\pi} \log^+ |f(re^{i\theta})| \, d\theta < \infty \,.$$

Here, if x is a positive real number, then

$$\log^+ x = \begin{cases} 0 & \text{if} \quad x \le 1 \\ \log x & \text{if} \quad x > 1 \,. \end{cases}$$

Prove that a function in the Nevanlinna class has radial boundary limits almost everywhere.

5. Refer to Exercise 4 for terminology. Prove that if f is in the Nevanlinna class and if $\{p_j\}$ are the zeros of f, then

$$\sum_{j=1}^{\infty} 1 - |p_j| < \infty .$$

6. There is a natural sense in which the space $H^\infty(D)$ is the "limit" of the spaces $H^p(D)$ as $p \to \infty$. Formulate and prove a statement to this effect.

7. If $f \in H^p(D)$ for some $0 < p \le \infty$ and if the radial boundary limit function f^* vanishes on a set of positive measure, then f must be identically 0. Prove this assertion.

8. Prove that every $f \in H^p(D)$ may be written as the sum of two nonvanishing elements of $H^p(D)$ with comparable norm.

9. Carathéodory's theorem states that if a holomorphic function f on D, the disc, has nontangential limit 0 on a set $E \subseteq \partial D$ of positive linear measure, then $f \equiv 0$ on D.

(a) If h is *any* holomorphic function on D and if $f - h$ has nontangential limit 0 on a set $E \subseteq \partial D$ of positive linear measure, then what can you conclude about f (relative to h)?

(b) If f has limit 0 along some nontangential curve at each point of a set $E \subseteq \partial D$ of positive linear measure, then what can you conclude about f?

10. Let $E \subseteq \partial D$ be an F_σ of first category and ϕ a continuous function on D. F. Bagemihl and W. Seidel [KOO] proved that there exists a holomorphic function f on D such that $\lim_{r \to 1^-} |\phi(r\zeta) - f(r\zeta)| = 0$ for all $\zeta \in E$. Give an example to show that the Bagemihl–Seidel theorem cannot hold if "radial convergence" is replaced by nontangential convergence. Conversely, give an example to show that Carathéodory's result of Exercise 9 cannot hold if nontangential convergence is replaced by radial convergence (feel free to use Bagemihl–Seidel here).

11. Prove the following classical result of Hardy and Littlewood (see Goluzin [GOL, p. 411]).

Theorem: Let $f : D \to \mathbb{C}$ be holomorphic and bounded. Let $0 < \alpha < \infty$. Then $f \in \Lambda_\alpha(D)$ if and only if there is an $\mathbb{N} \ni k > \alpha$ and a $C_k > 0$ such that

$$|f^{(k)}(z)| \le C_k(1 - |z|)^{\alpha - k} , \qquad \text{all} \quad z \in D. \qquad (*)$$

Prove that $(*)$ holds for one $k > \alpha$ if and only if it holds for all $k > \alpha$.

12. This problem is philosophically related to the preceding one. The classical result, once again, is due to G. H. Hardy and J. E. Littlewood [GOL].

Theorem: Let $f : D \to \mathbb{C}$ be holomorphic. Let $0 < p < 1$. If $f' \in H^p(D)$, then $f \in H^q(D)$, where $q = p/(1-p)$. If $f' \in H^1$, then $f \in A(D)$; indeed, $f|_{\partial D}$ is absolutely continuous. [Here $A(D)$ is the continuous functions on $\overline{D}$ that are holomorphic on D.]

It is even possible to prove results for $p > 1$: if $f' \in H^p, 1 < p \leq \infty$, then $f \in \Lambda_\alpha(D)$, where $\alpha = 1 - 1/p$. Although the reader should have no difficulty verifying the assertion for $p = \infty$, he may encounter trouble with $0 < p < \infty$. Nevertheless, for $1 < p < \infty$, the problem is accessible if we use some Sobolev space ideas:

Identify f with its boundary function f^*. Prove that the hypotheses imply that $(d/d\theta)f^*(e^{i\theta}) \in L^p([0, 2\pi])$ in the sense of distributions. Prove that, for $p > 1$, this implies that $f^* \in \Lambda_{1-1/p}(\partial D)$. Hence $f \in \Lambda_{1-1/p}(\bar{D})$. When $p = 1$, matters are even simpler.

13. Construct an example of a bounded analytic function f on the disk, f not identically zero, such that for almost every P in the boundary of the disk there is a tangential sequence $\{z_j\}$ in D approaching P along which f tends to 0. Here a sequence in D is "tangential" if it escapes every nontangential approach region. Consider variants of this result: can you replace the sequence $\{z_j\}$ by a curve at each point? If you replace "bounded" by H^p, then can you strengthen the example? A good reference for this sort of result is I. Priwalow [PRI] and Collingwood and Lohwater [COL].

14. In Proposition 6.1.19, where do the factors $-\overline{p}_j/|p_j|$ come from?

Part II

Real and Harmonic Analysis

7

The Cauchy–Riemann Equations

Genesis and Development

Certainly every student of complex analysis learns of the Cauchy–Riemann equations

$$\frac{\partial u}{\partial x} = \frac{\partial v}{\partial y} \,,$$

$$\frac{\partial u}{\partial y} = -\frac{\partial v}{\partial x} \,.$$

These identities, which follow directly from the definition of complex derivative, give an important connection between the real and complex parts of a holomorphic function. Certainly conformality, harmonicity, and many other fundamental ideas are effectively explored by way of the Cauchy–Riemann equations.

Less familiar to most students, and an idea that has come into its own only in the past forty years, is the idea of studying the *inhomogeneous Cauchy–Riemann equations*. The key is to introduce the operators

$$\frac{\partial}{\partial z} = \frac{1}{2}\left(\frac{\partial}{\partial x} - i\frac{\partial}{\partial y}\right) \qquad \text{and} \qquad \frac{\partial}{\partial \overline{z}} = \frac{1}{2}\left(\frac{\partial}{\partial x} + i\frac{\partial}{\partial y}\right) \,.$$

One checks that a continuously differentiable function h is holomorphic on a domain Ω if and only if

$$\frac{\partial h}{\partial \overline{z}} \equiv 0 \quad \text{on } \Omega \,.$$

By means of this language, one is led to consider the *inhomogeneous* partial differential equation

$$\frac{\partial u}{\partial \overline{z}} = f \,. \tag{7.0.1}$$

There are explicit formulas for solving the equation (7.0.1), and they turn out to be effective tools for constructing holomorphic functions and creating analytic objects. In the present chapter we explore this circle of ideas. Later in the book, we shall see equation (7.0.1) applied in Tom Wolff's dramatic solution of the corona problem.

A good understanding of multivariable calculus is all that is required for an appreciation and understanding of the ideas in this chapter.

7.1 Introduction

Every basic course in complex variables treats the Cauchy–Riemann equations. For a continuously differentiable function $f = u + iv$, these are the equations

$$\frac{\partial u}{\partial x} = \frac{\partial v}{\partial y}, \qquad \frac{\partial u}{\partial y} = -\frac{\partial v}{\partial x}\,.$$

The basic result in the subject is that if $f = u+iv$ is continuously differentiable on the domain Ω, then f is holomorphic if and only if it satisfies these partial differential equations.

It is sometimes useful to introduce the notation

$$\frac{\partial}{\partial z} = \frac{1}{2}\left(\frac{\partial}{\partial x} - i\frac{\partial}{\partial y}\right)\;, \quad \frac{\partial}{\partial \overline{z}} = \frac{1}{2}\left(\frac{\partial}{\partial x} + i\frac{\partial}{\partial y}\right)\,.$$

It is a simple matter to see that a continuously differentiable $f = u + iv$ on Ω is holomorphic if and only if $\partial f/\partial \overline{z} \equiv 0$ on Ω. To wit,

$$\begin{aligned} 0 &= \frac{\partial f}{\partial \overline{z}} \\ &= \frac{1}{2}\left(\frac{\partial}{\partial x} + i\frac{\partial}{\partial y}\right)(u+iv) \\ &= \frac{1}{2}\left(\frac{\partial u}{\partial x} - \frac{\partial v}{\partial y}\right) + \frac{i}{2}\left(\frac{\partial u}{\partial y} + \frac{\partial v}{\partial x}\right)\,. \end{aligned}$$

Of course this identity can hold if and only if the real and imaginary parts vanish identically. The Cauchy–Riemann equations result.

Less familiar to most students is the fact that there is considerable utility in studying the *inhomogeneous equation*

$$\frac{\partial u}{\partial \overline{z}} = f(z)\,, \tag{7.1.1}$$

for a given function f. The equation has considerable intrinsic interest, but it also can be useful for constructing holomorphic functions.

7.2 Solution of the Inhomogeneous Cauchy–Riemann Equations

We begin by recalling a version of Stokes's theorem expressed in complex notation (this idea was previewed in the proof of Proposition 1.3.2). We have already defined $\partial/\partial z$ and $\partial/\partial\overline{z}$. We also define the companion notations $dz = dx + idy$ and $d\overline{z} = dx - idy$. Then we have this standard result from calculus:

Theorem 7.2.1. *Let Ω be a bounded domain with C^1 boundary. Let $\omega = \alpha(z)dz$ be a 1-form with coefficient that is continuously differentiable on $\overline{\Omega}$. Then*

$$\int_{\partial\Omega} \omega = \iint\limits_{\Omega} \frac{\partial\alpha}{\partial\overline{z}} d\overline{z} \wedge dz \,. \tag{7.2.1}$$

For a general 1-form $\lambda = a(z)dz + b(z)d\overline{z}$, it is useful to write

$$\begin{aligned} d\lambda &= \partial\lambda + \overline{\partial}\lambda \\ &= \left[\frac{\partial a}{\partial z}\, dz \wedge dz + \frac{\partial b}{\partial z}\, dz \wedge d\overline{z}\right] + \left[\frac{\partial a}{\partial \overline{z}}\, d\overline{z} \wedge dz + \frac{\partial b}{\partial \overline{z}}\, d\overline{z} \wedge d\overline{z}\right] \\ &= \frac{\partial a}{\partial \overline{z}}\, d\overline{z} \wedge dz + \frac{\partial b}{\partial z}\, dz \wedge d\overline{z} \,. \end{aligned}$$

With this notation, formula (7.2.1) in the theorem becomes

$$\int_{\partial\Omega} \omega = \iint\limits_{\Omega} \overline{\partial}\omega \,.$$

Now we prove a generalized version of the Cauchy integral formula. Note that it is valid for essentially *all* functions—not just holomorphic functions.

Corollary 7.2.2. *If $\Omega \subseteq \mathbb{C}$ is a bounded domain with C^1 boundary and if $f \in C^1(\overline{\Omega})$, then, for any $z \in \Omega$,*

$$f(z) = \frac{1}{2\pi i} \int_{\partial\Omega} \frac{f(\zeta)}{\zeta - z} d\zeta - \frac{1}{2\pi i} \iint\limits_{\Omega} \frac{\big(\partial f(\zeta)/\partial\overline{\zeta}\big)}{\zeta - z} d\overline{\zeta} \wedge d\zeta.$$

Proof. Fix $z \in \Omega$ and choose $\epsilon > 0$ so that $\overline{D}(z,\epsilon) \subseteq \Omega$. Set $\Omega_\epsilon = \Omega \setminus \overline{D}(z,\epsilon)$. We apply Stokes's theorem to the form

$$\omega(\zeta) = \frac{f(\zeta)d\zeta}{\zeta - z}$$

on the domain Ω_ϵ. Note here that ω has no singularity on $\overline{\Omega_\epsilon}$.

Thus

$$\int_{\partial\Omega_\epsilon} \omega(\zeta) = \iint\limits_{\Omega_\epsilon} \overline{\partial}\omega\,.$$

Writing this all out gives

$$= \iint\limits_{\Omega_\epsilon} \frac{\partial f(\zeta)/\partial\overline{\zeta}}{\zeta - z}\, d\overline{\zeta} \wedge d\zeta\,. \tag{7.2.2}$$

Observe that we have reversed the orientation on the second integral on the left because the disk is *inside* the region Ω.

Now, as $\epsilon \to 0^+$ the integral on the right tends to

$$\iint\limits_{\Omega} \frac{\partial f(\zeta)/\partial\overline{\zeta}}{\zeta - z}\, d\overline{\zeta} \wedge d\zeta\,.$$

The first integral on the left does not depend on ϵ. The second integral on the left requires a little analysis:

$$\int_{\partial D(z,\epsilon)} \frac{f(\zeta)}{\zeta - z} d\zeta = \int_0^{2\pi} \frac{f(z+\epsilon e^{it})}{\epsilon e^{it}} \cdot i\epsilon e^{it}\, dt = i \int_0^{2\pi} f(z+\epsilon e^{it})\, dt\,.$$

Now the last expression tends, as $\epsilon \to 0^+$, to $2\pi i f(z)$.

Putting all this information into equation (7.2.2) yields

$$f(z) = \frac{1}{2\pi i} \int_{\partial\Omega} \frac{f(\zeta)}{\zeta - z}\, d\zeta - \frac{1}{2\pi i} \iint\limits_{\Omega} \frac{\partial f/\partial\overline{\zeta}}{\zeta - z}\, d\overline{z} \wedge dz$$

as desired. □

Corollary 7.2.3. *With hypotheses as in Corollary 7.2.2, and the additional assumption that* $\overline{\partial} f = 0$ *on* Ω, *we have*

$$f(z) = \frac{1}{2\pi i} \int_{\partial\Omega} \frac{f(\zeta)}{\zeta - z} d\zeta.$$

Remark 7.2.4. Corollary 7.2 is the familiar Cauchy integral formula from analysis of one complex variable.

Now we derive an explicit integral solution formula for the inhomogeneous Cauchy–Riemann equations. We will see that the holomorphicity of the one-dimensional Cauchy kernel will play an important role in the derivation of this formula. First we give a careful definition of the support of a form:

Definition 7.2.5. If λ is a differential form on $\mathbb{R}^2$ or $\mathbb{C}$, then the *support* of λ, written $\operatorname{supp}\lambda$, is the complement of the union of all open sets on which all the coefficients of λ vanish identically.

Theorem 7.2.6. *Let $\psi \in C^1_c(\mathbb{C})$, the continuously differentiable functions with compact support. The function defined by*

$$u(\zeta) = -\frac{1}{2\pi i}\int \frac{\psi(\xi)}{\xi - \zeta} d\bar\xi \wedge d\xi = -\frac{1}{\pi}\int \frac{\psi(\xi)}{\xi-\zeta} dA(\xi) \tag{7.2.3}$$

satisfies

$$\overline{\partial} u(\zeta) = \frac{\partial u}{\partial \bar\zeta}(\zeta) d\bar\zeta = \psi(\zeta)\, d\bar\zeta. \tag{7.2.4}$$

Proof. Let $D(0,R)$ be a large disk that contains the support of ψ. Then

$$\begin{aligned}
\frac{\partial u}{\partial \bar\zeta}(\zeta) &= -\frac{1}{2\pi i}\frac{\partial}{\partial\bar\zeta}\int_{\mathbb{C}} \frac{\psi(\xi)}{\xi-\zeta} d\bar\xi \wedge d\xi \\
&= -\frac{1}{2\pi i}\frac{\partial}{\partial\bar\zeta}\int_{\mathbb{C}} \frac{\psi(\xi+\zeta)}{\xi} d\bar\xi \wedge d\xi \\
&= -\frac{1}{2\pi i}\int_{\mathbb{C}} \frac{\frac{\partial\psi}{\partial\bar\xi}(\xi+\zeta)}{\xi} d\bar\xi \wedge d\xi \\
&= -\frac{1}{2\pi i}\int_{D(0,R)} \frac{\frac{\partial\psi}{\partial\bar\xi}(\xi)}{\xi-\zeta} d\bar\xi \wedge d\xi .
\end{aligned}$$

By Corollary 7.2.2, this last equals

$$\psi(\zeta) - \frac{1}{2\pi i}\int_{\partial D(0,R)} \frac{\psi(\xi)}{\xi-\zeta} d\xi = \psi(\zeta) - 0 = \psi(\zeta).$$

Here we have used the support condition on ψ. This is the result that we wish to prove. □

7.3 Development and Application of the $\bar\partial$ Equation

We begin by proving a refined regularity statement for the equation

$$\frac{\partial}{\partial \bar z} u = \psi\, . \tag{7.3.1}$$

Proposition 7.3.1. *Let Ω be a bounded domain and let $k \in \{0, 1, 2, \dots\}$. If $\psi \in C^k_c(\Omega)$, then the solution u of equation* (7.3.1) *that we constructed in Theorem 7.2.6 satisfies $u \in C^k(\Omega)$.*

Proof. Let us begin with $k = 1$. So we assume that $\psi \in C^1_c(\Omega)$, which simply means that the form ψ has C^1 coefficients with compact support in Ω. Now we calculate that

$$\begin{aligned}\frac{\partial}{\partial x}u(z) &= \frac{\partial}{\partial x}\left[-\frac{1}{\pi}\int\frac{\psi(\zeta)}{\zeta - z}dA(\zeta)\right]\\ &= \frac{\partial}{\partial x}\left[-\frac{1}{\pi}\int\frac{\psi(\zeta - z)}{\zeta}dA(\zeta)\right]\\ &= \left[\frac{1}{\pi}\int\frac{\frac{\partial\psi}{\partial x}(\zeta - z)}{\zeta}dA(\zeta)\right].\end{aligned}$$

Now the function $\partial\psi/\partial x$ lies in $C_c(\Omega)$ by hypothesis. And the function $1/\zeta$ is integrable, as a simple application of polar coordinates shows. Thus the convolution in the last expression defines a *continuous* function of z. So we have verified that $\partial u/\partial x \in C_c(\Omega)$. A similar argument shows that $\partial u/\partial y \in C_c(\Omega)$. Thus $u \in C^1(\Omega)$, as was required.

The result for higher k follows by a simple induction. The proof is therefore complete. □

Remark 7.3.2. In fact something stronger is true: If $\psi \in C^k$ then $u \in C^{k+1-\epsilon}$ for any $\epsilon > 0$. One can make an even sharper statement if one uses precise function spaces that have been designed for this purpose. More information on these matters may be found in [KRA2].

Corollary 7.3.3. *Let Ω be a bounded domain. If $\psi \in C_c^\infty(\Omega)$, then the solution u of equation* (7.3.1) *that we constructed in Theorem 7.2.6 satisfies $u \in C^\infty(\Omega)$.*

Corollary 7.3.4. *Consider the partial differential equation*

$$\frac{\partial}{\partial\bar{z}}u = \psi \qquad (7.3.2)$$

on the bounded domain Ω. Assume that $\psi \in C_c^k(\Omega)$ for some $k \in \{0, 1, 2, \ldots\}$. Then any solution v of the equation (7.3.2) *will lie in $C^k(\Omega)$.*

Proof. Let u be the solution constructed in Theorem 7.2.6. We proved in Proposition 7.3.1 that $u \in C^k(\Omega)$. But notice that

$$\frac{\partial}{\partial\bar{z}}[u(z) - v(z)] = \psi(z) - \psi(z) \equiv 0\,.$$

Thus $v(z) = u(z) + h(z)$, where h is holomorphic on Ω. It follows, therefore, that $v \in C^k(\Omega)$. □

Now an important point must be made. Thus far, we have studied the solution of the $\bar\partial$ equation (7.3.1) only when the data function ψ is *compactly supported* in Ω. In practice, when we actually apply the $\bar\partial$ equation to problems in function theory, this sort of result is not adequate. In fact we need to know how to solve the equation (7.3.1) when the data function ψ lies in $C^k(\Omega)$, *but ψ is not assumed to have compact support in Ω.* The next theorem addresses this need:

Theorem 7.3.5. *Let Ω be a domain in $\mathbb{C}$ and let $k \in \{0, 1, 2, \ldots\}$. If $\psi \in C^k(\Omega)$, then there exists a C^k function u on Ω such that $\partial u/\partial \overline{z} = \psi$.*

Proof. Begin by writing $\Omega = \cup_j K_j$, with each $K_j \subset \overset{\circ}{K}_{j+1}$ and each K_j compact and simply connected and the closure of its interior for $j = 1, 2, \ldots$. Now let $\eta_j \in C_c^\infty(\Omega)$ be such that $\operatorname{supp} \eta_j \subseteq K_{j+1}$ and $\eta_j \equiv 1$ on K_j.

Now set

$$\psi = \sum_j \psi_j \equiv \eta_1 \psi + \sum_{j=2}^{\infty} (\eta_j - \eta_{j-1}) \psi \, .$$

Then of course each ψ_j lies in $C_c^k(\Omega)$. Hence there is a function $u_j \in C^k(\Omega)$ such that $\partial u_j/\partial \overline{z} = \psi_j$. Notice that each u_j is holomorphic on $\overset{\circ}{K}_{j-1}$ for $j = 2, 3, \ldots$.

Now we apply Runge's theorem (see [GRK1] or [RUD2]) to the function u_j with respect to $K_{j-1} \subseteq \Omega$ for $j = 2, 3, \ldots$. In this way we find a holomorphic v_j on Ω such that $\sup_{K_{j-1}} |u_j - v_j| < 2^{-j}$. Set $u = \sum_\ell (u_\ell - v_\ell)$. Then it is easy to see that the series converges uniformly on each K_j. Moreover, it is obvious that $\partial u/\partial \overline{z} = \psi$.

Finally observe that, because the supremum norm of a holomorphic function on an open set dominates the C^k norm on a compact subset (use the Cauchy estimates), it follows that $u \in C^k(\Omega)$. That completes our proof. □

Corollary 7.3.6. *Let Ω be a domain in $\mathbb{C}$. If $\psi \in C^\infty(\Omega)$, then there exists a C^∞ function u on Ω such that $\partial u/\partial \overline{z} = \psi$.*

Proof. Obvious by inspection of the proof of the theorem. □

Now we may build on everything we have done to give a new proof of the Mittag–Leffler theorem of classical function theory using the $\bar{\partial}$ problem.

Theorem 7.3.7 (Mittag–Leffler). *Let $\Omega \subseteq \mathbb{C}$ be a domain and let $\{p_j\}_{j=1}^\infty \subseteq \Omega$ be a discrete set (i.e., with no interior accumulation point). For each j, let f_j be a function that is meromorphic in a small neighborhood of p_j (which neighborhood contains none of the other p_ℓ's) with a single pole at p_j. Then there exists a meromorphic function f on Ω such that f is analytic on $\Omega \setminus \{p_j\}_{j=1}^\infty$ and such that $f - f_j$ is analytic near p_j for each j.*

Proof. For each j, let U_j be a small open disk about p_j that contains none of the other p_ℓ's. Set $U_0 \equiv \Omega \setminus \{p_j\}_{j=1}^\infty$. Then the open sets $\{U_j\}_{j=0}^\infty$ cover Ω. Set $f_0 \equiv 1$. Let φ_ℓ be a C^∞ partition of unity subordinate to this covering. Let us say that φ_ℓ is supported on $U_{j(\ell)}$.

For each $k = 0, 1, 2, \ldots$, set

$$h_k \equiv \sum_{\ell=0}^{\infty} \varphi_\ell \cdot \left(f_k - f_{j(\ell)}\right) .$$

Observe that

$$\psi(z) \equiv \frac{\partial h_k}{\partial \overline{z}}(z) \quad \text{for } z \in U_k \tag{7.3.3}$$

is a well-defined C^∞ function on all of Ω. This is so because

$$\frac{\partial h_k}{\partial \overline{z}} = \sum_{\ell=0}^{\infty} \frac{\partial \varphi_\ell}{\partial \overline{z}} \cdot (f_k - f_{j(\ell)})$$

and

$$\sum_{\ell=0}^{\infty} \frac{\partial \varphi_\ell}{\partial \overline{z}} \equiv 0\,.$$

Now, using Corollary 7.3.6, we let $u \in C^\infty(\Omega)$ be such that $\partial u/\partial \overline{z} = \psi$ on Ω.

Finally, we set

$$f(z) = f_k(z) - h_k(z) + u(z) \qquad \text{for } z \in \Omega_k\,.$$

The reader may check that this is a well-defined meromorphic function on Ω that satisfies all our requirements. □

Problems for Study and Exploration

1. Write an explicit solution for the equation

$$\frac{\partial u}{\partial \overline{z}} = z\,.$$

What is the most general solution of this equation?

2. Let φ be a C^1 function of compact support such that $\iint \varphi\, dx\, dy \neq 0$. Show that the equation

$$\frac{\partial u}{\partial \overline{z}} = \varphi$$

has no solution with compact support.

3. Find the solution of the equation

$$\frac{\partial u}{\partial \overline{z}} = \overline{z}$$

that is orthogonal on the disk to all L^2 holomorphic functions in the sense that

$$\iint_D [\overline{z}] \cdot \overline{[h(z)]}\, dA(z) = 0$$

for all $h \in L^2(D)$ holomorphic. Is this solution unique?

4. Repeat Exercise 3 with $\overline{z}$ on the right-hand side of the equation replaced by z.

5. Calculate

$$\left(\frac{\partial}{\partial \overline{z}}\right)\left(\frac{\partial}{\partial \overline{z}}\right)^* + \left(\frac{\partial}{\partial \overline{z}}\right)^*\left(\frac{\partial}{\partial \overline{z}}\right)$$

in a suitable sense (here * denotes the adjoint). Show that the result is the Laplace operator.

6. Verify directly that

$$\frac{\partial}{\partial \overline{z}}\overline{z} = 1 \, , \qquad \frac{\partial}{\partial \overline{z}} z = 0 \, ,$$
$$\frac{\partial}{\partial z}\overline{z} = 0 \, , \qquad \frac{\partial}{\partial z} z = 1 \, .$$

7. If f is holomorphic and nonvanishing, then verify that $|f|^p$ is subharmonic for any $p > 0$. Do this using the differential operators studied in this chapter.

8. If f is holomorphic and nonvanishing, then verify that $\log|f|$ is harmonic.

9. Let $C(z,\zeta) = 1/(z-\zeta)$ be the Cauchy kernel. Verify that in the sense of distributions,

$$\frac{\partial}{\partial \overline{\zeta}} C(z,\zeta) = \delta_z \, ,$$

where δ_z is the Dirac delta mass at z.

10. Use the result of Exercise 9 to create a "distribution" proof of the Cauchy integral formula.

8

The Green's Function and the Poisson Kernel

Genesis and Development

Every smoothly bounded domain in the complex plane has a Green's function. The Green's function is fundamental to the Poisson integral, the theory of harmonic functions, and to the broad panorama of complex function theory.

The Green's function contains information about the geometry of its domain, and it also contains information about the harmonic analysis of the domain. We can rarely calculate the Green's function explicitly, but we can obtain enough qualitative information so that it is a potent tool.

The Green's function is so basic that its existence and essential properties depend only on multivariable calculus. The theory of harmonic measure, the construction of the Poisson kernel, the solution of the Dirichlet problem, and the rudiments of potential theory all depend on the Green's function. We shall explore all these additional ideas later in the book.

We develop the material in this chapter partly for its own interest—it provides an introduction to the Laplacian and the Green's function and the Poisson kernel. But we will also make good use of these results elsewhere in the book, particularly in our study of conformal mappings and the corona problem. We have already used the Green's function in our study of the Bergman kernel in Chapter 1.

8.1 The Laplacian and Its Fundamental Solution

The partial differential operator

$$\triangle = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} = 4\frac{\partial}{\partial \overline{z}}\frac{\partial}{\partial z} = 4\frac{\partial}{\partial z}\frac{\partial}{\partial \overline{z}}$$

is called the *Laplacian*. A twice continuously differentiable function u on a domain Ω is termed *harmonic* if $\triangle u \equiv 0$ on Ω.

It is a topic of particular interest to solve the following "boundary value problem." Let φ be a continuous function on $\partial\Omega$. Consider the conditions (on a function $u \in C^2(\Omega) \cap C(\overline{\Omega})$)

$$\begin{cases} \triangle u = 0 \text{ on } \quad \Omega\,, \\ u = \varphi \quad \text{on } \partial\Omega\,. \end{cases} \tag{8.1.1}$$

This is the *Dirichlet problem*. If Ω is a thin, heat-conducting metal plate and φ is a continuous temperature distribution on $\partial\Omega$, then the solution u of the system (8.1.1) represents the resulting steady-state heat distribution on Ω. For us, the Dirichlet problem is a useful device for constructing harmonic functions.

O. Perron has given us a method for solving the Dirichlet problem. It produces this solution as the supremum of subharmonic functions whose boundary limits are less than or equal to φ. We cannot provide the details here, but refer the reader to [AHL2] or [GRK1]. A typical theorem will be stated in a moment. First we need a definition.

Definition 8.1.1. Let $\Omega \subseteq \mathbb{C}$ be a domain and $P \in \partial\Omega$. We say that P has a *barrier* if there is a point $Q \notin \overline{\Omega}$ such that the line segment $\overline{PQ}$ intersects $\overline{\Omega}$ only in the point P. See Figure 8.1.

Theorem 8.1.2. *Let $\Omega \subseteq \mathbb{C}$ be a domain. Assume that each point of $\partial\Omega$ has a barrier. Then the Dirichlet problem can always be solved on Ω.*

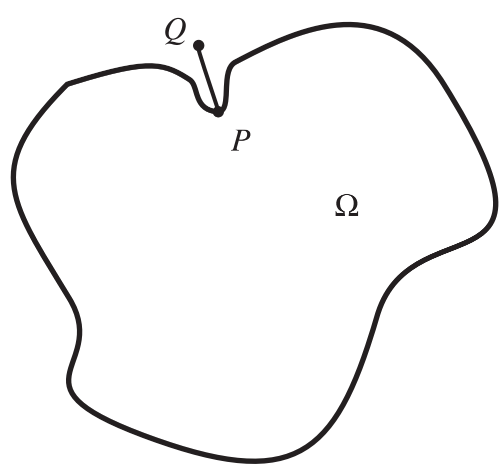

Fig. 8.1. A barrier.

We refer the reader to [AHL2, p. 251] and [GRK1, p. 237] for details of this theorem and its proof. It is worth noting that if Ω is a bounded domain with C^1 boundary, then certainly each point of $\partial\Omega$ has a barrier. Just use the implicit function theorem.

In fact for many domains there is another approach to the Dirichlet problem that will be of particular interest for us. This method involves the Poisson kernel and the Green's function.

On the disk—which is a very special domain indeed—there are ad hoc methods for deriving the Poisson kernel. Let us outline one of these now.

Suppose that u is a harmonic function that, for convenience, we assume is harmonic on a neighborhood of the closed disk $\overline{D}$. We may write $u\big|_{\partial D}$ (using the theory of Fourier series) as

$$u(e^{i\theta}) = \sum_{j=-\infty}^{\infty} a_j e^{ij\theta} .$$

Here of course

$$a_j = \frac{1}{2\pi} \int_0^{2\pi} u(e^{it}) e^{-ijt}\, dt .$$

See [KRA4] for details.

It follows that on the disk,

$$u(re^{i\theta}) = \sum_{j=-\infty}^{\infty} a_j r^{|j|} e^{ij\theta} .$$

Writing this out, we find that

$$\begin{aligned}
u(re^{i\theta}) &= \sum_{j=-\infty}^{\infty} \left[\frac{1}{2\pi} \int_0^{2\pi} u(e^{it}) e^{-ijt}\, dt \right] r^{|j|} e^{ij\theta} \\
&= \frac{1}{2\pi} \int_0^{2\pi} \left[\sum_{j=-\infty}^{\infty} r^{|j|} e^{ij(\theta-t)} \right] u(e^{it})\, dt \\
&= \frac{1}{2\pi} \int_0^{2\pi} \left[1 + \sum_{j=1}^{\infty} (re^{i(\theta-t)})^j + \sum_{j=1}^{\infty} (re^{-(i(\theta-t))})^j \right] u(e^{it})\, dt \\
&= \frac{1}{2\pi} \int_0^{2\pi} \left[1 + 2\mathrm{Re} \sum_{j=1}^{\infty} (re^{i(\theta-t)})^j \right] u(e^{it})\, dt \\
&= \frac{1}{2\pi} \int_0^{2\pi} \left[1 + 2\mathrm{Re} \left(\frac{1}{1 - re^{i(\theta-t)}} - 1 \right) \right] u(e^{it})\, dt
\end{aligned}$$

$$= \frac{1}{2\pi} \int_0^{2\pi} \left\{ 1 + 2 \left[\frac{r\cos(\theta - t) - r^2}{1 - 2r\cos(\theta - t) + r^2} \right] \right\} u(e^{it})\, dt$$
$$= \frac{1}{2\pi} \int_0^{2\pi} \frac{1 - r^2}{1 - 2r\cos(\theta - t) + r^2} u(e^{it})\, dt\,.$$

We conclude that the Poisson kernel for the unit disk D is

$$P(r, \theta) = \frac{1}{2\pi} \cdot \frac{1 - r^2}{1 - 2r\cos\theta + r^2}\,.$$

Now we wish to say a few words about how to derive the Poisson kernel for a more general domain—say a bounded domain in $\mathbb{C}$ with C^2 boundary. First we need Green's theorem from classical potential theory:

Theorem 8.1.3. *Let $\Omega \subseteq \mathbb{R}^2$ be a domain with C^2 boundary. Let ds denote arc length measure on $\partial\Omega$. Let ν be the unit outward normal vector field on $\partial\Omega$. Then, for any functions $u, v \in C^2(\overline{\Omega})$, we have*

$$\int_{\partial\Omega} \Big[u(\nu v) - v(\nu u) \Big]\, ds = \int_{\Omega} (u\Delta v - v\Delta u)\, dA.$$

We shall not prove Green's theorem, but refer the reader to [KRA1], where it is shown that Green's theorem is a corollary of Stokes's theorem. Any good calculus book (see [BLK]) will also have a thorough discussion of Green's theorem.

Now we systematically develop the theory of the Green's function and the Poisson kernel. It will be convenient in these discussions to momentarily forget the complex structure of $\mathbb{C}$ and instead think of ourselves as living in $\mathbb{R}^2$.

Proposition 8.1.4. *The fundamental solution for the Laplacian on $\mathbb{R}^2$ is given by*

$$\Gamma(x) = (2\pi)^{-1} \log|x|\,.$$

More precisely, for any $\phi \in C_c^\infty(\mathbb{R}^2)$, it holds that

$$\int (\triangle\phi)\Gamma\, dA = \phi(0).$$

Remark 8.1.5. This result is so important that it merits discussion. The conclusion of the proposition says that $\triangle\Gamma = \delta_0$, the Dirac delta mass, in the weak (or distribution) sense. In particular, $\triangle\Gamma(x) = 0$ when $x \neq 0$. This assertion may also be verified directly using calculus.

It is a straightforward exercise for the reader to derive from the proposition the statement that if $\phi \in C_c^\infty(\mathbb{R}^2)$, then $u = \Gamma * \phi$ satisfies $\triangle u = \phi$. Thus Γ is the kernel for a solution operator of the Laplace equation.

Proof of the Proposition. Fix $\phi \in C_c^\infty(\mathbb{R}^2)$ and a large disk $D(0,R) \subseteq \mathbb{R}^2$ such that $\operatorname{supp}\phi \subseteq D(0,R)$. Let $0 < \epsilon < R$ and apply Green's theorem on the domain $\Omega_\epsilon \equiv D(0,R) \setminus \overline{D}(0,\epsilon)$ with $u = \phi$ and $v = \Gamma$. Then, since $\triangle\Gamma(x) = 0$ for $x \neq 0$, we see that

$$\begin{aligned}\int_{\Omega_\epsilon} (\triangle\phi)\Gamma\, dA = \int_{\Omega_\epsilon} (\triangle\phi)\Gamma\, dA - \int_{\Omega_\epsilon} \phi\cdot(\triangle\Gamma)\, dA \\ = \int_{\partial D(0,R)} \{(\nu\phi)\cdot\Gamma - \phi\cdot\nu\Gamma\}\, ds \\ + \int_{\partial D(0,\epsilon)} \{(\nu\phi)\cdot\Gamma - \phi\cdot\nu\Gamma\}\, ds \\ = 0 - \phi(0)\int_{\partial D(0,\epsilon)} \nu\Gamma\, ds \\ + \left[\int_{\partial D(0,\epsilon)} (\nu\phi)\cdot\Gamma(x) + \int_{\partial D(0,\epsilon)} \{\phi(0) - \phi(x)\}\cdot\nu\Gamma(x)\, ds(x)\right].\end{aligned}$$

Let us examine the two expressions inside the square brackets. For the first integral, notice that $\nu\phi$ is bounded and Γ is of size $\log\epsilon$. Thus the integral is of size $C\cdot\epsilon\cdot\log\epsilon$, and that tends to zero. For the second integral, $\phi(0) - \phi(x)$ is of size $C\cdot\epsilon$ and $\nu\Gamma$ is of size C/ϵ. So the integrand is bounded and the integral is of size $C\cdot\epsilon$, which tends to zero. All in all then,

$$\int_{\Omega_\epsilon} (\triangle\phi)\Gamma\, dA = -\phi(0)\int_{\partial D(0,\epsilon)} \nu\Gamma\, ds + o(1) \quad \text{as} \quad \epsilon \to 0^+ .$$

We have used here the facts that $|\Gamma(x)| = c\cdot|\log|x||$ and the circumference of $\partial D(0,\epsilon)$ is $c'\cdot\epsilon$. Letting $\epsilon \to 0^+$, and noting that the outward normal vector field ν to $\partial\Omega_\epsilon$ at points of $\partial D(0,\epsilon)$ is oriented *oppositely* to ν when thought of as the outward normal vector field to $\partial D(0,\epsilon)$ itself (see Figure 8.2), gives

$$\begin{aligned}\int (\triangle\phi)\Gamma\, dx &= \lim_{\epsilon\to 0} \phi(0)\frac{1}{2\pi}\int_{\partial D(0,\epsilon)} |x|^{-1}\, ds(x) \\ &= \lim_{\epsilon\to 0} \phi(0)\cdot\frac{1}{2\pi}\cdot(2\pi\epsilon)\cdot\frac{1}{\epsilon} \\ &= \phi(0)\end{aligned}$$

as desired. □

Lemma 8.1.6. *If u is harmonic on $\Omega \equiv D(x_0,r) \subseteq \mathbb{R}^2$ and is C^2 on $\overline{\Omega}$, then*

$$\int_{\partial D(x_0,r)} \nu u\, ds = 0.$$

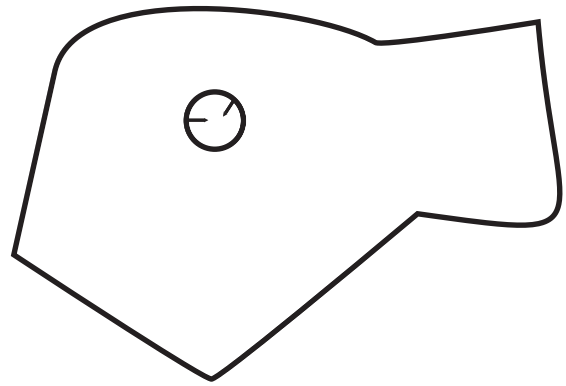

Fig. 8.2. Outward normal vector field.

Proof. Apply Green's theorem with the functions u and $v \equiv 1$ on the domain Ω. The details of the proof may be found in [KRA1, Chapter 1]. $\square$

Theorem 8.1.7. *Let u be harmonic on a domain $\Omega \subseteq \mathbb{R}^2$. Suppose that $\overline{D}(P,r) \subseteq \Omega$. Then*

$$\frac{1}{2\pi r}\int_{\partial D(P,r)} u(x)\,ds(x) = u(P).$$

Proof. Let $0 < \epsilon < r$ and apply Green's theorem on the region $\Omega_\epsilon \equiv D(P,r) \setminus \overline{D}(P,\epsilon)$ with u the given harmonic function and $v = \Gamma(x-P)$. The result is

$$\begin{aligned} 0 &= \int_{\Omega_\epsilon} u \triangle v - v \triangle u \, dA \\ &= \left\{ \int_{\partial D(P,r)} u\nu\Gamma(x-P)\,ds + \int_{\partial D(P,\epsilon)} u\nu\Gamma(x-P)\,ds \right\} \\ &\qquad - \left\{ \int_{\partial D(P,r)} (\nu u)\Gamma(x-P)\,ds + \int_{\partial D(P,\epsilon)} (\nu u)\Gamma(x-P)\,ds \right\}. \end{aligned}$$

Now the second expression in braces is equal to zero by Lemma 8.1.6 since $\Gamma(x-P)$ is constant on $\partial D(P,r)$ and on $\partial D(P,\epsilon)$. Thus we have

$$0 = \frac{1}{2\pi}\left[r^{-1}\int_{\partial D(P,r)} u\,ds \; - \; \epsilon^{-1}\int_{\partial D(P,\epsilon)} u\,ds \right].$$

The second expression inside the brackets tends, as $\epsilon \to 0^+$, to

$$2\pi u(P).$$

We conclude that

$$\frac{1}{2\pi r} \int_{\partial D(P,r)} u(x)\, ds(x) = u(P). \quad \square$$

The next result is the classical maximum principle.

Corollary 8.1.8. *Let u be a nonconstant, real-valued harmonic function on a domain $\Omega \subseteq \mathbb{R}^2$. Then, for every $x \in \Omega$, it holds that*

$$u(x) < \sup_{y \in \Omega} u(y).$$

Remark 8.1.9. There are several useful ways to restate the maximum principle. One of these is the contrapositive statement: if u is real-valued harmonic on Ω and there is an $x \in \Omega$ such that $u(x) = \sup_{y \in \Omega} u(y)$, then u is constant.

A slightly weaker statement is this: Assume that Ω is bounded. If u is continuous on $\overline{\Omega}$ and harmonic on Ω, then $\max_{\overline{\Omega}} u = \max_{\partial\Omega} u$. This statement has a strong intuitive appeal.

Proof of the Corollary. Let $A = \sup_{y \in \Omega} u(y)$. We may as well suppose that A is finite or else there is nothing to prove. Let $S = \{w \in \Omega : u(w) = A\}$. We will show that S is either empty or all of Ω.

If S is not empty, let $x \in S$. There then exists a number $r_0 > 0$ such that $\overline{D}(x, r_0) \subseteq \Omega$. Then, for every $0 < r < r_0$, we have from the mean value property that

$$\begin{aligned} A &= \frac{1}{s(\partial D(x,r))} \int_{\partial D(x,r)} A(y)\, ds(y) \\ &\geq \frac{1}{s(\partial D(x,r))} \int_{\partial D(x,r)} u(y)\, ds(y) \\ &= u(x) \\ &= A. \end{aligned}$$

It follows that u is identically equal to A on $\partial D(x,r)$ for all $0 \leq r < r_0$, hence on $D(x, r_0)$. Thus S is open. But S is clearly closed—by its definition—since u is continuous. Since S is not empty and (by assumption) Ω is connected, we conclude that $S = \Omega$. Hence $u \equiv A$. This completes the proof. $\square$

Corollary 8.1.10. *If u_1, u_2 are harmonic on $\Omega \subset\subset \mathbb{R}^2$, continuous on $\overline{\Omega}$, and if $u_1|_{\partial\Omega} = u_2|_{\partial\Omega}$, then $u_1 \equiv u_2$ on $\overline{\Omega}$.*

Exercises for the Reader

1. Prove that, if u is harmonic on $\Omega \subseteq \mathbb{R}^2$, $P \in \Omega$, and $\overline{D}(P,r) \subseteq \Omega$, then

$$\frac{1}{A(D(P,r))} \int_{D(P,r)} u(x)\, dA(x) = u(P). \qquad \diamondsuit$$

2. Prove the following trivial variant of Theorem 8.1.7. Under the same hypotheses,

$$\frac{1}{2\pi} \int_{\partial D(0,1)} u(P + ry)\, ds(y) = u(P). \qquad \diamondsuit$$

8.2 The Green's Function and Consequences

Definition 8.2.1 (The Green's Function). Let $\Omega \subseteq \mathbb{R}^2$ be a bounded domain with C^2 boundary. A function $G : (\Omega \times \overline{\Omega}) \setminus \{\text{diagonal}\} \to \mathbb{R}$ is the *Green's function* (for the Laplacian) on Ω if:

(1) G is C^2 on $\Omega \times \Omega \setminus \{\text{diagonal}\}$ and, for any small $\epsilon > 0$, is $C^{2-\epsilon}$ up to $(\Omega \times \overline{\Omega}) \setminus \{\text{diagonal}\}$.
(2) $\triangle_y G(x,y) = 0$ for $x \neq y$, $x \in \Omega$, $y \in \Omega$.
(3) For each fixed $x \in \Omega$ the function $G(x,y) + \Gamma(y - x)$ is harmonic as a function of $y \in \Omega$ (even at the point x).
(4) $G(x,y)\big|_{y \in \partial\Omega} = 0$ for each fixed $x \in \Omega$.

Remark 8.2.2. The reader should check that Properties **(1)**–**(4)** uniquely determine the function G. Notice that we have not yet asserted that Green's functions exist; this assertion is the content of Proposition 8.2.3 below.

We also have not said much about the dependence of G on the x variable. In fact $G(x,y)$ is symmetric in the variables x and y (Exercise, or see [KRA1, Chapter 1]).

We can now see that Green's functions always exist.

Proposition 8.2.3. *Let $\Omega \subseteq \mathbb{R}^2$ be a bounded domain with C^2 boundary. Then Ω has a Green's function.*

Proof. Fix $x \in \Omega$. Let $f_x(y) = -\,\Gamma(y - x)\big|_{y \in \partial\Omega}$. Let $F_x(y)$ be the unique solution to the Dirichlet problem on Ω with boundary data $f_x(y)$. Set $G(x,y) = -\Gamma(y - x) - F_x(y)$. Then Properties **(2)**–**(4)** of the Green's function follow immediately. The smoothness assertion (Property **1**) follows from regularity properties of elliptic boundary value problems, and we shall not provide the details (see the books [MOR], [BJS], or [KRA2]). $\square$

The reader should not now be surprised to learn that the Green's function, together with Green's formula, gives us a new integral formula.

