Graduate Texts in Mathematics

John B. Conway

Functions of One Complex Variable II

Springer-Verlag

John B. Conway

Functions of One Complex Variable II

With 15 Illustrations

Springer-Verlag
New York Berlin Heidelberg London Paris
Tokyo Hong Kong Barcelona Budapest

John B. Conway
Department of Mathematics
University of Tennessee
Knoxville, TN 37996-1300
USA

Mathematics Subjects Classifications (1991): 03-01, 31A05, 31A15

Library of Congress Cataloging-in-Publication Data
Conway, John B.
Functions of one complex variable II / John B. Conway.
p. cm. — (Graduate texts in mathematics ; 159)
Includes bibliographical references (p. —) and index.
ISBN 0-387-94460-5 (hardcover : acid-free)
1. Functions of complex variables. I. Title. II. Title: Functions of one complex variable 2. III. Title: Functions of one complex variable two. IV. Series.
QA331.7.C365 1995
515'.93—dc20 95-2331

Printed on acid-free paper.

Production managed by Laura Carlson; manufacturing supervised by Jeffrey Taub.
Photocomposed pages prepared from the author's LATEX file.
Printed and bound by R.R. Donnelley & Sons, Harrisonburg, VA.
Printed in the United States of America.

9 8 7 6 5 4 3 2 1

ISBN 0-387-94460-5 Springer-Verlag New York Berlin Heidelberg

To The Memory of my Parents

Cecile Marie Boudreaux

and

Edward Daire Conway, Jr.

Preface

This is the sequel to my book *Functions of One Complex Variable I*, and probably a good opportunity to express my appreciation to the mathematical community for its reception of that work. In retrospect, writing that book was a crazy venture.

As a graduate student I had had one of the worst learning experiences of my career when I took complex analysis; a truly bad teacher. As a non-tenured assistant professor, the department allowed me to teach the graduate course in complex analysis. They thought I knew the material; I wanted to learn it. I adopted a standard text and shortly after beginning to prepare my lectures I became dissatisfied. All the books in print had virtues; but I was educated as a modern analyst, not a classical one, and they failed to satisfy me.

This set a pattern for me in learning new mathematics after I had become a mathematician. Some topics I found satisfactorily treated in some sources; some I read in many books and then recast in my own style. There is also the matter of philosophy and point of view. Going from a certain mathematical vantage point to another is thought by many as being independent of the path; certainly true if your only objective is getting there. But getting there is often half the fun and often there is twice the value in the journey if the path is properly chosen.

One thing led to another and I started to put notes together that formed chapters and these evolved into a book. This now impresses me as crazy partly because I would never advise any non-tenured faculty member to begin such a project; I have, in fact, discouraged some from doing it. On the other hand writing that book gave me immense satisfaction and its reception, which has exceeded my grandest expectations, makes that decision to write a book seem like the wisest I ever made. Perhaps I lucked out by being born when I was and finding myself without tenure in a time (and possibly a place) when junior faculty were given a lot of leeway and allowed to develop at a slower pace—something that someone with my background and temperament needed. It saddens me that such opportunities to develop are not so abundant today.

The topics in this volume are some of the parts of analytic function theory that I have found either useful for my work in operator theory or enjoyable in themselves; usually both. Many also fall into the category of topics that I have found difficult to dig out of the literature.

I have some difficulties with the presentation of certain topics in the literature. This last statement may reveal more about me than about the state of the literature, but certain notions have always disturbed me even though experts in classical function theory take them in stride. The best example of this is the concept of a multiple-valued function. I know there are ways to make the idea rigorous, but I usually find that with a little work it isn't necessary to even bring it up. Also the term multiple-valued function violates primordial instincts acquired in childhood where I was sternly taught that functions, by definition, cannot be multiple-valued.

The first volume was not written with the prospect of a second volume to follow. The reader will discover some topics that are redone here with more generality and originally could have been done at the same level of sophistication if the second volume had been envisioned at that time. But I have always thought that introductions should be kept unsophisticated. The first white wine would best be a Vouvray rather than a Chassagne-Montrachet.

This volume is divided into two parts. The first part, consisting of Chapters 13 through 17, requires only what was learned in the first twelve chapters that make up Volume I. The reader of this material will notice, however, that this is not strictly true. Some basic parts of analysis, such as the Cauchy-Schwarz Inequality, are used without apology. Sometimes results whose proofs require more sophisticated analysis are stated and their proofs are postponed to the second half. Occasionally a proof is given that requires a bit more than Volume I and its advanced calculus prerequisite. The rest of the book assumes a complete understanding of measure and integration theory and a rather strong background in functional analysis.

Chapter 13 gathers together a few ideas that are needed later. Chapter 14, "Conformal Equivalence for Simply Connected Regions," begins with a study of prime ends and uses this to discuss boundary values of Riemann maps from the disk to a simply connected region. There are more direct ways to get to boundary values, but I find the theory of prime ends rich in mathematics. The chapter concludes with the Area Theorem and a study of the set $\mathcal{S}$ of schlicht functions.

Chapter 15 studies conformal equivalence for finitely connected regions. I have avoided the usual extremal arguments and relied instead on the method of finding the mapping functions by solving systems of linear equations. Chapter 16 treats analytic covering maps. This is an elegant topic that deserves wider understanding. It is also important for a study of Hardy spaces of arbitrary regions, a topic I originally intended to include in this volume but one that will have to await the advent of an additional volume.

Chapter 17, the last in the first part, gives a relatively self contained treatment of de Branges's proof of the Bieberbach conjecture. I follow the approach given by Fitzgerald and Pommerenke [1985]. It is self contained except for some facts about Legendre polynomials, which are stated and explained but not proved. Special thanks are owed to Steve Wright and Dov Aharonov for sharing their unpublished notes on de Branges's proof of the Bieberbach conjecture.

Chapter 18 begins the material that assumes a knowledge of measure theory and functional analysis. More information about Banach spaces is used here than the reader usually sees in a course that supplements the standard measure and integration course given in the first year of graduate study in an American university. When necessary, a reference will be given to Conway [1990]. This chapter covers a variety of topics that are used in the remainder of the book. It starts with the basics of Bergman spaces, some material about distributions, and a discourse on the Cauchy transform and an application of this to get another proof of Runge's Theorem. It concludes with an introduction to Fourier series.

Chapter 19 contains a rather complete exposition of harmonic functions on the plane. It covers about all you can do without discussing capacity, which is taken up in Chapter 21. The material on harmonic functions from Chapter 10 in Volume I is assumed, though there is a built in review.

Chapter 20 is a rather standard treatment of Hardy spaces on the disk, though there are a few surprising nuggets here even for some experts.

Chapter 21 discusses some topics from potential theory in the plane. It explores logarithmic capacity and its relationship with harmonic measure and removable singularities for various spaces of harmonic and analytic functions. The fine topology and thinness are discussed and Wiener's criterion for regularity of boundary points in the solution of the Dirichlet problem is proved.

This book has taken a long time to write. I've received a lot of assistance along the way. Parts of this book were first presented in a pubescent stage to a seminar I presented at Indiana University in 1981-82. In the seminar were Greg Adams, Kevin Clancey, Sandy Grabiner, Paul McGuire, Marc Raphael, and Bhushan Wadhwa, who made many suggestions as the year progressed. With such an audience, how could the material help but improve. Parts were also used in a course and a summer seminar at the University of Tennessee in 1992, where Jim Dudziak, Michael Gilbert, Beth Long, Jeff Nichols, and Jeff vanEeuwen pointed out several corrections and improvements. Nathan Feldman was also part of that seminar and besides corrections gave me several good exercises. Toward the end of the writing process I mailed the penultimate draft to some friends who read several chapters. Here Paul McGuire, Bill Ross, and Liming Yang were of great help. Finally, special thanks go to David Minda for a very careful reading of several chapters with many suggestions for additional references and exercises.