Theorem 8.2.4 (The Poisson Integral Formula). *Let $\Omega \subseteq \mathbb{R}^2$ be a bounded domain with C^2 boundary. Let ν represent the unit outward normal vector field on $\partial\Omega$. Let the Poisson kernel on Ω be defined by*

$$P(x,y) = -\nu_y G(x,y).$$

If $u \in C(\overline{\Omega})$ is also harmonic on Ω, then

$$u(x) = \int_{\partial\Omega} P(x,y)u(y)\,ds(y) \quad \text{for all } x \in \Omega.$$

Proof. First assume that $u \in C^2(\overline{\Omega})$. Fix $x \in \Omega$ and a positive number $\epsilon < \operatorname{dist}(x, \partial\Omega)$. We apply Green's formula on $\Omega_\epsilon \equiv \Omega \setminus \overline{D}(x,\epsilon)$ with $u(y)$ the given harmonic function and $v(y) = G(x,y)$. Then

$$\begin{aligned}
&\int_{\partial\Omega_\epsilon} (\nu_y G(x,y))\, u(y)\, ds(y) - \int_{\partial\Omega_\epsilon} G(x,y)(\nu_y u(y))\, ds(y) \\
&= \int_{\Omega_\epsilon} (\triangle_y G(x,y))\, u(y)\, dA(y) - \int_{\Omega_\epsilon} G(x,y)(\triangle_y u(y))\, dA(y).
\end{aligned}$$

Since $G(x,\cdot)$ and $u(\cdot)$ are harmonic, the right side vanishes. Using the definition of $G(x,y)$ from the proof of Proposition 8.2.3, we may rewrite the last equation as

$$\begin{aligned}
&\int_{\partial\Omega_\epsilon} \nu_y G(x,y) u(y)\, ds(y) \\
&= \int_{\partial\Omega} G(x,y)\nu_y u(y)\, ds(y) - \int_{\partial D(x,\epsilon)} \Gamma(x-y)\nu_y u(y)\, ds(y) \\
&\quad - \int_{\partial D(x,\epsilon)} F_x(y)\nu_y u(y)\, ds(y).
\end{aligned}$$

The first term on the right vanishes since $G(x,\cdot)\big|_{\partial\Omega} = 0$. The second vanishes because $\Gamma(x-\cdot)\big|_{\partial D(x,\epsilon)}$ is constant and because of Lemma 8.1.6. The last term vanishes as $\epsilon \to 0^+$ since the integrand is bounded.

Thus we have

$$\int_{\partial\Omega_\epsilon} (\nu_y G(x,y))\, u(y)\, ds(y) = o(1)\,,$$

or

$$\int_{\partial\Omega} (\nu_y G(x,y))\, u(y)\, ds(y) = -\int_{\partial D(x,\epsilon)} (\nu_y G(x,y))\, u(y)\, ds(y) + o(1).$$

Recall that the unit outward normal to $\partial\Omega_\epsilon$ at a point $y \in \partial D(x,\epsilon)$ is the negative of the unit outward normal to $\partial D(x,\epsilon)$ at y. With $P(x,y) = -\nu_y G(x,y)$, we may now write

$$\begin{aligned}\int_{\partial\Omega} P(x,y)u(y)\,ds(y) &= \int_{\partial D(x,\epsilon)} \nu_y G(x,y)u(y)\,ds(y) + o(1)\\ &= -\int_{\partial D(x,\epsilon)} \nu_y \Gamma(x-y)u(y)\,ds(y)\\ &\quad -\int_{\partial D(x,\epsilon)} \nu_y F_x(y)u(y)\,ds(y) + o(1).\end{aligned}$$

The second term on the right side of the last equation is $o(1)$ as $\epsilon \to 0^+$ because the integrand is bounded. A now familiar computation (see the proof of Proposition 8.1.4) then shows that the first term on the right tends to $u(x)$. Letting $\epsilon \to 0^+$ we conclude that

$$u(x) = \int_{\partial\Omega} P(x,y)u(y)\,ds(y).$$

This is the Poisson integral formula when $u \in C^2(\overline{\Omega})$.

To eliminate the hypothesis that $u \in C^2(\overline{\Omega})$, we exhaust Ω by relatively compact, smoothly bounded subdomains Ω_j and apply the preceding result to u on $\overline{\Omega}_j$. A limiting argument then completes the proof. Details are left to the interested reader. □

Corollary 8.2.5. *For each fixed $y \in \partial\Omega, P(x,y)$ is harmonic in x.*

Proof. Of course this follows from the previously noted fact that $G(x,y) = G(y,x)$.

An alternative proof may be obtained by noting that $\int_{\partial\Omega} P(x,y)\phi(y)\,ds(y)$ is a harmonic function of x for every $\phi \in C(\partial\Omega)$. Now choose a sequence ϕ_j such that $\phi_j\,ds$ converges to δ_y (the Dirac mass at y) in the weak-$*$ topology. □

8.3 Calculation of the Poisson Kernel

The Poisson kernel can almost never be computed explicitly. However, we *can* compute it for the disk. Elsewhere in the present book we have indicated how to determine the Poisson kernel for the disk by summing a Fourier series. Now we calculate the kernel using the paradigm of Theorem 8.2.4. We begin with an important geometric transformation:

Definition 8.3.1 (The Kelvin Inversion). The *Kelvin inversion* on the domain $\mathbb{R}^2 \setminus \{0\}$ is the map

$$K : \mathbb{R}^2 \setminus \{0\} \to \mathbb{R}^2 \setminus \{0\}$$

given by $K(x) = x/|x|^2$. Notice that $K = K^{-1}$.

Proposition 8.3.2. *Let $\Omega \subseteq \mathbb{R}^2$ and $0 \notin \Omega$. If u is harmonic on Ω, then*

$$x \mapsto u \circ K(x)$$

is harmonic on $K(\Omega)$.

Proof. Exercise for the reader (just compute). □

Theorem 8.3.3. *The Poisson kernel for the disk $\Omega = D$ is given by*

$$P(x,y) = \frac{1}{2\pi}\frac{1-|x|^2}{|x-y|^2}\,,$$

or

$$P_r(e^{i\theta}) = \frac{1}{2\pi}\cdot\frac{1-r^2}{1-2r\cos\theta + r^2}\,.$$

Proof. If we can construct the Green's function for the disk, then the rest is straightforward. A glance at the proof of the existence of the Green's function shows that the main obstacle is to construct F_x for $x \in D$. Fix $x \in D$. Notice that the Kelvin inversion on D is the map $y \mapsto y/|y|^2$. Set

$$F_x(y) = -\Gamma(x - K(y)) - \log|y| = -\frac{1}{2\pi}\log\left|x - \frac{y}{|y|^2}\right| - \log|y|\,.$$

Now $\Gamma(x-y)$ is plainly harmonic on $\mathbb{R}^2 \setminus \{x\}$; hence $F_x(y)$ is harmonic on a neighborhood of $\overline{D} \setminus \{0\}$. The presence of the (perhaps surprising) term $-\log|y|$ is to cancel the singularity at the origin. It may now be checked directly that in fact F_x is smooth at 0, so it follows from the continuity of the derivative that F_x is harmonic on a neighborhood of $\overline{D}$. Since the Kelvin transform K fixes ∂D, it follows that

$$F_x(\,\cdot\,)|_{\partial D} = -\Gamma(x - \cdot)|_{\partial D}\,.$$

Thus F_x is the function we seek, and the Green's function for D must be

$$\begin{aligned} &-\Gamma(x-y) - F_x(y) \\ &= -\frac{1}{2\pi}\left\{\log|x-y| - \log\left|x - \frac{y}{|y|^2}\right| - \log|y|\right\}. \end{aligned}$$

Finally, by Theorem 8.2.4, the Poisson kernel for the disk is

$$\begin{aligned} P(x,y) &= -\nu_y G(x,y) \\ &= \frac{1}{2\pi}\sum_{j=1}^{2} y_j \frac{\partial}{\partial y_j}\left\{\log|x-y| \;-\log\left|x - \frac{y}{|y|^2}\right| - \log|y|\right\} \\ &= \frac{1}{2\pi}\sum_{j=1}^{2} y_j \left\{\frac{1}{2}|x-y|^{-2}\cdot 2(y_j - x_j)\right. \\ &\quad \left. -\frac{1}{2}\left|x - \frac{y}{|y|^2}\right|^{-2}\cdot\sum_{\ell=1}^{2} 2\left(\frac{y_\ell}{|y|^2} - x_\ell\right)\left(\frac{-2y_j y_\ell}{|y|^4} + \frac{\delta_{j\ell}}{|y|^2}\right) - \frac{1}{2}\cdot\frac{2y_j}{|y|^2}\right\}. \end{aligned}$$

When $y \in \partial D$ this yields

$$\begin{aligned} P(x,y) &= \frac{1}{2\pi} \cdot |x-y|^{-2} \left\{ |y|^2 - x \cdot y - \sum_{\ell=1}^{2} \sum_{j=1}^{2} y_j (y_\ell - x_\ell)(-2y_j y_\ell + \delta_{j\ell}) \right\} - 1 \\ &= \frac{1}{2\pi} \cdot |x-y|^{-2} \{1 - x \cdot y + 2 - 1 - 2(x \cdot y) + x \cdot y\} - 1 \\ &= \frac{1}{2\pi} \cdot \frac{1-|x|^2}{|x-y|^2}. \end{aligned}$$

This is the desired result. The expression in polar coordinates is an immediate consequence. □

Proposition 8.3.4. *The Poisson kernel for $D \subseteq \mathbb{R}^2$ has the properties*

1. $P(x,y) \geq 0.$
2. $\int_{\partial D} P(x,y)\, ds(y) = 1,$ *all* $x \in D.$
3. *For any* $\delta > 0,$ *any fixed* $\zeta_0 \in \partial D,$

$$\lim_{D \ni x \to \zeta_0} \int_{|\zeta_0 - y| > \delta} P(x,y)\, ds(y) = 0.$$

Proof. Inequality **1** is obvious. Also, by the reproducing property of P and the fact that $P \geq 0$, we have

$$1 = \int_{\partial D} 1 \cdot P(x,y)\, ds(y) = \|P(x,\cdot)\|_{L^1(\partial D,\, ds)}.$$

This is property **2**. As for property **3**, choose $0 < \epsilon < \delta/4$. Then $|\zeta_0 - y| > \delta$ and $|x - \zeta_0| < \delta^2 \epsilon/8$ imply that

$$P(x,y) = \frac{1}{2\pi} \frac{1-|x|^2}{|x-y|^N} \leq \frac{1}{2\pi} \cdot \frac{2 \cdot \delta^2 \epsilon/8}{(\delta/2)^2} < \epsilon$$

uniformly in $|\zeta_0 - y| > \delta$. The result follows. □

Theorem 8.3.5 (The Poisson Integral for the Disk). *Let $D \subseteq \mathbb{R}^2$ be the unit disk and $f \in C(\partial D)$. Then the function*

$$F(x) = \begin{cases} \int_{\partial D} P(x,y) f(y)\, ds(y) & \text{if } x \in D\ , \\ f(x) & \text{if } x \in \partial D\,, \end{cases}$$

is harmonic on D and continuous on $\overline{D}$.

Proof. We already know that the Poisson kernel is harmonic in the x variable by Corollary 8.2.5 (or, with the explicit formula on the disk, one can check this directly). Therefore, by differentiating under the integral sign, one sees that F is harmonic on D. Since F is obviously continuous (indeed smooth) inside D, and also is continuous on ∂D, it remains to check that if $\zeta \in \partial D$, then

$$\lim_{D \ni x \to \zeta} Pf(x) = f(\zeta).$$

Let $\epsilon > 0$. Since f is uniformly continuous on ∂D we may find a $\delta > 0$ such that if $\xi, \zeta \in \partial D$ and $|\xi - \zeta| \leq \delta$ then, $|f(\zeta) - f(\xi)| < \epsilon$. With this δ fixed, we write

$$\begin{aligned} Pf(x) - f(\zeta) &= \int_{\partial D} P(x,\xi) f(\xi)\, ds(\xi) - f(\zeta) \\ &= \int_{\partial D} P(x,\xi)[f(\xi) - f(\zeta)]\, ds(\xi). \end{aligned}$$

Here we have used part **2** of Proposition 8.3.4.

But this last expression equals

$$\begin{aligned} &\int_{\partial D \cap \{|\zeta - \xi| \leq \delta\}} P(x,\xi)[f(\xi) - f(\zeta)]\, ds(\xi) \\ &\quad + \int_{\partial D \cap \{|\zeta - \xi| > \delta\}} P(x,\xi)[f(\xi) - f(\zeta)]\, ds(\xi) \\ &\equiv I + II. \end{aligned}$$

Now $|I| \leq \epsilon \int_{\partial D} P(x,\xi)\, ds(\xi) = \epsilon$ by the choice of δ. On the other hand,

$$|II| \leq 2 \sup_{\partial D} |f| \int_{\partial D \cap \{|\zeta - \xi| > \delta\}} P(x,\xi)\, ds(\xi) \to 0$$

as $x \to \zeta$ by part **3** of Proposition 8.3.4. This is what we wished to prove. □

Exercises for the Reader

1. Give another proof of Theorem 8.3.3, using Theorem 8.1.2 and Theorem 8.2.4, that involves no computation. This proof will be valid for any domain Ω with C^2 boundary.

2. Give a proof of Theorem 8.3.3 that uses the mean value property and conformal invariance.

3. Examine the proof of Theorem 8.3.3 to see that $Pf(r\zeta)$ tends to $f(\zeta)$ *uniformly* in $\zeta \in \partial D$ when $f \in C(\partial D)$.

Remark 8.3.6. Let $\Omega \subseteq \mathbb{R}^2$ be any domain with C^2 boundary. It follows from the maximum principle that, for $y \in \partial\Omega$ and x near y, $G(x,y) > 0$. Hence, by the Hopf lemma (see Lemma 5.3.1), we conclude that $P(x,y) > 0$. Therefore, for each $x \in \Omega$, the argument in Proposition 8.3.4 shows that $\|P(x,\cdot)\|_{L^1(\partial\Omega,\, ds)} = 1$. Thus, for $\phi \in C(\partial\Omega)$, the functional

$$\phi \mapsto \int_{\partial\Omega} P(x,y)\phi(y)\,ds(y)$$

is bounded. From this, Theorem 8.2.4, and the maximum principle, we have the next result.

Proposition 8.3.7. *The Poisson kernel for a C^2 domain Ω is uniquely determined by the property that it is positive and solves the Dirichlet problem for $\triangle$.*

Problems for Study and Exploration

1. Prove that the Green's function is always a symmetric function of its arguments: $G(z,\zeta) = G(\zeta,z)$.
2. Provide the details of the proof of Proposition 8.3.7.
3. Let Ω be *any* bounded domain in $\mathbb{C}$ with C^2 boundary. Let $P(x,y)$ be the Poisson kernel for Ω. It is a fact—not so easy to prove—that there are positive constants c and C (depending on Ω) such that

$$c \cdot \frac{\delta(x)}{|x-y|^2} \le P(x,y) \le C \cdot \frac{\delta(x)}{|x-y|^2}\,.$$

   Use the fact that the Poisson kernel is related to the Green's function, together with the result of Remark 8.3.6, to prove the left-hand inequality.
4. Use the method of conformal mapping to derive a formula for the Poisson kernel of the upper halfplane $U = \{z \in \mathbb{C} : \operatorname{Im} z > 0\}$.
5. This exercise outlines a method to construct the Poisson kernel that is analogous to the Bergman kernel construction, by way of a basis, in Chapter 2. Let us consider the L^2 functions on the unit circle of the complex plane that are boundary functions of holomorphic functions on the disk. This space is commonly called H^2, and can also be characterized as those L^2 functions on the unit circle that have no Fourier coefficients of negative index. The inner product on this space is

$$\langle f,g\rangle = \frac{1}{2\pi}\int_0^{2\pi} f(e^{i\theta})\cdot\overline{g(e^{i\theta})}\,d\theta\,.$$

   (a) An orthogonal basis for this space is given by the functions

$$\psi_j(\zeta) = \zeta^j\,.$$

   Calculate the norm of each ψ_j and produce an orthonormal basis $\{\varphi_j\}_{j=0}^{\infty}$.

   (b) Prove that the basis from part **(a)** is complete.

(c) Calculate the sum

$$S(z,\zeta) = \sum_{j=0}^{\infty} \varphi_j(z)\overline{\varphi_j(\zeta)}\,.$$

This is called the *Szegő kernel*.

(d) Now define a new kernel, called the *Poisson–Szegő kernel*, by the formula

$$\mathcal{P}(z,\zeta) = \frac{|S(z,\zeta)|^2}{S(z,z)}\,.$$

What is the Poisson–Szegő kernel for the unit disk?

6. Imitate the argument in Exercise 5 to derive the Poisson–Szegő kernel for the upper halfplane $U = \{z \in \mathbb{C} : \mathrm{Im}\, z > 0\}$.

7. Use the Hopf lemma to derive a version of the maximum principle for harmonic functions.

8. Write an explicit formula for the Green's function of the upper halfplane $U = \{z \in \mathbb{C} : \mathrm{Im}\, z > 0\}$.

9. Without some regularity hypothesis on the boundary of the domain, the Dirichlet problem may fail to have a solution. Examine the domain $\Omega = \{z \in \mathbb{C} : 0 < |z| < 1\}$. Show that the Dirichlet problem with boundary data

$$\varphi(\zeta) = \begin{cases} 0 & \text{if} \quad \zeta = 0 \\ 1 & \text{if} \quad |\zeta| = 1 \end{cases}$$

has no solution.

10. What is the Poisson kernel for the strip

$$\Omega = \{\zeta \in \mathbb{C} : -1 < \mathrm{Im}\, \zeta < 1\}\,?$$

9

Harmonic Measure

Genesis and Development

Harmonic measure is a device for estimating harmonic functions on a domain. It has become an essential tool in potential theory and in studying the corona problem. It is useful in studying the boundary behavior of conformal mappings, and it tells us a great deal about the boundary behavior of holomorphic functions and solutions of the Dirichlet problem. All these are topics that will be touched on in the present book.

From the point of view of function algebras (which we discuss in Chapters 4 and 11), harmonic measure may be thought of as a *representing measure* for a multiplicative linear functional. It is a device that bridges the techniques of hard analysis, measure theory, functional analysis, and operator algebras.

A good grounding in real analysis will be helpful for an appreciation of the ideas in this chapter.

9.1 Introductory Remarks

Let $U \subseteq \mathbb{C}$ be a domain, that is, a connected open set. A twice continuously differentiable function u on U is said to be *harmonic* if it satisfies the partial differential equation

$$\triangle u \equiv \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} \right) u(x, y) = 0 \tag{9.1.1}$$

at all points of U. The partial differential operator $\triangle$ is called the *Laplacian*, and the differential equation (9.1.1) is called *Laplace's* equation. We now review some of the key elementary ideas connected with this fundamental equation.

It is a matter of great interest to solve the "first boundary-value problem" associated to the Laplacian. This problem is also known as *Dirichlet's problem.* We quickly review its essential features. Let f be a continuous function on ∂U. We seek a function u that is

(i) continuous on $\overline{U}$,
(ii) twice continuously differentiable on U,
(iii) harmonic on U,
(iv) satisfies $u\big|_{\partial U} = f$.

Thus u is to be the harmonic continuation of f to the interior of U.

The Dirichlet problem has both mathematical and physical interest. If U is a thin metal plate, formed of heat-conducting material, and if f represents an initial heat distribution on the boundary of the plate, then the solution u of the Dirichlet problem represents the steady-state heat distribution in the interior. We have treated the Dirichlet problem in detail in the last chapter.

It is worthwhile to take a moment and consider the special case when U is the unit disk $D = D(0,1)$. We may give a heuristic treatment of the Dirichlet problem as follows:

(a) In case $f(e^{i\theta}) = e^{ij\theta}$, $j = 0, 1, 2, \ldots$, then by inspection we see that $u(re^{i\theta}) = r^j e^{ij\theta}$, or $u(z) = z^j$, is the solution of the Dirichlet problem;
(b) In case $f(e^{i\theta}) = e^{ij\theta}$, $j = -1, -2, \ldots$, then by inspection we see that $u(re^{i\theta}) = r^{|j|} e^{ij\theta}$, or $u(z) = \overline{z}^{|j|}$, is the solution of the Dirichlet problem;
(c) In case $f(e^{i\theta}) = \sum_{j=-K}^{K} a_j e^{ij\theta}$, then, by linearity, the corresponding solution of the Dirichlet problem is

$$u(re^{i\theta}) = \sum_{j=0}^{K} a_j r^j e^{ij\theta} + \sum_{j=-K}^{-1} a_j r^{|j|} e^{ij\theta} .$$

Now the elementary theory of Fourier series (see [KRA3] or [RUD1]) tells us that any continuous function on the boundary of D may be uniformly approximated by trigonometric polynomials as in **(c)**. Thus we should be able to obtain the corresponding solution of the Dirichlet problem.

Proceeding formally, suppose that $f \sim \sum_{j=-\infty}^{\infty} \widehat{f}(j) e^{ij\theta}$, where

$$\widehat{f}(j) = \frac{1}{2\pi} \int_0^{2\pi} f(t) e^{-ijt}\, dt$$

is the standard Fourier coefficient (see [KRA4]). Then the corresponding solution of the Dirichlet problem should be

$$\begin{aligned} u(re^{i\theta}) &= \sum_{j=-\infty}^{\infty} \widehat{f}(j) r^{|j|} e^{ij\theta} \\ &= \sum_{j=-\infty}^{\infty} \frac{1}{2\pi} \int_0^{2\pi} f(t) e^{-ijt}\, dt \cdot r^{|j|} e^{ij\theta} \\ &= \frac{1}{2\pi} \int_0^{2\pi} f(t) \cdot \left[\sum_{j=-\infty}^{\infty} r^{|j|} e^{ij(\theta - t)} \right] . \end{aligned}$$

The expression in brackets is merely a double geometric series, and it is easily summed. What we find is that

$$u(re^{i\theta}) = \frac{1}{2\pi} \int_0^{2\pi} f(t) \cdot \frac{1 - r^2}{1 - 2r\cos(\theta - t) + r^2}\, dt . \tag{9.1.2}$$

The expression

$$P(r, \theta) = \frac{1}{2\pi} \frac{1 - r^2}{1 - 2r\cos(\theta - t) + r^2}$$

is called the *Poisson kernel*, and the formula (9.1.2) is known as the *Poisson integral formula*.

For a general domain, one can almost never know the Poisson kernel explicitly. However, one can (with sufficiently powerful tools) obtain useful qualitative information. Here is a sample fact that is used frequently in the theory of partial differential equations and harmonic analysis.

Proposition 9.1.1. *Let Ω be a bounded domain in $\mathbb{C}$ with C^2 boundary. Then the Poisson kernel $P(z, \zeta)$ for Ω satisfies*

$$c \cdot \frac{d(z)}{|z - \zeta|^2} \le P(z, \zeta) \le C \cdot \frac{d(z)}{|z - \zeta|^2}$$

for suitable positive constants c, C and for $d(z)$ the distance of z to $\partial\Omega$.

Proof. See [KRA1], [KRA7]. □

It is important also to observe that the Dirichlet problem may be solved for more general boundary data than a continuous function. In fact if the boundary function is L^∞, or even L^1, then its Poisson integral certainly makes sense. This fact will be crucial in what follows. Even without the Poisson integral (and the Poisson integral certainly will not exist in general on a domain with nonrectifiable boundary) then we can still consider the Dirichlet problem with, say, piecewise continuous boundary data. Such a Dirichlet problem is solved by the Perron method.

9.2 The Idea of Harmonic Measure

Now let $\Omega \subseteq \mathbb{C}$ be a domain with boundary consisting of finitely many Jordan curves (we call such a domain a *Jordan domain*). Let E be a finite union of arcs in $\partial\Omega$ (we allow an entire connected boundary component—a simple, closed curve—to be one of these arcs). Then the *harmonic measure* of E at the point $z \in \Omega$ with respect to Ω is the value at z of the bounded harmonic function ω on Ω with boundary limit 1 at points of E and boundary limit 0 at points of $\partial\Omega \setminus E$ (except possibly at the endpoints of the arcs that make up E). We denote the harmonic measure by $\omega(z, \Omega, E)$.

The first question to ask about harmonic measure is that of existence and uniqueness: does harmonic measure always *exist*? If it does exist, is it unique? The answer to both these queries is "yes". Let us see why.

Fix a Jordan domain Ω. It is a classical result (see [AHL2] as well as our Section 4.4) that there is a conformal mapping $\Phi : \Omega \to U$ of Ω to a domain U bounded by finitely many circles. If $F \subseteq \partial U$ consists of finitely many circular arcs, then it is obvious that the harmonic measure of F exists—it would just be the Poisson integral of the characteristic function of F, and that Poisson integral can be shown to exist by elementary conformal mapping arguments (i.e., mapping to the disk). Now a classic result of Carathéodory (see Section 5.1) guarantees that Φ and its inverse extend continuously to the respective boundaries, so that the extended function is a homeomorphism of the closures. Thus we may pull the harmonic measure back from U to Ω via Φ.

As for uniqueness: If Ω is bounded, then uniqueness follows from the standard maximum principle. For the general case, we need an extended maximum principle that is due to Lindelöf.

Proposition 9.2.1. *Let Ω be a domain whose boundary is not a finite set. Let u be a real-valued harmonic function on Ω, and assume that there is a real constant $M > 0$ such that*

$$u(z) \leq M \qquad \text{for } z \in \Omega\,.$$

Suppose that m is a real constant and that

$$\limsup_{z \to \zeta} u(z) \leq m \tag{9.2.1}$$

for all except possibly finitely many points $\zeta \in \partial\Omega$. Then $u(z) \leq m$ for all $z \in \Omega$.

Remark 9.2.2. Observe that any Jordan domain certainly satisfies the hypotheses of the proposition. Furthermore, by classical results coming from potential theory (see [AHL2]), any domain on which the Dirichlet problem is solvable will satisfy the hypotheses of the theorem. Note in particular that Ω certainly need not be bounded.

We conclude by noting that this result is a variant of the famous Phragmen–Lindelöf theorem (see [RUD2]).

Proof of Proposition 9.2.1. First assume that Ω is bounded. We will remove this extra hypothesis later. Let the diameter of Ω be d. Let the exceptional boundary points (at which (9.2.1) does not hold) be called $\zeta_1, \ldots, \zeta_k$. Let $\epsilon > 0$. We then may apply the ordinary textbook maximum principle (see [GRK1]) to the auxiliary function

$$h(z) = u(z) + \epsilon \sum_{j=1}^{k} \log \frac{|z - \zeta_j|}{d} \,.$$

Notice that h (instead of u) satisfies the hypotheses of the proposition at *every boundary point.* So $h \le m$ on all of Ω. Then we let $\epsilon \to 0$ and the result follows.

Now consider the case of $\overline{\Omega}$ unbounded. If $\overline{\Omega}$ has an exterior point (i.e., a point in the interior of the complement of the closure of Ω), then we may apply an inversion and reduce to the case in the preceding paragraph.

Finally, if $\overline{\Omega}$ has no exterior point, then let $R > |\zeta_j|$ for all $j = 1, \ldots, k$. Let

$$\Omega_1 = \{z \in \Omega : |z| < R\}$$

and

$$\Omega_2 = \{z \in \Omega : |z| > R\} \,.$$

These may not be domains (i.e., they could be disconnected), but they are certainly open sets. Let

$$S = \{z \in \Omega : |z| = R\} \,.$$

If $u \le m$ on S, then we apply the result of the first paragraph on Ω_1 and the result of the second paragraph on Ω_2.

If it is *not* the case that $u \le m$ on S then u will have a maximum N on S with $N > m$. But this will of course be a maximum for u on all of Ω. Since $u \le m$ at the ends of the arcs of S, it follows that u actually achieves the maximum N on S. Therefore, by the usual maximum principle, u is identically equal to the constant N. But then the boundary condition (9.2.1) cannot obtain. This is a contradiction. □

It is important to realize that, if Ω has reasonably nice boundary, then $\omega(z, \Omega, E)$ is nothing other than the Poisson integral of the characteristic function of E. [We of course used this fact in our discussion of the existence of harmonic measure.] In any event, by the maximum principle the function ω takes real values between 0 and 1. Therefore $0 < \omega(z, \Omega, E) < 1$.

One of the reasons that harmonic measure is important is that it is a conformal invariant. Essential to this fact is Carathéodory's theorem that a conformal map of Jordan domains will extend continuously to the closures. We now formulate the invariance idea precisely.

Proposition 9.2.3. *Let Ω_1, Ω_2 be domains with boundaries consisting of Jordan curves in $\mathbb{C}$ and $\varphi : \Omega_1 \to \Omega_2$ a conformal map. If $E_1 \subseteq \partial\Omega_1$ is an arc and $z \in \Omega_1$, then*

$$\omega(z, \Omega_1, E_1) = \omega(\varphi(z), \Omega_2, \varphi(E_1)) .$$

Here $\varphi(E_1)$ is well defined by Carathéodory's theorem.

Proof. Let h denote the harmonic function $\omega(\varphi(z), \Omega_2, \varphi(E_1))$. Then $h \circ \varphi$ equals harmonic measure for $\varphi(E_1)$ on Ω_1 (by Carathéodory's theorem). □

Remark 9.2.4. It is worth recording here an important result of F. and M. Riesz (see [KOO, p. 72]). Let $\Omega \subseteq \mathbb{C}$ be a Jordan domain bounded by a rectifiable boundary curve. Let $\varphi : D \to \Omega$ be a conformal mapping. If $E \subseteq \partial D$ has Lebesgue linear measure zero, then $\varphi(E)$ also has Lebesgue linear measure zero. See also Theorem 9.6.2 below.

9.3 Some Examples

The harmonic measure helps us to understand the growth and value distribution of harmonic and holomorphic functions on Ω. It has become a powerful analytic tool. We begin to understand harmonic measure by first calculating some examples.

Example 9.3.1. Let Ω be the upper halfplane. Let E be an interval $[-T, T]$ on the real axis, centered at 0. Let us calculate $\omega(z, \Omega, E)$ for $z = x + iy \cong (x, y) \in \Omega$.

It is a standard fact (see [GRK1]) that the Poisson kernel for Ω is

$$P(x, y) = \frac{1}{\pi} \cdot \frac{y}{x^2 + y^2} .$$

Here we use the traditional real notation for the kernel. In other words, the harmonic function ω that we seek is given by

$$\omega(x + iy) = \frac{1}{\pi} \int_{-\infty}^{\infty} \chi_{[-T,T]}(t) \cdot \frac{y}{(x-t)^2 + y^2} \, dt = \frac{1}{\pi} \int_{-T}^{T} \frac{y}{(x-t)^2 + y^2} \, dt .$$

Now it is a simple matter to rewrite the integral as

$$\begin{aligned} \omega(x + iy, \Omega, E) &= \frac{1}{\pi} \frac{1}{y} \int_{-T}^{T} \frac{1}{\left(\frac{x-t}{y}\right)^2 + 1} \, dt \\ &= -\frac{1}{\pi} \mathrm{Tan}^{-1} \left(\frac{x-t}{y} \right) \Big|_{-T}^{T} \\ &= \frac{1}{\pi} \mathrm{Tan}^{-1} \left(\frac{x+T}{y} \right) - \frac{1}{\pi} \mathrm{Tan}^{-1} \left(\frac{x-T}{y} \right) . \end{aligned}$$

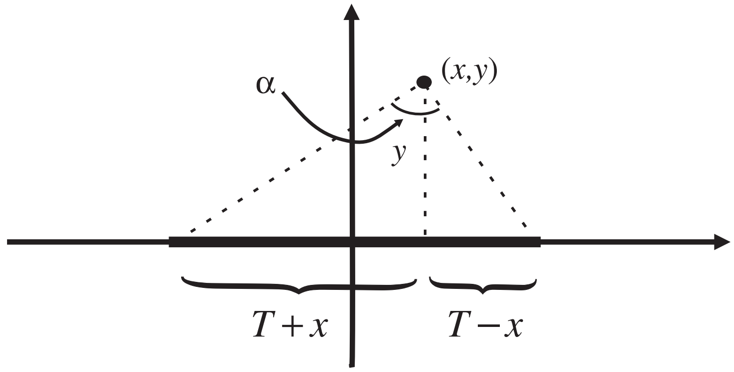

Fig. 9.1. The value of $\omega(z, \Omega, E)$.

The reader may check for himself that this function is harmonic, tends to 1 as (x, y) approaches $E \subseteq \partial\Omega$, and tends to 0 as (x, y) approaches $\partial\Omega \setminus E$.

Glancing at Figure 9.1, we see that the value of $\omega(z, \Omega, E)$ is simply α/π, where α is the angle subtended at the point $z = (x, y)$ by the interval E.

More generally, if E is any bounded, closed interval in the real line, then the harmonic measure $\omega(z, \Omega, E)$ will be $1/\pi$ times the angle subtended at the point $z = (x, y)$ by the interval E. But we can say more. If now E is the finite disjoint union of closed, bounded intervals,

$$E = I_1 \cup \cdots \cup I_k \,,$$

then the harmonic measure of E is just the sum of the harmonic measures $\omega(z, \Omega, I_j)$ for each of the individual intervals. So the harmonic measure at z is just $1/\pi$ times the sum of the angles subtended at $z = (x, y)$ by each of the intervals I_j.

Example 9.3.2. Let Ω be the unit disk. Let E be an arc of the circle with central angle α. Then, for $z \in \Omega$,

$$\omega(z, \Omega, E) = \frac{2\theta - \alpha}{2\pi} \,,$$

where θ is the angle subtended by E at z. See Figure 9.2.

It turns out to be convenient to first treat the case where E is the arc from $-i$ to i. See Figure 9.3.

We may simply calculate the Poisson integral of $\chi_{(-\pi/2,\pi/2)}$, where of course the argument of this characteristic function is the angle in radian measure on the circle. Thus, for $z = re^{i\lambda}$,

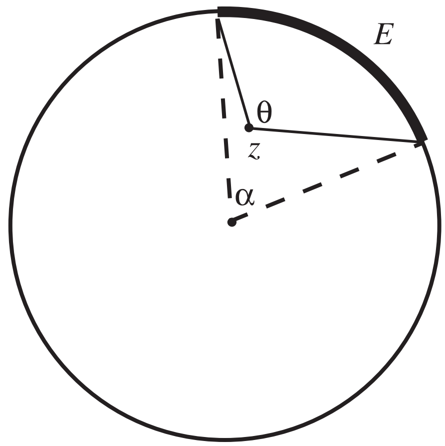

Fig. 9.2. The angle subtended by E at z.

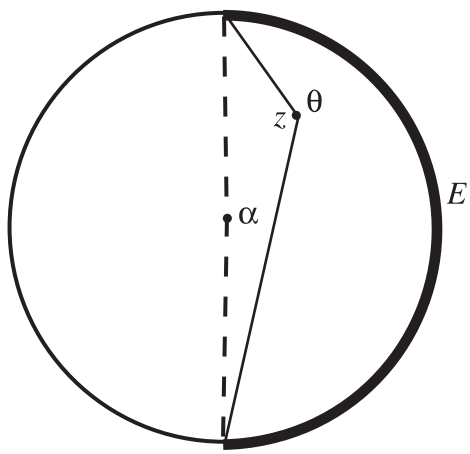

Fig. 9.3. The case E is the arc from $-i$ to i.

$$\begin{aligned}\omega(z, \Omega, E) &= \frac{1}{2\pi} \int_{-\pi/2}^{\pi/2} \frac{1 - r^2}{1 - 2r\cos(\lambda - t) + r^2}\, dt \\ &= \frac{1 - r^2}{2\pi} \int_{-\pi/2}^{\pi/2} \frac{dt}{(1 + r^2) - 2r\cos(\lambda - t)} .\end{aligned}$$

At this point it is useful to recall from calculus (and this may be calculated using the Weierstrass $w = \tan x/2$ substitution) that

$$\int \frac{dx}{a+b\cos x} = \frac{2}{\sqrt{a^2-b^2}}\mathrm{Tan}^{-1}\left(\sqrt{\frac{a-b}{a+b}}\tan\frac{x}{2}\right).$$

Therefore

$$\omega(z,\Omega,E) = \frac{1}{\pi}\left[\mathrm{Tan}^{-1}\left(\frac{1+r}{1-r}\tan\frac{\lambda+\frac{\pi}{2}}{2}\right) - \mathrm{Tan}^{-1}\left(\frac{1+r}{1-r}\tan\frac{\lambda-\frac{\pi}{2}}{2}\right)\right].$$

Some easy but tedious manipulations now allow us to rewrite this last expression as

$$\begin{aligned}\omega(z,\Omega,E) = \frac{1}{\pi}\Bigg[&\mathrm{Tan}^{-1}\left(\left(\frac{1+r}{1-r}\right)\cdot\left(\frac{\sin\lambda+1}{\cos\lambda}\right)\right)\\ &-\mathrm{Tan}^{-1}\left(\left(\frac{1+r}{1-r}\right)\cdot\left(\frac{\sin\lambda-1}{\cos\lambda}\right)\right)\Bigg]. \qquad (9.3.1)\end{aligned}$$

Recall, however, that

$$\tan(\alpha-\beta) = \frac{\tan\alpha-\tan\beta}{1+\tan\alpha\tan\beta};$$

hence

$$\mathrm{Tan}^{-1}a - \mathrm{Tan}^{-1}b = \mathrm{Tan}^{-1}\left(\frac{a-b}{1+ab}\right). \qquad (9.3.2)$$

Applying this simple idea to (9.3.1) yields

$$\omega(z,\Omega,E) = \frac{1}{\pi}\mathrm{Tan}^{-1}\left(\frac{\left(\frac{1+r}{1-r}\right)\left(\frac{\sin\lambda+1}{\cos\lambda}\right) - \left(\frac{1+r}{1-r}\right)\left(\frac{\sin\lambda-1}{\cos\lambda}\right)}{1+\left(\frac{1+r}{1-r}\right)\left(\frac{\sin\lambda+1}{\cos\lambda}\right)\cdot\left(\frac{1+r}{1-r}\right)\left(\frac{\sin\lambda-1}{\cos\lambda}\right)}\right).$$

Elementary simplifications finally lead to

$$\omega(z,\Omega,E) = \frac{1}{\pi}\mathrm{Tan}^{-1}\left(\frac{1-r^2}{2r\cos\lambda}\right).$$

Now it is helpful to further rewrite this last expression (again using (9.3.2)) as

$$\frac{2\left[\mathrm{Tan}^{-1}\left(\frac{1-r\sin\lambda}{r\cos\lambda}\right) + \mathrm{Tan}^{-1}\left(\frac{1+r\sin\lambda}{r\cos\lambda}\right)\right] - \pi}{2\pi}$$

(remembering, of course, that $\pi/2 - \mathrm{Tan}^{-1}(\gamma) = \mathrm{Tan}^{-1}(1/\gamma)$). But this last is just the formula

$$\frac{2\theta-\alpha}{2\pi},$$

for the special angle $\alpha = \pi$, that was enunciated at the start of the example.

For the case of general E, we may first suppose that E is an arc centered at the point $1 \in \partial D$. Second, we may reduce the general case to the one just calculated by using a Möbius transformation.

Example 9.3.3. Suppose that Ω is an annulus with radii $0 < r_1 < r_2 < \infty$. Then one may verify by inspection that if E is the outer boundary circle of the annulus, then

$$\omega(z, \Omega, E) = \frac{\log(|z|/r_1)}{\log(r_2/r_1)} .$$

In general it is a tricky business to calculate exactly the harmonic measure for a given region Ω and a given E. But one may often obtain useful estimates. The ensuing discussion will bear out this point.

9.4 Hadamard's Three Circles Theorem

Hadamard's three circles theorem (sometimes called the "three lines theorem") is an important sharpening of the classical maximum principle. It has proved useful in various parts of analysis, notably in proving the Riesz–Thorin interpolation theorem for linear operators. Here we shall give a thorough treatment of the three circles theorem from the point of view of harmonic measure. Afterwards we shall discuss the result of Riesz and Thorin.

We begin by treating some general comparison principles regarding the harmonic measure. It will facilitate our discussion to first introduce a slightly more general concept of harmonic measure.

Let $\Omega \subseteq \mathbb{C}$ be, as usual, a domain. Let A be a closed set in the extended plane $\widehat{\mathbb{C}}$. Let E denote that part of $\partial(\Omega \setminus A)$ that lies in A. [In what follows, we will speak of pairs (Ω, A).] Then $\omega(z, \Omega \setminus A, E)$ will be called *the harmonic measure of A with respect to Ω*, assuming that the geometry is simple enough that we can compute this number component-by-component (of $\Omega \setminus A$). Generally speaking, in the present chapter, we assume that all boundaries that arise consist of finitely many Jordan curves. When it is needed, we assume that when we analyze sets like $[\widetilde{\Omega} \setminus \widetilde{A}] \cap \widetilde{A}$, then there will be only finitely many boundary arcs. This standing hypothesis is made so that we can readily apply Lindelöf's Proposition 9.2.1. Often this standing hypothesis will go unspoken.

The first new comparison tool is called the *majorization principle.*

Theorem 9.4.1. *Consider two pairs (Ω, A) and $(\widetilde{\Omega}, \widetilde{A})$. Let*

$$f : \Omega \setminus A \to \widetilde{\Omega}$$

be holomorphic. Assume furthermore that if $\Omega \ni z \to A$, then $\widetilde{\Omega} \ni f(z) \to \widetilde{A}$. Then

$$\omega(z, \Omega, A) \leq \omega(f(z), \widetilde{\Omega}, \widetilde{A})$$

for $z \in f^{-1}(\widetilde{\Omega} \setminus \widetilde{A})$.

Remark 9.4.2. Of course we consider only holomorphic f because we want a mapping that preserves harmonic functions under composition.

Proof of Theorem 9.4.1. Let us abbreviate

$$\omega = \omega(z, \Omega, A) \qquad \text{and} \qquad \widetilde{\omega} = (f(z), \widetilde{\Omega}, \widetilde{A})\,.$$

Now we apply the maximum principle to $\omega - \widetilde{\omega}$ on a connected component V of $f^{-1}(\widetilde{\Omega} \setminus \widetilde{A})$. As z approaches the boundary of V, either z tends to a boundary point of Ω that is not on A, or else $f(z)$ tends to $\widetilde{A}$. In either of these circumstances, $\limsup_{z \to \partial V} (\omega - \widetilde{\omega}) \leq 0$ *except* when $f(z)$ tends to an endpoint of the boundary arcs of $\widetilde{\Omega} \setminus \widetilde{A}$ that lie on $\widetilde{A}$. Since there are only finitely many such points, Lindelöf's Proposition 9.2.1 tells us that the maximum principle still remains valid. We therefore conclude that $\omega \leq \widetilde{\omega}$ on all of $f^{-1}(\widetilde{\Omega} \setminus \widetilde{A})$. □

Corollary 9.4.3. *The function $\omega(z, \Omega, A)$ increases if either Ω increases or A increases. That is to say, if $\Omega \subseteq \Omega^*$ and $A \subseteq A^*$, then*

$$\omega(z, \Omega, A) \leq \omega(z, \Omega^*, A) \qquad \textit{and} \qquad \omega(z, \Omega, A) \leq \omega(z, \Omega, A^*)\,.$$

Remark 9.4.4. This result is reminiscent of the "extension principle" that we treated in Section 6.2 during our study of the Lindelöf principle.

Proof of the Corollary. We prove the first statement and leave the second for the reader.

We apply the theorem with $f : \Omega \setminus A \to \Omega^*$ being the identity mapping. The result is now immediate. □

Now let Ω^* be an open disk of radius $R > 0$ and let A^* be a smaller closed disk of radius $0 < r < R$. One may see by inspection that

$$\omega(f(z), \Omega^*, A^*) = \frac{\log(R/|f(z)|)}{\log R/r}\,.$$

We do not yet say what f is, and in practice there is considerable latitude. Nonetheless, we see immediately that the function ω is identically equal to 1 when $f(z) \in \partial A^*$ and identically equal to 0 when $f(z) \in \partial \Omega^*$. And of course it is harmonic.

Now we have our first result where the quantitative properties of the harmonic measure play a decisive role.

Theorem 9.4.5. *Let f be a holomorphic function on a domain Ω. Let A be a closed set. If $|f(z)| < M$ in Ω and $|f(z)| \leq m < M$ on A, then, for $0 \leq \theta \leq 1$,*

$$|f(z)| \leq m^\theta M^{1-\theta}$$

at points z where $\omega(z, \Omega, A) \geq \theta$.

Remark 9.4.6. This result is of fundamental philosophical importance. It shows how the harmonic measure is a device for interpolating information about the function f.

Proof of Theorem 9.4.5. Of course $f : \Omega \to D(0, M)$. We take $\widetilde{A} = D(0, m)$, $A = f^{-1}(\widetilde{A})$, and $\widetilde{\Omega} = f(\Omega)$. Then we apply the previous theorem and the remark following. So

$$\omega(z, \Omega, A) \le \omega(f(z), \widetilde{\Omega}, \widetilde{A}) = \frac{\log(M/|f(z)|)}{\log M/m} .$$

At a point z for which $\omega(z, \Omega, A) \ge \theta$ we have

$$\theta \le \frac{\log(M/|f(z)|)}{\log M/m} .$$

This inequality is equivalent to the desired conclusion. □

If we take Ω and A in this last theorem to be annuli, then we can draw an important and precise conclusion. Namely, we have this version of the three circles theorem.

Theorem 9.4.7 (Hadamard). *Let f be a holomorphic function on an annulus $\mathcal{A} = \{z \in \mathbb{C} : c < |z| < C\}$. For $c < \tau < C$ we let*

$$M(\tau) \equiv \max_{|z|=\tau} |f(z)| .$$

If $c < r < \rho < R < C$, then we have

$$\log M(\rho) \le \frac{\log R - \log \rho}{\log R - \log r} \cdot \log M(r) + \frac{\log \rho - \log r}{\log R - \log r} \cdot \log M(R) .$$

[*This inequality says quite plainly that* $\log M(s)$ *is a convex function of* $\log s$.]