On the technical side, Stephanie Stacy and Shona Wolfenbarger worked diligently to convert the manuscript to TEX. Jinshui Qin drew the figures in the book. My son, Bligh, gave me help with the index and the bibliography.

In the final analysis the responsibility for the book is mine.

John B Conway
University of Tennessee

Contents of Volume II

Contents of Volume I

Chapter 13

Return to Basics

In this chapter a few results of a somewhat elementary nature are collected. These will be used quite often in the remainder of this volume.

§1 Regions and Curves

In this first section a few definitions and facts about regions and curves in the plane are given. Some of these may be familiar to the reader. Indeed, some will be recollections from the first volume.

Begin by recalling that a region is an open connected set and a simply connected region is one for which every closed curve is contractible to a point (see 4.6.14). In Theorem 8.2.2 numerous statements equivalent to simple connectedness were given. We begin by recalling one of these equivalent statements and giving another. Do not forget that $\mathbb{C}_\infty$ denotes the extended complex numbers and $\partial_\infty G$ denotes the boundary of the set G in $\mathbb{C}_\infty$. That is, $\partial_\infty G = \partial G$ when G is bounded and $\partial_\infty G = \partial G \cup \{\infty\}$ when G is unbounded.

It is often convenient to give results about subsets of the extended plane rather than about $\mathbb{C}$. If something was proved in the first volume for a subset of $\mathbb{C}$, but it holds for subsets of $\mathbb{C}_\infty$ with little change in the proof, we will not hesitate to quote the appropriate reference from the first twelve chapters as though the result for $\mathbb{C}_\infty$ was proved there.

1.1 Proposition. *If G is a region in $\mathbb{C}_\infty$, the following statements are equivalent.*

(a) *G is simply connected.*

(b) *$\mathbb{C}_\infty \setminus G$ is connected*

(c) *$\partial_\infty G$ is connected.*

Proof. The equivalence of (a) and (b) has already been established in (8.2.2). In fact, the equivalence of (a) and (b) was established without assuming that G is connected. That is, it was only assumed that G was a simply connected open set; an open set with every component simply connected. The reader must also pay attention to the fact that the connectedness of G will not be used when it is shown that (c) implies (b). This will be used when it is shown that (b) implies (c).

So assume (c) and let us prove (b). Let F be a component of $\mathbb{C}_\infty \setminus G$; so F is closed. It follows that $F \cap \text{cl}\, G \neq \emptyset$ (cl denotes the closure operation in $\mathbb{C}$ while cl_∞ denotes the closure in the extended plane.) Indeed, if it were the case that $F \cap \text{cl}\, G = \emptyset$, then for every z in F there is an $\varepsilon > 0$ such that $B(z;\varepsilon) \cap G = \emptyset$. Thus $F \cup B(z;\varepsilon) \subseteq \mathbb{C}_\infty \setminus G$. But $F \cup B(z;\varepsilon)$ is connected. Since F is a component of $\mathbb{C}_\infty \setminus G$, $B(z;\varepsilon) \subseteq F$. Since z was an arbitrary point, this implies that F is an open set, giving a contradiction. Therefore $F \cap \text{cl}\, G \neq \emptyset$.

Let $z_0 \in F \cap \text{cl}\, G$; so $z_0 \in \partial_\infty G$. By (c) $\partial_\infty G$ is connected, so $F \cup \partial_\infty G$ is a connected set that is disjoint from G. Therefore $\partial_\infty G \subseteq F$ since F is a component of $\mathbb{C}_\infty \setminus G$. What we have just shown is that every component of $\mathbb{C}_\infty \setminus G$ must contain $\partial_\infty G$. Hence there can be only one component and so $\mathbb{C}_\infty \setminus G$ is connected.

Now assume that condition (b) holds. So far we have not used the fact that G is connected; now we will. Let $U = \mathbb{C}_\infty \setminus \text{cl}_\infty G$. Now $\mathbb{C}_\infty \setminus U = \text{cl}_\infty G$ and $\text{cl}_\infty G$ is connected. Since we already have that (a) and (b) are equivalent (even for non-connected open sets), U is simply connected. Thus $\mathbb{C}_\infty \setminus \partial_\infty G = G \cup U$ is the union of two disjoint simply connected sets and hence must be simply connected. Since (a) implies (b), $\partial_\infty G = \mathbb{C}_\infty \setminus (G \cup U)$ is connected. □

1.2 Corollary. *If G is a region in $\mathbb{C}$, then the map $F \to F \cap \partial_\infty G$ defines a bijection between the components of $\mathbb{C}_\infty \setminus G$ and the components of $\partial_\infty G$.*

Proof. If F is a component of $\mathbb{C}_\infty \setminus G$, then an argument that appeared in the preceding proof shows that $F \cap \partial_\infty G \neq \emptyset$. Also, since $\partial_\infty G \subseteq \mathbb{C}_\infty \setminus G$, any component C of $\partial_\infty G$ that meets F must be contained in F. It must be shown that two distinct components of $\partial_\infty G$ cannot be contained in F.

To this end, let $G_1 = \mathbb{C}_\infty \setminus F$. Since G_1 is the union of G and the components of $C_\infty \setminus G$ that are distinct from F, G_1 is connected. Since $\mathbb{C}_\infty \setminus G_1 = F$, a connected set, G_1 is simply connected. By the preceding proposition, $\partial_\infty G_1$ is connected. Now $\partial_\infty G_1 \subseteq \partial_\infty G$. In fact for any point z in $\partial_\infty G_1$, $\emptyset \neq B(z;\varepsilon) \cap (\mathbb{C}_\infty \setminus G_1) \subseteq B(z;\varepsilon) \cap (\mathbb{C}_\infty \setminus G)$. Also if $B(z;\varepsilon) \cap G = \emptyset$, then $B(z;\varepsilon) \subseteq \mathbb{C}_\infty \setminus G$ and $B(z;\varepsilon) \cap F \neq \emptyset$; thus $z \in \text{int}\, F$, contradicting the fact that $z \in \partial_\infty G_1$. Thus $\partial_\infty G_1 \subseteq \partial_\infty G$. Therefore any component of $\partial_\infty G$ that meets F must contain $\partial_\infty G_1$. Hence there can be only one such component of $\partial_\infty G$. That is, $F \cap \partial_\infty G$ is a component of $\partial_\infty G$.

This establishes that the map $F \to F \cap \partial_\infty G$ defines a map from the components of $\mathbb{C}_\infty \setminus G$ to the components of $\partial_\infty G$. The proof that this correspondence is a bijection is left to the reader. □

Recall that a *simple closed curve* in $\mathbb{C}$ is a path $\gamma : [a,b] \to \mathbb{C}$ such that $\gamma(t) = \gamma(s)$ if and only if $t = s$ or $|s - t| = b - a$. Equivalently, a simple closed curve is the homeomorphic image of $\partial\mathbb{D}$. Another term for a simple closed curve is a *Jordan curve.* The Jordan Curve Theorem is given here, but a proof is beyond the purpose of this book. See Whyburn [1964].

1.3 Jordan Curve Theorem. *If γ is a simple closed curve in $\mathbb{C}$, then $\mathbb{C} \setminus \gamma$ has two components, each of which has γ as its boundary.*

Clearly one of the two components of $\mathbb{C} \setminus \gamma$ is bounded and the other is unbounded. Call the bounded component of $\mathbb{C} \setminus \gamma$ the *inside* of γ and call the unbounded component of $\mathbb{C} \setminus \gamma$ the *outside* of γ. Denote these two sets by ins γ and out γ, respectively.