Proof. Let

$$\Omega = \{z \in \mathbb{C} : r < |z| < R\}$$

and let

$$A = \{z \in \mathbb{C} : r \le |z| \le r + \epsilon\} \text{ , some } \epsilon > 0 .$$

We assume that f is holomorphic on Ω, $|f| \le M$ on Ω, and $|f| \le m$ on A. The preceding theorem tells us that

$$|f(z)| \le m^\theta M^{1-\theta} \tag{9.4.1}$$

on the set where $\omega(z, \Omega, A) \ge \theta$, that is, on the set where

$$\frac{\log(|z|/R)}{\log((r + \epsilon)/R)} \ge \theta .$$

[Note that the roles of $r + \epsilon$ and R are reversed from their occurrence in Example 9.3.3, just because now we are creating harmonic measure for the inside circle of the annulus.]

In particular, inequality (9.4.1) holds when $|z| = (r+\epsilon)^\theta R^{1-\theta}$. We take

$$\theta = \frac{\log R - \log \rho}{\log R - \log(r+\epsilon)}$$

and

$$|z| = (r+\epsilon)^\theta R^{1-\theta} \equiv \rho_\epsilon .$$

Then (9.4.1) translates to

$$M(\rho_\epsilon) \le M(r+\epsilon)^\theta \cdot M(R)^{1-\theta} .$$

Letting $\epsilon \to 0^+$ and taking the logarithm of both sides yields the desired result. □

We note that Ahlfors [AHL2], from which our exposition derives, likes to express the conclusion of this last result as

$$\det \begin{pmatrix} 1 & 1 & 1 \\ \log r & \log \rho & \log R \\ \log M(r) & \log M(\rho) & \log M(R) \end{pmatrix} \ge 0 .$$

We close this section by formulating the three-lines version of Hadamard's theorem. It is proved from Theorem 9.4.7 simply with conformal mapping, and we leave the details to the interested reader.

Theorem 9.4.8. *Let f be a holomorphic function on the strip $S = \{z \in \mathbb{C} : 0 < \operatorname{Re} z < 1\}$. For $0 < x < 1$ we let*

$$M(x) \equiv \max_{\operatorname{Re} z = x} |f(z)| .$$

If $0 < a < \rho < b < 1$, then we have

$$\log M(\rho) \le \frac{\log b - \log \rho}{\log b - \log a} \cdot \log M(a) + \frac{\log \rho - \log a}{\log b - \log a} \cdot \log M(b) .$$

[*This inequality says quite plainly that $\log M(s)$ is a convex function of $\log s$.*]

9.5 A Discussion of Interpolation of Linear Operators

Marcel Riesz discovered the idea of interpolating linear operators in the following context. Perhaps the most important linear operator in all of analysis is the *Hilbert transform*

$$H : f \mapsto \int_{\mathbb{R}} \frac{f(t)}{x-t}\, dt .$$

This operator arises naturally in the study of the boundary behavior of conjugates of harmonic functions on the disk, in the existence and regularity theory for the Laplacian, in the summability theory of Fourier series, and more generally in the theory of singular integral operators. [We provide detailed discussion of these ideas, in context, in Chapter 10 below.] It had been an open problem for some time to show that H is a bounded operator on L^p, $1 < p < \infty$.

It turns out that the boundedness on L^2 is easy and follows from Plancherel's theorem in Fourier analysis. Riesz cooked up some extremely clever tricks to derive boundedness on L^p when p is an even integer. [These ideas are explained in detail in [KRA4] and also in Section 10.3.] He needed to find some way to derive therefrom the boundedness on the "intermediate" L^p spaces. Now we introduce enough language to explain precisely what the concept of "intermediate space" means.

Let X_0, X_1, Y_0, Y_1 be Banach spaces. Intuitively, we will think of a linear operator T such that

$$T : X_0 \to Y_0$$

continuously and

$$T : X_1 \to Y_1$$

continuously. We wish to posit the existence of spaces X_θ and Y_θ, $0 \leq \theta \leq 1$ such that these new spaces are natural "intermediaries" of X_0, Y_0, X_1, Y_1 and, furthermore, that

$$T : X_\theta \to Y_\theta$$

continuously. There is interest in knowing how the norm of T acting on X_θ depends on the norms of T acting on X_0 and X_1.

One rigorous method for approaching this situation is as follows (see [BEL], [STW], [KAT] for our inspiration). Let X_0, Y_0, X_1, Y_1 be given as in the last paragraph. Suppose that

$$T : X_0 \cap X_1 \longrightarrow Y_0 \cup Y_1$$

is a linear operator with the properties that

(i) $\|Tx\|_{Y_0} \leq C_0 \|x\|_{X_0}$ for all $x \in X_0 \cap X_1$;
(ii) $\|Tx\|_{Y_1} \leq C_0 \|x\|_{X_1}$ for all $x \in X_0 \cap X_1$.

Then we want to show that there is a collection of norms $\| \quad \|_{X_\theta}$ and $\| \quad \|_{Y_\theta}$, $0 < \theta < 1$, such that

$$\|Tx\|_{Y_\theta} \leq C_\theta \|x\|_{X_\theta} .$$

So the problem comes down to how to construct the "intermediate norms" $\| \quad \|_{X_\theta}$ and $\| \quad \|_{Y_\theta}$ from the given norms $\| \quad \|_{X_0}$, $\| \quad \|_{Y_0}$, $\| \quad \|_{X_1}$, and $\| \quad \|_{Y_1}$.

There are in fact a number of paradigms for effecting the indicated construction. Most prominent among these are the "real method" (usually attributed to Lyons and Peetre—see [LYP]) and the "complex method" (usually attributed to Calderón—see [CAL]). In this text we shall concentrate on the complex method, which of course was inspired by the ideas of Riesz that were described at the beginning of this section.

Instead of considering the general paradigm for complex interpolation, we shall concentrate on the special case that was of interest to Riesz and Thorin. We now formulate our main theorem.

Theorem 9.5.1. *Let $1 \le p_0 < p_1 \le \infty$ and $1 \le q_0 < q_1 \le \infty$. Suppose that*

$$T : L^{p_0}(\mathbb{R}^2) \cap L^{p_1}(\mathbb{R}^2) \longrightarrow L^{q_0}(\mathbb{R}^2) \cup L^{q_1}(\mathbb{R}^2)$$

is a linear operator satisfying

$$\|Tf\|_{L^{q_0}} \le C_0 \|f\|_{L^{p_0}}$$

and

$$\|Tf\|_{L^{q_1}} \le C_0 \|f\|_{L^{p_1}} .$$

Define p_θ and q_θ by

$$\frac{1}{p_\theta} = (1-\theta)\cdot\frac{1}{p_0} + \theta\cdot\frac{1}{p_1}$$

and

$$\frac{1}{q_\theta} = (1-\theta)\cdot\frac{1}{q_0} + \theta\cdot\frac{1}{q_1} ,$$

with $0 \le \theta \le 1$. Then T is a bounded operator from L^{p_θ} to L^{q_θ} and

$$\|Tf\|_{L^{q_\theta}} \le C_0^{1-\theta} C_1^{\theta} \|f\|_{L^{p_\theta}} .$$

Proof. Fix a nonzero function f that is continuous and with compact support in $\mathbb{R}^2$. We will prove an a priori inequality for this f, and then extend to general f at the end. Now certainly $f \in L^{p_0} \cap L^{p_1}$. We consider the holomorphic function

$$H : z \mapsto \frac{\left(\int_{\mathbb{R}^2} |Tf(x)|^{(1-z)q_0 + zq_1}\, dx\right)^{(1-z)/q_0 + z/q_1}}{\left(\int_{\mathbb{R}^2} |f(x)|^{(1-z)p_0 + zp_1}\, dx\right)^{(1-z)/p_0 + z/p_1}}$$

on the strip $S = \{z \in \mathbb{C} : 0 < \operatorname{Re} z < 1\}$. We define

$$M(s) = \sup_{\operatorname{Re} z = s} |H(z)| , \ 0 < s < 1 .$$

Notice that

$$\begin{aligned} M(0) &= \sup_{t\in\mathbb{R}} \left| \frac{\int_{\mathbb{R}^2} |Tf(x)|^{(1-it)q_0 + itq_1}\, dx^{(1-it)/q_0 + it/q_1}}{\int_{\mathbb{R}^2} |f(x)|^{(1-it)p_0 + itp_1}\, dx^{(1-it)/p_0 + it/p_1}} \right| \\ &\le \sup_{t\in\mathbb{R}} \frac{\int_{\mathbb{R}^2} |Tf(x)|^{q_0}\, dx^{1/q_0}}{\int_{\mathbb{R}^2} |f(x)|^{p_0}\, dx^{1/p_0}} \\ &\le C_0 . \end{aligned}$$

A similar calculation shows that

$$M(1) \le C_1 \,.$$

Obviously this is grist for the three-lines theorem. We may conclude that $M(s) \le C_0^{1-s} \cdot C_1^s$, $0 < s < 1$.

Now let p_z, q_z be given by

$$\frac{1}{p_z} = (1-z) \cdot \frac{1}{p_0} + z \cdot \frac{1}{p_1}$$

and

$$\frac{1}{q_z} = (1-z) \cdot \frac{1}{q_0} + z \cdot \frac{1}{q_1} \,.$$

Now let $z = s + it$. Then

$$\frac{\|Tf\|_{L^{q_z}}}{\|f\|_{L^{p_z}}} \le M(s) \le C_0^{1-s} C_1^s \,.$$

In other words,

$$\|Tf\|_{L^{q_s}} \le C_0^{1-s} C_1^s \|f\|_{L^{p_s}} \,.$$

This is the desired conclusion for a function f that is continuous with compact support.

The result for general $f \in L^{p_\theta}$ can be achieved with a simple approximation argument. □

9.6 The F. and M. Riesz Theorem

One of the classic results of function theory was proved by the brothers F. and M. Riesz. It makes an important statement about the absolute continuity of harmonic measure. We begin with a preliminary result that captures the essence of the theorem. Our exposition here follows the lead of [GARM, Ch. 6].

For $0 < p < \infty$ we define the *Hardy space*

$$H^p(D) = \left\{ f \text{ holomorphic on } D : \sup_{0<r<1} \left(\int_0^{2\pi} |f(re^{i\theta})|^p \, d\theta \right)^{1/p} \equiv \|f\|_{H^p} < \infty \right\} .$$

These spaces are generalizations of the more classical space $H^\infty(D)$ of bounded holomorphic functions. See Chapter 6 for a more thorough discussion of these spaces.

The most important fact about a function in a Hardy space is that it has a boundary function. We enunciate that result here, and refer the reader to Chapter 6 for the details.

Theorem 9.6.1. *Let $0 < p \le \infty$ and let $f \in H^p(D)$. Then there is a function $f^* \in L^p(\partial D)$ such that*

(9.6.1) $\lim_{r \to 1^-} f(re^{i\theta}) = f^*(e^{i\theta})$ *for almost every* $\theta \in [0, 2\pi)$.

(9.6.2) *Let $f_r(e^{i\theta}) = f(re^{i\theta})$ for $0 < r < 1$. Assume that $0 < p < \infty$. Then* $\lim_{r \to 1^-} \|f - f_r\|_{L^p(\partial D)} = 0$.

In the following discussions we shall consider rectifiable curves. Basically, a rectifiable curve is a curve with finite length. The technical definition is that the curve is locally the Lipschitz image of the unit interval (see [FED]). But context will make it clear that the intuitive notion of rectifiability will suffice for our purposes.

Theorem 9.6.2. *Let Ω be a domain such that $\gamma = \partial\Omega$ is a Jordan curve. Let*

$$\varphi : D \to \Omega$$

be a conformal map. Then the curve γ is rectifiable if and only if $\varphi' \in H^1$. In case $\phi' \in H^1$, then we have

$$\|\varphi'\|_{H^1} = \text{length}(\gamma) = \mathcal{H}^1(\gamma)\,. \tag{9.6.3}$$

Here $\mathcal{H}^1$ is one-dimensional Hausdorff measure.

Remark 9.6.3. It is known (see [KRA1] and our Chapter 6) that a function in H^1 has a boundary limit function that is in $L^1(\partial D)$. Thus the hypothesis of the theorem says essentially that φ is absolutely continuous on the boundary of D. It makes sense then that φ would preserve length on the boundary. Theorem 9.6.2 is essentially equivalent to the result of F. and M. Riesz that we described in Remark 9.2.4.

Proof of Theorem 9.6.2. Once again we invoke Carathéodory's theorem; thus we know that φ and its inverse extend continuously and univalently to their respective boundaries. Let us assume that $\varphi' \in H^1$. Let

$$0 = \theta_0 < \theta_1 < \theta_2 < \cdots < \theta_k = 2\pi$$

be a partition of $[0, 2\pi]$. Then of course

$$\begin{aligned}\sum_{j=1}^{k} \left|\varphi(e^{i\theta_j}) - \varphi(e^{i\theta_{j-1}})\right| &= \lim_{r \to 1^-} \sum_{j=1}^{k} \left|\varphi(re^{i\theta_j}) - \varphi(re^{i\theta_{j-1}})\right| \\ &= \lim_{r \to 1^-} \sum_{j=1}^{k} \left| \int_{\theta_{j-1}}^{\theta_j} \varphi'(re^{i\theta}) i r e^{i\theta}\, d\theta \right| \\ &\le \|\varphi'\|_{H^1}\,. \end{aligned} \tag{9.6.4}$$

But plainly the length of γ is the supremum, over all partitions of $[0, 2\pi)$, of the left-hand side of (9.6.4). We conclude that γ is rectifiable and

$$\text{length}(\gamma) \le \|\varphi'\|_{H^1}.$$

For the converse, assume that γ is rectifiable. If $0 < r < 1$ is fixed, then let $\gamma_r = \varphi(\{z \in \mathbb{C} : |z| = r\})$. Let $\epsilon > 0$. Now choose a partition $\{\theta_0, \theta_1, \ldots, \theta_k\}$ of the interval $[0, 2\pi)$ as before so that

$$\sum_{j=1}^{k} \left|\varphi(re^{i\theta_j}) - \varphi(re^{i\theta_{j-1}})\right| \ge \text{length}(\gamma_r) - \epsilon .$$

We write

$$\eta(z) = \sum_{j=1}^{k} \left|\varphi(ze^{i\theta_j}) - \varphi(ze^{i\theta_{j-1}})\right| .$$

Then η, being the sum of absolute values of holomorphic functions, is subharmonic. By Carathéodory's theorem, η is continuous on $\overline{D}$. Hence

$$\sup_D \eta(z) = \sup_\theta \eta(e^{i\theta}) \le \text{length}(\gamma) .$$

We conclude that

$$\int_0^{2\pi} |\varphi'(re^{i\theta})|\, d\theta = \text{length}(\gamma_r) \le \eta(r) + \epsilon \le \text{length}(\gamma) + \epsilon .$$

It follows that $\phi' \in H^1$ and equality (9.6.3) is valid. We conclude by noting that, for a rectifiable curve, the ordinary notion of length and the one-dimensional Hausdorff measure are the same. □

The reason that we include Theorem 9.6.2 in the present chapter—apart from its general aesthetic interest—is that it can be interpreted in the language of harmonic measure. Let notation be as in Theorem 9.6.2. To wit, let $\gamma = \partial\Omega$ be a rectifiable Jordan curve as above and let $F \subseteq \gamma$ be a subcurve. Let $F = \varphi(E)$ and $\alpha = \varphi(0)$. Then Carathéodory's theorem, Example 9.3.2, and Proposition 9.2.3 tell us that

$$\omega(\alpha, \Omega, F) = \omega(0, D, E) = \frac{1}{2\pi}|E| ,$$

where the $|\ \cdot\ |$ denotes arc length. Since F is an arc, we see that Theorem 9.6.1 and the proof of (9.6.3) tell us that

$$\mathcal{H}^1(F) = \mathcal{H}^1(\varphi(E)) = \lim_{r\to 1^-} \int_E |\varphi'(re^{i\theta})|\, d\theta = \int_E |\varphi'(e^{i\theta})|\, d\theta . \tag{9.6.5}$$

Of course if (9.6.5) holds for arcs, then, by passing to unions and intersections, we see that it holds for Borel sets F. Thus we derive the important conclusion

$$\omega(\alpha, \Omega, F) = 0 \Rightarrow \mathcal{H}^1(F) = 0\,.$$

Conversely, since (by a standard uniqueness theorem—see [KOO]) $|\{\theta : |\varphi'(e^{i\theta})| = 0\}| = 0$, we see that

$$\mathcal{H}^1(F) = 0 \Rightarrow \omega(\alpha, \Omega, F) = 0\,.$$

What we have proved, then, is that when $\gamma = \partial\Omega$ is rectifiable then harmonic measure for Ω and linear measure on γ are mutually absolutely continuous. We now summarize this result in a formally enunciated theorem.

Theorem 9.6.4 (F. and M. Riesz, 1916). *Let Ω be a simply connected planar domain such that $\gamma = \partial\Omega$ is a rectifiable Jordan curve. Assume that*

$$\varphi : D \to \Omega$$

is conformal. Then $\varphi' \in L^1(\partial D)$. For any Borel set $E \subseteq \partial D$ it holds that

$$\mathcal{H}^1(\varphi(E)) = \int_E |\varphi'(e^{i\theta}|\, d\theta\,.$$

Also, for any Borel set $F \subseteq \partial\Omega$ and any point $\alpha \in \Omega$,

$$\omega(\alpha, \Omega, F) = 0 \Leftrightarrow \mathcal{H}^1(F) = 0\,.$$

We conclude by noting that Theorems 9.6.2 and 9.6.4 are equivalent, just because formula (9.6.3) is valid precisely when $\varphi' \in H^1$.

Problems for Study and Exploration

1. Let $\Omega = \{z \in \mathbb{C} : -1 < \operatorname{Im} z < 1\}$ and let $E \subseteq \partial\Omega$ be the set $E = \{t + i : -\infty < t < \infty\}$. Calculate the harmonic measure of E with respect to Ω. Can you relate this problem to Example 9.3.3?
2. Let $\Omega = \{z \in \mathbb{C} : -1 < \operatorname{Im} z < 1\}$ and let $E \subseteq \partial\Omega$ be the set $E = \{t + i : -1 < t < 1\}$. Calculate the harmonic measure of E with respect to Ω.
3. Let U be the upper halfplane and let $E = \{x + i0 : x \leq 0\}$. Calculate the harmonic measure of E with respect to U.
4. Use translation and subtraction, together with the result of Exercise 3, to calculate the harmonic measure of an interval in the boundary of the upper halfplane. Compare with Example 9.3.1.
5. Kakutani's theorem from 1944 (see [KAK]) asserts that the harmonic measure of a set $E \subseteq \partial\Omega$ evaluated at a point $z \in \Omega$ is equal to the probability that a Brownian traveler in Ω starting at z will exit the domain at a point of E. Look up the basics of Brownian motion (for example, in the books [DUR] or [NEL]) and explain what Kakutani's theorem means. Use the theorem to give an elegant solution to Exercise 3.

6. Let D be the unit disk. A set $E \subseteq \partial D$ has harmonic measure identically 0 with respect to D. What can you conclude about E?

7. What does harmonic measure—on the disk let us say—have to do with the Poisson kernel? Give a detailed answer.

8. Let $\Omega \subseteq \mathbb{C}$ be a domain and let $E, E' \subseteq \partial\Omega$. Suppose that

$$\omega(z, \Omega, E) \geq \omega(z, \Omega, E')$$

for all $z \in \Omega$. What can you conclude about E, E'?

9. Refer to Exercise 5. Suppose that $E, E' \subseteq \partial\Omega$ and assume that $E \subseteq E'$. Is there any relationship between $\omega(z, \Omega, E)$ and $\omega(z, \Omega, E')$?

10. Refer to Exercises 5 and 9. Suppose that $E, E' \subseteq \partial\Omega$. What relationships are there, if any, among $\omega(z, \Omega, E)$, $\omega(z, \Omega, E')$, $\omega(z, \Omega, E \cap E')$, and $\omega(z, \Omega, E \cup E')$?

11. Let D be the disk. Prove that, if $\omega(z, D, E)$ is not identically 0, then $\omega(0, D, E)$ is not 0.

12. Refer to Exercise 11. Prove that, if $\omega(z, \Omega, E)$ is not identically 0, then in fact it is never 0.

10

Conjugate Functions and the Hilbert Transform

Genesis and Development

A major theme of analysis in the early twentieth century was the study of convergence of Fourier series. There are two basic types of convergence: pointwise convergence and norm convergence. The study of norm convergence gives rise rather quickly to the study of the Hilbert transform.

Specifically, let f be an L^p function on the circle and let $S_N f$ be the partial sums of its Fourier series. It is a fact that $S_N f \to f$ in L^p norm *for all such* f if and only if the Hilbert transform is bounded in norm on L^p. This statement alone was a triumph of the very new subject of functional analysis.

It is straightforward to see that the Hilbert transform is bounded on L^2. In fact L^2 convergence of Fourier series can be proved by many different means; the most convenient and elegant is to use elementary Hilbert space theory. But the situation for $p \neq 2$ is difficult.

It was Marcel Riesz who, in 1926, was finally able to establish the L^p boundedness of the Hilbert transform for $1 < p < \infty$. One of the bonuses of his stunning proof is that he invented the subject of interpolation of linear operators.

We shall touch on all these topics in the present chapter. A very basic familiarity with functional analysis will be helpful here.

10.0 Introduction

One of the great sagas of twentieth-century analysis is the development of the connection between the boundary behavior of holomorphic functions on the disk, the existence of conjugate harmonic functions, and how both these phenomena are controlled and explained by Fourier series and the Hilbert transform. In the present chapter we will tell a good part of this story.

The presentation here will be partly complex function theory, partly the theory of integral operators, and partly harmonic analysis. There is also some functional analysis thrown into the mix. It is the stepping stone to a lot of good analysis, much of which is still studied today.

10.1 Discovering the Hilbert Transform

We saw in Section 6.1 that a function f in the Hardy class $H^p(D)$ on the disk may be identified in a natural way with its boundary function, which we continue to call f^*. Fix attention for the moment on $p = 1$.

If $\phi \in L^1(\partial D)$ and is real-valued, then we may define a harmonic function u on the disk by

$$u(re^{i\theta}) = \frac{1}{2\pi} \int_0^{2\pi} \phi(e^{i\psi}) \frac{1 - r^2}{1 - 2r\cos(\theta - \psi) + r^2} \, d\psi.$$

Of course this is just the usual Poisson integral of ϕ. As was proved in Section 6.1, the function ϕ is the "boundary function" of u in a natural manner. Let v be the harmonic conjugate of u on the disk (we may make the choice of v unique by demanding that $v(0) = 0$). Thus $h \equiv u + iv$ is holomorphic. We may then ask whether the function v has a boundary limit function $\widetilde{\phi}$.

To see that $\widetilde{\phi}$ exists, we reason as follows: Suppose that the original function ϕ is nonnegative (any real-valued ϕ is the difference of two such functions, so there is no loss in making this additional hypothesis). Since the Poisson kernel is positive, it follows that $u > 0$. Now consider the holomorphic function

$$F = e^{-u-iv}.$$

The positivity of u implies that F is bounded. Thus $F \in H^\infty$. By Theorem 6.1.23, we may conclude that F has radial boundary limits at almost every point of ∂D. Unraveling our notation (and thinking a moment about the ambiguity caused by multiples of 2π), we find that v itself has radial boundary limits almost everywhere. We define thereby the function $\widetilde{\phi}$.

Of course the function h can be expressed (up to an additive factor of 1/2 and a multiplicative factor of 1/2) as the Cauchy integral of ϕ. The real part of the Cauchy kernel is the Poisson kernel (again up to a multiplicative and additive factor of 1/2—see the calculation following), so it makes sense that the real part of h on D converges back to ϕ. By the same token, the imaginary part of h is the integral of ϕ against the imaginary part of the Cauchy kernel, and it will converge to $\widetilde{\phi}$. It behooves us to calculate the imaginary part of the Cauchy kernel.

In the Cauchy integrand $(1/2\pi i)[1/(\zeta - z)]d\zeta$, we set $z = re^{i\theta}$ and $\zeta = e^{i\psi}$. Then

$$\begin{aligned}\frac{1}{2\pi i}\frac{1}{\zeta - z}\,d\zeta &= \frac{-i\overline{\zeta}\,d\zeta}{2\pi\overline{\zeta}(\zeta - z)} \\ &= \frac{d\psi}{2\pi(1 - re^{i(\theta-\psi)})} = \frac{(1 - re^{-i(\theta-\psi)})d\psi}{2\pi|1 - re^{i(\theta-\psi)}|^2} \\ &= \frac{[1 - r\cos(\theta-\psi)] + i[r\sin(\theta-\psi)]}{2\pi|1 - re^{i(\theta-\psi)}|^2}d\psi\,. \end{aligned} \qquad (10.1.1)$$

For convenience, let us set $t = \theta - \psi$. Then the imaginary part of the Cauchy kernel, as just normalized and calculated, equals

$$\widetilde{P}_r(e^{it}) \equiv \frac{1}{2\pi} \cdot \frac{r\sin t}{1 - 2r\cos t + r^2}.$$

[If we denote the real part of (10.1.1) by P_r^*, then a quick calculation shows that $P_r^* - 1/[4\pi] = [1/2]\cdot P_r$, where P_r is the usual Poisson kernel.] We conclude that

$$v(re^{i\theta}) = 2 \cdot \widetilde{P}_r * \phi(e^{i\theta}).$$

If we formally let $r \to 1^-$ (and suppress some nasty details—see [KAT]), then we find that $\widetilde{\phi}$ is just the convolution of ϕ with the kernel

$$\begin{aligned}\widetilde{k}(t) &\equiv 2 \cdot \frac{\sin t}{2 - 2\cos t} \\ &= 2 \cdot \cdot \frac{2\sin t/2\cos t/2}{2[1 - \cos^2 t/2 + \sin^2 t/2]} \\ &= \cot(t/2). \end{aligned}$$

The operator

$$H : \phi \mapsto \phi * \cot(t/2) \qquad (10.1.2)$$

is the Hilbert transform from the theory of Fourier series (see Chapter 1 of [KRA4]). Its study is motivated by questions of norm convergence of classical Fourier series.

10.2 The Modified Hilbert Transform

We first note that, in practice, people do not actually look at the operator consisting of convolution with $\cot \frac{t}{2}$. This kernel is a transcendental function, and is tedious to handle. Thus what we do instead is to look at the operator

$$\mathcal{H} : f \mapsto \text{P.V.}\, \frac{1}{2\pi}\int_{-\pi}^{\pi} f(x - t) \cdot \frac{2}{t}\,dt. \qquad (10.2.1)$$

Here P.V. stands for "principal value." This idea is discussed in greater detail in the next section; it is a way to make sense of a formally nonconvergent integral. Clearly the kernel $2/t$ is much easier to think about than $\cot \frac{t}{2}$. It is also homogeneous of degree -1, a fact that will prove significant when we adopt a broader point of view later. But what gives us the right to replace the complicated integral (10.1.2) by the apparently simpler integral (10.2.1)?

Let us examine the difference

$$I(t) \equiv \cot\left(\frac{t}{2}\right) - \frac{2}{t}, \quad 0 < |t| < 2\pi,$$

which we extend to $\mathbb{R} \setminus 2\pi\mathbb{Z}$ by 2π-periodicity. Using the Taylor expansions of sine and cosine near the origin (and exploiting Landau's notation), we may write

$$\begin{aligned} I(t) &= \frac{1+\mathcal{O}(t^2)}{t/2+\mathcal{O}(t^3)} - \frac{2}{t} \\ &= \frac{2+\mathcal{O}(t^2)}{t+\mathcal{O}(t^3)} - \frac{2}{t} \\ &= \frac{2}{t} \cdot \left[\frac{1+\mathcal{O}(t^2)}{1+\mathcal{O}(t^3)} - 1\right] \\ &= \frac{2}{t} \cdot \mathcal{O}(t^2) \\ &= \mathcal{O}(t). \end{aligned}$$

Thus the difference between the two kernels under study is (better than) a bounded function. In particular, it is in every L^p class. So we think of

$$\begin{aligned} \text{P.V.}\, \frac{1}{2\pi}\int_{-\pi}^{\pi} f(x-t)\cot\left(\frac{t}{2}\right) dt &= \text{P.V.}\, \frac{1}{2\pi}\int_{-\pi}^{\pi} f(x-t)\cdot\frac{2}{t}\, dt \\ &\quad + \text{P.V.}\, \frac{1}{2\pi}\int_{-\pi}^{\pi} f(x-t)\cdot I(t)\, dt \\ &= \mathcal{H}f(x) + \frac{1}{2\pi}\int_{-\pi}^{\pi} f(t) I(x-t)\, dt \\ &\equiv J_1 + J_2. \end{aligned}$$

By Schur's lemma (see [STW]), the integral J_2 is trivial to study: The integral operator

$$\phi \mapsto \phi * I$$

is bounded on every L^p space. Thus, in order to study the Hilbert transform, it suffices for us to study the integral in (10.2.1). In practice, harmonic analysts study the integral in (10.2.1) and refer to it as the "(modified) Hilbert transform" without further comment. To repeat, the beautiful fact is that the original Hilbert transform is bounded on a given L^p space *if and only if* the new (modified) transform (10.2.1) is bounded on that same L^p. [In practice it is convenient to forget about the "2" in the numerator of the kernel in (10.2.1).]

10.3 The Hilbert Transform and Fourier Series

In this section we will explain the connection of the Hilbert transform with the theory of Fourier series.

The Hilbert transform H is one of the most important linear operators in all of mathematical analysis. First, it is the key to all convergence questions for the partial sums of Fourier series. Second, it is a paradigm for all singular integral operators on Euclidean space. Third, the analogue of the Hilbert transform on the line is uniquely determined by its invariance properties with respect to the groups that act naturally on 1-dimensional Euclidean space.

We first must review the key concepts from Fourier series. From a theoretical point of view, it is convenient (see [KRA4]) to think of Fourier analysis as being done on the circle group $\mathbb{T} \equiv \mathbb{R}/2\pi\mathbb{Z}$. But in practice we work on the interval $[0, 2\pi)$ with ordinary Lebesgue measure. Notice that this interval contains one element from each equivalence class in $\mathbb{T}$.

Definition 10.3.1. Let f be an integrable function on $\mathbb{T}$. For $j \in \mathbb{Z}$, we define

$$\widehat{f}(j) = a_j \equiv \frac{1}{2\pi} \int_0^{2\pi} f(t) e^{-ijt} \, dt.$$

We call $\widehat{f}(j) = a_j$ the j^{th} *Fourier coefficient* of f.

The first fact we notice about Fourier coefficients is this:

Lemma 10.3.2. *If f is integrable then, for each j,*

$$|\widehat{f}(j)| \le \|f\|_{L^1} .$$

Proof. Immediate from the definition of $\widehat{f}(j)$. □

One of the most profound, but truly elementary, facts about Fourier coefficients is the next result of Riemann and Lebesgue about the asymptotic vanishing of these terms:

Lemma 10.3.3. *Let f be integrable. Then*

$$\lim_{j \to \pm\infty} |\widehat{f}(j)| = 0 .$$

Proof. First consider the case when f is a trigonometric polynomial, which means that

$$f(t) = \sum_{j=-M}^{M} a_j e^{ijt}.$$

Then $\widehat{f}(j) = 0$ as soon as $|j| \ge M$. That proves the result for trigonometric polynomials.

Now let f be any integrable function. Let $\epsilon > 0$. Choose a trigonometric polynomial p such that $\|f - p\|_{L^1} < \epsilon$ (use the Stone-Weierstrass theorem [RUD2]). Let N be the degree of the trigonometric polynomial p and let $|j| > N$. Then

$$\begin{aligned} |\widehat{f}(j)| &\le \left| [f-p]\widehat{\ }(j) \right| + |\widehat{p}(j)| \\ &\le \|f-p\|_{L^1} + 0 \\ &< \epsilon. \end{aligned}$$

This proves the result. □

In the subject of Fourier series, it is convenient to build a factor of $1/2\pi$ into our integrals. We have already seen this feature in the definition of the Fourier coefficients. But we will also let

$$\|f\|_{L^p(\mathbb{T})} \equiv \left[\frac{1}{2\pi} \int_0^{2\pi} |f(e^{it})|^p \, dt \right]^{1/p}, \quad 1 \le p < \infty.$$

The fundamental issue of Fourier analysis is this: We introduce the formal expression

$$Sf \sim \sum_{j=-\infty}^{\infty} \widehat{f}(j) e^{ijt}. \tag{10.3.1}$$

We call the expression *formal*, because we do not know whether the series converges; if it does converge, we do not know whether it converges to the function f.[1]

Hot debate over the summing of Fourier series was brought to a halt by P. Dirichlet in 1828; for the first time in the context of Fourier series, he defined what was meant for a series to *converge*. And he proved a version of our Theorem 10.3.5 below. Given our current perspective, it is clear that the first step in such a program is to define the notion of a partial sum.

Definition 10.3.4. Let f be an integrable function on $\mathbb{T}$ and let the formal Fourier series of f be as in (10.3.1). We define the N^{th} *partial sum* of f to be the expression

$$S_N f(x) = \sum_{j=-N}^{N} \widehat{f}(j) e^{ijx}.$$

We say that the Fourier series *converges* to f at the point x if

$$S_N f(x) \to f(x) \qquad \text{as } N \to \infty$$

in the sense of convergence of ordinary sequences of complex numbers.

[1] Recall here the theory of Taylor series from calculus: the Taylor series for a typical C^∞ function g generally does not converge, and when it does converge it does not typically converge to the function g.

It is most expedient to begin our study of summation of Fourier series by finding an integral formula for $S_N f$. Thus we write

$$\begin{aligned} S_N f(x) &= \sum_{j=-N}^{N} \widehat{f}(j) e^{ijx} \\ &= \sum_{j=-N}^{N} \frac{1}{2\pi} \int_0^{2\pi} f(t) e^{-ijt}\, dt\; e^{ijx} \\ &= \frac{1}{2\pi} \int_0^{2\pi} \left[\sum_{j=-N}^{N} e^{ij(x-t)} \right] f(t)\, dt. \end{aligned} \tag{10.3.2}$$

We need to calculate the sum in brackets; for that will be a universal object associated to the summation process S_N, and unrelated to the particular function f that we are considering.

Now

$$\sum_{j=-N}^{N} e^{ijs} = e^{-iNs} \sum_{j=0}^{2N} e^{ijs} = e^{-iNs} \sum_{j=0}^{2N} \left[e^{is} \right]^j. \tag{10.3.3}$$

The sum on the right is a geometric sum, and we may instantly write a formula for it (as long as $s \neq 0 \bmod 2\pi$):

$$\sum_0^{2N} [e^{is}]^j = \frac{e^{i(2N+1)s} - 1}{e^{is} - 1}.$$

Substituting this expression into (10.3.3) yields

$$\begin{aligned} \sum_{j=-N}^{N} e^{ijs} &= e^{-iNs}\, \frac{e^{i(2N+1)s} - 1}{e^{is} - 1} \\ &= \frac{e^{i(N+1)s} - e^{-iNs}}{e^{is} - 1} \\ &= \frac{e^{i(N+1)s} - e^{-iNs}}{e^{is} - 1} \cdot \frac{e^{-is/2}}{e^{-is/2}} \\ &= \frac{e^{i(N+1/2)s} - e^{-i(N+1/2)s}}{e^{is/2} - e^{-is/2}} \\ &= \frac{\sin(N + \frac{1}{2})s}{\sin \frac{1}{2} s}. \end{aligned}$$

We see that we have derived a closed formula (no summation signs) for the relevant sum. In other words, using (10.3.2), we now know that

$$S_N f(x) = \frac{1}{2\pi} \int_0^{2\pi} \frac{\sin \left[N + \frac{1}{2}\right](x-t)}{\sin \frac{x-t}{2}} f(t)\, dt.$$

The expression

$$D_N(s) = \frac{\sin \left[N + \frac{1}{2}\right] s}{\sin \frac{s}{2}}$$

is called the *Dirichlet kernel*. It is the fundamental object in any study of the summation of Fourier series. In summary, our formula is

$$S_N f(x) = \frac{1}{2\pi} \int_0^{2\pi} D_N(x-t) f(t)\, dt$$

or (after a change of variable—where we exploit the fact that our functions are periodic to retain the same limits of integration)

$$S_N f(x) = \frac{1}{2\pi} \int_0^{2\pi} D_N(t) f(x-t)\, dt.$$

For now, we notice that

$$\begin{aligned}
\frac{1}{2\pi} \int_0^{2\pi} D_N(t)\, dt &= \frac{1}{2\pi} \int_0^{2\pi} \sum_{j=-N}^{N} e^{ijt}\, dt \\
&= \sum_{j=-N}^{N} \frac{1}{2\pi} \int_0^{2\pi} e^{ijt}\, dt \\
&= \frac{1}{2\pi} \int_0^{2\pi} e^{i0t}\, dt \\
&= 1.
\end{aligned}$$

Departing from many popular studies, we will begin our work with the following not-terribly-well-known theorem of Dirichlet:

Theorem 10.3.5. *Let f be an integrable function on $\mathbb{T}$ and suppose that f is differentiable at x. Then $S_N f(x) \to f(x)$.*

An immediate corollary of the theorem is that the Fourier series of a differentiable function converges to that function at *every* point. Contrast this result with the situation for Taylor series!

Proof of the Theorem. We examine the expression $S_N f(x) - f(x)$:

$$\begin{aligned}
|S_N f(x) - f(x)| &= \left| \frac{1}{2\pi} \int_0^{2\pi} D_N(t) f(x-t)\, dt - f(x) \right| \\
&= \left| \frac{1}{2\pi} \int_0^{2\pi} D_N(t) f(x-t)\, dt - \frac{1}{2\pi} \int_0^{2\pi} D_N(t) f(x)\, dt \right| .
\end{aligned}$$

Notice that something very important has transpired in the last step: we used the fact that $\frac{1}{2\pi}\int_0^{2\pi} D_N(t)\,dt = 1$ to rewrite the simple expression $f(x)$ (which is constant with respect to the variable t) in an interesting fashion; this step will allow us to combine the two expressions inside the absolute value signs.

Thus we have

$$|S_N f(x) - f(x)| = \left| \frac{1}{2\pi} \int_0^{2\pi} D_N(t)[f(x-t) - f(x)]\,dt \right| .$$

We may translate f so that $x = 0$, and (by periodicity) we may perform the integration from $-\pi$ to π (instead of from 0 to 2π). Thus our integral is

$$P_N \equiv \left| \frac{1}{2\pi} \int_{-\pi}^{\pi} D_N(t)[f(t) - f(0)]\,dt \right| .$$

Note that another change of variable has allowed us to replace $-t$ by t. Now fix $\epsilon > 0$ and write

$$\begin{aligned} P_N \leq & \left\{ \left| \frac{1}{2\pi} \int_{-\pi}^{-\epsilon} D_N(t)[f(t) - f(0)]\,dt \right| \right. \\ & \quad + \left. \left| \frac{1}{2\pi} \int_{\epsilon}^{\pi} D_N(t)[f(t) - f(0)]\,dt \right| \right\} \\ & \quad + \left| \frac{1}{2\pi} \int_{-\epsilon}^{\epsilon} D_N(t)[f(t) - f(0)]\,dt \right| \\ \equiv & \ I + II . \end{aligned}$$

We may note that

$$\sin(N + 1/2)t = \sin Nt \cos t/2 + \cos Nt \sin t/2$$

and thus rewrite I as

$$\begin{aligned} & \left| \frac{1}{2\pi} \int_{-\pi}^{-\epsilon} \sin Nt \left[\cos \frac{1}{2} t \cdot \frac{f(t) - f(0)}{\sin \frac{t}{2}} \right] dt \right. \\ & \quad \left. + \frac{1}{2\pi} \int_{-\pi}^{-\epsilon} \cos Nt \left[\sin \frac{1}{2} t \cdot \frac{f(t) - f(0)}{\sin \frac{t}{2}} \right] dt \right| \\ & \quad + \left| \frac{1}{2\pi} \int_{\epsilon}^{\pi} \sin Nt \left[\cos \frac{1}{2} t \cdot \frac{f(t) - f(0)}{\sin \frac{t}{2}} \right] dt \right. \\ & \quad \left. + \frac{1}{2\pi} \int_{\epsilon}^{\pi} \cos Nt \left[\sin \frac{1}{2} t \cdot \frac{f(t) - f(0)}{\sin \frac{t}{2}} \right] dt \right| . \end{aligned}$$

These four expressions are all analyzed in the same way, so let us look at the first of them. The expression

$$\left[\cos \frac{1}{2} t \cdot \frac{f(t) - f(0)}{\sin \frac{t}{2}} \right] \chi_{[-\pi,-\epsilon]}(t)$$

is an integrable function (because t is bounded from zero on its support). Call it $g(t)$. Then our first integral may be written as

$$\left| \frac{1}{2\pi} \frac{1}{2i} \int_{-\pi}^{\pi} e^{iNt} g(t)\, dt - \frac{1}{2\pi} \frac{1}{2i} \int_{-\pi}^{\pi} e^{-iNt} g(t)\, dt \right| .$$

Each of these last two expressions is ($1/2i$ times) the $\pm N^{\text{th}}$ Fourier coefficient of the integrable function g. The Riemann–Lebesgue lemma tells us that, as $N \to \infty$, they tend to zero. The other three parts of I are handled in the same way. That takes care of I.

The analysis of II is similar, but slightly more delicate. First observe that

$$f(t) - f(0) = \mathcal{O}(t).$$

[Here $\mathcal{O}(t)$ is Landau's notation for an expression that is not greater than $C \cdot |t|$.] More precisely, the differentiability of f at 0 tells us immediately that $[f(t) - f(0)]/t \to f'(0)$, hence that $|f(t) - f(0)| \le C \cdot |t|$ for t small.

Thus

$$II = \left| \frac{1}{2\pi} \int_{-\epsilon}^{\epsilon} \frac{\sin \left[N + \frac{1}{2}\right] t}{\sin \frac{t}{2}} \cdot \mathcal{O}(t)\, dt \right| .$$

Regrouping terms, as we did in our estimate of I, we see that

$$II \le \left| \frac{1}{2\pi} \int_{-\epsilon}^{\epsilon} \sin Nt \left[\cos \frac{t}{2} \cdot \frac{\mathcal{O}(t)}{\sin \frac{t}{2}} \right] dt \right| + \left| \frac{1}{2\pi} \int_{-\epsilon}^{\epsilon} \cos Nt \left[\sin \frac{t}{2} \cdot \frac{\mathcal{O}(t)}{\sin \frac{t}{2}} \right] dt \right| .$$

The expressions in brackets are integrable functions (in the first instance, because $\mathcal{O}(t)$ cancels the singularity that would be induced by $\sin[t/2]$), and (as before) integration against $\cos Nt$ or $\sin Nt$ amounts to calculating a $\pm N^{\text{th}}$ Fourier coefficient. As $N \to \infty$, these tend to zero by the Riemann–Lebesgue lemma.

To summarize, our expression P_N tends to 0 as $N \to \infty$. That is what we wished to prove. □

Now we define the Hilbert transform—from the Fourier series perspective—and see what it says about the summation of Fourier series.

We begin by defining the Hilbert transform as a multiplier operator. Indeed, let $\mathbf{h} = \{h_j\}$, with $h_j = -i \operatorname{sgn} j$; here the convention is that

$$\operatorname{sgn} x = \begin{cases} -1 \text{ if } x < 0 \\ \ \ 0 \text{ if } x = 0 \\ \ \ 1 \text{ if } x > 0 . \end{cases}$$

Then

$$Hf \equiv \mathcal{M}_{\mathbf{h}} f ,$$

where $\mathcal{M}_{\mathbf{h}}$ is the so-called "multiplier operator" that assigns to f the new function with Fourier series

$$\sum_{j=-\infty}^{\infty} h_j \widehat{f}(j) e^{ijx}\,.$$

So defined, the Hilbert transform has the following connection with the partial sum operators:

$$\begin{aligned}\chi_{[-N,N]}(j) &= \frac{1}{2}\big[1+\operatorname{sgn}(j+N)\big] - \frac{1}{2}\big[1+\operatorname{sgn}(j-N)\big] \\ &\qquad + \frac{1}{2}\left[\chi_{\{-N\}}(j) + \chi_{\{N\}}(j)\right] \\ &= \frac{1}{2}\big[\operatorname{sgn}(j+N) - \operatorname{sgn}(j-N)\big] \\ &\qquad + \frac{1}{2}\left[\chi_{\{-N\}}(j) + \chi_{\{N\}}(j)\right].\end{aligned}$$

Therefore

$$S_N f(e^{it}) = ie^{-iNt} H[e_N f] - ie^{iNt} H[e_{-N} f] + \frac{1}{2}\left[\mathcal{P}_{-N} f + \mathcal{P}_N f\right], \qquad (10.3.4)$$

where $\mathcal{P}_j$ is the orthogonal projection onto the space spanned by e^{ijx}. Here e_N is the operator that shifts the Fourier series of f by N (i.e., multiplies f by e^{iNx}).

To understand this last equality, let us examine a piece of it. Let us look at the linear operator corresponding to the multiplier

$$m(j) \equiv \operatorname{sgn}(j+N).$$

Let $f(t) \sim \sum_{j=-\infty}^{\infty} \widehat{f}(j) e^{itj}$. Then

$$\begin{aligned}\mathcal{M}_m f(t) &= \sum_j \operatorname{sgn}(j+N) \widehat{f}(j) e^{ijt} \\ &= \sum_j \operatorname{sgn}(j) e^{-iNt} \widehat{f}(j-N) e^{ijt} \\ &= ie^{-iNt} \sum_j (-i)\operatorname{sgn}(j) \widehat{f}(j-N) e^{ijt} \\ &= ie^{-iNt} \sum_j (-i)\operatorname{sgn}(j) \big(e_N f\big)\widehat{\ }(j) e^{ijt} \\ &= ie^{-iNt} H[e_N f](t).\end{aligned}$$

This is of course precisely what is asserted in the first half of the right-hand side of (10.3.4). The other parts of (10.3.4) may be understood similarly.