Note that if γ is a rectifiable Jordan curve, so that the winding number $n(\gamma; a)$ is defined for all a in $\mathbb{C} \setminus \gamma$, then $n(\gamma; a) \equiv \pm 1$ for a in ins γ while $n(\gamma; a) \equiv 0$ for a in out γ. Say γ is *positively oriented* if $n(\gamma; a) = 1$ for all a in ins γ. A curve γ is *smooth* if γ is a continuously differentiable function and $\gamma'(t) \neq 0$ for all t. Say that γ is a *loop* if γ is a positively oriented smooth Jordan curve.

Here is a corollary of the Jordan Curve Theorem

1.4 Corollary. *If γ is a Jordan curve,* ins γ *and* $(\text{out}\,\gamma) \cup \{\infty\}$ *are simply connected regions.*

Proof. In fact, $\mathbb{C}_\infty \setminus \text{ins } \gamma = \text{cl}_\infty(\text{out } \gamma)$ and this is connected by the Jordan Curve Theorem. Thus ins γ is simply connected by Proposition 1.1. Similarly, out $\gamma \cup \{\infty\}$ is simply connected. □

A *positive Jordan system* is a collection $\Gamma = \{\gamma_1, \ldots, \gamma_m\}$ of pairwise disjoint rectifiable Jordan curves such that for all points a not on any γ_j, $n(\Gamma; a) \equiv \sum_{j=1}^m n(\gamma; a) = 0$ or 1. Let out $\Gamma \equiv \{a \in \mathbb{C} : n(\Gamma; a) = 0\}$ = the *outside* of Γ and let ins $\Gamma \equiv \{a \in \mathbb{C} : n(\Gamma; a) = 1\}$ = the *inside* of Γ. Thus $\mathbb{C} \setminus \Gamma = \text{out } \Gamma \cup \text{ins } \Gamma$. Say that Γ is smooth if each curve γ_j in Γ is smooth.

Note that it is not assumed that ins Γ is connected and if Γ has more than one curve, out Γ is never connected. The boundary of an annulus is an example of a positive Jordan system if the curves on the boundary are given appropriate orientation. The boundary of the union of two disjoint closed annuli is also a positive Jordan system, as is the boundary of the union of two disjoint closed disks.

If X is any set in the plane and A and B are two non-empty sets, say that X *separates* A *from* B if A and B are contained in distinct components of the complement of X. The proof of the next result can be found on page 34 of Whyburn [1964].

1.5 Separation Theorem. *If K is a compact subset of the open set U, $a \in K$, and $b \in \mathbb{C}_\infty \setminus U$, then there is a Jordan curve γ in U such that γ is disjoint from K and γ separates a from b.*

In the preceding theorem it is not possible to get that the point a lies in ins γ. Consider the situation where U is the open annulus ann$(0; 1, 3)$,

$K = \{z; |z| = 3/2\}$, $a = 3/2$, and $b = 0$.

1.6 Corollary. *The curve γ in the Separation Theorem can be chosen to be smooth.*

Proof. Let $\Omega = \text{ins}\,\gamma$ and for the moment assume that $a \in \Omega$. The other case is left to the reader. Let $K_0 = K \cap \Omega$. Since $\gamma \cap K = \emptyset$, it follows that K_0 is a compact subset of Ω that contains a. Since Ω is simply connected, there is a Riemann map $\tau : \mathbb{D} \to \Omega$. By a compactness argument there is a radius r, $0 < r < 1$, such that $\tau(r\mathbb{D}) \supseteq K_0$. Since U is open and $\gamma \subseteq U$, r can be chosen so that $\tau(r\partial\mathbb{D}) \subseteq U$. Let σ be a parameterization of the circle $r\partial\mathbb{D}$ and consider the curve $\tau \circ \sigma$. Clearly $\tau \circ \sigma$ separates a from b, is disjoint from K, and lies inside U. □

Note that the proof of the preceding corollary actually shows that γ can be chosen to be an analytic curve. That is, γ can be chosen such that it is the image of the unit circle under a mapping that is analytic in a neighborhood of the circle. (See §4 below.)

1.7 Proposition. *If K is a compact connected subset of the open set U and b is a point in the complement of U, then there is a loop γ in U that separates K and b.*

Proof. Let $a \in K$ and use (1.6) to get a loop γ that separates a and b. Let Ω be the component of the complement of γ that contains a. Since $K \cap \Omega \neq \emptyset$, $K \cap \gamma = \emptyset$, and K is connected, it must be that $K \subseteq \Omega$. □

The next result is used often. A proof of this proposition can be given starting from Proposition 8.1.1. Actually Proposition 8.1.1 was not completely proved there since the statement that the line segments obtained in the proof form a finite number of closed polygons was never proved in detail. The details of this argument are combinatorially complicated. Basing the argument on the Separation Theorem obviates these complications.

1.8 Proposition. *If E is a compact subset of an open set G, then there is a smooth positively oriented Jordan system Γ contained in G such that $E \subseteq \text{ins}\,\Gamma \subseteq G$.*

Proof. Now G can be written as the increasing union on open sets G_n such that each G_n is bounded and $\mathbb{C} \setminus G_n$ has only a finite number of components (7.1.2). Thus it suffices to assume that G is bounded and $\mathbb{C}\setminus G$ has only a finite number of components, say $K_0, K_1, \ldots, K_n$ where K_0 is the unbounded component.

It is also sufficient to assume that G is connected. In fact if $U_1, U_2, \ldots$ are the components of G, then $\{U_m\}$ is an open cover of E. Hence there is a finite subcover. Thus for some integer m there are compact subsets E_k of U_k, $1 \leq k \leq m$, such that $E = \bigcup_1^m E_k$. If the proposition is proved under the additional assumption that G is connected, this implies there is a smooth positively oriented Jordan system Γ_k in U_k such that $E_k \subseteq \text{ins}\,\Gamma \subseteq U_k$; let $\Gamma = \bigcup_1^m \Gamma_k$. Note that since $\text{cl}\,(\text{ins}\,\Gamma_k) = \Gamma_k \cup \text{ins}\,\Gamma_k \subseteq U_k$, $\text{cl}\,(\text{ins}\,\Gamma_k) \cap \text{cl}\,(\text{ins}\,\Gamma_i) = \emptyset$ for $k \neq i$. Thus Γ is also a positively oriented smooth Jordan system in G and $E \subseteq \text{ins}\,\Gamma = \bigcup_1^m \text{ins}\,\Gamma_k \subseteq G$.

Let $\varepsilon > 0$ such that for $0 \leq j \leq n$, $(K_j)_\varepsilon \equiv \{z : \text{dist}(z, K_j) \leq \varepsilon\}$ is disjoint from E as well as the remainder of these inflated sets. Also pick a point a_0 in $\text{int}K_0$. By Proposition 1.7 for $1 \leq j \leq n$ there is a smooth Jordan curve γ_j in $\{z : \text{dist}(z, K_j) < \varepsilon\}$ that separates a_0 from K_j. Note that a_0 belongs to the unbounded component of the complement of $\{z : \text{dist}(z, K_j) < \varepsilon\}$. Thus $K_j \subseteq \text{ins}\,\gamma$; and $a_0 \in \text{out}\,\gamma_j$. Give γ_j a negative orientation so that $n(\gamma_j : z) = -1$ for all z in K_j.