We shall see that the Hilbert transform is bounded on L^2. In fact one can see this result immediately from Bessel's inequality in the theory of Hilbert space. We shall give an independent proof below that H is bounded on L^p for $1 < p < \infty$. Similar remarks apply to the projection operators $\mathcal{P}_j$, just because they are Hilbert space projections. Taking these boundedness assertions for granted, we now re-examine equation (10.3.4). Multiplication by a complex exponential does not change the size of an L^p function (in technical language, it is an *isometry* of L^p). So (10.3.4) tells us that S_N is a sum and difference of compositions of operators, all of which are bounded on L^p. And the norm is plainly bounded independent of N. If we assume that H is bounded on L^p, $1 < p < \infty$, then an easy density argument[2] shows that that norm convergence holds in L^p for $1 < p < \infty$. We now state this as a theorem:

Theorem 10.3.6. *Let $1 < p < \infty$ and assume that the Hilbert transform H is bounded on L^p. Let $f \in L^p(\mathbb{T})$. Then $\|S_N f - f\|_{L^p} \to 0$ as $N \to \infty$. Explicitly,*

$$\lim_{N\to\infty} \left[\int_{\mathbb{T}} |S_N f(x) - f(x)|^p \, dx \right]^{1/p} = 0.$$

The converse statement is true as well (see [KAT] or [KRA4] for the details).

It is useful in the study of the Hilbert transform to be able to express it explicitly as an integral operator. The next lemma is of great utility in this regard.

Lemma 10.3.7. *If the Fourier multiplier $\Lambda = \{\lambda_j\}_{j=-\infty}^{\infty}$ induces a bounded linear operator $\mathcal{M}_\Lambda$ on L^p, then the operator is given by a convolution kernel K. In other words,*

$$\mathcal{M}_\Lambda f(x) = f * K(x) = \frac{1}{2\pi} \int_0^{2\pi} f(t) K(x - t) \, dt.$$

That convolution kernel is specified by the formula

$$K(e^{it}) = \sum_{j=-\infty}^{\infty} \lambda_j e^{it}.$$

[*In practice, the sum that defines this kernel may have to be interpreted using a summability technique, or using distribution theory, or both.*]

Proof. A rigorous proof of this lemma would involve a digression into distribution theory and the Schwartz kernel theorem. We refer the interested reader to either [STW] or [SCH]. □

[2] To wit, first note that if p is a trigonometric polynomial, $p(x) = \sum_{j=-K}^{K} a_j e^{ijx}$, then surely $S_N p \to p$ in norm—just because when $N > k$ then $S_N p = p$. Then the uniform norm estimate on the partial summation operators allows one to pass to the conclusion that $S_N f \to f$ for any $f \in L^p$.

In practice, when we need to get our hands on the kernel associated to a multiplier, we will just calculate it. So this lemma will play a tacit role in our work.

If we apply Lemma 10.3.7 directly to the multiplier for the Hilbert transform, we obtain the formal series

$$\sum_{j=-\infty}^{\infty} -i \cdot \operatorname{sgn} j \cdot e^{ijt}.$$

Of course the terms of this series do not tend to zero, so this series does not converge in any conventional sense. Instead we use Abel summation (i.e., summation with factors of $r^{|j|}$, $0 \le r < 1$) to interpret the series: For $0 \le r < 1$ let

$$k_r(e^{it}) = \sum_{j=-\infty}^{\infty} -ir^{|j|} \cdot \operatorname{sgn} j \cdot e^{ijt}.$$

The sum over the positive indices is

$$\begin{aligned} -i\sum_{j=1}^{\infty} r^j \cdot e^{ijt} &= -i\sum_{j=1}^{\infty} \left[re^{it}\right]^j \\ &= -i\left[\frac{1}{1-re^{it}} - 1\right] \\ &= \frac{-ire^{it}}{1-re^{it}}. \end{aligned}$$

Similarly, the sum over negative indices can be calculated to be equal to

$$\frac{ire^{-it}}{1-re^{-it}}.$$

Adding these two pieces yields that

$$\begin{aligned} k_r(e^{it}) &= \frac{-ire^{it}}{1-re^{it}} + \frac{ire^{-it}}{1-re^{-it}} \\ &= \frac{-ir[e^{it}-e^{-it}]}{|1-re^{it}|^2} \\ &= \frac{2r\sin t}{|1-re^{it}|^2} \\ &= \frac{2r\sin t}{1+r^2-2r\cos t} \\ &= \frac{2r\cdot 2\cdot \sin\frac{t}{2}\cos\frac{t}{2}}{(1+r^2-2r)+2r(1-\cos^2\frac{t}{2}+\sin^2\frac{t}{2})} \\ &= \frac{4r\sin\frac{t}{2}\cos\frac{t}{2}}{(1+r^2-2r)+2r(2\sin^2\frac{t}{2})}. \end{aligned}$$

We formally let $r \to 1^-$ to obtain the kernel

$$k(e^{it}) = \frac{\sin\frac{t}{2}\cos\frac{t}{2}}{\sin^2\frac{t}{2}} = \cot\frac{t}{2}\,. \tag{10.3.5}$$

This is the standard formula for the kernel of the Hilbert transform, indeed it is the formula that we obtained in our consideration of conjugate harmonic functions in Section 10.1. It should be noted that we suppressed various subtleties about the validity of Abel summation in this context, as well as issues concerning the fact that the kernel k is not integrable. For the full story, consult [KAT].

The discussion up to now explicitly exhibits the connection between conjugate harmonic functions on the unit disk D and summation of Fourier series—for both processes give rise to the kernel $k(t) = \cot[t/2]$. Of course the connection is much more profound than that; much of the study of modern complex function theory, function algebras, and Fourier series hinges on this connection. For further details of the saga, see [GAR], [HOF], [KRA4].

We resolve the nonintegrability problem for the integral kernel k in (10.3.5) by using the so-called *Cauchy principal value*; the Cauchy principal value is denoted by P.V. and will now be defined. Thus we usually write

$$\text{P.V.}\,\frac{1}{2\pi}\int_{-\pi}^{\pi} f(x-t)\cot\left(\frac{t}{2}\right)\,dt,$$

and we interpret this to mean

$$\lim_{\epsilon\to 0^+}\frac{1}{2\pi}\int_{\epsilon<|t|\le\pi} f(x-t)\cot\left(\frac{t}{2}\right)\,dt. \tag{10.3.6}$$

Observe in (10.3.6) that, for $\epsilon > 0$ fixed, $\cot(t/2)$ is actually *bounded* on the domain of integration. Therefore the integral in (10.3.6) makes sense, by Hölder's inequality, as long as $f \in L^p$ for some $1 \le p \le \infty$. The deep question is whether the limit exists, and whether that limit defines an L^p function.

We now reiterate the most fundamental fact about the Hilbert transform in the language of integral operators:

Theorem 10.3.8 (M. Riesz). *The operator*

$$Hf(x) \equiv \text{P.V.}\,\frac{1}{2\pi}\int_{-\pi}^{\pi} f(x-t)\cot\left(\frac{t}{2}\right)\,dt$$

is defined on $L^p(\mathbb{T})$, $1 \le p \le \infty$. It is bounded on L^p when $1 < p < \infty$, but is unbounded on L^1 and L^∞.

We will prove Theorem 10.3.8 in Section 10.4. The significance of Theorem 10.3.8 is given by Theorem 10.3.9 (see also Theorem 10.3.6, which anticipated this result):

Theorem 10.3.9. *Let $1 \le p \le \infty$. Norm-convergent partial summation of Fourier series is valid in the L^p topology if and only if the integral operator in (10.3.6) is bounded on L^p.*

An immediate corollary of Theorems 10.3.8 and 10.3.9 is that norm-convergent partial summation of Fourier series is valid in L^p if and only if $1 < p < \infty$.

The result just enunciated is fundamental to the study of Fourier series. But it also holds great philosophical significance in the modern history of analysis. For it shows that we may reduce the study of the (infinitely many) partial sums of the Fourier series of a function to the study of a *single* integral operator. The device for making this reduction is—rather than study one function at a time—to study an entire space of functions at once. Many of the basic ideas in functional analysis—including the uniform boundedness principle, the open mapping theorem, and the Hahn–Banach theorem—grew out of questions of Fourier analysis.

10.4 Proof of Theorem 10.3.8

Now we shall prove that the Hilbert transform is bounded on $L^p(\mathbb{T})$, $1 < p < \infty$. We will present an argument due to S. Bochner. This will allow us to make good use of the Riesz–Thorin interpolation theorem that we proved in Section 9.5.

Proposition 10.4.1. *The Hilbert transform is bounded on $L^p(\mathbb{T})$ when $p = 2k$ is a positive, even integer.*

Proof. Let f be a continuous, real function on $[0, 2\pi)$. We normalize f (by subtracting off a constant) so that $\int f dx = 0$. Let u be its Poisson integral, so u is harmonic on the disk D and vanishes at 0. Let v be that harmonic conjugate of u on D such that $v(0) = 0$. Then $h = u + iv$ is holomorphic and $h(0) = 0$.

Fix $0 < r < 1$. Now we write

$$\begin{aligned}
0 = 2\pi h^p(0) &= \int_0^{2\pi} h^{2k}(re^{i\theta})\, d\theta \\
&= \int_0^{2\pi} [u(re^{i\theta}) + iv(re^{i\theta})]^{2k}\, d\theta \\
&= \int_0^{2\pi} u^{2k}\, d\theta + i\binom{2k}{1} \int u^{2k-1} v\, d\theta \\
&\quad - \binom{2k}{2} \int u^{2k-2} v^2\, d\theta + \cdots \\
&\quad + i^{2k-1} \binom{2k}{2k-1} \int u v^{2k-1}\, d\theta + i^{2k} \int v^{2k}\, d\theta\,.
\end{aligned}$$

We rearrange the last equality as

$$\int_0^{2\pi} v^{2k}\,d\theta \le \binom{2k}{2k-1}\int_0^{2\pi} |uv^{2k-1}|\,d\theta + \binom{2k}{2k-2}\int_0^{2\pi}|u^2v^{2k-2}|\,d\theta + \cdots$$

$$+\binom{2k}{2}\int_0^{2\pi}|u^{2k-2}v^2|\,d\theta + \binom{2k}{1}\int_0^{2\pi}|u^{2k-1}v|\,d\theta + \int_0^{2\pi}|u^{2k}|\,d\theta\,.$$

We apply Hölder's inequality to each composite term on the right—using the exponents $2k/j$ and $2k/[2k-j]$ on the j^{th} term, for $j = 1, 2, \ldots, 2k-1$. It is convenient to let $S = \left[\int u^{2k}\,d\theta\right]^{1/2k}$ and $T = \left[\int v^{2k}\,d\theta\right]^{1/2k}$ and we do so. The result is

$$T^{2k} \le \binom{2k}{2k-1}ST^{2k-1} + \binom{2k}{2k-2}S^2T^{2k-2} + \cdots + \binom{2k}{2}S^{2k-2}T^2 + \binom{2k}{1}S^{2k-1}T + S^{2k}\,.$$

Now define $U = T/S$ and rewrite the inequality as

$$U^{2k} \le \binom{2k}{2k-1}U^{2k-1} + \binom{2k}{2k-2}U^{2k-2} + \cdots + \binom{2k}{2}U^2 + \binom{2k}{1}U + 1\,.$$

Divide through by U^{2k-1} to obtain

$$U \le \binom{2k}{2k-1} + \binom{2k}{2k-2}U^{-1} + \cdots + \binom{2k}{2}U^{-2k+3} + \binom{2k}{1}U^{-2k+2} + U^{-2k+1}\,.$$

If $U \ge 1$, then it follows that

$$U \le \binom{2k}{2k-1} + \binom{2k}{2k-2} + \cdots + \binom{2k}{2} + \binom{2k}{1} + 1 \le 2^{2k}\,.$$

We conclude, therefore, that

$$\|v\|_{L^{2k}} \le 2^k \|u\|_{L^{2k}}\,.$$

But of course the function $v(re^{i\theta})$ is the Hilbert transform of $u(re^{i\theta})$. The proof is therefore complete. □

Theorem 10.4.2. *The Hilbert transform is bounded on L^p, $1 < p < \infty$.*

Proof. We know that the Hilbert transform is bounded on L^2, L^4, L^6, ... We may immediately apply the Riesz–Thorin theorem (Section 9.5) to conclude that the Hilbert transform is bounded on L^p for $2 \le p \le 4$, $4 \le p \le 6$, $6 \le p \le 8$, etc. In other words, the Hilbert transform is bounded on L^p for $2 \le p < \infty$.

Now let $f \in L^p$ for $1 < p < 2$. Let φ be any element of $L^{p/[p-1]}$ with norm 1. Notice that $2 < p/[p-1] < \infty$. Then

$$\begin{aligned}
\int Hf \cdot \varphi d\theta &= \int \left[\int f(\psi) \cot \frac{\theta - \psi}{2} \, d\psi \right] \varphi(\theta) \, d\theta \\
&= \iint \varphi(\theta) \cot \frac{\theta - \psi}{2} \, d\theta f(\psi) \, d\psi \\
&= - \int \left[\int \varphi(\theta) \cot \frac{\psi - \theta}{2} \, d\theta \right] f(\psi) \, d\psi \\
&= - \int H\varphi(\psi) f(\psi) \, d\psi \, .
\end{aligned}$$

Using Hölder's inequality together with the fact that we know that the Hilbert transform is bounded on $L^{p/[p-1]}$, we may bound the right-hand side by the expression $C\|\varphi\|_{L^{p/[p-1]}} \|f\|_{L^p} \le \|f\|_{L^p}$. Since this estimate holds for any such choice of φ, the result follows. □

We complete our consideration of the Hilbert transform by treating what happens on the spaces L^1 and L^∞.

Proposition 10.4.3. *Norm summability for Fourier series fails in both L^1 and L^∞.*

Proof. It suffices for us to show that the modified Hilbert transform (as defined in Section 10.2) fails to be bounded on L^1 and fails to be bounded on L^∞. In fact the following lemma will cut the job by half:

Lemma 10.4.4. *If the modified Hilbert transform $\mathcal{H}$ is bounded on L^1, then it is bounded on L^∞.*

Proof. Let f be an L^∞ function. Then

$$\|\mathcal{H}f\|_{L^\infty} = \sup_{\substack{\phi \in L^1 \\ \|\phi\|_{L^1} = 1}} \left| \int \mathcal{H}f(x) \cdot \phi(x) \, dx \right| = \sup_{\substack{\phi \in L^1 \\ \|\phi\|_{L^1} = 1}} \left| \int f(x) (\mathcal{H}^* \phi)(x) \, dx \right| .$$

But an easy formal argument (as in the proof of Theorem 10.4.2) shows that

$$\mathcal{H}^* \phi = -\mathcal{H}\phi.$$

Here $\mathcal{H}^*$ is the adjoint of $\mathcal{H}$. [In fact a similar formula holds for *any* convolution operator—exercise.] Thus the last line gives

$$\begin{aligned}\|\mathcal{H}f\|_{L^\infty} &= \sup_{\substack{\phi\in L^1\\ \|\phi\|_{L^1}=1}} \left|\int f(x)\mathcal{H}\phi(x)\,dx\right| \\ &\le \sup_{\substack{\phi\in L^1\\ \|\phi\|_{L^1}=1}} \|f\|_{L^\infty}\cdot\|\mathcal{H}\phi\|_{L^1} \\ &\le \sup_{\substack{\phi\in L^1\\ \|\phi\|_{L^1}=1}} \|f\|_{L^\infty}\cdot C\|\phi\|_{L^1} \\ &= C\cdot\|f\|_{L^\infty}.\end{aligned}$$

Here C is the norm of the modified Hilbert transform acting on L^1. We have shown that if $\mathcal{H}$ is bounded on L^1, then it is bounded on L^∞. That completes the proof. □

Remark 10.4.5. In fact the proposition that we just proved is true in considerable generality. We may replace L^1, L^∞ by any two conjugate spaces $L^p, L^{p'}$ with $1/p + 1/p' = 1$. And we may replace the Hilbert transform by any convolution operator. Details are left for the interested reader.

Now let us resume the proof of Proposition 10.4.3. Let $f = \chi_{[0,a]}$, where a is a small positive number and χ denotes the characteristic function of the indicated interval. We may calculate the (modified) Hilbert transform of f by hand and see that

$$\mathcal{H}f(x) = c\cdot\log\frac{|x|}{|x-a|}.$$

In particular, $\mathcal{H}f$ is an *unbounded* function. Therefore the Hilbert transform is not bounded on L^∞. [It is not difficult to calculate that $\|D_N\|_{L^1} \approx \log N$, where D_N is the Dirichlet kernel. This fact already suggests that the Hilbert transform is not bounded on L^∞.] By Lemma 10.4.4, we may also conclude that boundedness of the Hilbert transform fails on L^1.

It follows from the arguments in the last paragraph that norm summability for Fourier series fails in L^1 and L^∞. □

Problems for Study and Exploration

1. Prove that the Hilbert transform maps C^∞ to C^∞. You will first need to formulate a precise statement of the result.
2. What is the Hilbert transform of the function $\cos\theta$?
3. What is the Hilbert transform of the function $|\theta|$?
4. The Hilbert transform is an isometry of L^2. Explain this statement and prove it.
5. The study of an integral operator of the form

$$f \mapsto f * K$$

is relatively straightforward when the kernel K is integrable, or even q^{th}-power integrable. But the kernel of the Hilbert transform is *not* integrable to any power. Explain.

6. Prove that the Hilbert transform is a self-adjoint operator in a suitable sense.

7. If f is an L^2 function on the real line, then we define the Hilbert transform in that context to be

$$Hf(x) = \text{P.V.} \int_{\mathbb{R}} \frac{f(t)}{x-t}\, dt\,.$$

Here we evaluate the *principal value* (i.e., PV) by integrating over a deleted neighborhood $|x-t| > \epsilon$ and then letting $\epsilon \to 0$. Using a little distribution theory, calculate that

$$\widehat{Hf}(\xi) = -i\,\text{sgn}(\xi) \cdot \widehat{f}(\xi)\,.$$

8. Prove that the Hilbert H transform is *idempotent* in the sense that $H \circ H = \pm\text{id}$.

9. Prove that if $k \in L^1$, then the operator $f \mapsto f * k$ is bounded on L^p, $1 \le p \le \infty$.

10. Let

$$1 + \frac{1}{q} = \frac{1}{p} + \frac{1}{r}\,.$$

Show that if $k \in L^r$, then the operator $f \mapsto f * k$ maps L^p to L^q for a suitable range of p, q, r. How does this result generalize that of Exercise 9? See [STE] for further details.

11. Use the theory of the Hilbert transform developed here to show that if we give an alternative definition of partial summation operator by

$$\widetilde{S}_N f(x) = \sum_{j=-\varphi(N)}^{\psi(N)} \widehat{f}(j) e^{ijx}\,,$$

with φ, ψ tending monotonically to $+\infty$ with N, then the same theory of Fourier series results. That is to say, a Fourier series converges (pointwise or in norm) using the S_j as defined in the text if and only if it converges using the $\widetilde{S}_j$.

11

Wolff's Proof of the Corona Theorem

Genesis and Development

The traditional definition of a *function algebra* is that it is any closed subalgebra of the continuous functions on a compact Hausdorff space. For us—at least at the beginning—the relevant compact Hausdorff space is the circle $\mathbb{T}$. Classically, an important function algebra has been $A(D)$—the functions continuous on $\overline{D}$ and holomorphic on D. [We call $A(D)$ the *disk algebra*.] Each such function can be identified with its restriction to the circle. And any such restriction has Fourier series with no coefficients of negative index. So this is clearly a subspace. It also follows by inspection that it is a subalgebra, and is closed.

The study of function algebras has expanded to include other spaces—especially spaces that are amenable to a uniform norm. One of the most important of these is $H^\infty(D)$. Of course any bounded, holomorphic function on the disk may be identified with an L^∞ function on the circle $\mathbb{T}$. So $H^\infty(D)$ may be thought of as a closed subalgebra of $L^\infty(\mathbb{T})$.

We gain deeper insight into the structure of $H^\infty(D)$, or of any function algebra, by studying its space of multiplicative linear functionals (i.e., ring homomorphisms). For $H^\infty(D)$, it is clear that point evaluation is a multiplicative linear functional. That is, if $p \in D$ is a fixed point, then

$$H^\infty(D) \ni f \mapsto f(p)$$

is a multiplicative linear functional. It is not hard to see using the Banach–Alaoglu theorem that the point evaluations cannot be all the multiplicative linear functionals on $H^\infty(D)$. In fact there must be uncountably many more of them.

It is a metatheorem that an explicit description of all the multiplicative linear functionals on $H^\infty(D)$ is beyond our grasp. But we could ask whether the point evaluations are somehow typical. For example, are they *dense* (in the appropriate topology, which turns out to be the weak-$*$ topology). The answer to this question is "yes," and it is provided by Lennart Carleson's very deep *corona theorem*.

One of the striking results in analysis from the late 1970s was Tom Wolff's new proof of the corona theorem. It uses important ideas from harmonic analysis that were introduced by Fefferman and Stein, and gives us an entirely new way to look at the matter.

The reader of this chapter will want to know a little functional analysis, and be comfortable with hard estimates.

11.1 Introductory Remarks

In the early 1940s, I. M. Gelfand introduced the idea of a Banach algebra. This is a Banach space that also has a multiplicative structure. The theory of Banach algebras has turned out to be a remarkably accessible, flexible, and powerful tool. One of the early triumphs of the theory was a brief and easy proof of Norbert Wiener's powerful theorem that the reciprocal of a nonvanishing function with absolutely convergent Fourier series also has absolutely convergent Fourier series.

In the 1950s and 1960s the theory of Banach algebras really took off. There was particular interest in *function algebras*, that is to say, closed uniform subalgebras of the continuous functions on a compact Hausdorff space. Of special interest were various spaces of holomorphic functions. For example the space of functions continuous on the closure $\overline{D}$ of the unit disk in $\mathbb{C}$ and holomorphic on the interior D can be thought of as a uniform subalgebra of the continuous functions on the unit circle (the boundary of D). We denote this important Banach algebra by $A(D)$.

In a related vein, the algebra of bounded holomorphic functions on D, denoted by $H^\infty(D)$, is a significant and much-studied subalgebra of $L^\infty(\partial D)$.

Gelfand taught us that the deeper structure of a Banach algebra $\mathcal{A}$ lives in its space of multiplicative linear functionals. These are the ring homomorphisms from $\mathcal{A}$ to $\mathbb{C}$. It turns out that such a multiplicative linear functional is automatically a bounded linear functional with norm precisely 1.

To see this, we first note that we shall always assume that the Banach algebra $\mathcal{A}$ has a unit 1 (to learn about the extra ideas and arguments that are needed when $\mathcal{A}$ does *not* have a unit, consult [RIC]). Now observe that, if φ is a multiplicative linear functional, then $\varphi(1) = \varphi(1 \cdot 1) = \varphi(1) \cdot \varphi(1)$ hence $\varphi(1) = 1$. Next, if $x \in \mathcal{A}$ and λ is a complex constant with $|\lambda| > \|x\|$ then certainly $(\lambda - x)^{-1}$ exists. Indeed, the inverse is given by the Neumann series:

$$(\lambda - x)^{-1} = \frac{1}{\lambda}\left(1 - \frac{x}{\lambda}\right)^{-1} = \frac{1}{\lambda}\sum_{j=0}^{\infty}\left(\frac{x}{\lambda}\right)^j .$$

The series clearly converges in norm. Thus we see that

$$(\lambda - x) \cdot (\lambda - x)^{-1} = 1 \, .$$

For φ a multiplicative linear functional, we thus see that

$$\varphi(\lambda - x) \cdot \varphi((\lambda - x)^{-1}) = 1 \, .$$

As a result, $\varphi(\lambda - x) \neq 0$, hence $\varphi(x) \neq \lambda$.

We conclude that $|\varphi(x)| \leq |\lambda|$ when $|\lambda| > \|x\|$. Therefore $|\varphi(x)| \leq \|x\|$. Hence the norm of the linear functional φ is less than or equal to 1. Since $\varphi(1) = 1$, we see in fact that $\|\varphi\| = 1$. The set of multiplicative linear functionals on $\mathcal{A}$ is thus a closed subset of the closed unit ball (indeed, the unit sphere) in the dual of $\mathcal{A}$ (thought of as a Banach *space*). Hence, by the Banach–Alaoglu theorem (see [RUD3]), it is weak-$*$ compact.

Now if φ is a multiplicative linear functional on the Banach algebra $\mathcal{A}$, then the kernel of φ is plainly a closed ideal $\mathbf{m}$. Since, by the first fundamental isomorphism of algebra, $\mathcal{A}/\mathbf{m} = \mathbb{C}$, it can be argued that $\mathbf{m}$ is a *maximal* ideal.

Conversely, if $\mathbf{m}$ is a maximal ideal in $\mathcal{A}$, then it is easy to see that the closure of $\mathbf{m}$ is also a proper ideal. In fact, if $x \in \mathcal{A}$ and $\|1 - x\| < 1$, then x^{-1} exists in $\mathcal{A}$ by the usual Neumann series argument. If $x \in \mathbf{m}$, then it would follow that $1 = x \cdot x^{-1} \in \mathbf{m}$. So $\mathbf{m}$ is not an ideal. In conclusion, $\|1 - x\| \geq 1$ for all $x \in \mathbf{m}$. As a result, the closure of $\mathbf{m}$ is still a proper ideal. It follows that $\mathcal{A}/\mathbf{m}$ is a Banach algebra over $\mathbb{C}$. In fact it is a field because the ideal $\mathbf{m}$ is maximal. It is a theorem of Gelfand that $\mathcal{A}/\mathbf{m}$ must therefore be isomorphic to $\mathbb{C}$. We conclude that the canonical homomorphism

$$\varphi : \mathcal{A} \mapsto \mathcal{A}/\mathbf{m} \cong \mathbb{C}$$

is a multiplicative linear functional. See [RUD2] for the details of these arguments.

Put in slightly different words, if $x \in \mathcal{A}$, then the isomorphism $\mathcal{A}/\mathbf{m} \cong \mathbb{C}$ tells us that there is a unique complex number λ such that $x - \lambda \in \mathbf{m}$. Then $\varphi(x) = \lambda$.

In summary, although it would make perfect sense to denote the space of all multiplicative linear functionals of the Banach algebra $\mathcal{A}$ by $\mathcal{A}^*$ (for example), we instead think of the space of multiplicative linear functionals as the *maximal ideal space* of $\mathcal{A}$ and denote it by $\mathbf{M}$. Equipped with the weak-$*$ topology, $\mathbf{M}$ is a compact Hausdorff space.

One of Gelfand's lovely observations is that we may think of each element x of $\mathcal{A}$ as a continuous function on $\mathbf{M}$ just by

$$\mathbf{M} \ni \mathbf{m} \mapsto \mathbf{m}(x) \, .$$

Thus x corresponds to a continuous function on the compact space $\mathbf{M}$. It is clear that different elements $x \in \mathcal{A}$ correspond to different continuous functions on $\mathbf{M}$. Thus Gelfand's result is that *any* Banach algebra is an algebra of continuous functions on a compact Hausdorff space. Because of this identification, we often write $x(\mathbf{m})$ rather than $\mathbf{m}(x)$.

11.2 The Banach Algebra H^∞

It is our intention in the present chapter to learn some of the deeper properties of the Banach algebra H^∞, the bounded analytic functions on the unit disk in $\mathbb{C}$. The maximal ideal space $\mathbf{M}$ of H^∞ is an extremely complicated object that is still intensely studied today. But it has a subset consisting of extremely simple elements. Namely, if $z \in D$ then the point evaluation functional

$$\alpha_z : H^\infty \ni f \mapsto f(z)$$

is a multiplicative linear functional. It of course corresponds to the maximal ideal

$$\mathbf{m}_z = \{f \in H^\infty : f(z) = 0\}\,.$$

These maximal ideals are very easy to understand. It is natural to ask whether $\{\mathbf{m}_z : z \in D\}$ is *dense* in $\mathbf{M}$ (the maximal ideal space of H^∞) in the weak-$*$ topology. The speculation that this question has an affirmative answer became known as the *corona problem*. Namely, if the set of point evaluations were not dense in $\mathbf{M}$, then it would be said that there was a "corona" of additional multiplicative linear functionals.

Lennart Carleson [CAR] proved the density of $\{\mathbf{m}_z\}$ in $\mathbf{M}$ in 1962 (see [CAR]). His celebrated result is now known as the *corona theorem*. Carleson's original combinatorial/geometric proof is quite delicate and complicated. It has the virtue of being applicable to many situations in geometric function theory, and it is an important part of our arsenal of weapons in the study of complex analysis (see [GAR]). But there is a much more accessible proof, due to Tom Wolff, that we shall present here. Wolff created this beautiful argument when he was still a graduate student in Berkeley, studying under Don Sarason. He never published the proof, and we are all indebted to Paul Koosis [KOO] for the elegant presentation that we are about to explain. It may be noted that J. Garnett and N. Varopoulos also contributed some useful simplifications to the proof given here.

11.3 Statement of the Corona Theorem

In fact the corona theorem has two important formulations, one of which has already been indicated. We shall enunciate them both now, and prove that they are indeed equivalent.

First Statement of the Corona Theorem: The point evaluation mappings are weak-$*$ dense in the maximal ideal space $\mathbf{M}$ of $H^\infty(D)$.

Second Statement of the Corona Theorem: If $f_1, f_2, \ldots, f_k$ are elements of $H^\infty(D)$ and if there is a $\delta > 0$ such that

$$\max_j |f_j(z)| \geq \delta \tag{11.3.1}$$

for all $z \in D$, then there exist $g_1, g_2, \ldots, g_k \in H^\infty(D)$ such that

$$f_1 g_1 + f_2 g_2 + \cdots + f_k g_k \equiv 1 \tag{11.3.2}$$

on the disk.

Proof that First Statement Implies Second Statement: Let $f_1, f_2, \ldots, f_k \in H^\infty$ satisfy (11.3.1). If there do *not* exist $g_1, g_2, \ldots, g_k \in H^\infty$ that satisfy $f_1 g_1 + \cdots + f_k g_k \equiv 1$ on D, then the set

$$A = \{f_1 g_1 + f_2 g_2 + \cdots + f_k g_k : g_1, g_2, \ldots, g_k \in H^\infty\}$$

is a proper ideal in H^∞. Of course this ideal is contained in some maximal ideal $\mathbf{m}$. But then $f_j(\mathbf{m}) = 0$ for $j = 1, \ldots, k$. Since we are assuming that the First Statement of the Corona is true, we now know that there is a net of points $z_\alpha \in D$ such that $f_j(z_\alpha) \to f_j(\mathbf{m}) = 0$ for $j = 1, \ldots, k$. But this contradicts (11.3.1). Thus $g_1, \ldots, g_k$ exist as claimed. □

Proof that Second Statement Implies First Statement: Fix an ideal $\mathbf{m} \in \mathbf{M}$ and, seeking a contradiction, suppose that there is no net z_α of points in D such that z_α tends weak-$*$ to $\mathbf{m}$. By the definition of the weak-$*$ topology (see [RUD3]), there must exist $h_1, \ldots, h_k \in H^\infty$ and a number $\delta > 0$ such that, for each $z \in D$, at least one of the inequalities

$$|h_j(z) - h_j(\mathbf{m})| > \delta \tag{11.3.3}$$

must hold. Set $f_j(z) = h_j(z) - h_j(\mathbf{m})$. Of course $h_j(\mathbf{m})$ is just a complex scalar. Then $f_j \in H^\infty$. And of course each $f_j(\mathbf{m}) = 0$. But, by (11.3.3), for each $z \in D$ there is an index j such that $|f_j(z)| \geq \delta > 0$.

As a result, (11.3.1) and (11.3.2) tell us that there exist $g_1, g_2, \ldots, g_k$ such that $f_1 g_1 + f_2 g_2 + \cdots + f_k g_k \equiv 1$. In particular,

$$f_1(\mathbf{m}) g_1(\mathbf{m}) + f_2(\mathbf{m}) g_2(\mathbf{m}) + \cdots + f_k(\mathbf{m}) g_k(\mathbf{m}) \equiv 1 \,.$$

But we already know that each $f_j(\mathbf{m}) = 0$, so that is a contradiction. □

We shall in fact prove the second formulation of the corona theorem. We produce the functions g_j by solving a dual extremal problem. It is worth noting that statement (11.3.2) without any bounds on the g_j's is relatively straightforward. The *Koszul complex* (see [KKRA1]) is a useful algebraic device for producing the g_j's.

11.4 Carleson Measures

Let $P \in \partial D$ and $h > 0$. A *Carleson region* in D is a set of the form

$$\mathcal{C}_{P,h} = \{z \in D : |z - P| < h\} . \tag{11.4.1}$$

We know that the unit disk D is conformally equivalent to the upper halfplane U. In that context, for $Q = (Q_1, 0) \in \partial U$ and $h > 0$, a Carleson region is defined to be

$$\mathcal{D}_{Q,h} = \{z = x + iy \in U : |x - Q_1| < h, 0 < y < h\} . \tag{11.4.2}$$

We certainly do not claim that the Cayley transform maps the Carleson regions defined in (11.4.1) to the Carleson regions defined in (11.4.2). But the two ideas are comparable in the sense that the Cayley transform of $\mathcal{C}_{P,h}$ lies in $\mathcal{D}_{Q,3h}$ and the inverse Cayley transform of $\mathcal{D}_{Q,h}$ lies in $\mathcal{C}_{P,3h}$. See Figure 11.1. In practice, we think of a Carleson region as being $2h$ units wide and h units tall. We leave the details of all these assertions to the reader.

Let μ be a positive measure on the unit disk $D \subseteq \mathbb{C}$. We say that μ is a *Carleson measure* if there is a constant $K > 0$ such that, for any $P \in \partial D$ and $h > 0$,

$$\mu(\mathcal{C}_{P,h}) \leq K \cdot h .$$

We sometimes call K the "Carleson constant" of the measure μ. This definition is based on a simple geometric condition. It is important because of its connection (due to Hörmander) to complex function theory.

Theorem 11.4.1. *Let μ be a Carleson measure on the disk D. Then, for any $f \in H^1(D)$,*

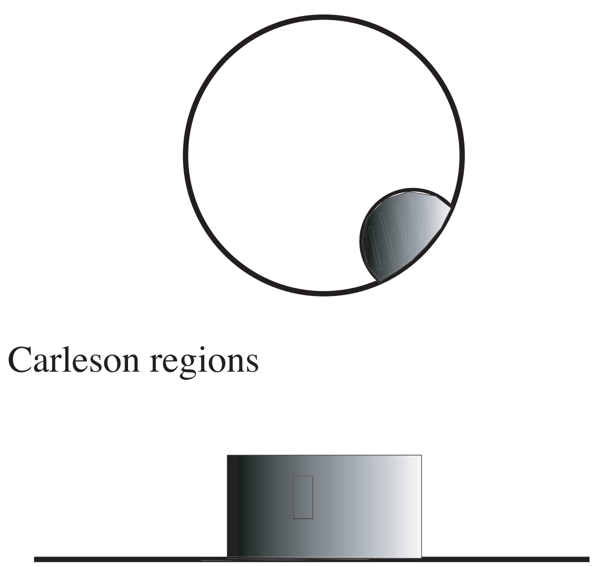

Fig. 11.1. Shapes of Carleson regions.

$$\iint_D |f(z)|\, d\mu(z) \le C \cdot \|f\|_{H^1(D)}\,. \qquad (11.4.3)$$

Here the constant C is independent of the choice of f, and depends on the constant K in the definition of Carleson measure.

The converse is true as well: If a positive measure μ on the disk satisfies (11.4.3) for all $f \in H^1(D)$, then μ is a Carleson measure.

Proof. First assume (11.4.3). Fix a point $P \in \partial D$ and $h > 0$. For notational simplicity we shall take $P = 1 + i0$. Define

$$f(z) = \frac{h}{(z - 1 - h)^2}\,.$$

It is straightforward to check that $f \in H^1(D)$. Indeed,

$$\begin{aligned}\int_{-\pi}^{\pi} \frac{h}{|e^{i\theta} - 1 - h|^2}\, d\theta &\le \int_{-\pi}^{\pi} \frac{h}{h^2 + \sin^2\theta}\, d\theta \\ &\le \int_{-\pi}^{\pi} \frac{h}{h^2 + \theta^2/4}\, d\theta \\ &\le \int_{-\pi/[2h]}^{\pi/[2h]} \frac{2}{1+t^2}\, dt \le 2\pi\,.\end{aligned}$$

Notice also that $|f(z)| \ge 1/[9h]$ for $z \in \mathcal{C}_{1,h}$. Thus, if (11.4.3) holds, then

$$\frac{1}{9h}\mu(\mathcal{C}_{1,h}) \le \iint_U |f(z)|\, d\mu(z) \le C \cdot 2\pi\,.$$

That is,

$$\mu(\mathcal{C}_{1,h}) \le 18\pi C \cdot h\,.$$

This proves that μ is a Carleson measure with the constant K equaling $18\pi C$.

For the converse, we assume that μ is a Carleson measure. Let $f \in H^1(D)$. Replacing $f(z)$ with $f(rz)$ for some fixed $0 < r < 1$, we may assume that f is continuous on $\overline{D}$. Subtracting a constant from f if necessary, we may assume that $f(0) = 0$. Now we must establish inequality (11.4.3) for a constant C that is independent of the choice of f.

We shall use the idea of distribution function. If g is a measurable function on a measure space (X, μ) and $\alpha > 0$, then we let $m_g(\alpha) = \mu(\{x \in X : |g(x)| > \alpha\})$. Observe that

$$\begin{aligned}\int_0^\infty pt^{p-1} m_g(t)dt &= \int_0^\infty pt^{p-1} \mu\{x : |g(x)| > t\}\, dt \\ &= \int_0^\infty pt^{p-1} \int_{\{x:|g(x)|>t\}} d\mu(x)\, dt\end{aligned}$$

$$
\begin{aligned}
&= \int_X \int_0^{|g(x)|} p \cdot t^{p-1}\, dt d\mu(x) \\
&= \int_X |g(x)|^p \, d\mu(x) \\
&= \|g\|_{L^p}^p \,.
\end{aligned}
$$

Our strategy now is to prove that if $f \in H^1(D)$, then

$$m_f(t) \leq C \cdot m_{f^*}(t)\,.$$

In this inequality we think of f on the left as a function on the disk D, and we calculate the distribution function with respect to the measure μ. On the right,

$$f^*(e^{it}) = \sup_{z \in \Gamma_\alpha(e^{it})} |f(z)|$$

is a function on the boundary ∂D; and we calculate the distribution function with respect to ordinary Lebesgue measure on ∂D. Integration of both sides will then yield the inequality

$$\int_D |f(z)|\, d\mu(z) \leq C \cdot \int_{\partial D} f^*(e^{i\theta})\, d\theta\,.$$

We will show momentarily that

$$\|f^*\|_{L^1} \leq C' \|f\|_{H^1}\,. \tag{11.4.4}$$

Thus the proof will then be complete.

In fact we already proved in Section 6.1 that, if $p \geq 1$, if φ is a function on ∂D and if u is its harmonic extension to D, then

$$\|u^*\|_{L^p} \leq C \cdot \|\varphi\|_{L^p}\,. \tag{11.4.5}$$

If instead $f \in H^p(D)$, $0 < p < 1$, then (again see Section 6.1) we may factor $f = B \cdot F$, where B is a Blaschke product and F is nonvanishing. We define $V(z) = F(z)^{p/2}$. This is a holomorphic, hence harmonic, function on the disk that satisfies a standard square-integrability condition. So the preceding result for u applies now to V and we obtain the desired inequality (11.4.5).

For $\lambda > 0$ we now set

$$E_\lambda = \{z \in \partial D : f^*(z) \leq \lambda\}$$

and

$$O_\lambda = \{z \in \partial D : f^*(z) > \lambda\}\,.$$

Thus $m_{f^*}(\lambda) = |O_\lambda|$ (where bars here denote Lebesgue measure). Now we need the following important geometric observation:

Let $z_0 \in E_\lambda$; then define

$$\overline{S}_{z_0} = \{z \in D : |z - z_0| \leq 2(1 - |z|)\} .$$

Since $f^*(z_0)$ is defined by taking a supremum over S_{z_0}, we must conclude that $|f(z)| \leq \lambda$ for all $z \in S_{z_0}$. By complementation,

$$\Omega_\lambda \equiv \{z \in D : |f(z)| > \lambda\}$$

must be contained in

$$\Omega'_\lambda \equiv D \setminus \bigcup_{z_0 \in E_\lambda} \overline{S}_{z_0} .$$

Now it is clear from the semicontinuity of f^* that O_λ is the disjoint union of open intervals J_k in ∂D. Say that J_k has length $2r_k$. For each index k, let Δ_k be the "triangular region" erected above J_k as in Figure 11.2. Analytically, the region Δ_k is given by

$$1 - |z| < \frac{1}{2} \cdot \min\{|z - e^{ir_k} z_0|, |z - e^{-ir_k} z_0|\} .$$

What is crucial to the argument is that the "triangles" Δ_k are in effect the components of the complement of the regions $\overline{S}_{z_0}$. More precisely,

$$\Omega'_\lambda = \bigcup_k \Delta_k .$$

As a result, $\Omega_\lambda \subseteq \bigcup_k \Delta_k$. We may conclude that

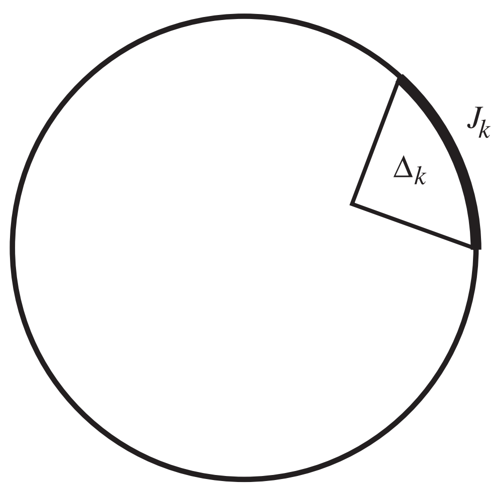

Fig. 11.2. The triangular region Δ_k.

$$m_f(\lambda) = \mu(\Omega_\lambda) \le \sum_k \mu(\Delta_k)\,.$$

But each Δ_k lives inside a Carleson region of roughly the same size, so we may conclude by the fact that μ is a Carleson measure that $\mu(\Delta_k) \le K \cdot |J_k|$. Therefore

$$m_f(\lambda) \le K \sum_k |J_k| = K \cdot |O_\lambda| = K \cdot m_{f^*}(\lambda)\,.$$

Using our calculation with the distribution function, we may now conclude that

$$\int_D |f(z)|\, d\mu(z) \le K \cdot \int_{\partial D} |f(e^{i\theta})|\, d\theta\,.$$

We know from Section 6.1 that the right-hand side is equivalent to $\|f\|_{H^1}$. So the result is proved. □

11.5 A Key Technical Lemma

For a smooth function g, we will find it useful to have the notation

$$\partial g = \frac{\partial g}{\partial z} = \frac{1}{2}\left(\frac{\partial}{\partial x} - i\frac{\partial}{\partial y}\right)$$

and

$$\overline{\partial} g = \frac{\partial g}{\partial \overline{z}} = \frac{1}{2}\left(\frac{\partial}{\partial x} + i\frac{\partial}{\partial y}\right)\,.$$

See the discussion in Chapter 7.

The purpose of this section is to prove the following result.

Lemma 11.5.1. *Let $R > 1$. Let h be a C^∞ function on the disk $D(0,R) \equiv \{z \in \mathbb{C} : |z| < R\}$. Assume that, in the disk $D(0,R)$, the measures*

$$\left(|z| \log \frac{1}{|z|}\right) |h(z)|^2\, dx\, dy$$

and

$$\left(|z| \log \frac{1}{|z|}\right) |\partial h(z)|\, dx\, dy$$

are Carleson measures with Carleson constants A and B respectively. Then there is a C^∞ function v on a disk $D(0,R')$ with $R' > 1$ such that

$$\overline{\partial} v(z) = h(z) \quad \text{on } D(0,R')$$

and

$$|v(e^{i\theta})| \le 9(\sqrt{A} + B)\,.$$

Of course we already know that we can solve the $\overline{\partial}$ problem—see Section 7.2. Wolff's special contribution is the estimate for the solution in terms of the Carleson constants. The proof will be broken into several lemmas, and will fill out the remainder of the section.

Proof of the Lemma. We may multiply h by a cutoff function that is identically equal to 1 in a neighborhood of $\overline{D}(0,1)$ and thus arrange for h to be C^∞ with compact support. As already indicated, we certainly know (by Theorem 7.2.6) that there is a function v_0, defined on all of $\mathbb{C}$, that satisfies $\overline{\partial}v_0 = h$. By differentiating under the integral sign, we may certainly confirm that v_0 is C^∞.

We need to modify v_0 in order to achieve the indicated estimates. We do so by adding a *holomorphic function* f to v_0. Of course this holomorphic function f will be annihilated by $\overline{\partial}$, so $v_0 + f$ will still be a solution to our $\overline{\partial}$ equation. Notice that, if there is an f in H^∞ that does the job, then $f_r(z) \equiv f(rz)$ will also do the job for r sufficiently close to, but less than, 1. So we can restrict attention to functions f that are continuous on $\overline{D}$ and holomorphic on D. The space of such functions is commonly denoted by $A(D)$ and is called the *disk algebra*. Our job, in effect, is to calculate the distance $\|v_0 - A(D)\|_\infty$.

We wish to use some functional analysis.