Note that $U = \mathbb{C} \setminus K_0$ is a simply connected region since its complement in the extended plane, K_0, is connected. Let $\tau : \mathbb{D} \to U$ be a Riemann map. For some r, $0 < r < 1$, $V = \tau(r\mathbb{D})$ contains $E \cup \bigcup_1^n K_j$ and $\partial V \subseteq \text{int}(K_0)_\varepsilon$. Let $\gamma_0 = \partial V$ with positive orientation. Clearly $E \cup \bigcup_1^n K_j \subseteq \text{ins}\,\gamma_0$ and $a_0 \in \text{out}\,\gamma_0$.

It is not difficult to see that $\Gamma = \{\gamma_0, \gamma_1, \ldots, \gamma_n\}$ is a smooth Jordan system contained in G. If $z \in K_j$ for $1 \leq j \leq n$, then $n(\Gamma, z) = n(\gamma_j, z) + n(\gamma_0, z) = -1 + 1 = 0$. Now $a_0 \in \text{out}\,\Gamma$; but the fact that $\Gamma \subseteq G$ and K_0 is connected implies that $K_0 \subseteq \text{out}\,\Gamma$. It follows that $\text{ins}\,\Gamma \subseteq G$.

On the other hand, if $z \in E$, then $z \in \text{out}\,\gamma_j$ for $1 \leq j \leq n$ and $z \in \text{ins}\,\gamma_0$. Thus $E \subseteq \text{ins}\,\Gamma$. □

1.9 Corollary. *Suppose G is a bounded region and $K_0, \ldots, K_n$ are the components of $\mathbb{C}_\infty \setminus G$ with ∞ in K_0. If $\varepsilon > 0$, then there is a smooth Jordan system $\Gamma = \{\gamma_0, \ldots, \gamma_n\}$ in G such that:*

(a) *for $1 \leq j \leq n$, $K_j \subseteq \text{ins}\,\gamma_j$;*

(b) *$K_0 \subseteq \text{out}\,\gamma_0$;*

(c) *for $0 \leq j \leq n$, $\gamma_j \subseteq \{z : \text{dist}(z, K_j) < \varepsilon\}$.*

Proof. Exercise. □

1.10 Proposition. *An open set G in $\mathbb{C}$ is simply connected if and only if for every Jordan curve γ contained in G, $\text{ins}\,\gamma \subseteq G$.*

Proof. Assume that G is simply connected and γ is a Jordan curve in G. So $\mathbb{C}_\infty \setminus G$ is connected, contains ∞, and is contained in $\mathbb{C}_\infty \setminus \gamma$. Therefore the Jordan Curve Theorem implies that $\mathbb{C} \setminus G \subseteq \text{out}\,\gamma$. Hence, $\text{cl}\,(\text{ins}\,\gamma) = \mathbb{C} \setminus \text{out}\,\gamma \subseteq G$.

Now assume that G contains the inside of any Jordan curve that lies in G. Let σ be any closed curve in G; it must be shown that σ is homotopic to 0 in G. Let $\varepsilon > 0$ be chosen so that $(\sigma)_\varepsilon \subseteq G$ and pick a point b in the unbounded component of the complement of $(\sigma)_\varepsilon$. By Proposition 1.7 there is a Jordan curve γ in $\{z : \text{dist}(z, \sigma) < \varepsilon\}$ that separates the compact set σ and the point b. The unbounded component of the complement of $(\sigma)_\varepsilon$ must be contained in the outside of γ so that $b \in \text{out } \gamma$; thus $\sigma \subseteq \text{ins } \gamma$. But ins γ is simply connected (1.4) so that σ is homotopic to 0 in ins γ. But by assumption G contains ins γ so that σ is homotopic to 0 in G and G is simply connected. □

1.11 Corollary. *If γ and σ are Jordan curves with $\sigma \subseteq \text{cl}(\text{ins}\,\gamma)$, then* ins $\sigma \subseteq$ ins γ.

A good reference for the particular properties of planar sets is Newman [1964].

Exercises

1. Give a direct proof of Corollary 1.11 that does not depend on Proposition 1.10.

2. For any compact set E, show that E_ε has a finite number of components. If E is connected, show that E_ε is connected.

3. Show that a region G is simply connected if and only if every Jordan curve in G is homotopic to 0.

4. Prove Corollary 1.9.

5. This exercise seems appropriate at this point, even though it does not use the results from this section. The proof of this is similar to the proof of the Laurent expansion of a function with an isolated singularity. Using the notation of Corollary 1.9, show that if f is analytic in G, then $f = f_0 + f_1 + \cdots + f_n$, where f_j is analytic on $\mathbb{C}_\infty \setminus K_j$ $(0 \leq j \leq n)$ and $f_j(\infty) = 0$ for $1 \leq j \leq n$. Show that the functions are unique. Also show that if f is a bounded function, then each f_j is bounded.

§2 Derivatives and Other Recollections

In this section some notation is introduced that will be used in this book and some facts about derivatives and other matters will be recalled.

For any metric space X, let $C(X)$ denote the algebra of continuous functions from X into $\mathbb{C}$. If n is a natural number and G is an open subset of $\mathbb{C}$, let $C^n(G)$ denote the functions $f : G \to \mathbb{C}$ such that f has continuous partial derivatives up to and including the n-th order. $C^0(G) = C(G)$ and $C^\infty(G) =$ the infinitely differentiable functions on

G. If $0 \le n \le \infty$, $C_c^n(G)$ denotes those functions f in $C^n(G)$ with supp$f \equiv$ *support* of $f \equiv \text{cl}\,\{z \in G : f(z) \neq 0\}$ compact.

It is convenient to think of functions f defined on $\mathbb{C}$ as functions of the complex variables z and $\overline{z}$ rather than the real variables x and y. These two sets of variables are related by the formulas

$$\begin{aligned} z &= x + iy & \overline{z} &= x - iy \\ x &= \frac{z + \overline{z}}{2} & y &= \frac{z + \overline{z}}{2i}. \end{aligned}$$

Thus for a differentiable function f on an open set G, it is possible to discuss the derivatives of f with respect to z and $\overline{z}$. Namely, define

$$\begin{aligned} \partial f &= \frac{\partial f}{\partial z} = \frac{1}{2}\left(\frac{\partial f}{\partial x} - i\frac{\partial f}{\partial y}\right), \\ \overline{\partial} f &= \frac{\partial f}{\partial \overline{z}} = \frac{1}{2}\left(\frac{\partial f}{\partial x} - i\frac{\partial f}{\partial y}\right). \end{aligned}$$

These formulas can be justified by an application of the chain rule. A derivation of the formulas can be obtained by considering $dz = dx + idy$ and $d\overline{z} = dx - idy$ as a module basis for the complex differentials on G, expanding the differential of f, df, in terms of the basis, and observing that the formulas for ∂f and $\overline{\partial} f$ given above are the coefficients of dz and $d\overline{z}$, respectively.

The origin of this notation is the theory of functions of several complex variables, but it is very convenient even here. In particular, as an easy consequence of the Cauchy-Riemann equations, or rather a reformulation of the result that a function is analytic if and only if its real and imaginary parts satisfy the Cauchy-Riemann equations, we have the following.

2.1 Proposition. *A function $f : G \to \mathbb{C}$ is analytic if and only if $\overline{\partial} f = 0$.*

So the preceding proposition says that a function is analytic precisely when it is a function of z alone and not of $\overline{z}$.

With some effort (not to be done here) it can be shown that all the laws for calculating derivatives apply to ∂ and $\overline{\partial}$ as well. In particular, the rules for differentiating sums, products, and quotients as well as the chain rule are valid. The last is explicitly stated here and the proof is left to the reader.