Lemma 11.5.2. *Let X be a Banach space and $Y \subseteq X$ a subspace. If $x \in X$, then*

$$\|x - Y\| = \sup\{|\alpha(x)| : \alpha \in X^*, \|\alpha\| \le 1, \alpha(Y) = 0\}\,.$$

Proof. If $\alpha \in X^*$, $\|\alpha\| \le 1$, $\alpha(Y) = 0$, then

$$|\alpha(x)| = |\alpha(x - y)| \le \|x - y\| \quad \text{for any } y \in Y\,.$$

Taking the infimum over all such y gives

$$|\alpha(x)| \le \|x - Y\|\,.$$

Now taking the supremum over all such α gives

$$\sup\{|\alpha(x)| : \alpha \in X^*, \|\alpha\| \le 1, \alpha(Y) = 0\} \le \|x - Y\|\,.$$

For the converse, consider a functional β that equals 0 on Y and equals $\|x - Y\|$ on x. Then

$$\|x - Y\| \le \|x - 0\| = \|x\|$$

so that $\|\beta\| \le 1$. Apply the Hahn–Banach theorem to extend β to a functional $\widetilde{\beta}$ on all of X having the same norm. Then

$$\|x - Y\| = |\widetilde{\beta}(x)| \le \sup\{|\alpha(x)| : \alpha \in X^*, \|\alpha\| \le 1, \alpha(Y) = 0\}\,.$$

This completes the proof. □

With this lemma in mind, it will be useful for us to think of $A = A(D)$ as a subspace of $C(\partial D)$, the continuous functions on ∂D. Notice that the annihilator of A is naturally identified with $(C/A)^*$. By the lemma,

$$\|v_0 - A\|_\infty = \sup\{|\tau(v_0)| : \tau \in [C/A]^*, \|\tau\| \leq 1\}.$$

We must now identify $[C/A]^*$.

Lemma 11.5.3. *Let*

$$H_0^1(D) = H_0^1 = \{f \in H^1(D) : \int_{-\pi}^{\pi} f(e^{i\theta})\, d\theta = 0\}.$$

Observe that $H_0^1 = z \cdot H^1$. *Then*

$$[C/A]^* = H_0^1.$$

Proof. Of course, by the Riesz representation theorem, the dual of C is the space M of finite Borel measures on ∂D. It follows that the dual of C/A is

$$\left\{\mu \in M : \int_{-\pi}^{\pi} f(e^{i\theta})\, d\mu(\theta) = 0 \text{ for all } f \in A\right\}.$$

Thus, for μ to be in the dual of A we must have that

$$\int_{-\pi}^{\pi} e^{ij\theta}\, d\mu(\theta) = 0 \text{ for } j = 0, 1, 2, \ldots.$$

Now the F. and M. Riesz theorem (Section 9.6) applies and $d\mu(\theta) = g(e^{i\theta})\, d\theta$ for some $g \in L^1(-\pi, \pi)$. Since all the negative Fourier coefficients of g are zero, it follows that $g \in H^1$. And the case $j = 0$ of the last equality tells us that in fact $g \in H_0^1$.

Conversely, if $g \in H_0^1$, we may define a functional α on C/A by setting

$$\alpha(f + A) = \int_{-\pi}^{\pi} g(e^{i\theta}) f(e^{i\theta})\, d\theta.$$

This completes the proof. □

Putting together our two lemmas, we find that

$$\|v_0 - A\|_\infty = \sup\{|\tau(v_0)| : \tau \in H_0^1, \|\tau\| \leq 1\}.$$

In other words,

$$\|v_0 - A\|_\infty = \sup\left\{\left|\int_0^{2\pi} v_0(e^{i\theta}) f(e^{i\theta})\, d\theta\right| : f \in H_0^1, \|f\|_1 \leq 1\right\}. \quad (11.5.1)$$

By our earlier discussion, we may restrict attention to those $f \in H^1$ that have an analytic continuation to a disk $D(0, R_f)$ with $R_f > 1$.

Now we apply Green's theorem to the integral on the right-hand side of (11.5.1). This is a bit tricky, and we write out the details. To begin, we write

$$\int_0^{2\pi} v_0(e^{i\theta})f(e^{i\theta})\,d\theta = \int_{\partial D} [v_0(z)f(z)]\frac{\partial}{\partial\nu}\log|z| - \frac{\partial}{\partial\nu}[v_0(z)f(z)]\log|z|\,ds(z)\,. \tag{11.5.2}$$

Here $\partial/\partial\nu$ denotes the outward normal derivative. Notice that the normal derivative $[\partial/\partial\nu]\log|z|$ equals 1 on the boundary of the disk. And of course $\log|z| = 0$ on ∂D.

The trouble with applying Green's theorem with this integrand on the disk is that $\log|z|$ has a singularity at the origin. So, to be rigorous, we need to apply Green's theorem on $D' \equiv D \setminus D(0,\epsilon)$ for some small $\epsilon > 0$. It is easy to see that the integral over $\partial D(0,\epsilon)$ makes a negligible contribution (vanishing as $\epsilon \to 0^+$), and we shall ignore it. The result then is that (11.5.2) equals

$$\begin{aligned}
&\iint_{D'} [v_0(z)f(z)] \triangle \log|z| - \triangle[v_0(z)f(z)]\log|z|\,dA(z)\\
&= \iint_{D'} -\triangle\,[v_0(z)f(z)]\log|z|\,dA(z)\\
&= \iint_{D'} \triangle[v_0(z)f(z)]\log\frac{1}{|z|}\,dA(z)\,.
\end{aligned}$$

Now letting $\epsilon \to 0^+$ yields

$$\int_0^{2\pi} v_0(e^{i\theta})f(e^{i\theta})\,d\theta = \iint_D \triangle[v_0(z)f(z)]\log\frac{1}{|z|}\,dA(z)\,. \tag{11.5.3}$$

Since f is holomorphic, we may easily calculate that

$$\triangle[v_0 f] = 4\partial\overline{\partial}[v_0 f] = 4\overline{\partial}v_0\partial f + 4f \triangle v_0\,.$$

Also

$$\partial\bar{\partial}v_0 = \partial h\,.$$

Putting these two calculations together gives

$$\triangle[v_0 f] = 4hf' + 4f\partial h\,.$$

Now the integral in (11.5.3) becomes

$$4\iint_D f(z)\partial h(z)\log\frac{1}{|z|}\,dx\,dy + 4\iint_D f'(z)h(z)\log\frac{1}{|z|}\,dx\,dy \equiv I + II\,. \tag{11.5.4}$$

We may estimate

$$I \le 4 \iint_D \left| \frac{f(z)}{z} \right| \left\{ |\partial h(z)||z| \log \frac{1}{|z|} \right\} dx\, dy \, .$$

Since $f \in H^1_0$, $f^*(z) \equiv f(z)/z \in H^1$ and $\|f\|_{H^1} = \|f^*\|_{H^1}$, Lemma 11.5.1 tells us that $|\partial h(z)||z| \log[1/|z|]$ is a Carleson measure μ. So we may write the last expression as

$$4 \iint_D |f^*(z)| \, d\mu(z) \, .$$

Now our analytic characterization of Carleson measures (Theorem 11.4.1) tells us that this last is dominated by $4B\|f\|_{H^1}$.

In order to estimate II, it is necessary for us to do some more calculations related to Green's theorem. We present a result of P. Stein.

Lemma 11.5.4. *If f is a nonvanishing holomorphic function, then*

$$\triangle |f(z)| = \frac{|\partial f / \partial z|^2}{|f(z)|} \, .$$

Proof. This is a straightforward calculation. We have

$$\begin{aligned}
\triangle |f(z)| &= 4 \frac{\partial}{\partial z} \frac{\partial}{\partial \overline{z}} [f(z)\overline{f(z)}]^{1/2} \\
&= 4 \frac{\partial}{\partial z} \left[\frac{1}{2} (f(z)\overline{f(z)})^{-1/2} f(z) \frac{\partial \overline{f(z)}}{\partial \overline{z}} \right] \\
&= 2 \frac{\partial}{\partial z} \left[f(z)^{1/2} \overline{f(z)}^{-1/2} \frac{\partial \overline{f}(z)}{\partial \overline{z}} \right] \\
&= 2 \left[\frac{1}{2} f(z)^{-1/2} \frac{\partial f}{\partial z} \overline{f}^{-1/2} \frac{\partial \overline{f}}{\partial \overline{z}} \right] \\
&= \frac{1}{|f(z)|} \cdot \left| \frac{\partial f(z)}{\partial z} \right|^2 . \quad \square
\end{aligned}$$

Lemma 11.5.5. *Let $W(z) = |z|W'(z)$, with the function W' being twice continuously differentiable on the closed unit disk. Then*

$$\int_{-\pi}^{\pi} W(e^{i\theta}) \, d\theta = \iint_D \left(\log \frac{1}{|z|} \right) \triangle W(z) \, dx\, dy \, .$$

Proof. Exactly like the calculation following Lemma 11.5.3 above. □

Proposition 11.5.6. *Let f be holomorphic on a disk $D(0, R)$ for some $R > 1$. Assume that f has a simple zero at 0 and no other zeros in $D(0, R)$. Then*

$$\int_{-\pi}^{\pi} |f(e^{i\theta})|\, d\theta = \iint_D \left(\log \frac{1}{|z|}\right) \frac{|\partial f/\partial z|^2}{|f(z)|}\, dx\, dy\,.$$

Proof. Simply apply the result of the preceding lemma with $W(z) = |f(z)| = |z| \cdot |f(z)/z|$. □

We estimate II from (11.5.4) using Schwarz's inequality:

$$II \le 4\sqrt{\iint_D \left(\log \frac{1}{|z|}\right) \frac{|f'(z)|^2}{|f(z)|}\, dx\, dy} \;\times\; \sqrt{\iint_D \left|\frac{f(z)}{z}\right| |h(z)|^2 |z| \log \frac{1}{|z|}\, dx\, dy}\,.$$

Now observe that if $f \in H^1_0$ and f has a zero of order k at the origin, then we may write

$$f(z) = z^k \widetilde{f}(z) = z\widetilde{f}(z) + z \cdot (z^{k-1} - 1)\widetilde{f}(z)\,.$$

Thus we have rendered f as the sum of two elements of H^1_0, each of which has a simple zero at the origin. In what follows, then, we may assume that our function f has a simple zero at the origin. For such an f, we have by the preceding proposition that

$$\iint_D \left(\log \frac{1}{|z|}\right) \frac{|f'(z)|^2}{|f(z)|}\, dx\, dy = \|f\|_{H^1}\,.$$

That is the first expression that resulted from Schwarz's inequality.

As in our estimations of term I, we have that

$$\iint_D \left|\frac{f(z)}{z}\right| |h(z)|^2 |z| \log \frac{1}{|z|}\, dx\, dy \le A\|f\|_{H^1}$$

for $f \in H^1_0$ (because $|h(z)|^2|z| \log[1/|z|]$ is a Carleson measure and we know that $f(z)/z \in H^1$). That is the second expression that resulted from Schwarz's inequality.

For $f \in H^1_0$, holomorphic on $\overline{D}$, we now see that

$$\left| 4 \iint_D f'(z)h(z) \log \frac{1}{|z|}\, dx\, dy \right| \le 8\sqrt{A}\|f\|_{H^1}\,.$$

Combining our estimates for terms I and II, we find that

$$\left| \int_0^{2\pi} v_0(e^{i\theta}) f(e^{i\theta})\, d\theta \right| \le 8(\sqrt{A} + B)\|f\|_{H^1}\,.$$

This estimate is valid for $f \in H^1_0$ that are holomorphic on $\overline{D}$. But this is all that is required. We conclude that

$$\|v_0 - A\|_\infty \leq 8(\sqrt{A} + B),$$

and that ends the proof of Lemma 11.5.1. □

11.6 Proof of the Corona Theorem

For simplicity we shall just prove the case $k = 2$ of the second formulation of the corona theorem. This special case is algebraically simpler, but contains all the key analytical ideas. The reader who wants to see the full proof for any number of pieces of corona data should consult [KOO].

By replacing $f_1(z), f_2(z)$ with $f_1(rz), f_2(rz)$, we may assume that each f_j is holomorphic on a neighborhood of $\overline{D}$. Our standing hypothesis is that there is a $\delta > 0$ such that either

$$|f_1(z)| > \delta > 0 \quad \text{or} \quad |f_2(z)| > \delta > 0$$

for all $z \in \overline{D}$. Of course we may assume that $\|f_1\|_\infty \leq 1$ and $\|f_2\|_\infty \leq 1$.

Now select a C^∞ function U such that

(a) $U(w)$ depends only on $|w|$;
(b) $U(w) \equiv 0$ for $|w| \leq \delta/2$;
(c) $U(w) \equiv 1$ for $|w| \geq \delta$.
(d) $0 \leq U(w) \leq 1$ for all $w \in D$.

Our standing hypothesis on f_1, f_2 translates to

$$U\big(f_1(z)\big) + U\big(f_2(z)\big) \geq 1 \quad \text{for } z \in \overline{D}\,.$$

Set

$$\phi_j(z) = \frac{U(f_j(z))}{U(f_1(z)) + U(f_2(z))} \quad \text{for } j = 1, 2\,.$$

These functions ϕ_j are clearly C^∞ on some open disk containing $\overline{D}$ and $\phi_1(z) + \phi_2(z) \equiv 1$. In particular, $\phi_1(z) = 1$ if $\phi_2(z) = 0$ and vice versa. Furthermore, each ϕ_j equals zero on the set where $|f_j(z)| < \delta/2$.

We have the tautological equation

$$\frac{\phi_1}{f_1} \cdot f_1 + \frac{\phi_2}{f_2} \cdot f_2 \equiv 1\,.$$

The equation suggests that g_1 can be ϕ_1/f_1 and g_2 can be ϕ_2/f_2. However, ϕ_1/f_1 and ϕ_2/f_2 are *not holomorphic*. Thus we follow a paradigm of Leray and Serre and endeavor to find correction terms that will make them holomorphic.

In particular, we set

$$g_1 = \frac{\phi_1}{f_1} + vf_2$$

and

$$g_2 = \frac{\phi_2}{f_2} - vf_1\,.$$

We seek a smooth function v that will make g_1, g_2 holomorphic. This means that we must mandate

$$\bar{\partial} g_1 \equiv 0 \tag{11.6.1}$$

and

$$\bar{\partial} g_2 \equiv 0\,. \tag{11.6.2}$$

Note in passing that no matter what the choice of v, the equation $f_1 g_1 + f_2 g_2 \equiv 1$ will be satisfied.

Since $\bar{\partial} f_1 = \bar{\partial} f_2 = 0$, equations (11.6.1) and (11.6.2) translate to

$$\frac{\bar{\partial}\phi_1}{f_1} + f_2 \bar{\partial} v = 0 \quad \text{and} \quad \frac{\bar{\partial}\phi_2}{f_2} - f_1 \bar{\partial} v = 0\,.$$

We know that $\phi_1 + \phi_2 \equiv 1$ hence $\bar{\partial}\phi_1 + \bar{\partial}\phi_2 \equiv 0$. As a result, our equations combine into the single condition

$$\bar{\partial} v = \frac{\bar{\partial}\phi_2}{f_1 f_2}\,.$$

Now we must estimate $\bar{\partial}\phi_2$. Notice that, on the open set where $|f_1(z)| < \delta/2$, we have $\phi_2(z) \equiv 1$ hence $\bar{\partial}\phi_2(z) \equiv 0$. Likewise, on the open set where $|f_2(z)| < \delta/2$, $\phi_1(z) \equiv 0$ hence $\bar{\partial}\phi_1(z) \equiv 0$. As a result,

$$\left| \frac{\bar{\partial}\phi_2}{f_1 f_2} \right| \le \frac{4}{\delta^2} \, |\bar{\partial}\phi_2| \quad \text{on } D$$

and

$$h(z) = \frac{\bar{\partial}\phi_2(z)}{f_1(z) f_2(z)}$$

is a C^∞ function on a neighborhood of $\overline{D}$.

Our aim now is to apply the technical Lemma 11.5.1 from Section 11.5. We know that for $|z| < 1$,

$$|h(z)| \le \frac{4}{\delta^2} |\bar{\partial}\phi_2(z)| = \frac{4}{\delta^2} \frac{|U(f_1(z))\bar{\partial}[U(f_2(z))] - U(f_2(z))\bar{\partial}[U(f_1(z))]|}{|U(f_1(z)) + U(f_2(z))|^2}\,.$$

Notice that

$$\bar{\partial}[U(f_2(z))] = \frac{\partial U}{\partial \overline{z}}(f_2(z)) \cdot \frac{\partial \overline{f_2}}{\partial \overline{z}}(z) d\overline{z}\,.$$

Hence

$$\left|\bar{\partial}[U(f_2(z))]\right| \le C_\delta |f_2'(z)|\,.$$

A similar estimate holds for $|\bar{\partial}U(f_1(z))|$. As a result, since $U(f_1(z)) + U(f_2(z)) \ge 1$, we may establish the estimate

$$|h(z)|^2 |z| \log \frac{1}{|z|} \le 2C_\delta^2 \big(|f_1'(z)|^2 + |f_2'(z)|^2\big)|z| \log \frac{1}{|z|}\,. \tag{11.6.3}$$

We know that f_1, f_2 are bounded by 1 and are holomorphic, hence harmonic, on D. The next milestone in our argument is Lemma 11.6.2. Its proof, which is long and technical, is deferred to the end of the chapter. We first need a little preliminary material.

If F is a given function on ∂D, then let U_F be its Poisson integral to the disk. If $0 < \rho < 1$, then set $V(z) = U_F(\rho z)$. For F as given, define

$$\mathcal{N}(F) = \sup_{z \in D} \big(U_{F^2}(z) - (U_F(z))^2\big)^{1/2}\,.$$

Another useful norm in harmonic analysis is

$$\|\phi\|_* = \sup_I \frac{1}{|I|} \int_I |\phi(t) - \phi_I|\, dt\,.$$

Here ϕ_I is the average of ϕ on the interval I. The $\| \ \|_*$ norm is called the BMO norm, referring to the functions of bounded mean oscillation invented by John and Nirenberg [JON]. Functions of bounded mean oscillation are of interest in part because they form the dual space of H^1 (see [STE]).

The norm $\mathcal{N}$ is comparable to the "bounded mean oscillation" norm that has just been defined. But it is somewhat easier to work with. Our first key result about this new norm is as follows:

Lemma 11.6.1. *If ϕ is real-valued and 2π-periodic, then*

$$\|\phi\|_* \le K \cdot \mathcal{N}(\phi)\,,$$

where K is a constant that does not depend on ϕ.

Proof. Let I be any interval. Now Schwarz's inequality tells us that

$$\begin{aligned} \frac{1}{|I|} \int_I |\phi(t) - \phi_I|\, dt &\le \left(\frac{1}{|I|} \int_I [\phi(t) - \phi_I]^2\, dt \right)^{1/2} \\ &= \left(\frac{1}{|I|^2} \int_I \int_I [\phi(t) - \phi(s)]^2\, dt\, ds \right)^{1/2}\,. \end{aligned}$$

[For this last equality, multiply out the integrands.]

Now assume that $|I| \le 2\pi$. We may assume that $I = (-\alpha, \alpha)$. We have

$$U_{\phi^2}(r) - [U_\phi(r)]^2 = \frac{1}{8\pi^2} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} [\phi(s) - \phi(t)]^2 \cdot \frac{1 - r^2}{1 - 2r\cos s + r^2} \times \frac{1 - r^2}{1 - 2r\cos t + r^2} \, ds \, dt \, ,$$

where we let $r = 1 - \sin[\alpha/2]$.

With this r, and with $-\alpha \le t \le \alpha$, we see that

$$\frac{1 - r^2}{1 - 2r\cos t + r^2} = \frac{(1 + r)(1 - r)}{(1 - r)^2 + 4r\sin^2(t/2)} \ge \frac{1}{5\sin(\alpha/2)} \ge \frac{2}{5\alpha} \, .$$

As a result,

$$\begin{aligned} U_{\phi^2}(r) - [U_\phi(r)]^2 &\ge \frac{1}{50\pi^2\alpha^2} \int_{-\alpha}^{\alpha} \int_{-\alpha}^{\alpha} [\phi(s) - \phi(t)]^2 \, dsdt \\ &= \left(\frac{2}{5\pi}\right)^2 \cdot \frac{1}{2|I|^2} \int_I \int_I [\phi(s) - \phi(t)]^2 \, dsdt \, . \end{aligned}$$

If we combine this last inequality with the one from the start of the proof, we find that

$$\frac{1}{|I|} \int_I |\phi(t) - \phi_I| \, dt \le \frac{5\sqrt{2}\pi}{2} \mathcal{N}(\phi)$$

for all intervals I of length not exceeding 2π.

Now let $k > 0$ be an integer and assume that $2\pi k < |I| \le 2\pi(k + 1)$. Let $J \supseteq I$ be an interval of length $2\pi(k + 1)$. Then

$$\frac{1}{|I|} \int_I |\phi(t) - \phi_J| \, dt \le \frac{k + 1}{k} \cdot \frac{1}{|J|} \int_J |\phi(t) - \phi_J| \, dt \le \frac{2}{|J|} \int_J |\phi(t) - \phi_I| \, dt \, .$$

Of course ϕ is 2π-periodic, so it makes sense to write $\phi_J = [1/2\pi] \int_0^{2\pi} \phi(s) \, ds$. Also

$$\frac{1}{|J|} \int_J |\phi(t) - \phi_J| \, dt = \frac{1}{2\pi} \int_{-\pi}^{\pi} |\phi(t) - \Phi_{[-\pi,\pi]}| \, dt \, ,$$

and we know that this last does not exceed $[5\sqrt{2}\pi/2]\mathcal{N}(\phi)$.

Finally, it is straightforward to check that

$$\begin{aligned} \frac{1}{|I|} \int_I |\phi(t) - \phi_I| \, dt &\le \frac{2}{|I|} \int_I |\phi(t) - \phi_J| \, dt \\ &\le \frac{4}{|J|} \int_J |\phi(t) - \phi_J| \, dt \le 10\sqrt{2}\pi \cdot \mathcal{N}(\phi) \, . \end{aligned}$$

This shows that

$$\|\phi\|_* \le 10\sqrt{2}\pi\mathcal{N}(\phi) \, . \quad \square$$

Lemma 11.6.2. *Let $\mathcal{C}_{P,h} \subseteq D$ be a Carleson region. Let V be a harmonic function on a neighborhood of $\overline{D}$, $V = U_F$. Then*

$$\left(|z| \log \frac{1}{|z|}\right) (\nabla V \cdot \nabla V)\, dx\, dy$$

is a Carleson measure, and the Carleson constant is majorized by a constant multiple of $(\mathcal{N}(F))^2$.

The proof of Lemma 11.6.2 is deferred to the end of the chapter.

Applying the lemma to f_1, f_2, we see that the right-hand side of (11.6.3) determines a Carleson measure and hence

$$|h(z)|^2 |z| \log \frac{1}{|z|}\, dx\, dy$$

is a Carleson measure. The Carleson constant A_δ for this measure depends only on δ.

With the technical lemma of Section 11.5 in mind, we need now to identify one more Carleson measure. Begin by calculating

$$\partial h = \frac{\partial \bar{\partial} \phi_2}{f_1 f_2} - \frac{\bar{\partial} \phi_2}{f_1 f_2} \left(\frac{f_1'}{f_1} + \frac{f_2'}{f_2} \right) .$$

Notice that the second term on the right vanishes identically on the open set W where either $|f_1|$ or $|f_2|$ is less than $\delta/2$. On the complement of that set we see that this second term is bounded by

$$\frac{8}{\delta^3} \sup_w |\nabla U(w)| (|f_1'(z)| + |f_2'(z)|)^2 .$$

The first term also vanishes identically on W, and on ${}^cW$ this term is bounded by

$$\frac{1}{4|f_1(z) f_2(z)|} \triangle \left(\frac{U(f_2(z))}{U(f_1(z)) + U(f_2(z))} \right) .$$

But $\triangle f_1 = \triangle f_2 \equiv 0$, so this last expression involves only the derivatives f_1' and f_2'. In modulus this expression is bounded by

$$C_\delta (|f_1'(z)|^2 + |f_2'(z)|^2) .$$

Thus we see that

$$|\partial h(z)| |z| \log \frac{1}{|z|} \le K_\delta (|f_1'(z)|^2 + |f_2'(z)|^2) |z| \log \frac{1}{|z|} .$$

As a result, using Lemma 11.5.1, we find that

$$|\partial h(z)| |z| \log \frac{1}{|z|}\, dx\, dy$$

is a Carleson measure with Carleson constant B_δ depending only on δ.

Now our technical Lemma 11.5.1 gives us a $\bar{\partial}$-solution v that is C^∞ in some neighborhood of $\overline{D}$ such that $\bar{\partial} v = h$ on D and $|v(e^{i\theta})| \le 9(\sqrt{A_\delta} + B_\delta)$. The functions

$$g_1 = \frac{\phi_1}{f_1} + v f_2$$

and

$$g_2 = \frac{\phi_2}{f_2} - v f_1$$

are in H^∞; in fact they are both in $A(D)$. And they satisfy, for trivial algebraic reasons, $g_1 f_1 + g_2 f_2 \equiv 1$ on D.

Finally, for $z \in \overline{D}$,

$$\left| \frac{\phi_1(z)}{f_1(z)} \right| \le \frac{2}{\delta} \quad \text{and} \quad \left| \frac{\phi_2(z)}{f_2(z)} \right| \le \frac{2}{\delta} .$$

Hence, for $j = 1, 2$,

$$|g_j(e^{i\theta})| \le \frac{2}{\delta} + 9\left(\sqrt{A_\delta} + B_\delta\right) ,$$

that is,

$$\|g_j\|_\infty \le \frac{2}{\delta} + 9\left(\sqrt{A_\delta} + B_\delta\right) .$$

This proves the corona theorem for two pieces of corona data.

11.7 Proof of Lemma 11.6.1

Our job is to show that

$$\iint\limits_D \left(|z| \log \frac{1}{|z|} \right) (\nabla V \cdot \nabla V) |g(z)| \, dx \, dy \le K''(\mathcal{N}(F))^2 \|g\|_{H^1} .$$

We do this in two steps. The first is fairly straightforward.

Lemma 11.7.1. *For all $g \in H^1(D)$ and for some constant K' we have*

$$\iint\limits_{|z|<1/2} \left(|z| \log \frac{1}{|z|} \right) (\nabla V \cdot \nabla V) |g(z)| \, dx \, dy \le K'(\mathcal{N}(F))^2 \|g\|_{H^1} .$$

Proof. For $|z| \le 1/2$, we have the easy estimates $|z| \log 1/|z| \le 1/e$ and $|g(z)| \le (1/\pi)\|g\|_{H^1}$.

Now $V(z) = U_F(\rho z)$, where $0 < \rho < 1$, hence $(\nabla V)(z) = \nabla U_F(\rho z) = \nabla U_{F_1}(\rho z)$, where

$$F_1(\theta) = F(\theta) - \frac{1}{2\pi}\int_{-\pi}^{\pi} F(t)\,dt\,.$$

We may differentiate the Poisson integral formula directly to find that for $|z| \le 1/2$,

$$|\nabla U_{F_1}(\rho z)| \le \frac{4}{\pi}\int_{-\pi}^{\pi} |F_1(e^{it})|\,dt\,,$$

that is,

$$|\nabla U_{F_1}(\rho z)| \le \frac{8}{2\pi}\int_{-\pi}^{\pi} \left| F(\theta) - \frac{1}{2\pi}\int_{-\pi}^{\pi} F(t)\,dt \right| d\theta$$

for $|z| \le 1/2$.

Inspecting the proof of Lemma 11.6, we see that the expression on the right is majorized by $20\sqrt{2}\pi\mathcal{N}(F)$. Putting our estimates together, we find that

$$\iint\limits_{|z|<1/2} \left(|z|\log\frac{1}{|z|}\right)(\nabla V \cdot \nabla V)|g(z)|\,dx\,dy \le \frac{800\pi}{e}(\mathcal{N}(F))^2\|g\|_{H^1}\,. \quad \square$$

The main work in establishing our result is in the next lemma.

Lemma 11.7.2. *For any $g \in H^1(D)$ and for some numerical constant K'' we have*

$$\iint\limits_{1/2<|z|<1} \left(|z|\log\frac{1}{|z|}\right)(\nabla V \cdot \nabla V)|g(z)|\,dx\,dy \le K''(\mathcal{N}(F))^2\|g\|_{H^1}\,.$$

Of course combining the two lemmas gives the result. The remainder of this section will be devoted to the proof of the last lemma.

The key technical ingredient to establishing the last lemma is this next result

Lemma 11.7.3. *Let $0 < h < 1/2$ and $P = e^{i\theta_0} \in \partial D$. Let $\mathcal{C}_{P,h}$ be the corresponding Carleson region. Then there is a numerical constant C such that*

$$\iint\limits_{\mathcal{C}_{P,h}} \left(|z|\log\frac{1}{|z|}\right)(\nabla V \cdot \nabla V)\,dx\,dy \le C(\mathcal{N}(F))^2 \cdot h\,.$$

Proof. We begin by transforming the problem from the disk to the upper halfplane. Set

$$\zeta = i\,\frac{1-z}{1+z}\,.$$

Here z is the variable in the disk and ζ the corresponding variable in the upper halfplane U. Write $\zeta = \xi + i\eta$. Likewise put $V(z) = w(\zeta)$. We may as well suppose that V, w are real-valued.

For $1/2 \le |z| < 1$ we have the estimates

$$\begin{aligned}
|z|\log\frac{1}{|z|} &\le \frac{1}{2}\log\frac{1}{|z|^2} \\
&= \frac{(1-|z|^2)}{2}\big(1+\frac{1}{2}(1-|z|^2)^2+\frac{1}{3}(1-|z|^2)^3+\cdots\big) \\
&\le \frac{(1-|z|^2)}{2|z|^2} \\
&\le 2(1-|z|^2)\,.
\end{aligned}$$

Written in terms of $\zeta = \xi + i\eta$, we have

$$|z|^2 = \frac{\xi^2+(\eta-1)^2}{\xi^2+(\eta+1)^2}\,,$$

hence

$$1-|z|^2 = \frac{4\eta}{\xi^2+(\eta+1)^2} \le 4\eta\,.$$

In conclusion,

$$|z|\log\frac{1}{|z|} \le 8\eta \quad \text{for } \frac{1}{2} \le |z| < 1\,. \tag{11.7.1}$$

Now let $\mathcal{C}_{P,h}$ correspond, under the mapping $z \leftrightarrow \zeta$, to a Carleson region $\mathcal{C}'_{P',h'}$ in the upper halfplane U. It is a simple calculation to see that $\mathcal{C}'_{P',h'}$ is also a Carleson region—in the sense that it is about $2h$ units wide and h units tall. Since the mapping under consideration is conformal, the quantity we are trying to estimate, namely

$$\iint\limits_{\mathcal{C}_{P,h}} \left(|z|\log\frac{1}{|z|}\right)\left\{\left(\frac{\partial V}{\partial x}\right)^2+\left(\frac{\partial V}{\partial y}\right)^2\right\}dx\,dy\,,$$

is equal to

$$\iint\limits_{\mathcal{C}'_{P',h'}} \left(|z|\log\frac{1}{|z|}\right)\left\{\left(\frac{\partial w}{\partial \xi}\right)^2+\left(\frac{\partial w}{\partial \eta}\right)^2\right\}d\xi\,d\eta\,,$$

and this in turn is

$$\le \iint\limits_{\mathcal{C}'_{P',h'}} 8\eta\left(\left(\frac{\partial w}{\partial \xi}\right)^2+\left(\frac{\partial w}{\partial \eta}\right)^2\right)d\xi\,d\eta\,.$$

In the last estimate we have used (11.7.1).

We summarize our estimate as

$$\iint\limits_{\mathcal{C}_{P,h}} \left(|z|\log\frac{1}{|z|}\right)|\nabla V|^2\,dx\,dy \le 8\iint\limits_{\mathcal{C}'_{P',h}} \eta(w_\xi^2+w_\eta^2)\,d\xi\,d\eta\,.$$

Here of course the subscripts on w denote derivatives. Our job now is to show that the quantity on the right is dominated by $K''(\mathcal{N}(F))^2 \cdot h$.

Now we shall exploit a trick. The Carleson region $\mathcal{C}'_{P',h}$ is contained in a semicircular region $M_{ch/2} \equiv \{\zeta : |\zeta| < (c/2) \cdot h, \eta > 0\}$. Here we may take $c > 2$. For $\zeta \in \mathcal{C}'_{P',h}$, we may be sure that

$$1 - \frac{|\zeta|}{ch} > \frac{1}{4} .$$

As a result,

$$\iint_{\mathcal{C}'_{P',h}} \eta(w_\xi^2 + w_\eta^2)\, d\xi\, d\eta \le 4 \iint_{\substack{|\zeta|<ch \\ \eta>0}} \left(1 - \frac{|\zeta|}{ch}\right) \eta(w_\xi^2 + w_\eta^2)\, d\xi\, d\eta .$$

Note that the factor $(1 - |\xi|/[ch])$ vanishes on the boundary of the region of integration.

Of course w is harmonic because V is, and

$$\left(\frac{\partial^2}{\partial \xi^2} + \frac{\partial^2}{\partial \eta^2}\right) w^2 = 2(w_\xi^2 + w_\eta^2) .$$

So now we must estimate

$$J = 2 \iint_{\substack{|\zeta|<ch \\ \eta>0}} \left(1 - \frac{|\zeta|}{ch}\right) \eta \left(\frac{\partial^2 w^2}{\partial \xi^2} + \frac{\partial^2 w^2}{\partial \xi^2}\right) d\xi\, d\eta$$

in terms of h and $\mathcal{N}(F)$.

Referring to the definition of $\mathcal{N}(F)$, if we put $P(z) = U_{F^2}(z) - [U_F(z)]^2$ for $z \in D$, then we have

$$0 \le P(z) \le (\mathcal{N}(F))^2 .$$

We of course have $z = (i - \zeta)/(i + \zeta)$, so that

$$(w(\zeta))^2 = (U_F(\rho z))^2 = U_{F^2}(\rho z) - P(\rho z) ;$$

notice that the term $U_{F^2}(\rho z)$ on the right is harmonic. If we set $b(\zeta) = P(\rho z)$, with $0 < \rho < 1$ as usual, then

$$\left(\frac{\partial^2}{\partial \xi^2} + \frac{\partial^2}{\partial \eta^2}\right) w^2 = -\left(\frac{\partial^2 b}{\partial \xi^2} + \frac{\partial^2 b}{\partial \eta^2}\right) .$$

So the integral that we must estimate comes down to

$$J = -2 \iint_{\substack{|\zeta|<ch \\ \eta>0}} \left(1 - \frac{|\zeta|}{ch}\right) \eta \left(\frac{\partial^2 b}{\partial \xi^2} + \frac{\partial^2 b}{\partial \eta^2}\right) d\xi\, d\eta ,$$

where

$$0 \le b(\zeta) \le (\mathcal{N}(F))^2 .$$

Let us write

$$\triangle_\zeta = \frac{\partial^2}{\partial \xi^2} + \frac{\partial^2}{\partial \eta^2} .$$

Of course $\zeta = \xi + i\eta$. We apply Green's theorem in the usual manner to the last integral and find that

$$\begin{aligned} J = -2 \iint\limits_{\substack{|\zeta|<ch \\ \eta>0}} b(\zeta) \triangle_\zeta \left(\eta \left(1 - \frac{|\zeta|}{ch} \right) \right) d\xi \, d\eta \\ +2 \int_\Gamma \left[b(\zeta) \frac{\partial(\eta(1-(|\zeta|/[2h])))}{\partial \nu_\zeta} - \eta \left(1 - \frac{|\zeta|}{ch} \right) \frac{\partial b(\zeta)}{\partial \nu_\zeta} \right] |d\zeta| . \end{aligned}$$

Here Γ is the semicircular contour that runs counterclockwise around the boundary of the region $M_{ch/2}$ and differentiation in ν_ζ is directional differentiation in the direction of the outward unit normal.

The line integral around the contour Γ is negative. This is so because

(a) $\eta(1 - |\zeta|/[2h]) \equiv 0$ on Γ;

(b) $\eta(1 - |\zeta|/[2h]) > 0$ inside Γ;

hence

(c) $\partial(\eta(1 - |\zeta|/[2h])/\partial \nu_\zeta \le 0$ on Γ.

Therefore

$$b(\zeta)\partial(\eta(1 - |\zeta|/[2h])/\partial \nu_\zeta \le 0 \quad \text{on } \Gamma$$

and the nonpositivity of the line integral follows.

We thus have

$$J \le -2 \iint\limits_{\substack{|\zeta|<ch \\ \eta>0}} b(\zeta) \triangle_\zeta \left(\eta \left(1 - \frac{|\zeta|}{ch} \right) \right) d\xi \, d\eta .$$

We evaluate the integral using polar coordinates $\zeta = \sigma e^{i\psi}$. Thus

$$\triangle_\zeta = \frac{1}{\sigma} \frac{\partial}{\partial \sigma} \left(\sigma \frac{\partial}{\partial \sigma} (\,\cdot\,) \right) + \frac{1}{\sigma^2} \frac{\partial^2}{\partial \psi^2} .$$

Also

$$\eta \left(1 - \frac{|\zeta|}{ch} \right) = \left(1 - \frac{|\zeta|}{ch} \right) \sigma \sin \psi .$$

From this we easily derive

$$\triangle_\zeta \left(\eta \left(1 - \frac{|\zeta|}{ch} \right) \right) = -\frac{3}{2h} \sin \psi .$$

Substituting these calculations into the integral that estimates J, we find that

$$\begin{aligned} J &\leq \frac{3}{h}\int_0^{\pi}\int_0^{ch} b(\sigma e^{i\psi})\sigma \sin\psi \, d\sigma \, d\psi \\ &\leq \frac{3}{h}(\mathcal{N}(F))^2 \int_0^{\pi}\int_0^{ch} \sigma \sin\psi \, d\sigma \, d\psi = 12(\mathcal{N}(F))^2 h \, . \end{aligned}$$

Returning to the beginning of our calculation, we conclude now with

$$\iint\limits_{\mathcal{C}_{P,h}} \left(|z| \log \frac{1}{|z|} \right) |\nabla V|^2 \, dx \, dy \leq 32 J \leq 384(\mathcal{N}(F))^2 h \, ,$$

and this is what we wished to prove. □

Now combining this last estimate with our standard analytic characterization of Carleson measures, the proof of our lemma is complete.

Problems for Study and Exploration

1. Let $f_1(z) = z - 1/2$ and $f_2(z) = z + 1/2$. Find bounded holomorphic functions g_1, g_2 such that $f_1 g_1 + f_2 g_2 \equiv 1$ on the entire disk.

2. Formulate a "corona theorem" for C^∞ functions on the unit disk. Now prove it. This is a straightforward exercise with partitions of unity. Explain why this proof does not adapt to holomorphic function theory.

3. The corona theorem is relatively easy to prove if we do not require the solutions g_j to be bounded. Explain why this is so.

4. The hypothesis of the corona theorem is that

$$\sum_j |f_j| \geq \delta > 0 \, ,$$

and an examination of the proof of the corona theorem gives upper bounds on the g_j in terms of powers of $1/\delta$. Can you tell from Wolff's proof what powers might be involved? In fact the optimal estimate for the g_j is not known.

5. Suppose that the f_j given in the hypothesis of the corona theorem are all polynomials. May we conclude that the g_j provided by the corona theorem are polynomials? Why or why not?

6. Let $\{p_j\}$ be points in the unit disk and φ_{p_j} the corresponding point evaluation functionals. Apply the Banach–Alaoglu theorem to elicit a new linear functional Φ that is the weak-$*$ limit of a subsequence of the $\{\varphi_{p_j}\}$. Explain why Φ is *not* necessarily a point evaluation.

7. Refer to Exercise 6. If the sequence $\{p_j\}$ accumulates at an interior point of the disk, then the limit functional will in fact be the obvious point evaluation. If instead the sequence $\{p_j\}$ accumulates at a boundary point of the disk, then the limit functional will *not* be a point evaluation. If $\{p_j\}$ and $\{q_j\}$ are *distinct* sequences that accumulate at the same boundary point p, then will their Banach–Alaoglu limits be *different* linear functionals?

8. A multiplicative linear functional on H^∞ that is constructed as in Exercises 6, 7 is said to "live in the fiber over p." Show that any such fiber has uncountably many distinct elements.

Part III

Algebraic Topics

12

Automorphism Groups of Domains in the Plane

Genesis and Development

Felix Klein's *Erlangen program* lays out a blueprint for understanding a geometry by way of the mappings that preserve that geometry. This vision has become quite prevalent and powerful in modern approaches to the subject. Certainly Alexandre Grothendieck and Saunders Mac Lane carried this idea to new heights in their modern formulations of algebraic geometry and algebraic topology.

For complex function theory, Klein's idea may be implemented by means of the study of conformal mappings. While there is certainly value in studying individual mappings, there is even more to be had from studying *groups* of conformal mappings. Thus we are led to study the group of conformal self-maps (the *automorphism group*) of a planar domain. This idea interacts elegantly and effectively with the notion of invariant metric (see Chapter 1).

A very basic acquaintance with the group concept and with conformal mappings is all that is required to appreciate the ideas in the present chapter. The study of automorphism groups exhibits a fruitful interaction of algebra, analysis, and geometry. It points to many new directions in the subject.

12.1 Introductory Concepts

As mentioned earlier, Felix Klein taught us that a natural way to understand a geometric object is by way of the groups of transformations that act on that object. In complex analysis, the natural transformations (or, in more abstract language, the "morphisms") are the conformal (one-to-one, onto, holomorphic) mappings. If $\phi : \Omega_1 \to \Omega_2$ is such a mapping, then ϕ is a device for transplanting the complex analysis of Ω_1 to the complex analysis of Ω_2 (and vice versa).

If Ω is a fixed domain, then of particular interest are the conformal *self-maps* of Ω. The collection of such mappings forms a group under composition of mappings:

- The composition of two self-maps is another;
- Each self-map of Ω has an inverse that is also a self-map;
- The identity map is the group identity;
- The binary operation of composition of mappings is associative.

We call this group the *automorphism group* of Ω, denote it by $\mathrm{Aut}(\Omega)$, and call its elements *automorphisms*.

The following result of H. Cartan will be of considerable utility in our study of automorphisms:

Proposition 12.1.1. *Let $\Omega \subseteq \mathbb{C}$ be a bounded domain. Let $P \in \Omega$ and suppose that $\phi : \Omega \to \Omega$ satisfies $\phi(P) = P$. If $\phi'(P) = 1$, then ϕ is the identity.*

Proof. We may assume that $P = 0$. Expanding ϕ in a power series about $P = 0$, we have

$$\phi(z) = z + P_k(z) + O(|z|^{k+1}),$$

where P_k is the first nonvanishing monomial (of degree k) of order exceeding 1 in the Taylor expansion. Notice that

$$P_k(z) = \left(\frac{1}{k!}\frac{\partial^k}{\partial z^k}\phi(0)\right) z^k .$$

Defining $\phi^j(z) = \phi \circ \cdots \circ \phi$ (j times) we have

$$\begin{aligned} \phi^2(z) &= z + 2P_k(z) + O(|z|^{k+1}) , \\ \phi^3(z) &= z + 3P_k(z) + O(|z|^{k+1}) , \\ &\vdots \\ \phi^j(z) &= z + jP_k(z) + O(|z|^{k+1}). \end{aligned}$$

Choose disks $D(0,a) \subseteq \Omega \subseteq D(0,b)$. Then for $0 \le j \in \mathbb{Z}$ we know that $D(0,a) \subseteq \operatorname{dom}\phi^j \subseteq D(0,b)$. Therefore the Cauchy estimates imply that, for our index k, we have

$$j\left|\frac{\partial^k}{\partial z^k}\phi(0)\right| = \left|\frac{\partial^k}{\partial z^k}\phi^j(0)\right| \le \frac{b \cdot k!}{a^k}.$$

Letting $j \to \infty$ yields that $\phi^{(k)}(0) = 0$.

We conclude that $P_k = 0$; this contradicts the choice of P_k unless $\phi(z) \equiv z$. □

There are other ways to look at this result, some of them quite illuminating. Imagine that Ω is equipped with the Bergman metric. Because $\phi'(P) = 1$, we see that ϕ is invertible in a neighborhood of P. If γ is a geodesic emanating from P, then $\phi \circ \gamma$ will also be a geodesic emanating from P. In fact, since $\Phi(P) = P$ and $\phi'(P) = 1$, the curve $\phi \circ \gamma$ will be a geodesic emanating from the same point and pointing in the same direction. By uniqueness for first-order differential equations (Picard's theorem), $\phi \circ \gamma$ will coincide with γ—at least at points near P. Since this statement is true for *any* geodesic emanating from P, we may conclude that ϕ is the identity in a neighborhood of P. By analytic continuation, we see that $\phi(z) \equiv z$.

Now suppose that Ω is a fixed domain and $P \in \Omega$ a given point. Let ϕ_1, ϕ_2 be automorphisms of Ω such that $\phi_1(P) = \phi_2(P)$ and $\phi_1'(P) = \phi_2'(P)$. Then $\psi \equiv \phi_1 \circ \phi_2^{-1}$ has the property that $\psi : \Omega \to \Omega$, $\psi(P) = P$, and $\psi'(P) = 1$. By Cartan's result, we conclude that $\psi(z) \equiv z$ and hence $\phi_1(z) \equiv \phi_2(z)$. In conclusion, the mapping

$$\mathrm{Aut}(\Omega) \ni \phi \mapsto (\phi(P), \phi'(P))$$

is one-to-one. In effect, this mapping shows that the automorphism group may be identified with a subset of some Euclidean space (the target is $\mathbb{C} \times \mathbb{C}$, which is $\mathbb{R}^4$). In point of fact (we shall not provide the details here, but see [KOB]), the group $\mathrm{Aut}(\Omega)$ is a *Lie group*, which simply means that $\mathrm{Aut}(\Omega)$ is a group that is also a manifold, and the group operations are continuous (indeed, real analytic) in the topology of the manifold.[1]

One consequence of the discussion in the last paragraph is that, for a given Ω, the group $\mathrm{Aut}(\Omega)$ has a *dimension*. It is natural to wonder what the dimension of $\mathrm{Aut}(\Omega)$ tells us about Ω itself. We shall discuss that question in Section 12.3.