2.2 Chain Rule. *Let G be an open subset of $\mathbb{C}$ and let $f \in C^1(G)$. If Ω is an open subset of $\mathbb{C}$ such that $f(G) \subseteq \Omega$ and $g \in C^1(\Omega)$, then $g \circ f \in C^1(G)$ and*

$$\begin{aligned} \partial(g \circ f) &= [(\partial g) \circ f]\, \partial f + \left[(\overline{\partial} g) \circ f\right] \partial \overline{f} \\ \overline{\partial}(g \circ f) &= [(\partial g) \circ f]\, \overline{\partial} f + \left[(\overline{\partial} g) \circ f\right] \overline{\partial f}. \end{aligned}$$

So if a formula for a function f can be written in terms of elementary functions of z and $\overline{z}$, then the rules of calculus can be applied to calculate the derivatives of f to any order. The next result contains such a calculation.

2.3 Proposition.

(a) $\partial(\log|z|) = 1\{2z\}$ *and* $\overline{\partial}(\log|z|) = 1\{2\overline{z}\}$.

(b) $\overline{\partial f} = \overline{\partial}\,\overline{f}$.

(c) *If* Δ *is the Laplacian,* $\dfrac{\partial^2}{\partial x^2} + \dfrac{\partial^2}{\partial y^2} \equiv \left(\dfrac{\partial}{\partial x}\right)^2 + \left(\dfrac{\partial}{\partial y}\right)^2$, *then* $\Delta = 4\overline{\partial}\partial = 4\partial\overline{\partial}$.

Proof. For part (a), write $\log|z| = \frac{1}{2}\log|z|^2 = \frac{1}{2}\log(z\overline{z})$ and apply the chain rule. The remaining parts are left to the reader. □

Hence a function $u : G \to \mathbb{C}$ is harmonic if and only if $\partial\overline{\partial}u = 0$ on G. Therefore, u is harmonic if and only if ∂u is analytic. (Note that we are considering complex valued functions to be harmonic; in the first volume only real-valued functions were harmonic.)

For any function u defined on an open set, the n-th order derivatives of u are all the derivatives of the form $\partial^j\overline{\partial}^k u$, where $j + k = n$.

A polynomial in z and $\overline{z}$ is a function $p(z, \overline{z})$ of the form $\sum a_{jk} z^j \overline{z}^k$, where a_{jk} is a complex number and the summation is over some finite set of non-negative integers. The n-th degree term of $p(z, \overline{z})$ is the sum of all the terms $a_{jk} z^j \overline{z}^k$ with $j + k = n$. The polynomial $p(z, \overline{z})$ has degree n if it has no terms of degree larger that n.

It is advantageous to rewrite several results from advanced calculus with this new notation.

2.4 Taylor's Formula. *If* $f \in C^n(G)$, $n \geq 1$, *and* $B(a; R) \subseteq G$, *then there is a unique polynomial* $p(z, \overline{z})$ *in* z *and* $\overline{z}$ *of degree* $\leq n - 1$ *and there is a function* g *in* $C^n(G)$ *such that the following hold:*

(a) $f = p + g$;

(b) *each derivative of* g *of order* $\leq n - 1$ *vanishes at* a;

(c) *for each* z *in* $B(a; R)$ *there is an* s, $0 < s < 1$, *(*s *depends on* z*) such that*

$$g(z) = \frac{1}{n!} \sum_{k+j=n} \left[\partial^k \overline{\partial}^j f\right] (a + s(z - a)) \, (z - a)^k \, (\overline{z} - \overline{a})^j .$$

Thus for each z in $B(a;R)$

$$|g(z)| \leq \frac{|z-a|^n}{n!} \sum_{k+j=n} \max\left\{\left|\partial^k \bar{\partial}^j f(w)\right| : |w-a| \leq |z-a|\right\}.$$

2.5 Green's Theorem. *If Γ is a smooth positive Jordan system with $G = \text{ins } \Gamma$, $u \in C(cl\ G)$, $u \in C^1(G)$, and $\bar{\partial}u$ is integrable over G, then*

$$\int_\Gamma u = 2i \iint_G \bar{\partial} u.$$

While here, let us note that integrals with respect to area measure on $\mathbb{C}$ will be denoted in a variety of ways. $\iint_G$ is one way (if the variable of integration can be suppressed) and $\int_G f d\mathcal{A} = \int_G f(z) d\mathcal{A}(z)$ is another. Which form of expression is used will depend on the context and our purpose at the time. The notation $\int f\ d\mathcal{A}$ will mean that integration is to be taken over all of $\mathbb{C}$. Finally, χ_K denotes the characteristic function of the set K; the function whose value at points in K is 1 and whose value is 0 at points of the complement of K.

Using Green's Theorem, a version of Cauchy's Theorem that is valid for non-analytic functions can be obtained. But first a lemma is needed. This lemma will also be used later in this book. As stated, the proof of this lemma requires knowledge of the Lebesgue integral in the plane, a violation of the ground rules established in the Preface. This can be overcome by replacing the compact set K below by a bounded rectangle. This modified version only uses the Riemann integral, can be proved with the same techniques as the proof given, and will suffice in the proof of the succeeding theorem.

2.6 Lemma. *If K is a compact subset of $\mathbb{C}$, then for every z*

$$\int_K |z-\zeta|^{-1} d\mathcal{A}(\zeta) < \infty.$$

Proof. If $h(\zeta) = |\zeta|^{-1}$, then using a change of variables shows that

$$\begin{aligned} \int_K |z-\zeta|^{-1} d\mathcal{A}(\zeta) &= \int \chi_K(\zeta) h(z-\zeta) d\mathcal{A}(\zeta) \\ &= \int \chi_K(z-\zeta) h(\zeta) d\mathcal{A}(\zeta) \\ &= \int_{z-K} h(\zeta) d\mathcal{A}(\zeta). \end{aligned}$$

If R is sufficiently large that $z - K \subseteq B(0; R)$, then

$$\begin{aligned}\int_K |z-\zeta|^{-1} d\mathcal{A}(\zeta) &\le \int_{B(0;R)} |\zeta|^{-1} d\mathcal{A}(\zeta) \\ &= \int_0^r \int_0^{2\pi} d\theta dr \\ &= 2\pi R.\end{aligned}$$

□

2.7 The Cauchy-Green Formula. *If* Γ *is a smooth positive Jordan system,* $G = \text{ins}\ \Gamma$, $u \in C(cl\ G)$, $u \in C^1(G)$, *and* $\bar\partial u$ *is integrable on* G, *then for every* z *in* G

$$u(z) = \frac{1}{2\pi i}\int_\Gamma u(\zeta)(\zeta - z)^{-1} d\zeta - \frac{1}{\pi}\int_G (\zeta - z)^{-1}\bar\partial u\ d\mathcal{A}(\zeta).$$

Proof. Fix w in G and choose $\varepsilon > 0$ such that $\overline{B}(w;\varepsilon) \subseteq G$. Put $B_\varepsilon = B(w;\varepsilon)$ and $G_\varepsilon = G \setminus \text{cl}\ B_\varepsilon$. Now apply Green's Theorem to the function $(z-w)^{-1}u(z)$ and the open set G_ε. (Note that $\partial G_\varepsilon = \Gamma \cup \partial B_\varepsilon$ and, with proper orientation, ∂G_ε becomes a positive Jordan system.) On G_ε,

$$\bar\partial\left[(z-w)^{-1}u\right] = (z-w)^{-1}\bar\partial u$$

since $(z-w)^{-1}$ is an analytic function on G_ε. Hence

2.8 $$\int_\Gamma \frac{u(z)}{z-w} dz - \int_{\partial B_\varepsilon} \frac{u(z)}{z-w} dz = 2i \int_{G_\varepsilon} \frac{1}{z-w}\bar\partial u\ d\mathcal{A}(z).$$

But

$$\begin{aligned}\lim_{\varepsilon\to 0}\int_{\partial B_\varepsilon} \frac{u(z)}{z-w} dz &= \lim_{\varepsilon \to 0} i \int_0^{2\pi} u(w + \varepsilon e^{i\theta}) d\theta \\ &= 2\pi i u(w).\end{aligned}$$

Because $(z-w)^{-1}$ is locally integrable (Lemma 2.6) and bounded away from w and $\bar\partial u$ is bounded near w and integrable away from w, the limit of the right hand side of (2.8) exists. So letting $\varepsilon \to 0$ in (2.8) gives

$$\int_\Gamma \frac{u(z)}{z-w} dz - 2\pi i u(w) = 2i \int_G \frac{1}{z-w}\bar\partial u\ d\mathcal{A}(z).$$

□

Note that if, in the preceding theorem, u is an analytic function, then $\bar\partial u = 0$ and this become Cauchy's Integral Formula.

2.9 Corollary. *If $u \in C_c^1(\mathbb{C})$ and $w \in \mathbb{C}$, then*

$$u(w) = -\frac{1}{\pi}\int \frac{1}{z-w}\bar{\partial} u \, d\mathcal{A}(z).$$

There are results analogous to the preceding ones where the Laplacian replaces $\bar{\partial}$.

2.10 Lemma. *If K is a compact subset of the plane, then*

$$\int_K |\log |z|| \, d\mathcal{A}(z) < \infty.$$

Proof. If polar coordinates are used, then it is left to the reader to show that for any $R > 1$

$$\int_{|z|\le R} |\log |z|| \, d\mathcal{A}(z) = \pi R^2 \left(\log R - \frac{1}{2}\right) + \pi.$$

This proves the lemma. □

2.11 Theorem. *If $u \in C_c^2$ on the plane and $w \in \mathbb{C}$, then*

$$u(w) = \frac{1}{2\pi}\int \log|z-w| \Delta u \, d\mathcal{A}(z).$$

Proof. Let R be positive such that $\operatorname{supp}(u) \subseteq B(w;R)$ and for $\varepsilon > 0$ let $G_\varepsilon = \{z : \varepsilon < |z-w| < R\}$ and $\gamma_\varepsilon = \{z : |z-w| = \varepsilon\}$ with suitable orientation. Green's Theorem implies

$$\begin{aligned}
\int_{\gamma_\varepsilon} \partial u \log|z-w| dz &= 2i\int_{G_\varepsilon} \bar{\partial}\left[\log|z-w|\partial u\right] d\mathcal{A}(z) \\
&= 2i\int_{G_\varepsilon} (\partial\bar{\partial} u) \log|z-w| d\mathcal{A}(z) \\
&\quad +2i\int_{G_\varepsilon} \frac{1}{2(\bar{z}-\bar{w})}\partial u \, d\mathcal{A}(z) \\
&= \frac{i}{2}\int_{G_\varepsilon} (\Delta u)\log|z-w| d\mathcal{A}(z) \\
&\quad +i\int_{G_\varepsilon} \frac{\partial u}{\bar{z}-\bar{w}} d\mathcal{A}(z).
\end{aligned}$$

Now $\left|\int_{\gamma_\varepsilon} (\partial u)\log|z-w|dz\right| \le M\varepsilon\log\varepsilon$ for some constant M independent of ε. Hence the integral converges to 0 as $\varepsilon \to 0$. Since $(\bar{z}-\bar{w})^{-1}$ is locally integrable and ∂u has compact support, $\iint_{G_\varepsilon} [\partial u/(\bar{z}-\bar{w})]$ converges as

$\varepsilon \to 0$. By Corollary 2.9 and Proposition 2.3(b) this limit must be $-\pi\overline{u}(w)$. Since Δu is continuous and has compact support, combining this latest information with the above equations, the theorem follows. □

We end this section with some results that connect areas with analytic functions. The first result is a consequence of the change of variables formula for double integrals and the fact that if f is an analytic function, then the Jacobian of f considered as a mapping from $\mathbb{R}^2$ into $\mathbb{R}^2$ is $|f'|^2$ (see Exercise 2).

2.12 Theorem. *If f is a conformal equivalence between the open sets G and Ω, then*

$$\text{Area}(\Omega) = \iint_G |f'|^2.$$

2.13 Corollary. *If Ω is a simply connected region, $\tau : \mathbb{D} \to \Omega$ is a Riemann map, and $\tau(z) = \sum_n a_n z^n$ in $\mathbb{D}$, then*

$$\text{Area}(\Omega) = \iint_{\mathbb{D}} |\tau'|^2 = \pi \sum_n n|a_n|^2.$$

Proof. The first equality is a restatement of the preceding theorem for this special case. For the second equality, note that $\tau'(z) = \sum_n n a_n z^{n-1}$. So for $r < 1$,

$$\begin{aligned} \left|\tau'\left(re^{i\theta}\right)\right|^2 &= \left(\sum_n n a_n r^{n-1} e^{i(n-1)\theta}\right) \overline{\left(\sum_m m a_m r^{m-1} e^{i(m-1)\theta}\right)} \\ &= \sum_{m,n} mn\overline{a_m} a_n r^{m+n-2} e^{i(n-m)\theta} \end{aligned}$$

and this series converges uniformly in θ. Using polar coordinates to calculate $\iint_{\mathbb{D}} |\tau'|^2$ and the fact that $\int_0^{2\pi} e^{i(n-m)\theta} d\theta = 0$ for $n \neq m$, we get

$$\begin{aligned} \iint_{\mathbb{D}} |\tau'|^2 &= \sum_n n^2 |a_n|^2 (2\pi) \int_0^1 r^{2n-1} dr \\ &= (2\pi) \sum_n n^2 |a_n|^2 \frac{1}{2n} \\ &= \pi \sum_n n|a_n|^2. \end{aligned}$$

□

If f fails to be a conformal equivalence, a version of this result remains valid. Namely, $\iint_G |f'|^2$ is the area of $f(G)$ "counting multiplicities." This is made specific in the next theorem. The proof of this result uses some measure theory; in particular, the reader must know the Vitali Covering Theorem.

2.14 Theorem. *If $f : G \to \Omega$ is a surjective analytic function and for each ζ in $\Omega, n(\zeta)$ is the number of points in $f^{-1}(\zeta)$, then*

$$\int_G |f'|^2 d\mathcal{A} = \int_\Omega n(\zeta) d\mathcal{A}(\zeta).$$

Proof. Since f is analytic, $\{z : f'(z) = 0\}$ is countable and its complement in G is an open set with the same measure. Thus without loss of generality we may assume that f' never vanishes; that is, f is locally one-to-one. Thus for each z in G there are arbitraily small disks centered at z on which f is one-to-one. The collection of all such disks forms a Vitali cover of G. By the Vitali Covering Theorem there are a countable number of pairwise disjoint open disks $\{D_n\}$ such that f is one-to-one on each D_n and Area$(G \setminus \cup_n D_n) = 0$.