12.2 Noncompact Automorphism Groups

Before we can engage in a detailed analysis of the dimensions of automorphism groups, we must amass some other ideas and techniques.

We have already noted that the automorphism group is a topological group. In fact the topology we use, as is standard in complex analysis, is the topology of *uniform convergence on compact sets* (in point-set topology this is known as the *compact-open topology*). As we shall see in the developing ideas, it is a matter of some interest to see when the automorphism group is compact and when it is not. First we consider some examples.

Example 12.2.1. Let $D = \{z \in \mathbb{C} : |z| < 1\}$ and $A = \{z \in \mathbb{C} : 1/2 < |z| < 2\}$. Then D is the unit disk in the complex plane and A is an annulus. Let us

[1] These issues are related to Hilbert's fifth problem about the topology of Lie groups. See [BRO].

use the idea of the automorphism group to see that D and A cannot be conformally equivalent.[2]

Now we know from any complex analysis text (see, for example, [GRK1]), that the collection of conformal self-maps of the disk D consists of all maps of the form

$$\zeta \mapsto e^{i\tau} \cdot \frac{\zeta - a}{1 - \overline{a}\zeta},$$

where τ is a fixed real number and a is a complex number of modulus less than 1. On the other hand, the collection of conformal maps of the annulus A consists of all rotations $\zeta \mapsto e^{i\tau}\zeta$ together with the inversion $\lambda : \zeta \mapsto 1/\zeta$ (and of course compositions of these two types). Again see [GRK1] and our Theorem 1.5.1.

Let us consider D from the point of view of sequential compactness. Consider the elements $\alpha_j \in \mathrm{Aut}(D)$ given by

$$\alpha_j(\zeta) = \frac{\zeta - (1 - 1/j)}{1 - (1 - 1/j)\zeta}.$$

Then we see that

$$\lim_{j \to \infty} \alpha_j(\zeta) \equiv -1,$$

uniformly on compact sets. We conclude that the topological group $\mathrm{Aut}(D)$ is *not* compact; for we have produced a sequence of elements of $\mathrm{Aut}(D)$ that has no subsequence *converging to an element of* D.

By contrast, our explicit description of the automorphism group of A shows that $\mathrm{Aut}(A)$ is just two copies of the circle. Put more analytically, if β_j are elements of $\mathrm{Aut}(A)$, then either **(i)** infinitely many of them have a factor of the inversion λ or else **(ii)** infinitely many of them do not. Suppose the former. Then each β_j may be written $\beta_j = \lambda \circ e^{i\rho_j}$, where $0 \leq \rho_j \leq 2\pi$. Then it is easy to extract a subsequence that converges uniformly. As a result, $\mathrm{Aut}(A)$ is compact. [A similar argument applies in the second eventuality.]

Now if there were a conformal mapping $\psi : D \to A$, then the induced mapping

$$\begin{aligned} \psi_* : \mathrm{Aut}(D) &\to \mathrm{Aut}(A), \\ \phi &\mapsto \psi \circ \phi \circ \psi^{-1}, \end{aligned}$$

would be a topological group isomorphism. But of course these two groups *cannot* be isomorphic as topological groups just because one is compact and the other is not![3]

[2] It should be stressed that it is a priori clear that these two domains cannot be conformally equivalent because they are not even topologically equivalent. After all, D is simply connected while A is not. But our point is to see how automorphism groups can be used to obtain useful results.

[3] It is also worth noting that one automorphism group (for the annulus) is abelian while the other (for the disk) is not. Details are left as an exercise.

We conclude that D and A are not conformally equivalent.

Thus we see that the automorphism group can serve as a device for differentiating conformal equivalence and/or inequivalence of domains. In the context of the plane this device may seem somewhat artificial, just because so many other powerful tools (the Riemann mapping theorem, the uniformization theorem, tools from Riemann surface theory, etc.) are at our disposal.

In fact compactness/noncompactness is only one of many possible dialectics that we could use to see that the annulus and the disk cannot be conformally equivalent. We could also look at the *dimension* of the automorphism groups. As already noted, the automorphism group of the annulus is two circles, and has dimension 1. By contrast, the automorphism group of the disk is all maps of the form

$$\zeta \mapsto e^{i\tau} \cdot \frac{\zeta - a}{1 - \overline{a}\zeta} .$$

Plainly there are three free parameters (one for the real parameter τ and two for the complex parameter a) and the automorphism group of the disk therefore has dimension 3. As a result, the annulus and the disk cannot be conformally equivalent.

In fact we may apply this philosophy in other interesting circumstances. Part of the content of the uniformization theorem of Köbe (see Section 4.6) is that there are only three (conformally distinct) simply connected Riemann surfaces: the disk D, the plane $\mathbb{C}$, and the Riemann sphere $\widehat{\mathbb{C}}$. Let us discuss for a moment the automorphism groups of these objects.

As already noted, $\mathrm{Aut}(D)$ has dimension 3. The conformal self-mappings of the plane are those of the form

$$\zeta \mapsto a\zeta + b$$

for arbitrary complex constants $a \neq 0$ and b. Thus $\mathrm{Aut}(\mathbb{C})$ has dimension 4. The conformal self-mappings of the sphere are the *linear fractional transformations*

$$\zeta \mapsto \frac{a\zeta + b}{c\zeta + d} .$$

To eliminate redundancies, we write these (by dividing out by c or d, whichever is nonzero) as

$$\zeta \mapsto \frac{\alpha\zeta + \beta}{\gamma\zeta + 1} \quad \text{or} \quad \zeta \mapsto \frac{\alpha\zeta + \beta}{\zeta + \delta} .$$

Thus, plainly, the dimension[4] of $\mathrm{Aut}(\widehat{\mathbb{C}})$ is 6. In any event, the calculations in the last paragraph explain why the disk, the plane, and the sphere must

[4] In fact it is worth noting that the conformal self-maps of the plane are just those conformal self-mappings of the sphere that map ∞ to ∞. When we pass from mappings of the plane to mappings of the sphere, we have the extra latitude of moving the point at ∞ to another point on the sphere—and that gives two more degrees of freedom. That is why $\dim(\mathrm{Aut}(\widehat{\mathbb{C}})) = 6$ and $\dim(\mathrm{Aut}(\mathbb{C})) = 4$.

all be conformally distinct. [Of course the sphere must be distinct from the other two anyway because it is topologically inequivalent.] Note that a more classical proof of the conformal inequivalence of the disk and the plane follows from Liouville's theorem.

Given what we have seen so far, it is natural to wonder which planar domains have compact automorphism group. It is a classical fact (see [HEI1], [HEI2], or [FAK]) that any domain with at least two, but finitely many, holes has in fact a *finite* automorphism group. [We shall provide some heuristic justification for this assertion below. See also the discussion at the end of Section 12.5.] Thus the automorphism group will certainly be compact. It is natural, then, to conjecture that if a planar domain is *not* simply connected, then it will have compact automorphism group (and, conversely, if the planar domain *is* simply connected then it will have noncompact automorphism group).

Part of this statement is obvious: If Ω is a simply connected planar domain, not the entire plane, then the Riemann mapping theorem tells us that Ω is conformally equivalent to the disk D. But then the argument given in our first example shows that $\mathrm{Aut}(\Omega)$ will be topologically isomorphic to $\mathrm{Aut}(D)$. It follows, then, that $\mathrm{Aut}(\Omega)$ is noncompact.

The other part of the statement is *false*, as the following example shows.

Example 12.2.2. Let $D = \{\zeta \in \mathbb{C} : |\zeta| < 1\}$ as before. Let $\overline{d} = \{\zeta \in \mathbb{C} : |\zeta| \leq 1/10\}$. Define

$$\phi(\zeta) = \frac{\zeta - 1/2}{1 - (1/2)\zeta} .$$

For $1 \leq j \in \mathbb{Z}$, we let

$$\phi^j(\zeta) \equiv \underbrace{\phi \circ \phi \circ \cdots \circ \phi}_{j \text{ times}}(\zeta) .$$

If $j < 0$, $j \in \mathbb{Z}$, we let

$$\phi^j(\zeta) \equiv [\phi^{-1}]^{|j|}(\zeta) .$$

And, finally, we set

$$\phi^0(\zeta) = \zeta .$$

With this language, we define

$$\Omega = D \setminus \bigcup_{j=-\infty}^{\infty} \phi^j(\overline{d}) .$$

Figure 12.1 exhibits this domain. It is obviously infinitely connected. What is its automorphism group?

Certainly any ϕ^k is an automorphism of Ω, just because $\phi^k \circ \phi^j = \phi^{j+k}$. Also the mapping $\zeta \mapsto -\zeta$ is an automorphism of Ω. It is not difficult to show, though we shall not provide the details, that these are all the automorphisms of Ω.

Further note that, by dint of tedious hand calculation, we may actually calculate ϕ^j. In fact, for all j,

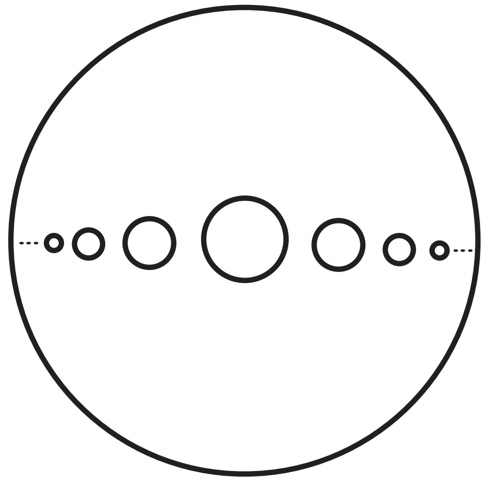

Fig. 12.1. The domain Ω.

$$\phi^j(\zeta) = \frac{\zeta - \frac{3^j-1}{3^j+1}}{1 - \frac{3^j-1}{3^j+1}\zeta} .$$

We see, therefore, that $\phi^j \to -1$, uniformly on compact sets, as $j \to +\infty$ (and likewise $\phi^j \to +1$, uniformly on compact sets, as $j \to -\infty$). In conclusion, $\mathrm{Aut}(\Omega)$ is noncompact.

The upshot of this last example is that the possession of holes is not the sole determinant of when a planar domain has compact automorphism group. In fact part of what makes this subject fascinating is that the positive results require a combination of algebra, complex function theory, and differential geometry. The next theorem answers the quandary into which we have fallen, and also illustrates this cross-fertilization of techniques.

Theorem 12.2.3. *Let $\Omega \subseteq \mathbb{C}$ be a bounded domain with C^1 boundary (i.e., the boundary consists of finitely many simple, closed, continuously differentiable curves). If Ω has noncompact automorphism group, then Ω is conformally equivalent to the unit disk.*

Proof. By the Riemann mapping theorem, it suffices to prove that Ω is simply connected. We establish this fact in several steps.

(1) There is a noncompact orbit. We claim that there must be a point $P \in \Omega$ and elements $\phi_j \in \mathrm{Aut}(\Omega)$ such that $\phi_j(P)$ accumulates at a boundary point $Q \in \partial\Omega$. If this were not the case, then for any collection $\{\psi_j\} \subseteq \mathrm{Aut}(\Omega)$ and any fixed $P \in \Omega$ it would hold that $\{\psi_j(P)\}$ lies in a fixed, compact subset $L \subseteq \Omega$. Of course $\mathrm{Aut}(\Omega)$ is a normal family (since Ω is bounded), hence there is a normally convergent subsequence ψ_{j_k}

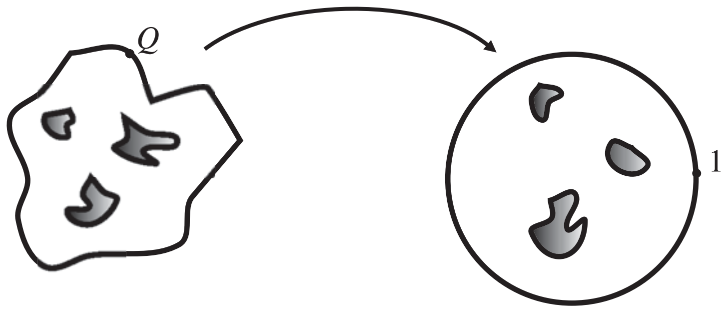

Fig. 12.2. Construction of a peaking function.

that converges to some limit holomorphic function ψ_0. Now the argument principle shows easily that this limit function ψ_0 is univalent. And the fact that all the $\psi_{j_k}(P)$ are trapped in L shows that the limit function maps Ω into Ω. Since similar reasoning applies to the inverse mappings, we may conclude that the limit mapping ψ_0 is in fact an automorphism. So we would see that $\mathrm{Aut}(\Omega)$ is compact. Since we assumed that in fact this is not the case, we may therefore conclude that the asserted points $P \in \Omega$ and $Q \in \partial\Omega$ and automorphisms ψ_j must exist.

(2) There is a peaking function at the point $Q \in \partial\Omega$. A "peaking function" is a function $\mu : \overline{\Omega} \to \overline{D}$ such that
- μ is continuous on $\overline{\Omega}$;
- μ is holomorphic on Ω;
- $\mu(Q) = 1$;
- $|\mu(\zeta)| < 1$ for all $\zeta \in \overline{\Omega} \setminus \{Q\}$.

Such a peaking function may be constructed by mapping Ω univalently, with a mapping ξ, into the unit disk so that the holes in Ω go to interior regions in the disk and the outer boundary of Ω goes to the boundary of the disk (see also our discussion in Section 4.5). Carathéodory's theorem (Section 5.1) tells us that the mapping extends continuously and univalently to the boundary. See Figure 12.2. Suppose that the boundary point $Q \in \partial\Omega$ gets mapped to $1 \in \partial D$. Now the function

$$f(\zeta) = \frac{\zeta + 1}{2}$$

is a peaking function for the point 1 in ∂D. Thus $\mu \equiv f \circ \xi$ is the peaking function that we seek for Ω at Q.

(3) If $K \subseteq \Omega$ is any compact subset and if ψ_j and P, Q are as in assertion (1) then there is a subsequence ψ_{j_k} such that $\psi_{j_k}(z) \to Q$ uniformly over $z \in K$. To see this, let ψ_j, P, Q be as in part **(1)**. Let μ be the peaking function as in part **(2)**. Consider the maps $g_j \equiv \mu \circ \psi_j$. Observe that each g_j is holomorphic and bounded by 1. By Montel's theorem, there is a subsequence g_{j_k} that converges uniformly on compact sets to a limit function g_0.

Of course $|g_0(\zeta)| \leq 1$ for all $\zeta \in \Omega$. But observe that

$$g_0(P) = \lim_{j\to\infty} g_j(P) = \lim_{j\to\infty} \mu(\psi_j(P)) = \mu(\lim_{j\to\infty} \psi_j(P)) = \mu(Q) = 1\,.$$

We have contradicted the maximum modulus principle (see [GRK1]) unless $g_0 \equiv 1$. But then this means that the ψ_{j_k} are converging, uniformly on compact sets, to the constant function with value Q. That is what has been claimed.

(4) There is a small, open disk $\mathcal{D}$ centered at Q such that $\Omega \cap \mathcal{D}$ is simply connected. In fact this is a straightforward application of the implicit function theorem (see [KRP]), and we leave the details to the reader. *Notice that this is the only step in the proof where we use the fact that the boundary curve is continuously differentiable.*

(5) The domain Ω is simply connected. Suppose not. Then there is a closed loop $\gamma : [0,1] \to \Omega$ that cannot be continuously deformed *in* Ω to a point. But of course the image curve of γ is a compact set C. Let $\mathcal{D}$ be the disk that we found in step **(4)**. By step **(3)**, there is a k so large that $\psi_{j_k}(C) \subseteq \mathcal{D} \cap \Omega$. But this means that $\psi_{j_k} \circ \gamma$ is a closed curve that lies entirely in the simply connected region $\mathcal{D} \cap \Omega$. Hence $\psi_{j_k} \circ \gamma$ can certainly be deformed to a point inside $\mathcal{D} \cap \Omega$. But of course the map ψ_j is a homeomorphism. Hence the image C of γ itself may deformed to a point in Ω. That is a contradiction.

(6) And now our proof is complete, for the simply connected domain Ω is conformally equivalent to the disk. This is what was claimed. □

We now have a fairly complete understanding of which domains in the complex plane have compact or noncompact automorphism group. In the next section we shall apply some of our new insights to the study of the dimension of the automorphism group.

We remark that the proof of Theorem 12.2.3 can be modified to give a new proof of the Riemann mapping theorem; thus one could, in principle, bypass the use of that theorem in proving Theorem 12.2.3.

12.3 The Dimension of the Automorphism Group

If a domain Ω has very many holes, then the automorphism group will be discrete (i.e., 0-dimensional). If the domain has just one hole, then the group will be 1-dimensional. The group can be (at least) 2-dimensional if and only if the domain is the disk or the entire plane or the punctured plane. The purpose of the present section is to flesh out these observations.

The Riemann mapping theorem says that a simply connected domain (no holes) that is not the entire plane is conformally equivalent to the unit disk. We will make use of a lovely theorem of classical function theory that says that a *finitely connected* domain (i.e., finitely many holes) is conformally equivalent

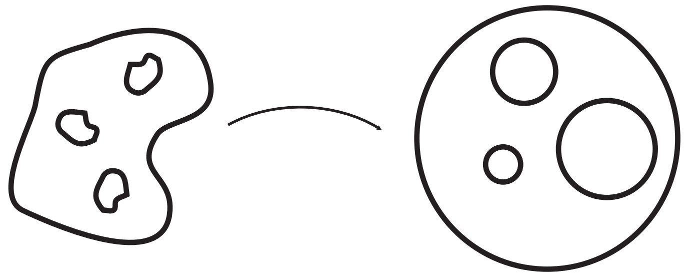

Fig. 12.3. A finitely connected domain is equivalent to the disk with finitely many smaller disks removed.

to the unit disk with finitely many smaller disks removed (Figure 12.3). The result is proved in detail in our Sections 4.2, 4.4. See also [AHL2].

In fact, He and Schramm [HES] have proved that if a domain has at most *countably many holes*, then it is conformally equivalent to the unit disk with at most countably many smaller disks removed.

If a domain (with C^1 boundary) has automorphism group with infinitely many elements, then $\mathrm{Aut}(\Omega)$ is either compact or noncompact. If it is noncompact, then Theorem 12.2.3 tells us that Ω is conformally equivalent to the disk. So we may as well assume that the automorphism group is compact. Thus the automorphism group, as a topological space, is compact; so it has an accumulation point. But then the group cannot be a 0-dimensional manifold (i.e., a discrete set). So it must be (at least) a 1-dimensional manifold.

Now the only compact 1-dimensional manifolds are the circle or a disjoint union of circles. For a domain with finitely many holes, the only way that multiple connected components of the automorphism group can occur is through inversion in one of the holes.[5]

Now assume, as above, that our domain is finitely connected. As previously noted, we may assume that it is a disk with finitely many smaller disks removed. We may use Schwarz reflection—with a little extra care—to extend any given automorphism λ across each boundary hole. We may repeat the reflection until the automorphism is extended entirely across the hole. Thus we now have an automorphism of the disk that fixes all the holes. If we ignore inversions, then the only automorphisms that we may consider are those that shuffle the holes. But if the automorphism group is 1-dimensional, then there is a 1-parameter subgroup φ_t of automorphisms. Fix a circle $\mathcal{C}$ about one of the holes—see Figure 12.4.

Then the mapping

$$t \mapsto \varphi_t(\mathcal{C})$$

[5] For example, in the simple case that the domain is the annulus $A = \{z \in \mathbb{C} : 1/2 < |z| < 2\}$, the automorphism group is the set of all rotations plus the inversion $z \mapsto 1/z$ plus compositions of the two. Topologically, this automorphism group is two circles. So the automorphism group has two connected components.

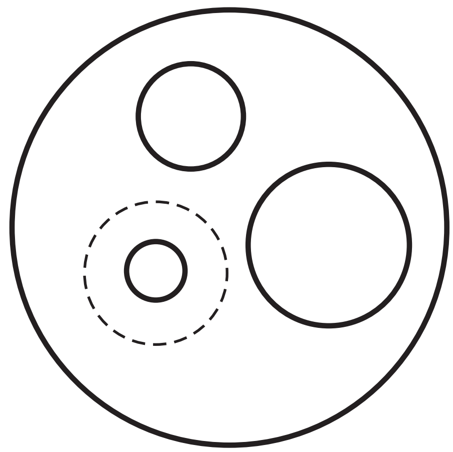

Fig. 12.4. A circle about one of the holes.

describes a continuously parametrized family of closed curves in Ω. It is impossible (by homotopy-theoretic considerations) for such a family to contain curves that encircle two different holes in the domain. So in fact our 1-parameter group of automorphisms must fix all the holes. But then there can be only one hole, because inversion in a hole would create new components of the complement, and the domain is thus (conformally equivalent to) an annulus.

Our heuristic arguments have shown that:

- The automorphism group of a domain with finitely many holes, but at least two, is finite.
- The only bounded domain with 1-dimensional automorphism group is the annulus.

Now which domains have 2-dimensional automorphism group? If Ω is bounded and $\dim(\mathrm{Aut}(\Omega)) = 2$, then the orbit of any point $P \in \Omega$ is two-dimensional.[6] It follows then that $\mathrm{Aut}(\Omega)$ is noncompact (assuming $\partial\Omega$ is C^1), and hence Ω is the disk. But we know the automorphism group of the disk explicitly: it is 3-dimensional. We conclude that there are no bounded domains with 2-dimensional automorphism group. Dimension 3 is the top dimension possible, and that occurs only when the domain is the disk.

What about unbounded domains? It is easy to see (refer to [GRK1]) that the automorphism group of the entire plane $\mathbb{C}$ is the set of all linear maps

$$z \mapsto az + b\,,$$

where $a \neq 0$ and b are arbitrary complex constants. Thus $\mathrm{Aut}(\mathbb{C})$ has dimension 4.

[6] Here the *orbit* of P is defined to be $\{\phi(P) : \phi \in \mathrm{Aut}(\Omega)\}$.

If instead $\Omega = \mathbb{C} \setminus \{0\}$—the punctured plane—then the automorphism group consists of rotations, dilations, and inversion ($z \mapsto 1/z$), and their compositions. Hence the group has dimension 2. In fact one can see that an automorphism of this Ω either takes the origin to the origin or takes the origin to infinity. In the first case, the point 0 is a "removable singularity" (by Riemann's theorem), and we can see that the mapping is the restriction to Ω of an automorphism of the plane. This makes it easy to determine the dimension of the group. Recall that we have already observed that there are no *bounded* domains with automorphism group of dimension 2.

12.4 The Iwasawa Decomposition

In Lie group theory there are important structure theorems that help us to understand the specific groups that we study. Sometimes in concrete situations these structure theorems take on special significance. We illustrate one of these results in the context of the automorphism group of the unit disk in the complex plane.

First let us formulate a simplified version of the abstract result:

Theorem 12.4.1. *Let G be a connected, semisimple Lie group. Then we may write*

$$G = K \cdot A \cdot N \,,$$

where K is a compact Lie group, A is an abelian Lie group, and N is a nilpotent Lie group.

We do not intend to prove the theorem here, but rather to discuss it. A good reference is [HEL]. Now the group we wish to study is $\mathrm{Aut}(D)$. For this group, it turns out that

$$K \text{ is the collection of automorphisms that fix the origin}\,.$$

The subgroups N and A are best understood by looking at the so-called *unbounded realization* of the disk.

That is, consider the mapping $\mathcal{C}$ given by

$$D \ni \zeta \mapsto i \cdot \frac{1-\zeta}{1+\zeta} \in U \,,$$

where U denotes the upper halfplane $U = \{\zeta \in \mathbb{C} : \mathrm{Im}\, \zeta > 0\}$. Then, as usual, we have the topological isomorphism

$$\mathrm{Aut}(D) \ni \psi \mapsto \varphi \circ \psi \circ \varphi^{-1} \in \mathrm{Aut}(U)\,.$$

Thus, if we can find the Iwasawa decomposition for $\mathrm{Aut}(U)$, then we will have found the decomposition for $\mathrm{Aut}(D)$.

You may calculate (exercise) that a linear fractional transformation

$$\varphi : \zeta \mapsto \frac{a\zeta + b}{c\zeta + d}$$

is an automorphism of the upper halfplane U if and only if (after normalization by multiplication by a constant in both the numerator and the denominator) a, b, c, d are real and $ad - bc > 0$. We wish to decompose

$$\varphi(\zeta) = \frac{a\zeta + b}{c\zeta + d} = \kappa \circ \alpha \circ \nu(\zeta)\,,$$

where $\nu \in N'$ has the form

$$\nu(\zeta) = \zeta + \mu$$

for some real μ; $\alpha \in A'$ has the form

$$\alpha(\zeta) = \lambda\zeta$$

for some positive λ; and $\kappa \in K'$ has the form

$$\kappa(\zeta) = \frac{\gamma\zeta + \delta}{-\delta\zeta + \gamma}$$

for some real γ, δ.

Notice that the elements of N' are left–right translations of U. These form an abelian, hence certainly nilpotent, group.[7] The elements of A' are dilations, thus forming an abelian group. Slightly less obvious is that the elements of K' fix the point $i \in U$, and hence form a compact group.

It is merely a tedious but elementary algebra problem to solve for $\mu, \lambda, \alpha, \beta$. They are uniquely determined, and we find that

$$\nu(\zeta) = \zeta + \left(\frac{dc + ba}{a^2 + c^2}\right)\,,$$

$$\alpha(\zeta) = \frac{a^2 + c^2}{ad - bc} \cdot \zeta\,,$$

and

$$\kappa(\zeta) = \frac{(a^2 d - abc)\zeta + (bc^2 - adc)}{(adc - bc^2)\zeta + (a^2 d - abc)}\,.$$

It is easy to verify that the positivity condition characterizing automorphisms of U is satisfied by the coefficients of this mapping κ.

We invite you to check that

$$\varphi(\zeta) = \kappa \circ \alpha \circ \nu(\zeta)\,.$$

[7] A group is said to be nilpotent if there is an integer k so that any commutator of order k in the group is the identity.

Now let us translate these three types of mappings back to the disk and see what the Iwasawa decomposition looks like in the original context in which the problem was posed above.

As we have already noted, the compact part of the group is easy. On the upper half space U, this piece K' is the subgroup of automorphisms that fix i. These are mappings of the form

$$\zeta \mapsto \frac{\gamma\zeta + \delta}{-\delta\zeta + \gamma}$$

with γ, δ real. The corresponding subgroup in the disk is the collection of all automorphisms that fix the origin. By the Schwarz lemma, these are just the rotations.

For the nilpotent and abelian pieces we must work a little harder.

Let $\nu' \in N'$. We must calculate $\mathcal{C}^{-1} \circ \nu' \circ \mathcal{C}$ to find the corresponding element of N. First of all,

$$\mathcal{C}^{-1}(\xi) = \frac{i - \xi}{i + \xi} .$$

A typical element of N' is $\xi \mapsto \nu'_\mu(\xi) = \xi + \mu$. Hence a corresponding typical element ν_μ of N will be

$$\begin{aligned}\nu_\mu(\zeta) &= \mathcal{C}^{-1} \circ \nu'_\mu \circ \mathcal{C}(\zeta)\\ &= \frac{i - \left(\frac{i(1-\zeta)}{1+\zeta} + \mu\right)}{i + \left(\frac{i(1-\zeta)}{1+\zeta} + \mu\right)}\\ &= -\left(\frac{\mu - 2i}{\mu + 2i}\right) \cdot \frac{\zeta - \frac{\mu}{-\mu+2i}}{1 - \overline{\left(\frac{\mu}{-\mu+2i}\right)}\zeta} .\end{aligned}$$

We can plainly see then that $\nu_\mu(\zeta)$ is the composition of a rotation and a Möbius transformation. Notice that $\lim_{\mu\to\pm\infty} \nu_\mu(\zeta) = -1$.

Now let $\alpha' \in A'$. We must calculate $\mathcal{C}^{-1} \circ \alpha' \circ \mathcal{C}$ to find the corresponding element of A. A typical element of A' is $\xi \mapsto \alpha'_\lambda(\xi) = \lambda\xi$ for $\lambda > 0$. Hence a corresponding typical element α_λ of A will be

$$\alpha_\lambda(\zeta) = \mathcal{C}^{-1} \circ \alpha'_\lambda \circ \mathcal{C}(\zeta) = \frac{i - \left(\lambda i \left(\frac{1-\zeta}{1+\zeta}\right)\right)}{i + \left(\lambda i \left(\frac{1-\zeta}{1+\zeta}\right)\right)} = \frac{\zeta - \frac{\lambda-1}{\lambda+1}}{1 - \zeta\left(\frac{\lambda-1}{\lambda+1}\right)} .$$

We see that $\alpha_\lambda(\zeta)$ is a Möbius transformation. Notice that $\lim_{\lambda\to 0} \alpha_\lambda(\zeta) = 1$ and $\lim_{\lambda\to+\infty} \alpha_\lambda(\zeta) = -1$.

Example 12.4.2. Let us consider whether $\mathrm{Aut}(D)$ has a subgroup that is group-theoretically isomorphic to the integers $\mathbb{Z}$. With the Iwasawa decomposition of the unbounded realization U this is obvious. For the translations

$$\nu_k'(\xi) = \xi + k\ , \quad k \in \mathbb{Z}\,,$$

form a subgroup with the required properties. Notice that this result is not entirely obvious if we work directly with the standard realization of elements of $\mathrm{Aut}(D)$ as the composition of rotations with Möbius transformations.

Likewise, we can see that $\mathrm{Aut}(D)$ has a subgroup that is isomorphic to the multiplicative group $\mathbb{R}^+$ of positive real numbers. This subgroup, looking at the Iwasawa decomposition of the automorphism group of the unbounded realization U, is just A itself, that is, the dilations

$$\alpha_\lambda'(\xi) = \lambda \cdot \xi\,.$$

The Iwasawa decomposition can be useful for understanding the orbits of points under the automorphism group action. Fix a point $P \in D$. Under the action of K (the automorphisms that fix 0), the point P is sent to all points of a circle centered at the origin. Under the action of N, the point P is sent to a curve that is the boundary of a horocycle (see [HIL]). Under the action of A, the point P is sent to a geodesic in the Poincaré–Bergman metric. We invite the reader to draw pictures (and to do the attendant calculations) to illustrate these points.

12.5 General Properties of Holomorphic Maps

Now we turn to a result of Lipman Bers that has a different kind of flavor. It characterizes domains in terms of a certain algebraic invariant.

Let $\Omega \subseteq \mathbb{C}$ be a domain. Let $\mathcal{O}(\Omega)$ denote the algebra of holomorphic functions from Ω to $\mathbb{C}$. Bers's theorem says, in effect, that the algebraic structure of $\mathcal{O}(\Omega)$ characterizes Ω. We begin our study by introducing a little terminology.

Definition 12.5.1. Let $\Omega \subseteq \mathbb{C}$ be a domain. A $\mathbb{C}$-algebra homomorphism $\varphi : \mathcal{O}(\Omega) \to \mathbb{C}$ is called a *character* of $\mathcal{O}(\Omega)$. If $c \in \mathbb{C}$, then the mapping

$$\begin{aligned} e_c : \mathcal{O}(\Omega) &\to \mathbb{C}\,, \\ f &\mapsto f(c)\,, \end{aligned}$$

is called a *point evaluation.* Every point evaluation is a character.

It should be noted that if $\varphi : \mathcal{O}(\Omega) \to \mathcal{O}(\widehat{\Omega})$ is not the trivial zero homomorphism, then $\varphi(1) = 1$. This follows because $\varphi(1) = \varphi(1 \cdot 1) = \varphi(1) \cdot \varphi(1)$. On any open set where the holomorphic function $\varphi(1)$ does not vanish, we find that $\varphi(1) \equiv 1$. The result follows by analytic continuation.

It turns out that every character of $\mathcal{O}(\Omega)$ is a point evaluation. This is the content of the next lemma.

Lemma 12.5.2. *Let φ be a character on $\mathcal{O}(\Omega)$. Then $\varphi = e_c$ for some $c \in \Omega$. Indeed, $c = \varphi(id) \in \Omega$. Here id is defined by $id(z) = z$.*

Proof. Let c be defined as in the statement of the lemma. Let $f(z) = z - c$. Then

$$\varphi(f) = \varphi(\mathrm{id}) - \varphi(c) = c - c = 0\,.$$

If it were not the case that $c \in \Omega$ then the function f would be a unit in $\mathcal{O}(\Omega)$. But then

$$1 = \varphi(f \cdot f^{-1}) = \varphi(f) \cdot \varphi(f^{-1}) = 0\,.$$

That is a contradiction. So $c \in \Omega$.

Now let $g \in \mathcal{O}(\Omega)$ be arbitrary. Then we may write

$$g(z) = g(c) + f(z) \cdot \widetilde{g}(z)\,,$$

where $\widetilde{g} \in \mathcal{O}(\Omega)$. Thus

$$\varphi(g) = \varphi(g(c)) + \varphi(f) \cdot \varphi(\widetilde{g}) = g(c) + 0 = g(c) = e_c(g)\,.$$

We conclude that $\varphi = e_c$, as was claimed. □

Now we may prove Bers's theorem.

Theorem 12.5.3. *Let Ω, $\widetilde{\Omega}$ be domains. Suppose that*

$$\varphi : \mathcal{O}(\Omega) \to \mathcal{O}(\widetilde{\Omega})$$

is a $\mathbb{C}$-algebra homomorphism. Then there exists one and only one holomorphic mapping $h : \widetilde{\Omega} \to \Omega$ such that

$$\varphi(f) = f \circ h \quad \textit{for all } f \in \mathcal{O}(\Omega)\,.$$

In fact, the mapping h is given by $h = \varphi(id_\Omega)$.

The homomorphism φ is bijective if and only if h is conformal, that is, a one-to-one and onto holomorphic mapping from $\widetilde{\Omega}$ to Ω.

Proof. Since we want the mapping h to satisfy $\varphi(f) = f \circ h$ for all $f \in \mathcal{O}(\Omega)$, it must in particular satisfy $\varphi(\mathrm{id}_\Omega) = \mathrm{id}_\Omega \circ h = h$. We take this as our definition of the mapping h.

If $a \in \widetilde{\Omega}$, then $e_a \circ \varphi$ is a character of $\mathcal{O}(\Omega)$. Thus our lemma tells us that $e_a \circ \varphi$ must in fact be a point evaluation on Ω. As a result,

$$e_a \circ \varphi = e_c\,, \quad \text{with } c = (e_a \circ \varphi)(\mathrm{id}_\Omega) = e_a(h) = h(a)\,.$$

Thus, if $f \in \mathcal{O}(\Omega)$, then

$$\varphi(f)(a) = e_a(\varphi \circ f) = (e_a \circ \varphi)(f) = e_{h(a)}(f) = f(h(a)) = (f \circ h)(a)$$

for all $a \in \widetilde{\Omega}$. We conclude that $\varphi(f) = f \circ h$ for all $f \in \mathcal{O}(\Omega)$.

For the last statement of the theorem, suppose that h is a one-to-one, onto conformal mapping of $\widetilde{\Omega}$ to Ω. If $g \in \mathcal{O}(\Omega)$, then set $f = g \circ h^{-1}$. It follows that $\varphi(f) = f \circ h = g$. Hence φ is onto. Likewise, if $\varphi(f_1) = \varphi(f_2)$, then $f_1 \circ h = f_2 \circ h$ hence, composing with h^{-1}, $f_1 \equiv f_2$. So φ is one-to-one. Conversely, suppose that φ is an isomorphism. Let $a \in \Omega$ be arbitrary. Then e_a is a character on $\mathcal{O}(\Omega)$; hence $e_a \circ \varphi^{-1}$ is a character on $\mathcal{O}(\widetilde{\Omega})$. By the lemma, there is a point $c \in \widetilde{\Omega}$ such that $e_a \circ \varphi^{-1} = e_c$. It follows that

$$e_a = e_c \circ \varphi .$$

Applying both sides to id_Ω yields

$$e_a(\mathrm{id}_\Omega) = (e_c \circ \varphi)(\mathrm{id}_\Omega) .$$

Unraveling the definitions gives

$$a = e_c(\mathrm{id}_\Omega \circ h) = h(c) .$$

Thus $h(c) = a$ and h is surjective. The argument in fact shows that the pre-image c is uniquely determined. So h is also one-to-one. □

As an application of Bers's theorem, we can see immediately that the disk $D = \{z \in \mathbb{C} : |z| < 1\}$ and the annulus $\mathcal{A} = \{z \in \mathbb{C} : 1/2 < |z| < 2\}$ are not conformally equivalent. For the algebra $\mathcal{O}(D)$ can be generated by 1 and z. But the algebra $\mathcal{O}(\mathcal{A})$ cannot be generated by 1 and just one other function (because natural generators for $\mathcal{O}(\mathcal{A})$ are $\{1, z, 1/z\}$ and it is impossible to come up with a shorter list). We leave the details of these assertions to the reader.

Recall from Section 4.5 the concept of essential boundary point. We say that a domain is *maximal* if each boundary point is essential. Given a domain Ω, there is a unique maximal $\Omega^* \supset \Omega$ (see [FIS, p. 66]). Now it is possible to prove the following result:

Theorem: *Let Ω_1, Ω_2 be two planar domains. Let $\varphi : \Omega_2^* \to \Omega_1^*$ be a one-to-one, onto holomorphic mapping. Then the mapping*

$$\begin{aligned} \Phi : H^\infty(\Omega_1) &\to H^\infty(\Omega_2) , \\ f &\mapsto f \circ \varphi \end{aligned} \tag{$*$}$$

defines a real algebra isomorphism of $H^\infty(\Omega_1)$ onto $H^\infty(\Omega_2)$ with $\Phi(i) = i$ (here i denotes the H^∞ function that is identically equal to i).

Conversely, if Φ is a real algebraic isomorphism of $H^\infty(\Omega_1)$ onto $H^\infty(\Omega_2)$ with $\Phi(i) = i$, then there is a one-to-one, onto mapping $\varphi : \Omega_2^* \to \Omega_1^*$ such that $(*)$ holds.

The proof of this result is not difficult (see [FIS, p. 67]). But we shall omit the details and move on to new directions. We now turn to some results about limits of automorphisms. The first was proved by H. Cartan in the 1930s. We begin with some preliminary remarks.

First, if f_j are holomorphic maps from Ω to Ω and if f_j converges uniformly on compact sets to a limit function f, and finally if f is not constant, then f is a holomorphic mapping from Ω to Ω. This assertion is immediate from the open mapping principle.

Second, if f_j are holomorphic mappings from Ω to Ω, if g_j are holomorphic functions on Ω, and if

(a) $\lim_{j\to\infty} f_j \equiv f$ is a holomorphic mapping from Ω to Ω,
(b) $\lim_{j\to\infty} g_j \equiv g$ is a holomorphic function on Ω,

then $\lim_{j\to\infty} g_j \circ f_j = g \circ f$. To see this, let $w_j \in \Omega$ be a sequence with limit $w \in \Omega$. Uniform convergence on compact sets gives us that $\lim_{j\to\infty} f_j(w_j) = f(w)$. By the same token, $\lim_{j\to\infty} g_j(f_j(w_j)) = g(f(w))$. We conclude that the sequence $g_j \circ f_j$ converges uniformly on compact sets to $g \circ f$.

Now we turn to Cartan's result.

Proposition 12.5.4. *Let Ω be a bounded domain and $\varphi_j \in \mathrm{Aut}(\Omega)$. Assume that the φ_j converge uniformly on compact sets to some function f. Then f is either itself an automorphism, or else f is constantly equal to some boundary point of Ω.*

Proof. Of course the limit function f will be holomorphic. Now we examine the sequence $g_j \equiv f_j^{-1}$. These form a bounded sequence of holomorphic functions on Ω. By Montel's theorem, there is a subsequence g_{j_k} that converges uniformly on compact subsets of Ω to some element $g \in \mathcal{O}(\Omega)$. We assert that

$$g'(f(w)) \cdot f'(w) \equiv 1 \quad \text{for all } w \in \Omega \text{ with } f(w) \in \Omega\,. \tag{12.5.1}$$

To see the assertion, observe that $g_j \circ f_j = \mathrm{id}$ and hence $g_j'(f_j(z)) \cdot f_j'(z) = 1$ for all j and all $z \in \Omega$. Thus we need only check that

$$\lim_{j\to\infty} g_j'(f_j(w)) = g'(f(w))$$

for all $w \in \Omega \cap f^{-1}(\Omega)$. This assertion holds, however, because $\lim_{j\to\infty} f_j(w) = f(w)$ and the sequence g_j' converges uniformly on compact sets to g'.

With our claim proved, we now note that, if f is not constant, then f is certainly a holomorphic map from Ω to Ω (by the first remark preceding the enunciation of our proposition). Also g is not constant by (12.5.1), and again our first remark implies that g is a holomorphic map from Ω to Ω. Since $g_j \circ f_j = \mathrm{id} = f_j \circ g_j$, our second remark now implies that

$$g \circ f = \mathrm{id} = f \circ g\,,$$

hence $f \in \mathrm{Aut}(\Omega)$.

But if f is constant then (12.5.1) tells us that $c \equiv f(\Omega)$ cannot be a point in Ω. It follows that $c \in \partial\Omega$. □

It is easy to see that, in the second conclusion of the proposition, the limit of the sequence of automorphisms cannot be an *isolated boundary point* of Ω. This follows because Ω is connected, and because of uniform convergence on compact sets.

A simple but good example to keep in mind for illustrating the proposition is $\Omega = D = \{z \in \mathbb{C} : |z| < 1\}$ and

$$f_j(z) = \frac{z - (1 - 1/j)}{1 - (1 - 1/j)z} .$$

Of course each f_j is an automorphism of the disk, but $\lim_{j\to\infty} f_j(z) \equiv -1 \in \partial D$.

The theory becomes particularly rich if we examine not just any sequence of automorphisms but rather the sequence of iterates of a single mapping or automorphism (we have already glimpsed this additional structure in our proof of Cartan's Proposition 12.1.1). In what follows, if $f : \Omega \to \Omega$ is a holomorphic mapping then we shall use the notation

$$f^j(z) = \underbrace{f \circ f \circ \cdots \circ f}_{j \text{ times}} \quad \text{for } j = 1, 2, \ldots .$$

Proposition 12.5.5. *Let f be a holomorphic mapping of the domain Ω to itself. Assume that some subsequence f^{j_k} converges uniformly on compact sets to a function $g \in \mathcal{O}(\Omega)$. We conclude that*

(a) *If $g \in \mathrm{Aut}(\Omega)$ then $f \in \mathrm{Aut}(\Omega)$.*
(b) *If g is not constant, then every convergent subsequence of the sequence $h_k \equiv f^{j_{k+1}-j_k}$ has the limit function id_Ω.*

Proof. First consider part **(a)**. If it were the case that $f(a) = f(b)$, some $a, b \in \Omega$, then it would follow that $f^j(a) = f^j(b)$ for all j and hence $g(a) = g(b)$. We must conclude that $a = b$ since $g \in \mathrm{Aut}(\Omega)$.

For surjectivity, we know that $f^{j_k}(\Omega) \subseteq f(\Omega)$ for every choice of index. It is immediate that $g(\Omega) \subseteq f(\Omega) \subseteq \Omega$, just by set theory. But we must have that $g(\Omega) = \Omega$ because g is an automorphism. Therefore $f(\Omega) = \Omega$. Assertion **(a)** is now proved.

We turn to **(b)**. By a previous remark, we certainly know that g is a holomorphic mapping from Ω to Ω. Let h be some subsequential limit of the h_k. Then, by an earlier remark, $f^{j_{k+1}} = h_k \circ f^{j_k}$ implies that $g = h \circ g$. Therefore h is the identity on $g(\Omega)$. But $g(\Omega)$ is open in Ω, so we know that $h \equiv \mathrm{id}_\Omega$. That completes the proof. □

One interesting consequence of this last proposition is that, if the iterates f^j of f converge uniformly on compact sets to a nonconstant function, then the limit function is the identity.

Now we have an important result of Cartan:

Theorem 12.5.6 (Cartan). *Let Ω be a bounded domain and let f be a holomorphic mapping of Ω to Ω. Suppose some subsequence f^{j_k} of the iterates of f converges uniformly on compact sets to a nonconstant function g. Then $f \in \mathrm{Aut}(\Omega)$.*

Proof. Montel's theorem tells us that $h_k \equiv f^{j_{k+1}-j_k}$ has a subsequence that converges uniformly on compact sets. By part **(b)** of the last proposition, we know that the limit of this subsequence is id_Ω. By part **(a)** of that proposition, $f \in \mathrm{Aut}(\Omega)$. □

The next result is philosophically related to the three-fixed-point theorem (Theorem 1.4.3) that we proved earlier.

Corollary 12.5.7. *Let Ω be a bounded domain and suppose that $f : \Omega \to \Omega$ is a holomorphic mapping with two distinct fixed points. Then f is an automorphism of Ω.*

Proof. Call the fixed points a and b with $a \neq b$. We assume, of course, that $a \neq b$. Montel's theorem tells us that there is a subsequence $\{f^{j_k}\}$ of the iterates that converges in Ω to a holomorphic function g. Certainly $g(a) = a$ and $g(b) = b$. Hence g is not constant. By the theorem, $f \in \mathrm{Aut}(\Omega)$. □

The next corollary is in part a reiteration of the theorem of Cartan with which we began this section. Compare it to the classical Schwarz lemma.