Put $\Lambda = \cup_n f(D_n) = f(\cup_n D_n)$. Because $f(\partial D_n)$ is a smooth curve, Area$(\Omega) =$ Area(Λ). For $1 \le k \le \infty$, let $\Lambda_k = \{\zeta \in \Lambda : n(\zeta) = k\}$; so Area$(\Lambda) = \sum_k$Area$(\Lambda_k)$. If $G_k = f^{-1}(\Lambda_k)$, then Theorem 2.12 implies

$$\begin{aligned} \int_{G_k} |f'|^2 d\mathcal{A} &= \sum_{n=1}^{\infty} \int_{D_n \cap G_k} |f'|^2 d\mathcal{A} \\ &= \sum_{n=1}^{\infty} \text{Area}(D_n \cap G_k) \\ &= k\text{Area}(\Lambda_k) \end{aligned}$$

since f is one-to-one on each D_n. Thus

$$\begin{aligned} \int_G |f'|^2 d\mathcal{A} &= \sum_{1 \le k \le \infty} \int_{G_k} |f'|^2 d\mathcal{A} \\ &= \sum_{1 \le k \le \infty} k \ \text{Area}(\Lambda_k) \\ &= \int_\Omega n(\zeta) \ d\mathcal{A}(\zeta). \end{aligned}$$

□

Exercises

1. Show that if K and L are compact subsets of $\mathbb{C}$, then there is a constant $M > 0$ such that $\int_K |z - \zeta|^{-1} d\mathcal{A}(\zeta) \le M$ for all z in L.

2. Show that if $f : G \to \mathbb{C}$ is an analytic function and we consider f as a function from the region G in $\mathbb{R}^2$ into $\mathbb{R}^2$, then the Jacobian of f is $|f'|^2$.

3. Let f be defined on $\mathbb{D}$ by $f(z) = \exp[(z+1)/(z-1))]$ and show that $\iint |f'(z)|^2 = \infty$. Discuss.

4. If G is a region and u is a real-valued harmonic function on G such that $\{z : u(z) = 0\}$ has positive area, then u is identically 0.

§3 Harmonic Conjugates and Primitives

In Theorem 8.2.2 it was shown that a region G in the plane has the property that every harmonic function on G has a harmonic conjugate if and only if G is simply connected. It was also shown that the simple connectivity of G is equivalent to the property that every analytic function on G has a primitive.

The above mentioned results neglect the question of when an individual harmonic function has a conjugate or an individual analytic function has a primitive. In this section these questions will be answered and it will be seen that even on an individual basis these properties are related.

We begin with an elementary result that has been used in the first volume without being made explicit. The proof is left to the reader.

3.1 Proposition. *If $f : G \to \mathbb{C}$ is an analytic function, then f has a primitive if and only if $\int_\gamma f = 0$ for every closed rectifiable curve γ in G.*

Another result, an easy exercise in the use of the Cauchy-Riemann equations, is the following.

3.2 Proposition. *If $u : G \to \mathbb{C}$ is a C^2 function, then u is a harmonic function on G if and only if $f = (u_x - iu_y)/2 = \partial u$ is an analytic function on G.*

It turns out that there is a close relation between the harmonic function u and the analytic function $f = \partial u$. Indeed, one function often can be studied with the help of the other. A key to this is the following computation. If γ is any closed rectifiable curve in G, then

3.3 $$\int_\gamma \partial u = \frac{i}{2}\int_\gamma (u_x dy - u_y dx).$$

In fact, $\int_\gamma f = \frac{1}{2}\int_\gamma (u_x - iu_y)(dx + idy) = \frac{1}{2}\int_\gamma (u_x dx + u_y dy) + \frac{i}{2}\int_\gamma (u_x dy - u_y dx)$ and $\int_\gamma (u_x dx + u_y dy) = 0$ since this is the integral of an exact differential.

We are now ready to present a direct relation between the existence of a harmonic conjugate and the existence of a primitive.

3.4 Theorem. *If G is a region in $\mathbb{C}$ and $u : G \to \mathbb{R}$ is a harmonic function, then the following statements are equivalent.*

(a) *The function u has a harmonic conjugate.*

(b) *The analytic function $f = \partial u$ has a primitive in G.*

(c) *For every closed rectifiable curve γ in G, $\int_\gamma (u_x dy - u_y dx) = 0$.*

Proof. By Proposition 3.1, (3.3) shows that (b) and (c) are equivalent.

(a) *implies* (b). If g is an analytic function on G such that $g = u + iv$, then the fact that the Cauchy-Riemann equations hold implies that $g' = u_x + iv_x = u_x - iu_y = 2f$.

(b) *implies* (a). Suppose g is an analytic function on G such that $g' = 2f$ and let U and V be the real and imaginary parts of g. Thus $g' = U_x + iV_x = 2f = u_x - iu_y$. It is now an easy computation to show that u and V satisfy the Cauchy-Riemann equations, and so V is a harmonic conjugate of u. □

For a function u the differential $u_x y - u_y dx$ is called the *conjugate differential* of u and is denoted $^*du$. Why? Suppose u is a harmonic function with a harmonic conjugate v. Using the Cauchy-Riemann equations the differential of v is $dv = v_x dx + v_y dy = -u_y dx + u_x dy = {}^*du$. So Theorem 3.4(c) says that a harmonic function u has a harmonic conjugate if and only if its conjugate differential $^*du$ is exact. (See any book on differential forms for the definition of an exact form.)

The reader might question whether Theorem 3.4 actually characterizes the harmonic functions that have a conjugate, since it merely states that this problem is equivalent to another problem of equal difficulty: whether a given analytic function has a primitive. There is some validity in this criticism, though this does not diminish the value of (3.4); it is a criticism of the result as it relates to the originally stated objective rather than any internal defect.

Condition (c) of the theorem says that to check whether a function has a conjugate you must still check an infinite number of conditions. In §15.1 below the reader will see that in the case of a finitely connected region this can be reduced to checking a finite number of conditions.

Here is a fact concerning the conjugate differential that will be used in the sequel. Recall that $\partial u/\partial n$ denotes the normal derivative of u with respect to the outwardly pointed normal to a given curve γ.

3.5 Proposition. *If u is a continuously differentiable function on the region G and γ is a closed rectifiable curve in G, then*

$$\frac{1}{\pi i}\int_\gamma \partial u = \frac{1}{2\pi}\int_\gamma {}^*du = \frac{1}{\pi i}\int_\gamma \frac{\partial u}{\partial n}|dz|.$$

Proof. The first equality is a rephrasing of (3.3) using the latest edition of the notation. The proof of the second equality is a matter of using the definitions of the relevant terms. This will not be used here and so the details are left to the reader. □

Exercises

1. If G is a region and $u : G \to \mathbb{R}$ is a continuous function, then u is harmonic if and only if for every a in G there is an $\delta > 0$ such that $u(a) = (2\pi)^{-1} \int u(a + re^{i\theta})d\theta$ for $r < \delta$. (A slight weakening of the fact that functions with the Mean Value Property are harmonic.)

2. If u is a real-valued function on G, show that $\partial u \, dz = \overline{\partial} u \, d\overline{z} = \frac{1}{2}(du + i^* du)$. Hence $\partial u \, dz + \overline{\partial} u \, d\overline{z} = du + i^* du$.

3. Prove that a region G is simply connected if and only if every complex valued harmonic function $u : G \to \mathbb{C}$ can be written as $u = g + \overline{h}$ for analytic functions g and h on G.

4. Let G be a region and $f : G \to \mathbb{C}$ an analytic function that never vanishes. Show that the following statements are equivalent. (a) There is an analytic branch of $\log f(z)$ on G (that is, an analytic function $g : G \to \mathbb{C}$ such that $\exp[g(z)] = f(z)$ for all z in G). (b) The function f'/f has a primitive. (c) For every closed rectifiable path γ in G, $\int_\gamma f'/f = 0$.