Corollary 12.5.8. *Let Ω be a bounded domain and $a \in \Omega$. Then $|f'(a)| \leq 1$ for all holomorphic mappings $f : \Omega \to \Omega$ that fix a. Let $\mathrm{Aut}_a(\Omega)$ consist of those automorphisms that fix the point a. Then we have*

$$\begin{aligned}\mathrm{Aut}_a(\Omega) = \{f : f \text{ is a holomorphic mapping from } \Omega \text{ to } \Omega\ , \\ f(a) = a\ ,\ |f'(a)| = 1\}\,.\end{aligned}$$

Proof. As usual, by Montel, there is a subsequence f^{j_k} of iterates that converges, uniformly on compact sets, to some holomorphic g. One may calculate that

$$\lim_{k\to\infty} [f^{j_k}]'(a) = \lim_{k\to\infty} [f'(a)]^{j_k} = g'(a)\,.$$

But this can be true only if $|f'(a)| \leq 1$ (otherwise the powers of the derivative would blow up).

In case $|f'(a)| = 1$, we see that $|g'(a)| = 1$. But then g is not constant, and Cartan's theorem tells us that $f \in \mathrm{Aut}(\Omega)$.

Conversely, suppose that $f \in \mathrm{Aut}(\Omega)$ and f fixes a. Then $f^{-1} \in \mathrm{Aut}_a(\Omega)$. We know already that $|f'(a)| \leq 1$. Likewise, $|1/f'(a)| = |(f^{-1})'(a)| \leq 1$, or $|f'(a)| \geq 1$. We conclude that $|f'(a)| = 1$. □

We saw some of the interest of $\mathrm{Aut}_a(\Omega)$ in the last corollary. Note that $\mathrm{Aut}_a(\Omega)$ is itself a group, known as the *isotropy subgroup of* a. By Montel's theorem, this subgroup is compact in the topology of uniform convergence on compact sets. Because an automorphism is an isometry in the Bergman metric of Ω, we find it useful to consider the mapping

$$\mathrm{Aut}_a(\Omega) \ni \varphi \mapsto \varphi'(a) \,. \tag{12.5.2}$$

Applying reasoning as in the last proof, or simply invoking the last corollary, we see that $|\varphi'(a)| = 1$. This mapping is also one-to-one (by reasoning that we have used before), just because an isometry is uniquely determined by where it takes a point (in this case a) and the derivative at that point.

Thus the mapping (12.5.2) is an injection of the compact group $\mathrm{Aut}_a(\Omega)$ into the unit circle in the complex plane. The image therefore must be a compact subgroup of the circle group. It is easy to see therefore that the image is either the entire circle or else a finite subgroup generated by some element of the form $e^{2\pi i/k}$, $k = 1, 2, \dots$.

An interesting corollary of the reasoning just presented—an interpretation really—is contained in the following result.

Proposition 12.5.9. *Let Ω be a bounded domain. Fix a point $a \in \Omega$. Let $f \in \mathrm{Aut}_a(\Omega)$. If $f'(a) > 0$, then $f = id_\Omega$.*

Proof. We know that $f'(a)$ lies on the circle. If $f'(a) > 0$, then $f'(a) = 1$. It follows now from Cartan's theorem that $f = \mathrm{id}_\Omega$. □

Now we bring the argument principle into play to study how the presence of topology adds rigidity to the automorphism group structure.

Proposition 12.5.10. *Let Ω be a domain. Let f_j be a sequence of functions holomorphic on Ω, and suppose that they converge uniformly on compact sets to a limit function f. Suppose that there is a closed path γ in Ω such that the intersection of the regions interior to $f_j \circ \gamma$ contains at least two points. Then f is nonconstant.*

Proof. Suppose not. If $f(z) \equiv a$, then there would be a point b in the region interior to the curve $f_j \circ \gamma$, $b \neq a$, such that

$$\frac{1}{2\pi i} \oint_\gamma \frac{f_j'}{f_j - b} \, d\zeta = \frac{1}{2\pi i} \oint_{f_j \circ \gamma} \frac{d\eta}{\eta - b}$$

is a nonzero integer for all but finitely many j. But the sequence $f_j'/(f_j - b)$ converges uniformly to 0 on γ. That is a contradiction. □

Now our main result about multiply connected domains is this:

Theorem 12.5.11. *Let Ω be a bounded, finitely connected domain in $\mathbb{C}$ with connectivity at least 2. Assume that each connected component of the complement of Ω has at least two points. Let f be a holomorphic self-map of Ω. Assume that every closed path γ in Ω that is not homologous to 0 in Ω has image under f that is also not homologous to 0 in Ω. Then $f \in \mathrm{Aut}(\Omega)$.*

Proof. As usual, we apply Montel's theorem to the sequence of iterates f^j. We find thereby a subsequence f^{j_k} that converges to some limit g. Because Ω has connectivity at least 2 (i.e., the complement has at least two connected components), there is a closed path γ in Ω that is not homologous to 0. Since $f^j \circ \gamma = f \circ (f^{j-1} \circ \gamma)$, we see that no path $f^{j_k} \circ \gamma$ can be homologous to 0 in Ω. Thus, for each k, there is a connected component K_k of the complement of Ω such that K_k lies in the interior region of the closed curve $f^{j_k} \circ \gamma$.

Since Ω has just finitely many holes (i.e., bounded connected components of the complement), we may suppose (by passing to a subsequence if necessary and renumbering the holes) that K_k lies in the region interior to $f^{j_k} \circ \gamma$ for each k. Since K_k has at least two points, we find from our preceding proposition that g is not constant. Now Theorem 12.5.6 tells us that f is an automorphism of Ω. □

In fact it is known (see [HEI1], [HEI2]) that a domain of finite connectivity at least three will have *finite* automorphism group; that is to say, the domain has only finitely many conformal self-maps (see our earlier discussion in Section 12.3). Heins even gave sharp upper bounds for the size of the automorphism group. Let Ω_k have k holes (i.e., Ω_k has connectivity $k+1$). Then the sharp upper bound $N(k)$ for the number of elements in $\text{Aut}(\Omega_k)$, $k \geq 2$, is

$$\begin{aligned} N(4) &= 12\,, \\ N(6) &= 24\,, \\ N(8) &= 24\,, \\ N(12) &= 60\,, \\ N(20) &= 60\,, \\ N(k) &= 2k \text{ for } k \neq 4, 6, 8, 12, 20\,. \end{aligned}$$

A very interesting open problem is to determine which finite groups arise as the automorphism groups of planar domains (there are some results for finitely connected regions). It is known that if G is a compact Lie group, then there is some smoothly bounded domain in some $\mathbb{C}^n$ with automorphism group equal to G. But it is difficult to say how large n must be in terms of elementary properties of the group G. See [BED], [SAZ] for details.

Problems for Study and Exploration

1. Calculate the automorphism group of a bounded domain with two holes. [*Hint:* Such a domain is conformally equivalent—see the results in Chapter 4— to an annulus with one more disjoint disk removed.]

2. Let

$$\Omega = \{\zeta \in \mathbb{C} : -1 < \operatorname{Im}\zeta < 1\} .$$

Give an explicit description of the automorphism group of Ω.

3. Let $\Omega_1 \subseteq \Omega_2 \subseteq \cdots$ be domains. Define $\Omega = \cup_j \Omega_j$. What can you say about the relationship of $\operatorname{Aut}(\Omega)$ with $\operatorname{Aut}(\Omega_j)$?

4. Give an example of a Riemann surface without boundary that has compact automorphism group.

5. Give an example of a Riemann surface—that is *not* a planar domain—that has noncompact automorphism group.

6. Calculate the automorphism group of the Riemann sphere.

7. What is the dimension of the automorphism group of the torus?

8. Let $\mathcal{A} = \{\zeta \in \mathbb{C} : 1/2 < |\zeta| < 2\}$. Calculate the isotropy group of the point i.

9. Refer to Exercise 2. Calculate the isotropy group of the origin in that domain.

10. Let Ω be the domain consisting of the box $\{\zeta \in \mathbb{C} : |\operatorname{Re}\zeta| < 2, |\operatorname{Im}\zeta| < 2\}$ with the four disks $\overline{D}(1+i, 0.1)$, $\overline{D}(1-i, 0.1)$, $\overline{D}(-1-i, 0.1)$, $\overline{D}(-1+i, 0.1)$ removed. Calculate the automorphism group of Ω. Now perturb one of the holes by 0.1. Show that the resulting domain has automorphism group consisting only of the identity.

13

Cousin Problems, Cohomology, and Sheaves

Genesis and Development

One of the most important things that we do in complex function theory is to construct holomorphic functions with specified properties. Given the way that we are prone to think, a natural way to effect this process is to perform some local construction and then to endeavor to extend the result to an entire domain Ω. The function theory of one complex variable is replete with methods for performing that "extension" process. Infinite products, analytic continuation, division problems, approximation theorems (Runge, Mergelyan), and the Cauchy–Riemann equations are just some of the devices that we have for taking a local construction and making it global.

But in fact there are many rather powerful and far-reaching techniques—which have become quite important in algebraic geometry and several complex variables—that can be used to good effect in one complex variable as well. Some of these devices are in fact not very well known in the context of one complex variable. In the present chapter we shall first introduce the Cousin problems—which are a poor man's version of sheaf cohomology—and show how they can be used to solve some very concrete and substantive problems that arise in classical function theory. We invite the reader, as an exercise, to study the more classical proofs of the results that we shall present.

We shall also include a brief introduction to sheaves and sheaf cohomology. Sheaves are a natural language for the study of Riemann surfaces and other analytic objects—both in one and several complex variables (see [GUN2], where the Riemann–Roch theorem for Riemann surfaces is formulated rather naturally in this language). They arise also in many other parts of mathematical analysis. Our treatment will give the reader a taste of the algebraic point of view in complex analysis.

13.1 The Cousin Problems

We shall motivate the form of the Cousin problems by first looking at a little combinatorial topology.

Example 13.1.1. Let

$$\Omega^0 = \{z \in \mathbb{C} : 1 < |z| \ < 2\}\,,$$
$$\Omega^1 = \{z \in \mathbb{C} : |z| < 2\}.$$

Consider the open coverings $\mathcal{U} = \{U_j\}$ of Ω^0 and $\mathcal{V} = \{V_j\}$ of Ω^1 that are illustrated in Figure 13.1.

We shall play the following game with these coverings: To each nonempty intersection $U_j \cap U_k$ we assign an integer g_{jk}, subject to the rules

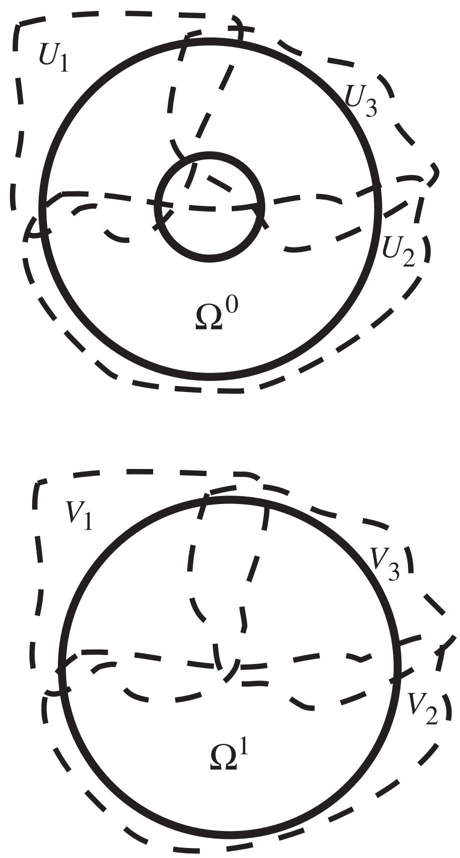

Fig. 13.1. Open coverings of Ω.

$$g_{jk} = -g_{kj} \,, \tag{13.1.1}$$

$$g_{ij} + g_{jk} + g_{ki} = 0 \quad \text{if} \quad U_i \cap U_j \cap U_k \neq \emptyset \,. \tag{13.1.2}$$

[We similarly assign integers h_{jk} to each nonempty intersection $V_j \cap V_k$.]

The question that we wish to consider is, "Will it always be possible to choose integers $g_i, i = 1, 2, 3, \ldots,$ such that $g_{jk} = g_k - g_j$ for all j, k [respectively, will it always be possible to choose integers $h_i, i = 1, 2, 3, \ldots,$ such that $h_{jk} = h_k - h_j$]? As motivation for this question, the reader should consider the problem of choosing a branch for $\log z$, z in the complex plane less the origin.

On the annulus, with cover $\{U_i\}$, the answer to our question is decidedly "no." For a counterexample, let $g_{12} = 1, g_{23} = 10$, and $g_{31} = 100$. Define the other g_{jk} according to (13.1.1). Observe that (13.1.2) is vacuous. If g_1, g_2, g_3 are chosen so that $g_2 - g_1 = 1$ and $g_3 - g_2 = 10$, then it must be that $g_3 - g_1 = 11$, and that is inconsistent with $g_{13} = -100$.

By contrast, our problem can always be solved on the disk Ω^1 with cover $\{V_i\}$ because the (now nonvacuous) condition (13.1.2) guarantees that the number g_{13} will be compatible with the conditions forced by $g_{12} = g_2 - g_1$ and $g_{23} = g_3 - g_2$. The contradiction that occurred on the annulus cannot arise.

The game we have been describing is an algebraic/combinatorial device for detecting the fact that the annulus has a hole and the disk does not. In some sense, the number of unsolvable problems of the form (13.1.1), (13.1.2) counts the number of holes in the region. The subject of algebraic topology (especially singular homology and cohomology theory) has its roots in considerations such as these (see [LEF1], [LEF2]).

We now set up a conceptual framework that will make an argument like this last one rather natural. Our first step in this program is to consider the Cousin problems. We enunciate them as follows:

FIRST COUSIN PROBLEM (Cousin I) Let $\Omega \subseteq \mathbb{C}$ be a domain. Let $\{U_i\}$ be an open covering of Ω. Suppose that for each U_j, U_k with nonempty intersection there is a holomorphic $g_{jk} : U_j \cap U_k \to \mathbb{C}$ satisfying

(a) $g_{jk} = -g_{kj}$;
(b) $g_{jk} + g_{k\ell} + g_{\ell j} = 0 \qquad \text{on} \qquad U_j \cap U_k \cap U_\ell$.

Problem: Find holomorphic functions g_j on U_j such that $g_{jk} = g_k - g_j$ on $U_j \cap U_k$ whenever this intersection is nonempty.

We also shall consider a multiplicative version of this problem:

SECOND COUSIN PROBLEM (Cousin II) Let $\Omega \subseteq \mathbb{C}$ be a domain. Let $\{U_i\}$ be an open covering of Ω. Suppose that for each U_j, U_k with nonempty intersection there is a *nonvanishing* holomorphic g_{jk} on $U_j \cap U_k$ such that

(a) $g_{jk} \cdot g_{kj} = 1$;
(b) $g_{jk} \cdot g_{k\ell} \cdot g_{\ell i} = 1 \qquad \text{on} \qquad U_j \cap U_k \cap U_\ell$.

Problem: Find nonvanishing holomorphic functions g_i on U_i such that $g_{jk} = g_k/g_j$ on $U_j \cap U_k$ whenever $U_j \cap U_k \neq \emptyset$.

Proposition 13.1.2. *The analogue of the First Cousin Problem for C^∞ (rather than holomorphic) functions always has a solution.*

Proof. Let $\{\phi_i\}$ be a C^∞ partition of unity on Ω subordinate to the covering $\{U_i\}$. So each ϕ_i is a C^∞ function supported in U_i and $\sum_i \phi_i = 1$ on Ω. See [MUN] for more on partitions of unity.

For each i, define

$$g_i(z) = \sum_\ell \phi_\ell(z) g_{\ell i}(z) , \qquad z \in U_i.$$

Notice that $g_i \in C^\infty(U_i)$. Then, on $U_j \cap U_k$, we have

$$\begin{aligned} g_k(z) - g_j(z) &= \sum_\ell \phi_\ell(z)\{g_{\ell k}(z) - g_{\ell j}(z)\} \\ &= \sum_\ell \phi_\ell(z) g_{jk}(z) \\ &= g_{jk}(z). \qquad \square \end{aligned}$$

Theorem 13.1.3. *If $\Omega \subseteq \mathbb{C}$ is a domain, then (holomorphic) Cousin I can always be solved on Ω.*

Proof. Let $\{\phi_i\}$ be a partition of unity subordinate to $\{U_i\}$. Define

$$h_i = \sum_\ell \phi_\ell g_{\ell i} \qquad \text{on} \quad U_i.$$

Of course, h_i may not be holomorphic. However, on $U_i \cap U_j$ we have

$$\frac{\partial}{\partial \bar{z}}(h_k - h_j) = \frac{\partial}{\partial \bar{z}} \sum_\ell \phi_\ell (g_{\ell k} - g_{\ell j}) = \frac{\partial}{\partial \bar{z}} \sum_\ell \phi_\ell (g_{jk}) = \frac{\partial}{\partial \bar{z}} g_{jk} = 0\,.$$

Therefore the form

$$f \equiv \frac{\partial}{\partial \bar{z}} h_i \qquad \text{on} \qquad U_i$$

is well defined and C^∞ on Ω. By Corollary 7.3.6, there is a $u \in C^\infty(\Omega)$ such that $\frac{\partial}{\partial \bar{z}} u = f$. Set

$$g_i = h_i - u \qquad \text{on} \quad U_i .$$

Then, on $U_j \cap U_k$, we have

$$g_k - g_j = h_k - h_j = g_{jk} .$$

Also, for each i, $\frac{\partial}{\partial \bar{z}} g_i = \frac{\partial}{\partial \bar{z}} h_i - \frac{\partial}{\partial \bar{z}} u = 0$ on U_i. Hence each g_i is holomorphic on U_i and we are done. □

Next we turn our attention to Cousin II.

Lemma 13.1.4. *Let $\Omega \subseteq \mathbb{R}^2$ be a simply connected domain. Let $f : \Omega \to \mathbb{C}$ be continuous and nonvanishing. Then there is a continuous g on Ω such that* $\exp g = f$. *If f is C^k, any $k \geq 0$, then g is C^k.*

Proof. By continuity, each $P \in \Omega$ has a neighborhood U_P on which f has a continuous logarithm. Fix $P_0 \in \Omega$ and let $\gamma : [0, 1] \to \Omega$ be a continuous Jordan curve with $\gamma(0) = P_0, \gamma(1) = P_0$. By a continuity argument, $f \circ \gamma(t)$ has a continuous logarithm for $0 \leq t < 1$ (that is, the set

$$S = \{s \in [0, 1) : \log(f \circ \gamma(t))$$
$$\text{is well defined and continuous for } 0 \leq t \leq s\}$$

is open, closed, and nonempty). Seeking a contradiction, we suppose that $\lim_{t \to 1^-} \log f \circ \gamma(t) \neq \log f \circ \gamma(0)$. Let $u(s, t)$ be a fixed-point homotopy of the curve γ with the constant curve at P_0. Thus

- u is continuous on $[0, 1] \times [0, 1]$;
- $u(0, t) = \gamma(t)$, all $t \in [0, 1]$;
- $u(1, t) \equiv P_0$, all $t \in [0, 1]$;
- $u(s, 0) = u(s, 1) = P_0$, all $s \in [0, 1]$.

The function

$$\rho(s) = \frac{1}{2\pi i} \left\{ \lim_{t \to 1^-} \log f(u(s, t)) - \log f(u(s, 0)) \right\}$$

is then a continuous, integer-valued function of s that satisfies $\rho(1) = 0$ and $\rho(0) \neq 0$. This is a contradiction. Thus in fact $\lim_{t \to 1^-} \log f \circ \gamma(t) = \log f \circ \gamma(0)$.

Now a standard argument shows that the set of all $P \in \Omega$ to which $\log f$ can be unambiguously propagated from the initial point P_0 is both open and closed. That completes the proof in the continuous, or C^0, case.

The C^k result follows by applying implicit differentiation to the equation $\exp g = f$. □

Lemma 13.1.5. *Let $\Omega \subseteq \mathbb{C}$ be simply connected and let $f : \Omega \to \mathbb{C}$ be a holomorphic and nonvanishing function. Then there is a holomorphic function g on Ω such that* $\exp g = f$.

Proof. By the preceding lemma, there is a C^1 function g that works. But then g satisfies the Cauchy–Riemann equations, as we see by implicit differentiation. □

Theorem 13.1.6. *Cousin II can be solved on any domain $\Omega \subseteq \mathbb{C}$.*

Proof. Let $\mathcal{U} = \{U_j\}$ be an open covering of Ω and let $\{g_{ij}\}$ be Cousin II data for this covering.

We first consider a special case.

Part I: Assume that each U_i is a disk. Use Lemma 13.1.4 to write each $g_{jk} = \exp h_{jk}$ (this is where we use the hypothesis that each U_i is a disk so that each $U_j \cap U_k$ is topologically trivial). Of course the h_{jk} are Cousin I data for the covering $\mathcal{U}$. Let $\{h_i\}$ be a solution of this Cousin I problem. Finally, set $g_i = \exp h_i$. That solves our original Cousin II problem.

Part II: In case the U_i are not all disks (hence not necessarily all topologically trivial), we let $\{\tilde{U}_j\}$ be a refinement of the covering $\{U_i\}$ such that each $\tilde{U}_j$ is a disk. Let $\rho : \mathbb{N} \to \mathbb{N}$ satisfy $\tilde{U}_i \subseteq U_{\rho(i)}$ for each i (ρ is called an *affinity function*). Define

$$\tilde{g}_{jk} : \tilde{U}_j \cap \tilde{U}_k \to \mathbb{C}$$

by $\tilde{g}_{jk}(z) = g_{\rho(j)\rho(k)}(z)$. Then $\{\tilde{g}_{jk}\}$ is a set of holomorphic Cousin II data for the covering $\{\tilde{U}_i\}$. Let $\{\tilde{g}_i\}$ be nonvanishing holomorphic functions that solve this Cousin II problem (as provided by Part I of the proof).

Now for any i, j, k, and $z \in U_i \cap \tilde{U}_j \cap \tilde{U}_k$ we have

$$\tilde{g}_k \tilde{g}_j^{-1} g_{\rho(k),i} g_{i,\rho(j)}(z) = \tilde{g}_k \tilde{g}_j^{-1} g_{\rho(k)\rho(j)}(z) = \tilde{g}_k \tilde{g}_j^{-1} \tilde{g}_{kj}(z) = 1\,.$$

Therefore

$$\tilde{g}_k g_{\rho(k),i}(z) = \tilde{g}_j g_{\rho(j),i}(z) \qquad \text{on} \quad U_i \cap \tilde{U}_j \cap \tilde{U}_k.$$

As a result,

$$g_i(z) \equiv \tilde{g}_k g_{\rho(k),i}(z) \qquad \text{on} \quad U_i \cap \tilde{U}_k$$

gives a well-defined nonvanishing holomorphic function on U_i.

Finally, for any j, k, ℓ, and $z \in \tilde{U}_\ell \cap U_j \cap U_k$, we have

$$g_k g_j^{-1}(z) = \tilde{g}_\ell g_{\rho(\ell),k}(z) \tilde{g}_\ell^{-1} g_{\rho(\ell),j}^{-1}(z) = g_{jk}(z).$$

This is the desired result. □

The next result is a typical application of Cousin II.

Theorem 13.1.7 (Weierstrass). *Let $\Omega \subseteq \mathbb{C}$ be any domain. Let $M \subseteq \Omega$ be a discrete set. Then there is a holomorphic f on all of Ω such that $M = \{z \in \Omega : f(z) = 0\}$.*

Proof. Let each $m \in M$ have a neighborhood U_m in Ω that excludes all other elements of M. Let $\{U_i\}$ be a locally finite subcover of M subordinate to $\{U_m\}_{m\in M}$. Let $U_0 = \Omega \setminus M$. Then U_0 is open in Ω. Let $f_i(z) = z - m_i$ on U_i and let f_0 on U_0 be identically equal to 1.

Define Cousin II data

$$g_{jk} = \frac{f_k}{f_j} \qquad \text{on} \quad U_j \cap U_k. \tag{13.1.3}$$

Since Cousin II is solvable on Ω, there exist nonvanishing holomorphic g_i on U_i such that

$$g_{jk} = \frac{g_k}{g_j} \qquad \text{on} \quad U_j \cap U_k\,. \tag{13.1.4}$$

Define

$$f(z) = \frac{f_i(z)}{g_i(z)} \qquad \text{on} \quad U_i.$$

Then (13.1.3) and (13.1.4) imply that f is well defined and holomorphic on Ω. Also $M = \{z \in \Omega : f(z) = 0\}$. □

Weierstrass's theorem was generalized by his student Mittag–Leffler. The result is the following incisive and widely-used theorem.

Theorem 13.1.8 (Mittag–Leffler). *Let $\Omega \subseteq \mathbb{C}$ be a domain and $M \subseteq \Omega$ a discrete set. For each $m \in M$ let p_m be the principal part of a meromorphic function at m. Then there is a meromorphic function h on all of Ω such that h/p_m is holomorphic near m for each m.*

Proof. The proof is nearly identical to that of the last result. Details are left for the reader. □

13.2 A Few Words About Sheaves

It is not out of place to talk about sheaves in a book on one complex variable. Certainly sheaves are the most natural language in which to formulate ideas about Riemann surfaces (see [GUN1], [GUN2]). Many of those ideas specialize down rather naturally to give very elegant formulations of ideas in the complex plane.

Let $\mathcal{F}$ be a topological space, X a Hausdorff space, and let $\pi : \mathcal{F} \to X$ be a continuous mapping satisfying

(13.2.1) π is surjective;
(13.2.2) If $f \in \mathcal{F}$ then there is a neighborhood $W \subseteq \mathcal{F}$ of f such that $\pi|_W$ is a homeomorphism.

Then the triple $(\mathcal{F}, X, \pi)$ is called a *sheaf*. The letter $\mathcal{F}$ alludes to the French *faisceau* (sheaf), honoring the French mathematician Jean Leray (1906–1998), who developed the idea. It was K. Oka (1901–1978) and H. Cartan (1904–) who first used sheaves in the theory of complex variables. The map π is called the sheaf *projection*. If $x \in X$, then the inverse image $\mathcal{F}_x \equiv \pi^{-1}(\{x\})$ is called the *stalk* over x. If $U \subseteq X$ and $\mu : U \to \mathcal{F}$ is a continuous map that satisfies $\pi \circ \mu(x) = x$ for all $x \in U$, then μ is called a *section* of $\mathcal{F}$ over U. The set of all sections of $\mathcal{F}$ over U is denoted by $\Gamma(U, \mathcal{F})$. Frequently the stalks are each equipped with an algebraic structure such as group, ring, or module. In such a case we speak of a sheaf of groups, rings, or modules.

We shall see that sheaves are a natural vehicle for passing from local information to global information. Local statements are usually formulated in terms of elements of $\Gamma(U, \mathcal{F}), U$ a small open set in X. Global statements are generally seen to be assertions about $\Gamma(X, \mathcal{F})$. There are several important elementary examples that should be kept in mind for the rest of the section.

Example 13.2.1. Let $X = M$ be a manifold. Define $\mathcal{F} = \mathbb{Z} \times M$ with the obvious topology. Let $\pi : \mathcal{F} \to M$ be given by $\pi(z, m) = m$. This is a *trivial* or *product* sheaf. If $U \subseteq M$, then $\mu : U \to \mathcal{F}$ given by $\mu(m) = (0, m)$ is a section of $\mathcal{F}$ over U. We call it the 0-section. Of course $\mathcal{F}$ is a sheaf of rings.

Example 13.2.2. Let $X = M$ be a paracompact manifold and $m_0 \in M$. Define

$$\mathcal{F} = (\mathbb{Z} \times \{m_0\}) \cup (\{0\} \times (M \setminus \{m_0\})) .$$

Topologize $\mathcal{F}$ as follows:

(i) For $(0, m) \in \mathcal{F}, m \neq m_0$, we let U be a neighborhood of m in M that is disjoint from m_0. Then $\{0\} \times U$ is a neighborhood of $(0, m)$ in M.
(ii) The only neighborhood of $(k, m_0) \in \mathcal{F}$ is the singleton $\{(k, m_0)\}$ itself.

Using the sets described in **(i)** and **(ii)** as a subbase, we may generate a topology for $\mathcal{F}$. Of course the projection is given by $\pi(k, m) = m$. Then π is surjective and is a local homeomorphism. This sheaf is known as the *skyscraper sheaf.* We encourage the reader to draw a picture. See also Figure 13.2.

Example 13.2.3. Let M be a paracompact C^∞ manifold. If $m \in M$ and $U \subseteq M$ is a neighborhood of m, then we define $C^\infty(m, U)$ to be the collection of all scalar-valued C^∞ functions defined on U. Let

$$C^\infty(m) = \bigcup_{U \ni m} C^\infty(m, U).$$

If $f, g \in C^\infty(m)$, then we define the relation $f \sim g$ if $f \equiv g$ on some small neighborhood of m. This is clearly an equivalence relation. The set $C^\infty_m \equiv C^\infty(m)/\sim$ is called *the ring of germs of C^∞ functions at m.*

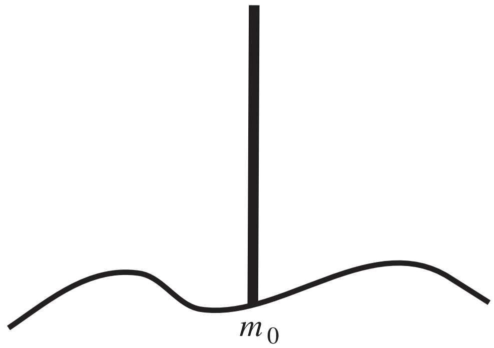

Fig. 13.2. The skyscraper sheaf.

If $f \in C^\infty(m)$ then we denote the residue class of f in C^∞_m by either $\gamma_m f$ or $[f]_m$, depending on the context. Our sheaf $\mathcal{F}$ will be

$$\mathcal{F} = \bigcup_{m \in M} C^\infty_m .$$

This is the *sheaf of germs of* C^∞ *functions* on M.

Topologize $\mathcal{F}$ as follows: If $F \in \mathcal{F}$, let f be a representative of F. Thus f is defined on an open $W \subseteq M$. Let an open neighborhood of F in $\mathcal{F}$ be given by

$$\{f, W\} \equiv \{\gamma_x f : x \in W\}.$$

The set of all such $\{f, W\}$ generates (as a subbase) the coarsest topology under which the projection $\pi : [f]_x \mapsto x$ is a local homeomorphism.

If a C^∞ function g on an open set $U \subseteq M$ is fixed, then the map

$$U \ni x \mapsto \gamma_x g \in \mathcal{F}$$

is a section of $\mathcal{F}$ over U. [Note, in particular, that a globally defined C^∞ function on M is just an element of $\Gamma(X, \mathcal{F})$.] Conversely, any section of $\mathcal{F}$ over an open set $U \subseteq M$ is induced by a C^∞ function on U. Thus the C^∞ functions on $U \subseteq M$ are just the sections of $\mathcal{F}$ over U.

The given topology on $\mathcal{F}$ is not Hausdorff. To see this, take $M = \mathbb{R}$ and $m = 0$ and consider the sheaf elements induced by the identically 0 function and by the function

$$f(x) = \begin{cases} 0 & \text{if } x \leq 0\,, \\ e^{-1/x^2} & \text{if } x > 0. \end{cases}$$

Remark 13.2.4. Naively, a sheaf looks like Figure 13.3. In the last example, each C^∞ function on an open set $U \subseteq M$ corresponds to a "layer" in $\mathcal{F}$. For each $x \in M$, the stalk over x corresponds to the germs of all C^∞ functions at x.

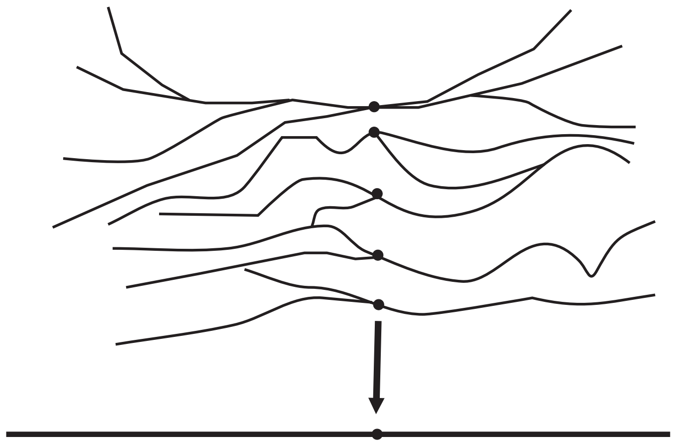

Fig. 13.3. Idea of a sheaf.

Exercise for the Reader: Describe C_x^∞ in terms of formal power series. As you do this, keep in mind the theorem of E. Borel that *any* formal power series about x is the Taylor series of some genuine C^∞ function defined in a neighborhood of x (see [KRP1]). Of course the series will not, in general, converge to that C^∞ function. ♢

The last example goes through in just the same way if we replace the C^∞ functions by either the C^k functions or the (smooth) differential forms of degree k (or bidegree (p, q) on a complex manifold) or the real analytic functions (when M is a real analytic manifold) or the holomorphic functions (when M is a complex analytic manifold). The sheaf of germs of holomorphic functions *is* Hausdorff (exercise).

Example 13.2.5. The constant sheaf $\mathbb{C} \times X$ is a sheaf over X if and only if $\mathbb{C}$ is equipped with the discrete topology (exercise). [*Hint:* Think about the property of local homeomorphism.]

Given a sheaf $(\mathcal{F}, X, \pi)$ of abelian groups (with group operations written *additively*) and a countable open covering $\mathcal{U} = \{U_i\}_{i=1}^\infty$ of X, we construct *cohomology groups* as follows. First, if $r \in \{0, 1, 2, \dots\}$, then an *r-cochain* with respect to $\mathcal{U}$ is a function f that assigns to each $(i_0, \dots, i_r) \in \mathbb{N}^{r+1}$ an element $f(i_0, \dots, i_r) \in \Gamma(U_{i_0} \cap \cdots \cap U_{i_r}, \mathcal{F})$, subject to the condition that f is an alternating function of its indices. Let $C^r(\mathcal{U}, \mathcal{F})$ denote the set of all r-cochains of $\mathcal{F}$ with respect to the cover $\mathcal{U}$. Notice that C^r is an abelian group in a natural way.[1]

Define the *coboundary operator*

$$\delta : C^r \to C^{r+1}$$

[1] Although we use the notation C^r in other contexts to denote spaces of smooth functions, there is little danger of confusion here.

by

$$(\delta f)(i_0, \ldots, i_{r+1}) = \sum_{j=0}^{r+1} (-1)^j f\left(i_0, \ldots, \widehat{i_j}, \ldots, i_{r+1}\right).$$

Here $\widehat{\ }$ denotes omission.

Lemma 13.2.6. *We have*

$$\delta^2 = 0.$$

Proof. Observe that $\delta^2 f$ is a sum of terms

$$f(i_0, \ldots, \widehat{i_j}, \ldots, \widehat{i_k}, \ldots, i_{r+2})$$

with $j < k$. When i_j is removed first, the signature is $(-1)^{j+(k-1)}$. If i_k is removed first then the signature is $(-1)^{k+j}$. By cancellation, the sum is zero. $\square$

Define $Z^r = Z^r(\mathcal{U}, \mathcal{F})$ to be those $f \in C^r$ such that $\delta f = 0$. Define $B^r = B^r(\mathcal{U}, \mathcal{F})$ to be those $g \in C^r$ such that $g = \delta f$ for some $f \in C^{r-1}$. Then Z^r and B^r are subgroups of C^r. We call B^r the group of *coboundaries* and Z^r the group of *cocycles*.

Notice that B^r is a subgroup of Z^r just because $\delta^2 = 0$. Thus the quotient

$$H^r(\mathcal{U}, \mathcal{F}) = H^r \equiv Z^r / B^r \ , \quad r \geq 1 \, ,$$

is well defined and is called the r^{th} *cohomology group of* X (with respect to the covering $\mathcal{U}$) *with coefficients in the sheaf* $\mathcal{F}$. By convention, since B^0 must clearly be $\emptyset$, we set $H^0 = Z^0$.

In more sophisticated language, the lemma tells us that we have the *semiexact sequence*

$$C^0 \xrightarrow{\delta} C^1 \xrightarrow{\delta} C^2 \xrightarrow{\delta} \cdots ,$$

where, at each C^r, $B^r = \text{Im}\,(\delta|_{C^{r-1}}) \subseteq \text{Ker}\,(\delta|_{C^r}) = Z^r$. We measure the extent to which this sequence fails to be exact at C^r by way of the cohomology group H^r.

Example 13.2.7. Let Ω be the annulus in $\mathbb{C}$ with $\mathcal{U} = \{U_1, U_2, U_3\}$ the covering (of Ω^0) indicated in the first example of Section 13.1. Let $\mathcal{F} = \mathbb{Z} \times \Omega$. Then $H^0(\mathcal{U}, \mathcal{F}) \equiv Z^0$ consists of those f on $I = \{1, 2, 3\}$ such that $f(i_0) - f(i_1) = 0$ on $U_{i_0} \cap U_{i_1}$ for all i_0, i_1. That is, f assigns to each U_i the same integer. In brief, $H^0 = Z^0 \cong \Gamma(\Omega, \mathcal{F}) \cong \mathbb{Z}$.

Now Z^1 consists of those f on $I \times I$ such that $f(i_0, i_1) - f(i_0, i_2) + f(i_1, i_3) = 0$ on $U_{i_0} \cap U_{i_1} \cap U_{i_2}$. [*Note:* This is vacuously satisfied since $U_1 \cap U_2 \cap U_3 = \emptyset$.] Thus $Z^1 = C^1$.

On the other hand, $Z^1 \ni f = \delta g$ for some $g \in C^0$ if and only if $f(i_0, i_1) = g(i_0) - g(i_1)$. [This simply says that the analogue of Cousin I can be solved for the sheaf $\mathcal{F}$.] In summary, $H^1 \equiv Z^1 / B^1 = C^1 / B^1$ counts the number of unsolvable Cousin I problems with data in the group $\mathbb{Z}$.

The 1-cocycle

$$f(1,2) = 0\,, \quad f(1,3) = 0\,, \quad f(2,3) = 1$$

is clearly not a coboundary. Also the 1-cocycle

$$g(1,2) = 1\,, \quad g(1,3) = 0\,, \quad g(2,3) = 0$$

is also not a coboundary. Finally, the 1-cocycle

$$h(1,2) = 0\,, \quad h(1,3) = -1\,, \quad h(2,3) = 0$$

is not a coboundary. However, notice that both $g - f$ and $h - f$ *are* coboundaries. Since f, g, h clearly generate all of Z^1 as a group, it follows that the residue of f alone in H^1 generates H^1. Thus H^1 is cyclic. Indeed, $H^1 \equiv \mathbb{Z}$.

Finally, $Z^r = B^r = C^r$ for $r \geq 2$ for vacuous reasons, hence $H^r = 0$ for $r \geq 2$.

Thus we see that the annulus Ω has a "one-dimensional" hole, and it is detected by $H^1(\mathcal{U}, \mathcal{F})$. The annulus has no holes of dimension two or higher.

Example 13.2.8. Let $\Omega, \mathcal{U} = \{U_i\}$ be as in the last example. However, let $\mathcal{F}$ be the sheaf of germs of C^k (k times continuously differentiable) functions. Then Z^0 consists of those $f(i)$ such that $f(i_0) = f(i_1)$ on $U_{i_0} \cap U_{i_1}$. We therefore see that $Z^0 = C^k(\Omega)$. Just as in Section 13.1, we may always use partitions of unity to solve Cousin I for C^k functions. Thus $H^1(\Omega, \mathcal{F}) = 0$. This is a manifestation of the fact that the sheaf of germs of C^k functions (or C^∞ functions) is *fine* (see E. Spanier [1]). For trivial reasons, $H^r = 0$ when $r \geq 2$.

Example 13.2.9. Let $\Omega, \mathcal{U}$ be as above. Let $\mathcal{F}$ be the sheaf of germs of holomorphic functions on Ω. As in the preceding example, $H^0(\Omega, \mathcal{F})$ is isomorphic to the group of holomorphic functions on Ω. By Theorem 13.1, Cousin I can always be solved on Ω, so $H^1 = 0$. Finally, $H^r = 0, r \geq 2$, for trivial reasons.

Example 13.2.10. Let $\Omega = \{z \in \mathbb{C} : |z| < 2\}$ and let $\mathcal{V}$ be the covering (of Ω^1) as in the first Example of Section 13.1. Let $\mathcal{F} = \mathbb{Z} \times \Omega$ as in Example 13.2.1. Then $H^0 = Z^0 = \mathbb{Z}$ as in Example 13.2.7.

Now $f \in C^1$ satisfies $\delta f = 0$ precisely when $f(1,2) - f(1,3) + f(2,3) = 0$ on $V_1 \cap V_2 \cap V_3$ (this condition is no longer vacuous). Also, $f = \delta g$ precisely when $f(i_0, i_1) = g(i_0) - g(i_1)$. As discussed in Section 13.1, it follows that $\delta f = 0$ implies that $f = \delta g$ for some $g \in C^0$. Hence $H^1 = A^1/B^1 = 0$. For trivial reasons, $H^r = 0$ when $r \geq 2$.

Exercise for the Reader: Let $\Omega \subseteq \mathbb{C}$ be an open set and $\mathcal{U} = \{U_j\}$ an open covering of Ω. Let $(\mathcal{F}, \Omega, \pi)$ be the sheaf of germs of holomorphic functions. Then $H^1(\mathcal{U}, \mathcal{F}) = 0$ if and only if every Cousin I problem for the covering $\mathcal{U}$ can be solved. $\diamondsuit$

Return now to the general setting of arbitrary Hausdorff X, sheaf of abelian groups $\mathcal{F}$, and covering $\mathcal{U}$. We need to see to what extent the cohomology groups depend on the covering $\mathcal{U}$. The hope is that, if the covering becomes sufficiently fine, then the cohomology groups are independent of the choice of covering. When formulated suitably carefully, this hope is fulfilled.

If $\mathcal{V} = \{V_i\}$ is a refinement of the covering $\mathcal{U} = \{U_j\}$, then we need to relate the cohomology computed with respect to $\mathcal{V}$ to that computed with respect to $\mathcal{U}$. In the end we hope to see that the cohomology stabilizes under the direct limit process for coverings.

To make the preceding paragraphs precise, we let $\mathcal{U} = \{U_j\}_{j\in J}$ be an open covering for X and $\mathcal{V} = \{V_i\}_{i\in I}$ be an open covering that refines $\mathcal{U}$. Let $\sigma : I \to J$ satisfy $V_i \subseteq U_{\sigma(i)}$ for all $i \in I$. We call σ an *affinity function.* Then σ induces a map σ^* of $C^r(\mathcal{U},\mathcal{F})$ to $C^r(\mathcal{V},\mathcal{F})$ as follows: Let $f \in C^r(\mathcal{U},\mathcal{F})$. Let $(i_0,\ldots,i_r) \in \mathbb{N}^{r+1}$. Define $\sigma^* f \in C^r(\mathcal{V},\mathcal{F})$ by

$$(\sigma^* f)(i_0,\ldots,i_r) = f(\sigma(i_0),\ldots,\sigma(i_r))|_{V_{i_0}\cap\cdots\cap V_{i_r}} .$$

Exercise for the Reader: Verify the following simple facts about σ^* :

(a) σ^* is a homomorphism;

(b) $\delta\sigma^* = \sigma^*\delta$;

(c) $\sigma^*(Z^r) \subseteq Z^r$ and $\sigma^*(B^r) \subseteq B^r$;

(d) By **(c)**, the induced map $\sigma^* : H^r(\mathcal{U},\mathcal{F}) \to H^r(\mathcal{V},\mathcal{F})$ is well defined. ♢

We call two r-cocycles with respect to a fixed covering *cohomologous* if they differ by a coboundary (i.e., if they are representatives of the same cohomology class in H^r).

The next lemma is technical but essential:

Lemma 13.2.11. *Let $\mathcal{U}$ be a covering of X and $\mathcal{V}$ a refinement of $\mathcal{U}$. Let σ and σ' be two affinity functions for $\mathcal{U}$ and $\mathcal{V}$. Then $\sigma^* = \sigma'^*$ on H^r.*

Proof. We proceed by means of a standard device of homological algebra known as a *chain homotopy.* Since the assertion is obvious for $r = 0$, we may assume that $r \geq 1$. Let $\Theta : C^r(\mathcal{U},\mathcal{F}) \to C^{r-1}(\mathcal{V},\mathcal{F})$ be given by

$$(\Theta f)(i_0,\ldots,i_{r-1}) = \sum_{\ell=0}^{r-1} f(\sigma(i_0),\ldots,\sigma(i_{\ell-1}),\sigma(i_\ell),\sigma'(i_\ell), \sigma'(i_{\ell+1}),\ldots,\sigma'(i_{r-1})).$$

[Notice that Θ is a sort of adjoint for δ.] We claim that

$$(\sigma')^*f - \sigma^* f = \Theta\delta f + \delta\Theta f. \tag{13.2.3}$$

This is the chain homotopy condition. Assuming that we have proved (13.2.3), notice that if $f \in C^r(\mathcal{U}, \mathcal{F}), \delta f = 0$, then $(\sigma')^*f - \sigma^* f = \delta\Theta f$. Thus $(\sigma')^* f$ and $\sigma^* f$ are cohomologous. The result follows.

It remains to prove (13.2.3). We compute that

$$\begin{aligned}
&\delta(\Theta f)(i_0, \ldots, i_r) = \sum_{k=0}^{r} (-1)^k (\Theta f)(i_0, \ldots, \widehat{i}_k, \ldots, i_r) \\
&= \sum_{k=0}^{r} (-1)^k \left\{ \sum_{\ell=0}^{k-1} (-1)^\ell f\left(\sigma(i_0), \ldots, \sigma(i_\ell), \sigma'(i_\ell), \ldots, \widehat{\sigma'(i_k)}, \ldots, \sigma'(i_r)\right) \right. \\
&\qquad \left. + \sum_{\ell=k}^{r-1} (-1)^\ell f\left(\sigma(i_0), \ldots, \widehat{\sigma(i_k)}, \ldots, \sigma(i_{\ell+1}), \sigma'(i_{\ell+1}), \ldots, \sigma'(i_r)\right) \right\} \\
&= \sum_{k=0}^{r} \sum_{\ell=0}^{k-1} (-1)^{k+\ell} f\left(\sigma(i_0), \ldots, \sigma(i_\ell), \sigma'(i_\ell), \ldots, \widehat{\sigma'(i_k)}, \ldots, \sigma'(i_r)\right) \\
&\qquad - \sum_{k=0}^{r} \sum_{p=k+1}^{r} (-1)^{p+k} f\left(\sigma(i_0), \ldots, \widehat{\sigma(i_k)}, \ldots, \sigma(i_p), \sigma'(i_p), \ldots, \sigma'(i_r)\right).
\end{aligned}$$

On the other hand,

$$\begin{aligned}
&\Theta(\delta f)(i_0, \ldots, i_r) \\
&= \sum_{\ell=0}^{r} (-1)^\ell \delta f(\sigma(i_0), \ldots, \sigma(i_\ell)\sigma'(i_\ell), \ldots, \sigma'(i_r)) \\
&= \sum_{\ell=0}^{r} (-1)^\ell \left\{ \sum_{k=0}^{\ell} (-1)^k f\big(\sigma(i_0), \ldots, \widehat{\sigma(i_k)}, \ldots, \sigma(i_\ell), \sigma'(i_\ell), \ldots, \sigma'(i_r)\big) \right. \\
&\qquad \left. + \sum_{k=\ell+1}^{r+1} (-1)^k f\big(\sigma(i_0), \ldots, \sigma(i_\ell), \sigma'(i_\ell), \ldots, \widehat{\sigma'(i_{k-1})}, \ldots, \sigma'(i_r)\big) \right\} \\
&= \sum_{\ell=0}^{r} \sum_{k=0}^{\ell} (-1)^{\ell+k} f\left(\sigma(i_0), \ldots, \widehat{\sigma(i_k)}, \ldots, \sigma(i_\ell), \sigma'(i_\ell), \ldots, \sigma'(i_r)\right) \\
&\qquad - \sum_{\ell=0}^{r} \sum_{p=\ell}^{r} (-1)^{p+\ell} f\left(\sigma(i_0), \ldots, \sigma(i_\ell), \sigma'(i_\ell), \ldots, \widehat{\sigma'(i_p)}, \ldots, \sigma'(i_r)\right) \\
&= \left\{ \sum_{\ell=0}^{r} \sum_{k=0}^{\ell-1} (-1)^{\ell+k} f\left(\sigma(i_0), \ldots, \widehat{\sigma(i_k)}, \ldots, \sigma(i_\ell), \sigma'(i_\ell), \ldots, \sigma'(i_r)\right) \right. \\
&\qquad \left. - \sum_{\ell=0}^{r} \sum_{p=\ell+1}^{r} (-1)^{p+\ell} f\left(\sigma(i_0), \ldots, \sigma(i_\ell), \sigma'(i_\ell), \ldots, \widehat{\sigma'(i_p)}, \ldots, \sigma'(i_r)\right) \right\}
\end{aligned}$$

$$+\left\{\sum_{\ell=0}^{r} f\left(\sigma(i_0),\ldots,\sigma(i_{\ell-1}),\sigma'(i_\ell),\ldots,\sigma'(i_r)\right)\right.$$
$$\left.-\sum_{\ell=0}^{r} f\big(\sigma(i_0),\ldots,\sigma(i_\ell),\sigma'(i_{\ell+1}),\ldots,\sigma'(i_r)\big)\right\}$$
$$= -\delta(\Theta f)(i_0,\ldots,i_r) + \{f(\sigma'(i_0),\ldots,\sigma'(i_r)) - f(\sigma(i_0),\ldots,\sigma(i_r))\}.$$

Here the first part of the last equality comes from renaming indices and switching the order of summation. Moreover, the equality of the expressions in braces holds because the series cancel. Now the last line is equal to

$$-\delta(\Theta f)(i_0,\ldots,i_r) + \{(\sigma')^* f(i_0,\ldots,i_r) - \sigma^* f(i_0,\ldots,i_r)\}.$$

It follows that

$$(\delta\Theta) f(i_0,\ldots,i_r) + (\Theta\delta) f(i_0,\ldots,i_r) = (\sigma')^* f(i_0,\ldots,i_r) - \sigma^* f(i_0,\ldots,i_r)$$

as desired. □

If $f \in Z^r(\mathcal{U},\mathcal{F})$, then it is convenient to adopt the standard notation $[f]$ to denote the equivalence class of f in $H^r = Z^r/B^r$. The lemma says that $\sigma^*([f]) = \sigma'^*([f])$ for all $[f] \in H^r(\mathcal{U},\mathcal{F})$ and any two affinity functions σ, σ'. If $\mathcal{V}$ is a refinement of $\mathcal{U}$, then f induces a well-defined cohomology class with respect to $\mathcal{V}$ (namely $\sigma^* f$ for *any* affinity function σ). Let

$$\mathcal{Z}^r(X,\mathcal{F}) \equiv \bigcup_{\text{coverings } \mathcal{U} \text{ of } X} Z^r(\mathcal{U},\mathcal{F}).$$

Let $f, g \in \mathcal{Z}^r$—say that $f \in Z^r(\mathcal{U},\mathcal{F}), g \in Z^r(\mathcal{V},\mathcal{F})$. We write $f \sim g$ provided that there is a common refinement $\mathcal{W}$ such that the induced cohomology classes of f and g with respect to $\mathcal{W}$ are equal. Then $\sim$ is an equivalence relation (exercise). The collection of equivalence classes is denoted by $H^r(X,\mathcal{F})$, the r^{th} cohomology of X with coefficients in the sheaf $\mathcal{F}$. In an obvious fashion, $H^r(X,\mathcal{F})$ is an abelian group. If $[f] \in H^r(\mathcal{U},\mathcal{F})$, let $[[f]]$ denote the residue class of $[f]$ in $H^r(X,\mathcal{F})$. Thus $[f] \mapsto [[f]]$ is a homomorphism of $H^r(\mathcal{U},\mathcal{F})$ into $H^r(X,\mathcal{F})$ for each r and each covering $\mathcal{U}$. The group $H^r(X,\mathcal{F})$ is the *direct limit* of the groups $H^r(\mathcal{U},\mathcal{F})$.

Each element of $H^r(X,\mathcal{F})$ comes from some covering $\mathcal{U}$. Thus the elements of $H^0(X,\mathcal{F})$ are in one-to-one correspondence with the global sections of $\mathcal{F}$. To understand $H^1(X,\mathcal{F})$, we need the next lemma:

Lemma 13.2.12. *Let $\mathcal{U}$ be an open covering of X and $\mathcal{V}$ a covering that is a refinement of $\mathcal{U}$. Let σ be an affinity function for $\mathcal{U}$ and $\mathcal{V}$. Then σ^* : $H^1(\mathcal{U},\mathcal{F}) \to H^1(\mathcal{V},\mathcal{F})$ is injective.*

Corollary 13.2.13. *For any covering $\mathcal{U}$ of X, the map $H^1(\mathcal{U},\mathcal{F}) \to H^1(X,\mathcal{F})$ is injective.*

Proof of Lemma 13.2.12. It suffices to show that $\operatorname{Ker}\sigma^* = [0]$. For this, we must prove that if $c \in Z^1(\mathcal{U}, \mathcal{F})$ satisfies $\sigma^* c \in B^1(\mathcal{V}, \mathcal{F})$, then $c \in B^1(\mathcal{U}, \mathcal{F})$. To see this, notice that the hypothesis $\sigma^* c \in B^1(\mathcal{V}, \mathcal{F})$ implies that $\sigma^* c = \delta\gamma$ for some $\gamma \in C^0(\mathcal{V}, \mathcal{F})$. With the notation $\mathcal{V} = \{V_i\}_{i \in I}, \mathcal{U} = \{U_j\}_{j \in J}$, we have for all $m, n \in I$ that

$$\gamma_n - \gamma_m = c_{\sigma(m)\sigma(n)} = c_{\sigma(m)j} - c_{\sigma(n)j} \qquad \text{on} \quad U_j \cap V_m \cap V_n$$

(here we have used the condition that $\delta c = 0$). In other words,

$$\gamma_n + c_{\sigma(n)j} = \gamma_m + c_{\sigma(m)j} \qquad \text{on} \quad U_j \cap V_m \cap V_n .$$

Hence the section $e_j \in \Gamma(U_j, \mathcal{F})$ given by

$$e_j = \gamma_n + c_{\sigma(n)j} \qquad \text{on} \quad U_j \cap V_n$$

is well defined. We now compute, on $U_k \cap U_\ell \cap V_n$,

$$e_\ell - e_k = (\gamma_n + c_{\sigma(n)\ell}) - (\gamma_n + c_{\sigma(n)k}) = c_{k\sigma(n)} + c_{\sigma(n)\ell} = c_{k\ell} .$$

Hence $c = \delta e$ or $c \in B^1(\mathcal{U}, \mathcal{F})$. □

It follows from Corollary 13.2.13 that $H^1(X, \mathcal{F}) = 0$ if and only if $H^1(\mathcal{U}, \mathcal{F}) = 0$ for some covering $\mathcal{U}$. Therefore Cousin I for the sheaf $\mathcal{F}$ is always solvable if and only if $H^1(X, \mathcal{F}) = 0$.

Unfortunately, there are no such simple descriptions of $H^p(X, \mathcal{F})$ in terms of $H^p(\mathcal{U}, \mathcal{F})$ for $p \geq 2$. On the other hand, the following important result of J. Leray often makes it unnecessary to pass to the direct limit in order to calculate with cohomology.

Theorem 13.2.14. *Let X be a paracompact manifold, $\mathcal{F}$ a sheaf over X, and $\mathcal{U} = \{U_i\}_{i \in I}$ an open covering on X such that $H^p(U_{i_0} \cap \cdots \cap U_{i_k}, \mathcal{F}) = 0$ for every $p \geq 1$ and every choice of $i_0, \ldots, i_p$. Then $H^p(\mathcal{U}, \mathcal{F}) = H^p(X, \mathcal{F})$ for all $p \geq 0$.*

We do not supply a proof of this theorem but instead refer the reader to [GUN1, p. 44]. It should be noted that, in many applications with X an open domain in Euclidean space, a covering by balls or disks will plainly satisfy the hypotheses of Leray's theorem. So the theorem is not difficult to use.

Our goal now is to develop a little homological algebraic machinery so as to give the reader the flavor of how sheaf cohomology is used in practice. We begin with the following exercise:

Exercise for the Reader: Let X be a C^∞ paracompact manifold. Let $\mathcal{E}$ be the sheaf of germs of C^∞ functions on Ω. Let $\mathcal{F}$ be a sheaf of $\mathcal{E}$-modules (i.e., each $\mathcal{F}_x$ is an $\mathcal{E}$-module). Then $H^p(\mathcal{U}, \mathcal{F}) = 0$ for every open covering $\mathcal{U} = \{U_i\}_{i \in I}$ and every $p > 0$. [*Hint*: This will be similar to the solution of

Cousin I for C^∞ functions. Let $\{\phi\}_{i\in I}$ be a partition of unity subordinate to $\mathcal{U}$. If $c \in Z^p(\mathcal{U},\mathcal{F})$ and $s = (s_0, s_1, \ldots, s_{p-1}) \in I^p$, then set $e(i) = \sum_k \phi_k c_{k,i}$. Then $\delta e = c$.] $\diamondsuit$

The *matière* of homological algebra is exact sequences. If A, B, C are abelian groups and

$$A \xrightarrow{\phi} B \xrightarrow{\psi} C$$

are homomorphisms, then we say that this sequence is *exact* at B if $\text{Image}\,\phi = \text{Ker}\,\psi$. If $C = \{0\}$, then "exactness" at B means that ϕ is surjective. If $A = \{0\}$, then exactness at B means that ψ is injective.

If $\mathcal{F}, \mathcal{G}, \mathcal{H}$ are sheaves (of abelian groups) over X and

$$0 \to \mathcal{F} \xrightarrow{\phi} \mathcal{G} \xrightarrow{\psi} \mathcal{H} \to 0$$

are homomorphisms (on each stalk), then "exactness" is defined in an obvious manner. These homomorphisms trivially induce homomorphisms

$$0 \to C^p(\mathcal{U},\mathcal{F}) \xrightarrow{\phi^*} C^p(\mathcal{U},\mathcal{G}) \xrightarrow{\psi^*} C^p(\mathcal{U},\mathcal{H}), \tag{13.2.4}$$

any $p > 0$, any open covering $\mathcal{U} = \{U_i\}_{i\in I}$. The induced sequence is also exact (exercise), except that the final map may not be surjective (which is why we have omitted the final 0).

Example 13.2.15. Let $\mathcal{F}$ be the constant sheaf $2\pi i\mathbb{Z} \times X$, where $X = \{z \in \mathbb{C} : 1 < |z| < 2\}$. Let $\mathcal{G}$ be the sheaf of germs of C^∞ functions over X. Let $\mathcal{H}$ be the sheaf of germs of nonvanishing C^∞ functions (under multiplication). In the diagram

$$0 \to \mathcal{F} \xrightarrow{\phi} \mathcal{G} \xrightarrow{\psi} \mathcal{H} \to 0\,, \tag{13.2.5}$$

we take ϕ to be the trivial injection and ψ the exponential map. Then the sequence is exact (because we check the assertion on the stalk level). But the induced exact sequence

$$0 \to C^0(\mathcal{U},\mathcal{F}) \xrightarrow{\phi^*} C^0(\mathcal{U},\mathcal{G}) \xrightarrow{\psi^*} C^0(\mathcal{U},\mathcal{H}) \tag{13.2.6}$$

is *not* surjective at the final stage for the covering $\mathcal{U} = \{X\}$ (or for *any* covering) because not every C^∞ function on the space X has a well-defined C^∞ logarithm.

This failure of exactness is both a source of difficulty and of rich theory. Indeed, the failure of an algebraic property of groups (in this case the property of exactness) can often be measured by a *resolution* of groups for which the property holds. In the present case, the resolution is given by an exact sequence of cohomology groups. We construct the resolution as follows:

$$
\begin{array}{ccccc}
0 & & 0 & & 0 \\
\downarrow & & \downarrow & & \downarrow \\
C^{p-1}(\mathcal{U},\mathcal{F}) & \xrightarrow{\delta} & C^{p}(\mathcal{U},\mathcal{F}) & \xrightarrow{\delta} & C^{p+1}(\mathcal{U},\mathcal{F}) \\
\downarrow \varphi^* & & \downarrow \varphi^* & & \downarrow \varphi^* \\
C^{p-1}(\mathcal{U},\mathcal{G}) & \xrightarrow{\delta} & C^{p}(\mathcal{U},\mathcal{G}) & \xrightarrow{\delta} & C^{p+1}(\mathcal{U},\mathcal{G}) \\
\downarrow \psi^* & & \downarrow \psi^* & & \downarrow \psi^* \\
C_0^{p-1}(\mathcal{U},\mathcal{H}) & \xrightarrow{\delta} & C_0^{p}(\mathcal{U},\mathcal{H}) & \xrightarrow{\delta} & C_0^{p+1}(\mathcal{U},\mathcal{H}) \\
\downarrow & & \downarrow & & \downarrow \\
0 & & 0 & & 0
\end{array}
$$

Figure 13.4

Instead of (13.2.4) we write

$$0 \to C^p(\mathcal{U},\mathcal{F}) \overset{\phi^*}{\to} C^p(\mathcal{U},\mathcal{G}) \overset{\psi^*}{\to} C_0^p(\mathcal{U},\mathcal{H}) \to 0, \qquad (13.2.7)$$

where C_0^p is precisely the subgroup of C^p that is the image of the map ψ^* in (13.2.4) induced by ψ. We have the commutative diagram given in Figure 13.4. The phrase "commutative diagram" means, for instance, that $\phi^* \circ \delta$ (where $\delta : C^{p-1}(\mathcal{U},\mathcal{F}) \to C^p(\mathcal{U},\mathcal{F})$ and $\phi^* : C^p(\mathcal{U},\mathcal{F}) \to C^p(\mathcal{U},\mathcal{G})$) equals $\delta \circ \phi^*$ (where $\delta : C^{p-1}(\mathcal{U},\mathcal{G}) \to C^p(\mathcal{U},\mathcal{G})$ and $\phi^* : C^{p-1}(\mathcal{U},\mathcal{F}) \to C^{p-1}(\mathcal{U},\mathcal{G})$) and likewise for all the other "boxes" in Figure 13.4. (The only possible way that one can write about this subject is to force the reader to verify the details.)

It follows that there are induced maps

$$0 \to Z^p(\mathcal{U},\mathcal{F}) \to Z^p(\mathcal{U},\mathcal{G}) \to Z_0^p(\mathcal{U},\mathcal{H}) \to 0$$

and

$$0 \to B^p(\mathcal{U},\mathit{\Gamma}) \to B^p(\mathcal{U},\mathcal{G}) \to B_0^p(\mathcal{U},\mathcal{H}) \to 0.$$

Hence the maps

$$
\begin{aligned}
0 \to H^0(\mathcal{U},\mathcal{F}) \to H^0(\mathcal{U},\mathcal{G}) &\to H_0^0(\mathcal{U},\mathcal{H})\,, \\
H^1(\mathcal{U},\mathcal{F}) \to H^1(\mathcal{U},\mathcal{G}) &\to H_0^1(\mathcal{U},\mathcal{H})\,,
\end{aligned}
\qquad (13.2.8)
$$

are well defined. We wish to construct homomorphisms that connect up the rows.

Proposition 13.2.16. *The commutative diagram (Figure 13.4) induces a natural homomorphism*

$$\delta^* : H_0^p(\mathcal{U}, \mathcal{H}) \to H^{p+1}(\mathcal{U}, \mathcal{F})$$

for every $p \geq 0$.

Proof. Let $[h] \in H_0^p(\mathcal{U}, \mathcal{H})$, and let h be a representative. Since ψ^* is surjective there is a $g \in Z^p(\mathcal{U}, \mathcal{G})$ such that $\psi^* g = h$. We claim that $[(\phi^*)^{-1}\delta g]$ makes sense and is uniquely determined as an element of $H^{p+1}(\mathcal{U}, \mathcal{F})$ by $[h]$.

Now $\psi^*\delta g = \delta\psi^* g = \delta h = 0$ whence, by exactness, $\delta g \in \phi^*(C^{p+1}(\mathcal{U}, \mathcal{F}))$. So our expression $[(\phi^*)^{-1}\delta g]$ makes sense but may not be uniquely determined. If $\delta g = \phi^* f$, then $\phi^*\delta f = \delta\phi^* f = \delta\delta g = 0$. Thus the injectivity of ϕ^* implies that $\delta f = 0$. So $\delta g \in \phi^*(Z^{p+1}(\mathcal{U}, \mathcal{F}))$.

It remains to check that when $[h] = 0$, that is, $h \in B_0^p(\mathcal{U}, \mathcal{H})$, then $(\phi^*)^{-1}\delta g \in B^{p+1}(\mathcal{U}, \mathcal{F})$. Now $h \in B_0^p(\mathcal{U}, \mathcal{H})$ implies that $h = \delta k$, some $k \in C_0^{p-1}(\mathcal{U}, \mathcal{H})$. The surjectivity of ψ^* yields an $m \in C_0^{p-1}(\mathcal{U}, \mathcal{G})$ with $\psi^* m = k$. But then $\delta m - g$ (with g as in the first paragraph) satisfies

$$\psi^*(\delta m - g) = \psi^*\delta m - \psi^* g = \delta\psi^* m - h = \delta k - h = 0.$$

By exactness, we have $\delta m - g \in \phi^*(C^p(\mathcal{U}, \mathcal{F}))$; say that $\delta m - g = \phi^*\mu$, some $\mu \in C^p(\mathcal{U}, \mathcal{F})$. Then

$$-\delta g = \delta(\delta m - g) = \delta\phi^*\mu = \phi^*\delta\mu \in \phi^*(B^{p+1}(\mathcal{U}, \mathcal{F})). \quad \square$$

Theorem 13.2.17 (The Snake Lemma). *The sequence*

$$\begin{aligned} 0 \to H^0(\mathcal{U}, \mathcal{F}) &\xrightarrow{\phi^*} H^0(\mathcal{U}, \mathcal{G}) \xrightarrow{\psi^*} H_0^0(\mathcal{U}, \mathcal{H}) \xrightarrow{\delta^*} H^1(\mathcal{U}, \mathcal{F}) \\ &\xrightarrow{\phi^*} H^1(\mathcal{U}, \mathcal{G}) \xrightarrow{\psi^*} H_0^1(\mathcal{U}, \mathcal{H}) \xrightarrow{\delta^*} H^2(\mathcal{U}, \mathcal{F}) \to \cdots \end{aligned}$$

is exact.

Proof. **Exactness at $H^p(\mathcal{U}, \mathcal{F})$:** Reviewing the proof of Proposition 13.2.16, we note that the image of δ^* in $H^p(\mathcal{U}, \mathcal{F})$ is precisely the set of cohomology classes determined by $Z^p(\mathcal{U}, \mathcal{F}) \cap (\phi^*)^{-1}\{B^p(\mathcal{U}, \mathcal{G})\}$. On the other hand, the kernel of ϕ^* in $H^p(\mathcal{U}, \mathcal{F})$ must be the subgroup of $(\phi^*)^{-1}\{B^p(\mathcal{U}, \mathcal{G})\}$ that lies in $Z^p(\mathcal{U}, \mathcal{F})$. (Note that exactness at the stage $H^0(\mathcal{U}, \mathcal{F})$ is trivial.)

Exactness at $H^p(\mathcal{U}, \mathcal{G})$: That Image $\phi^* \subseteq \operatorname{Ker}\psi^*$ is immediate. Conversely, if $[g] \in \operatorname{Ker}\psi^*$, then let $g \in Z^p$ be a representative for $[g]$. Then $\psi^* g \in B_0^p(\mathcal{U}, \mathcal{H})$, so $\psi^* g = \delta f$ for some $f \in C_0^{p-1}(\mathcal{U}, \mathcal{H})$. Referring again to Figure 13.4, we see that since ψ is surjective there is an $e \in C^{p-1}(\mathcal{U}, \mathcal{G})$ with $\psi^* e = f$. But then

$$\psi^*(g - \delta e) = \psi^* g - \psi^* \delta e = \delta f - \delta \psi^* e = \delta f - \delta f = 0.$$

Therefore $g - \delta e \in \operatorname{Ker} \psi^*$ *on the cochain level.* Thus there is an $h \in C^p(\mathcal{U}, \mathcal{F})$ with $\phi^* h = g - \delta e$. Also, $\phi^* \delta h = \delta \phi^* h = \delta(g - \delta e) = 0 - 0 = 0$. Since ϕ^* is injective, we conclude that $\delta h = 0$, hence $h \in Z^p$. Finally, $\phi^*[h] = [\phi^* h] = [g - \delta e] = [g]$ as desired. □

Exactness at $\boldsymbol{H^p(\mathcal{U}, \mathcal{H})}$: If $[g] \in H^p(\mathcal{U}, \mathcal{G})$, then $\delta^* \psi^*[g] = (\phi^*)^{-1} \delta[g]$ by definition of δ^*. But $\delta[g] = 0$ and ϕ^* is univalent. Hence $\delta^* \psi^*[g] = 0$.

Now suppose that $[h] \in H^p(\mathcal{U}, \mathcal{H})$ satisfies $\delta^*[h] = 0$. This means, according to the definition of δ^*, that there is a $g \in C^p(\mathcal{U}, \mathcal{G})$ with $\psi^* g = h$ and $\delta^*[h] = (\phi^*)^{-1} \delta[g] = 0$. Set $x = (\phi^*)^{-1} \delta g$. Then x is a coboundary, so $x = \delta y$ for some $y \in C^p(\mathcal{U}, \mathcal{F})$. Let $f = g - \phi^* y$. Then $\delta f = \delta g - \phi^* \delta y = \delta g - \phi^* x = 0$. Hence $f \in Z^p(\mathcal{U}, \mathcal{G})$. Finally, $\psi^*[f] = [\psi^* g - \psi^* \phi^* y] = [\psi^* g] = [h]$ as desired. □

We next wish to eliminate the need for the special groups H_0^p. If $\mathcal{U} = \{U_j\}_{j \in J}$ is an open covering of X and $\mathcal{V} = \{V_i\}_{i \in I}$ is another open covering that refines $\mathcal{U}$, let $\sigma : I \to J$ be an affinity function for these coverings. Then, as above, σ induces a homomorphism

$$\sigma^* : H_0^p(\mathcal{U}, \mathcal{H}) \to H_0^p(\mathcal{V}, \mathcal{H})$$

(where we continue to use the notation of the snake lemma).

Lemma 13.2.18. *Assume that X is a paracompact manifold. Let $\mathcal{U} = \{U_j\}_{j \in J}$ be an open covering of X and $g \in C^p(\mathcal{U}, \mathcal{H})$. Then there is a refinement $\mathcal{V} = \{V_i\}_{i \in I}$ of $\mathcal{U}$ and an affinity function $\sigma : I \to J$ such that $\sigma^* g \in C_0^p(\mathcal{V}, \mathcal{H})$.*

Proof. By passing to a refinement, we may suppose that $\mathcal{U}$ is locally finite. Also we may choose open sets $\{W_j\}_{j \in J}$ such that $\cup_j W_j = X$ and $\overline{W}_j \subseteq U_j$, each j. Now we claim that each $x \in X$ has an open neighborhood V_x satisfying the following conditions:

(13.2.9) If $s \in J^{p+1}$ and $x \in U_{s_0} \cap \cdots \cap U_{s_p} \equiv U_s$, then $V_x \subseteq U_s$ and if there is a $\mu \in \Gamma(V_x, \mathcal{G})$ such that $\psi \circ \mu = g$;
(13.2.10) If $x \in W_j$ then $V_x \subseteq W_j$;
(13.2.11) If $V_x \cap W_j \neq \emptyset$ then $V_x \subseteq U_j$.

To see that such V_x exist, note that the fact that the manifold is normal and that $\mathcal{U}$ is locally finite makes all the assertions but the second part of **(13.2.9)** trivial. But the latter follows from part **(13.2.2)** of the definition of sheaf. For each $x \in X$, let $\rho(x) \in J$ be such that $x \in W_{\rho(x)}$. By **(13.2.10)**, we know that $V_x \subseteq W_{\rho(x)}$. If $x_0, \ldots, x_p \in X$ and $V_{x_0} \cap \cdots \cap V_{x_p} \neq \emptyset$, then $V_{x_0} \cap W_{\rho(x_k)} \neq \emptyset$ for all $k = 0, \ldots, p$. Hence, by **(13.2.11)**, $V_{x_0} \subseteq U_{\rho(x_0)} \cap \cdots \cap U_{\rho(x_p)}$. It follows from **(13.2.9)** that $g_{\rho(x_0) \cdots \rho(x_p)} \in \Gamma(U_{x_0} \cap \cdots \cap U_{x_p}, \mathcal{G}) \subseteq \Gamma(V_{x_0} \cap \cdots \cap v_{x_p}, \mathcal{H})$ satisfies $g_{\rho(x_0) \cdots \rho(x_p)} = \psi \circ \mu$ for some $\mu \in \Gamma(V_{x_0} \cap \cdots V_{x_p}, \mathcal{G})$. Now let $\mathcal{V} = \{V_i\}$

be a locally finite refinement of the covering $\{V_x\}_{x\in X}$ of X. If $\sigma : I \to J$ is an affinity function such that $x \in V_i$ implies $\rho(x) = \sigma(i)$ then, by definition, $\sigma^* g \in C_0^p(\mathcal{V}, \mathcal{H})$. □

Corollary 13.2.19. *If X is a paracompact manifold, then the canonical map λ of the direct limit of $H_0^p(\mathcal{U}, \mathcal{H})$ into $H^p(X, \mathcal{H})$ is a surjective isomorphism.*

Proof. Denote the direct limit of $H_0^p(\mathcal{U}, \mathcal{H})$ by $H_0^p(X, \mathcal{H})$. The surjectivity is immediate from Lemma 13.2.18. For the injectivity, it suffices to observe that, if $\mathcal{U}$ is an open covering of X, then

$$C_0^p(\mathcal{U}, \mathcal{H}) \cap B^p(\mathcal{U}, \mathcal{H}) = B_0^p(\mathcal{U}, \mathcal{H}). \quad \square$$

Theorem 13.2.20. *If X is a paracompact manifold and $0 \to \mathcal{F} \to \mathcal{G} \to \mathcal{H} \to 0$ is an exact sequence of sheaves of abelian groups over X, then there is a long exact cohomology sequence*

$$\begin{aligned} 0 \to H^0(X, \mathcal{F}) \to H^0(X, \mathcal{G}) \to H^0(X, \mathcal{H}) \to H^1(X, \mathcal{F}) \\ \to H^1(X, \mathcal{G}) \to H^1(X, \mathcal{H}) \to \cdots . \end{aligned}$$

Proof. Combine Theorem 13.2.17 and Corollary 13.2.19. □

Remark: Notice that in Theorem 13.2.17 we did not specify what the maps are in the exact sequence. In applications of exact sequences, it does not matter what the specific maps are. What is important is the relationship specified by the exactness of the sequence.

Exercise for the Reader: Apply Theorem 13.2.20 to the sheaves that we discussed in Example 13.2.15. What conclusions can you draw about the exponential map? ♢

Exercise for the Reader: Give an example of a first Cousin problem on a Riemann surface that cannot be solved. To what extent does $H^1(\Omega, \mathcal{O})$ measure the number of unsolvable Cousin problems? ♢

Problems for Study and Exploration

1. Explain why the sheaf of germs of holomorphic functions is not fine.
2. Explain why a C^k function (or a holomorphic function, or a real analytic function) on a domain Ω is nothing other than a section of the appropriate sheaf.
3. Let $\Omega \subseteq \mathbb{C}$ be a domain and $M \subseteq \Omega$ a discrete set. Define a sheaf over Ω as follows: If $z \in \Omega \setminus M$ then let the stalk over z be the germs of *all* nonvanishing holomorphic functions. If $m \in M$ then the stalk over m is the ideal generated by the germ of $(z - m)$. The resulting sheaf is called the *ideal sheaf* of M. What is a suitable topology on M? Give a classical interpretation of what a global section of this sheaf would be.

4. Let $\mathcal{S}$ and $\mathcal{T}$ be sheaves of rings over a domain $\Omega \subseteq \mathbb{C}$. Define $\mathcal{S} \oplus \mathcal{T}$. Define $\mathcal{S} \otimes \mathcal{T}$.

5. Let $\mathcal{A}$ be the annulus $\{\zeta \in \mathbb{C} : 1 < |\zeta| < 2\}$. Consider the cohomology of $\mathcal{A}$ with coefficients in the sheaf of germs of holomorphic functions. Which cohomology class is the obstruction to calculating logarithms of holomorphic functions on $\mathcal{A}$? How does your answer change—if at all—when $\mathcal{A}$ is replaced by $D = \{\zeta \in \mathbb{C} : |\zeta| < 1\}$?

6. Let $\mathcal{S}$ be the sheaf of germs of holomorphic functions over the domain $\Omega_1 \subseteq \mathbb{C}$ and let $\mathcal{T}$ be the sheaf of germs of holomorphic functions over the domain $\Omega_2 \subseteq \mathbb{C}$. How is the sheaf $\mathcal{S} \times \mathcal{T}$ related to the sheaf of germs of holomorphic functions over $\Omega_1 \times \Omega_2$? [*Hint:* A function $F(z_1, z_2)$ is said to be holomorphic if it is holomorphic in each variable separately. Equivalently, it is holomorphic if it has a convergent power series expansion $\sum_{j,k} a_{j,k}(z_1 - p_1)^j(z_2 - p_2)^k$ about each point (p_1, p_2) of the domain.]

7. Explain the concept of analytic continuation—particularly the monodromy theorem—in the language of the sheaf of germs of holomorphic functions.

8. Using Exercise 7 as inspiration, define a Riemann surface using the language of sheaves.

9. Formulate the classical Weierstrass theorem, as discussed in the text, in the language of sheaf cohomology. Now prove it.

10. Formulate the classical Mittag–Leffler theorem, as discussed in the text, in the language of sheaf cohomology. Now prove it.

11. Formulate a version of the classical Mittag–Leffler theorem for compact Riemann surfaces. Use the language of sheaf cohomology. Now prove it.

Bibliography

[**AHL1**] L. Ahlfors, An extension of Schwarz's lemma, *Trans. Amer. Math. Soc.* 43 (1938), 359–364.

[**AHL2**] L. Ahlfors, *Complex Analysis*, 3rd ed., McGraw-Hill, New York, 1979.

[**AHL3**] L. Ahlfors, *Conformal Invariants*, McGraw-Hill, New York, 1973.

[**ALK**] G. Aladro and S. G. Krantz, A criterion for normality in $\mathbb{C}^n$, *J. Math. Anal. Appl.* 161(1991), 1–8.

[**APF**] L. Apfel, Localization properties and boundary behavior of the Bergman kernel, thesis, Washington University, 2003.

[**BAS**] F. Bagemihl and W. Seidel, Some boundary properties of analytic functions, *Math. Zeitschr.* 61(1954), 186–199.

[**BED**] E. Bedford and J. Dadok, Bounded domains with prescribed group of automorphisms, *Comment. Math. Helv.* 62(1987), 561–572.

[**BEK**] S. R. Bell and S. G. Krantz, Smoothness to the boundary of conformal mappings, *Rocky Mountain J. Math.* 17(1987), 23–40.

[**BEG**] C. Berenstein and R. Gay, *Complex Analysis and Special Topics in Harmonic Analysis*, Springer-Verlag, 1995.

[**BELL**] S. Bell and E. Ligocka, A simplification and extension of Fefferman's theorem on biholomorphic mappings, *Invent. Math.* 57(1980), 283–289.

[$\mathbf{BEL}_1$] J. Bergh and J. Löfström, *Interpolation Spaces: An Introduction*, Springer-Verlag, New York, 1973.

[$\mathbf{BEL}_2$] S. R. Bell, Ahlfors maps, the double of a domain, and complexity in potential theory and conformal mapping, *Journal d'Analyse Mathematique* 78(1999), 329–344.

[**BER**] S. Bergman, *The Kernel Function and Conformal Mapping*, Am. Math. Soc., Providence, RI, 1970.

[**BJS**] L. Bers, F. John, and M. Schechter, *Partial Differential Equations*, Interscience, New York, 1964.

[**BLK**] B. Blank and S. G. Krantz, *Calculus and Analytic Geometry*, Key Curriculum Press, Emeryville, CA, 2005.

[**BOO**] W. Boothby, *An Introduction to Differentiable Manifolds and Riemannian Geometry*, 2nd ed., Academic Press, New York, 2003.

[**BRE**] G. Bredon, *Introduction to Compact Transformation Groups*, Academic Press, New York, 1972.

[**BRO**] F. Browder, *Mathematical Developments Arising from the Hilbert Problems*, American Mathematical Society, Providence, RI, 1976.

[**BUK**] D. Burns and S. G. Krantz, Rigidity of holomorphic mappings and a new Schwarz lemma at the boundary, *Jour. of the A.M.S.* 7(1994), 661–676.

[**CAL**] A. P. Calderón, Intermediate spaces and interpolation, the complex method, *Studia Math.* 24(1964), 113–190.

[**CAR**] L. Carleson, Interpolation by bounded analytic functions and the corona problem, *Ann. of Math.* 76(1962), 547–559.

[**COL**] E. F. Collingwood and A. J. Lohwater, *Cluster Set Theory*, Cambridge University Press, Cambridge, 1966.

[**CON**] J. B. Conway, *Functions of One Complex Variable II*, Springer-Verlag, New York, 1995.

[**DUR**] R. Durrett, *Brownian motion and martingales in analysis*, Wadsworth International Group, Belmont, CA 1984.

[**EAH**] C. Earle and R. Hamilton, A fixed point theorem for holomorphic mappings, *Proc. Symp. Pure Math.*, Vol. XVI, 1968, 61–65.

[**EPS**] B. Epstein, *Orthogonal Families of Analytic Functions*, Macmillan, New York, 1965.

[**FAK**] H. Farkas and I. Kra, *Riemann Surfaces*, 2^{nd} ed., Springer-Verlag, New York, 1992.

[**FED**] H. Federer, *Geometric Measure Theory*, Springer-Verlag, New York and Berlin, 1969.

[**FIS**] S. Fisher, *Function Theory on Planar Domains*, John Wiley & Sons, New York, 1983.

[**FKKM**] B. Fridman, K.-T. Kim, S. G. Krantz, and D. Ma, On fixed points and determining sets for automorphisms, *Michigan Journal of Mathematics* 50(2002), 507–515.

[**FUKS**] B. A. Fuks, *Functions of a Complex Variable and Some of Their Applications*, Pergamon Press, London, 1961.

[**GAM**] T. Gamelin, *Complex Analysis*, Springer-Verlag, New York, 2001.

[**GAR**] J. B. Garnett, *Bounded Analytic Functions*, Academic Press, New York, 1981.

[**GARM**] J. B. Garnett and D. Marshall, *Harmonic Measure*, Cambridge University Press, Cambridge, 2005.

[**GOL**] Goluzin, *Geometric Theory of Functions of a Complex Variable*, American Mathematical Society, Providence, 1969.

[**GRK1**] R. E. Greene and S. G. Krantz, *Function Theory of One Complex Variable*, American Mathematical Society, Providence, RI, 2002.

[**GKM**] D. Gromoll, W. Klingenberg, and W. Meyer, Riemannsche Geometrie im Grossen, 2^{nd} ed., Lecture Notes in Mathematics, v. 55, Springer-Verlag, New York, 1975.

[**GUN1**] R. C. Gunning, *Riemann Surfaces*, Princeton University Press, Princeton, NJ, 1966.

[**GUN2**] R. C. Gunning, *Lectures on Vector Bundles over Riemann Surfaces*, Princeton University Press, Princeton, NJ, 1967.

[**HEI1**] M. Heins, A note on a theorem of Radó concerning the $(1, m)$ conformal maps of a multiply-connected region into itself, *Bull. Am. Math. Soc.* 47(1941), 128–130.

[**HEI2**] M. Heins, On the number of 1–1 directly conformal maps which a multiply-connected plane region of finite connectivity p (> 2) admits onto itself, *Bull. Am. Math. Soc.* 52(1946), 454–457.

[**HEL**] S. Helgason, *Differential Geometry and Symmetric Spaces*, Academic Press, New York, 1962.

[**HES**] Z. He and O. Schramm, Koebe uniformization and circle packing, *Ann. of Math.* 137(1993), 369–4013.

[**HIL**] E. Hille, *Analytic Function Theory*, 2^{nd} ed., Ginn and Co., Boston, 1973.

[**HIR**] M. Hirsch, *Differential Topology*, Springer-Verlag, New York, 1976.

[**HOF**] K. Hoffman, *Banach Spaces of Analytic Functions*, Prentice-Hall, Englewood Cliffs, NJ, 1962.

[**HUS**] D. Husemoller, *Fibre Bundles*, 2^{nd} ed., Springer-Verlag, New York, 1975.

[**JON**] F. John and L. Nirenberg, On functions of bounded mean oscillation, *Comm. Pure and Appl. Math.* 14(1961), 415–4213.

[**KAK**] S. Kakutani, Two-dimensional Brownian motion and harmonic functions, *Proc. Imp. Acad. Tokyo* 20(1944), 706–714.

[**KAT**] Y. Katznelson, *An Introduction to Harmonic Analysis*, John Wiley & Sons, New York, 1968.

[**KLI**] W. Klingenberg, *Riemannian Geometry*, 2^{nd} ed., de Gruyter Studies in Mathematics, Berlin, 1995.

[**KOB**] S. Kobayashi, *Hyperbolic Manifolds and Holomorphic Mappings*, Marcel Dekker, New York, 1970.

[**KON**] S. Kobayashi and Nomizu, *Foundations of Differential Geometry*, Vols. I and II, Wiley-Interscience, New York, 1963, 1969.

[**KOO**] P. Koosis, *Introduction to H_p Spaces*, 2^{nd} ed., Cambridge Tracts in Mathematics, v. 115, Cambridge University Press, Cambridge, 1998.

[**KRA1**] S. G. Krantz, *Function Theory of Several Complex Variables*, reprint of the 2^{nd} ed., American Mathematical Society, Providence, RI, 2001.

[**KRA2**] S. G. Krantz, *Partial Differential Equations and Complex Analysis*, CRC Press, Boca Raton, FL, 1992.

[**KRA3**] S. G. Krantz, *Real Analysis and Foundations*, CRC Press, Boca Raton, FL, 1992.

[**KRA4**] S. G. Krantz, *A Panorama of Harmonic Analysis*, Mathematical Association of America, Washington, D.C., 1999.

[**KRA5**] S. G. Krantz, *Complex Analysis: The Geometric Viewpoint*, A CARUS Monograph of the Mathematics Association of America, Washington, D.C., 1990.

[**KRA6**] S. G. Krantz, A compactness principle in complex analysis, *Division de Matematicas, Univ. Autonoma de Madrid Seminarios*, vol. 3, 1987, 171–194.

[**KRA7**] S. G. Krantz, Calculation and estimation of the Poisson kernel, *Jour. Math. Anal. and Applic.* 302(2005), 143–148.

[**KRP1**] S. G. Krantz and H. R. Parks, *A Primer of Real Analytic Functions*, 2^{nd} ed., Birkhäuser Boston, Boston, 2002.

[**LEV**] O. Lehto and Virtanen, Boundary behavior and normal meromorphic functions, *Acta Math.* 97(1957), 47–65.

[**LOS**] L. Loomis and S. Sternberg, *Advanced Calculus*, Addison-Wesley, Reading, MA, 1968.

[**LUR**] D. H. Luecking and L. A. Rubel, *Complex Analysis: A Functional Analysis Approach*, Springer-Verlag, New York, 1984.

[**LYP**] J. Lyons and J. Peetre, Sur une classe d'espaces d'interpolation, *Inst. Hautes Études Sci. Publ. Math.*, No. 19, 1964, 5–68.

[**MAR**] Markushevitch, *The Theory of Analytic Functions*, Mir Press, Moscow, 1983.

[**MIS**] D. Minda and G. Schober, Another elementary approach to the theorems of Landau, Montel, Picard and Schottky, *Complex Variables,* 2 (1983), 157–164.

[**MOR**] C. B. Morrey, *Multiple Integrals in the Calculus of Variations*, Springer, Berlin, 19613.

[**MUN**] J. R. Munkres, *Elementary Differential Topology*, Princeton University Press, Princeton, NJ, 1966.

[**LEF1**] S. Lefschetz, *Introduction to Topology*, Princeton University Press, Princeton, 1949.

[**LEF2**] S. Lefschetz, *Topology*, American Mathematical Society, Providence, 1930.

[**NEL**] E. Nelson, *Dynamical Theories of Brownian Motion*, Princeton University Press, Princeton, NJ, 1967.

[**OHS**] T. Ohsawa, A remark on the completeness of the Bergman metric, *Proc. Japan Acad. Ser. A Math. Sci.* 57(1981), 238–240.

[**OSS**] R. Osserman, A sharp Schwarz inequality on the boundary, *Proc. Amer. Math. Soc.* 128(2000), 3513–3517.

[**PAI**] P. Painlevé, thesis, Paris, 1887.

[**PRI**] I. Privalov, *Randeigenschaften Analytischer Funktionen*, Deutsch Verlag der Wissenschaften, Berlin, 1956.

[**REM**] R. Remmert, *Classical Topics in Complex Function Theory*, Springer-Verlag, New York, 1998.

[**RIC**] C. E. Rickart, *General Theory of Banach Algebras*, van Nostrand, Princeton, NJ, 1960.

[**ROB**] A. Robinson, Metamathematical problems, *J. Symbolic Logic* 38(1973), 500–5113.

[**RUD1**] W. Rudin, *Principles of Mathematical Analysis*, 3rd ed., McGraw-Hill, New York, 19713.

[**RUD2**] W. Rudin, *Real and Complex Analysis*, McGraw-Hill, New York, 1963.

[**RUD3**] W. Rudin, *Functional Analysis*, 2nd ed., McGraw-Hill, New York, 1991.

[**SAZ**] R. Saerens and W. Zame, The isometry groups of manifolds and the automorphism groups of domains, *Trans. Amer. Math. Soc.* 301(1987), 413–429.

[**SCH**] L. Schwartz, *Théorie des Distributions*, Hermann, Paris, 1951.

[**SIC**] J. Siciak, On series of homogeneous polynomials and their partial sums, *Annales Polonici Math.* 51(1990), 289–302.

[**SPA**] E. Spanier, *Algebraic Topology*, Springer-Verlag, New York, 1981.

[**STE**] E. M. Stein, *Harmonic Analysis*, Princeton University Press, Princeton, NJ, 1993.

[**STO**] Stoker, *Differential Geometry*, Wiley, New York, 1969.

[**STW**] E. M. Stein and G. Weiss, *Introduction to Fourier Analysis on Euclidean Spaces*, Princeton University Press, Princeton, NJ, 1971.

[**THO**] G. B. Thomas, *Calculus with Analytic Geometry*, 7th ed., Addison-Wesley, Reading, MA, 1988.

[**VEL**] J. A. Velling, thesis, Stanford University, 1985.

[**WOL**] J. Wolf, *Spaces of Constant Curvature*, 5th ed., Publish or Perish Press, Wilmington, DE, 1984.

[**WU**] H. H. Wu, Normal families of holomorphic mappings, *Acta Math.* 119(1967), 193–233.

[**YAU**] S. T. Yau, A generalized Schwarz lemma for Kähler manifolds, *Am. J. Math.* 100(1978), 197–204.

[**ZAL**] L. Zalcman, A heuristic principle in complex analysis, *Amer. Math. Monthly* 82(1975), 813–817.

Index