5. Let $r = p/q$ be a rational function, where p and q are polynomials without a common divisor. Let $a_1, \ldots, a_n$ be the distinct zeros of p with multiplicities $\alpha_1, \ldots, \alpha_n$ and let $b_1, \ldots, b_m$ be the distinct zeros of q with multiplicities $\beta_1, \ldots, \beta_m$. If G is an open set in $\mathbb{C}$ that contains none of the points $a_1, \ldots, a_n, b_1, \ldots, b_m$, show that there is an analytic branch of $\log r(z)$ if and only if for every closed rectifiable path γ in G,
$$0 = \sum_{j=1}^{n} \alpha_j n(\gamma; a_j) - \sum_{i=1}^{m} \beta_i n(\gamma : b_i).$$

§4 Analytic Arcs and the Reflection Principle

If Ω is a region and $f : \mathbb{D} \to \Omega$ is an analytic function, under what circumstances can f be analytically continued to a neighborhood of cl $\mathbb{D}$? This question is addressed in this section. But first, recall the Schwarz Reflection Principle (9.1.1) where an analytic function is extended across the real line provided it is real-valued on the line. It is probably no surprise that this can be generalized by extending functions across a circle; the details are given below. In this section more extensive formulations of the Reflection Principle are formulated. The relevant concept is that of an analytic arc. Before addressing this issue, we will concentrate on circles.

Suppose G is any region that does not include 0. If $G^\# \equiv \{1/\bar{z} : z \in G\}$, then $G^\#$ is the reflection of G across the unit circle $\partial\mathbb{D}$. If f is an analytic function on G, then $f^\#(\zeta) = \overline{f\left(\overline{\zeta^{-1}}\right)}$ defines an analytic function on $G^\#$. Similarly if G is any region and a is a point not in G, then for some radius $r > 0$

4.1 $$G^\# = \left\{a + \frac{r^2}{\bar{z} - \bar{a}} : z \in G\right\} = \left\{\zeta : a + \frac{r^2}{\bar{\zeta} - \bar{a}} \in G\right\}$$

is the reflection of G across the circle $\partial B(a;r)$. Note that $a \notin G^\#$ and $G^{\#\#} = G$. If f is an analytic function on G and $G^\#$ is as above, then

4.2 $$f^\#(\zeta) = \overline{f\left(a + \frac{r^2}{\bar{\zeta} - \bar{a}}\right)}$$

is analytic on $G^\#$. Here is one extension of the Reflection Principle.

4.3 Proposition. *If G is a region in $\mathbb{C}$, $a \notin G$, and $G = G^\#$, let $G_+ \equiv G \cap B(a;r)$, $G_0 \equiv G \cap \partial B(a;r)$, and $G_- \equiv G \cap [\mathbb{C} \setminus B(a;r)]$. If $f : G_+ \cup G_0 \to \mathbb{C}$ is a continuous function that is analytic on G_+, $f(G_0) \subseteq \mathbb{R}$, and $f^\# : G \to \mathbb{C}$ is defined by letting $f^\#(z) = f(z)$ for z in $G_+ \cup G_0$ and letting $f^\#(z)$ be defined as in (4.2) for z in G_-, then $f^\#$ is an analytic function on G. If f is one-to-one and* Im *f has constant sign, then $f^\#$ is a conformal equivalence.*

Proof. Exercise □

The restraint in the preceding proposition that f is real-valued on G_0 can also be relaxed.

4.4 Proposition. *If G is a region in $\mathbb{C}$, $a \notin G$, and $G = G^\#$, let G_+, G_-, G_0 be as in the preceding proposition. If $f : G_+ \cup G_0 \to \mathbb{C}$ is a continuous function that is analytic on G_+ and there is a point α not in $f(G_+)$ and a $\rho > 0$ such that $f(G_0) \subseteq \partial B(\alpha;\rho)$ less one point and if $f^\# : G \to \mathbb{C}$ is defined by letting $f^\#(z) = f(z)$ on $G_+ \cup G_0$ and*

$$f^\#(z) = \alpha + \frac{\rho^2}{\overline{f\left(a + \frac{r^2}{\bar{z} - \bar{a}}\right)} - \bar{\alpha}}$$

for z in G_-, then $f^\#$ is analytic. If f is one-to-one and $f(G_+)$ is contained entirely in either the inside or the outside of $B(\alpha;\rho)$, then $f^\#$ is a conformal equivalence.

Proof. Let T be a Möbius transformation that maps $\partial B(\alpha;\rho)$ onto $\mathbb{R}\cup\{\infty\}$ and takes the missing point to ∞; so $T\circ f$ satisfies the hypothesis of the preceding proposition. The rest of the proof is an exercise. □

Let $\mathbb{D}_+ \equiv \{z\in\mathbb{D} : \operatorname{Im} z > 0\}$.

4.5 Definition. If Ω is a region and L is a connected subset of $\partial\Omega$, then L is a *free analytic boundary arc* of Ω if for every ω in L there is a neighborhood Δ of ω and a conformal equivalence $h:\mathbb{D}\to\Delta$ such that:

(a) $h(0)=\omega$;

(b) $h(-1,1)=L\cap\Delta$;

(c) $h(\mathbb{D}_+)=\Omega\cap\Delta$.

Note that the above definition implies that $\Omega\cap\Delta$ is a simply connected region. The first result about free analytic boundary arcs is that every arc in $\partial\mathbb{D}$ is a free analytic boundary arc of $\mathbb{D}$, a welcome relief. Most of the proof is left to the reader. The symbol $\mathbb{H}_+$ is used to denote the upper half plane, $\{z : \operatorname{Im} z > 0\}$.

4.6 Lemma. *If $w\in\partial\mathbb{D}$ and $\varepsilon>0$, then there is a neighborhood V of w such that $V\subseteq B(w;\varepsilon)$ and there is conformal equivalence $h:\mathbb{D}\to V$ such that $h(0)=w$, $h(-1,1)=V\cap\partial\mathbb{D}$, and $h(\mathbb{D}_+)=V\cap\mathbb{D}$.*

Proof. It may be assumed that $w=1$. Choose α on $\partial\mathbb{D}$ such that $\operatorname{Im}\alpha>0$ and the circle orthogonal to $\partial\mathbb{D}$ that passes through α and $\overline{\alpha}$ lies inside $B(1;\varepsilon)$. Let h be the Möbius transformation that takes 0 to 1, 1 to α, and ∞ to -1. It is not hard to see that $h(\mathbb{R}_\infty)=\partial\mathbb{D}$ and $h(\mathbb{H}_+)=\mathbb{D}$. If $h(\partial\mathbb{D})=C$ and V is the inside circle C, then these fulfill the properties stated in the conclusion of the lemma. The details are left to the reader. □

The next lemma is useful, though its proof is elementary. It says that about each point in a free analytic boundary arc there is a neighborhood basis consisting of sets such as appear in the definition.

4.7 Lemma. *If L is a free boundary arc of Ω, $\omega\in L$, and U is any neighborhood of ω, then there is a neighborhood Δ of ω with $\Delta\subseteq U$ and a conformal equivalence $h:\mathbb{D}\to\Delta$ such that:*

(a) $h(0)=\omega$;

(b) $h(-1,1)=L\cap\Delta$;

(c) $h(\mathbb{D}_+)=\Omega\cap\Delta$,

Proof. According to the definition there is a neighborhood Λ of ω and a conformal equivalence $k:\mathbb{D}\to\Lambda$ with $k(0)=\omega$, $k(-1,1)=\Lambda\cap L$, and

$k(\mathbb{D}_+) = \Omega \cap \Lambda$. The continuity of k implies the existence of r, $0 < r < 1$, such that $k(r\mathbb{D}) \subseteq U$. Let $\Delta = k(r\mathbb{D})$ and define $h : \mathbb{D} \to \Delta$ by $h(z) = k(rz)$. It is left to the reader to check that h and Δ have the desired properties. $\square$

4.8 Theorem. *Let G and Ω be regions and let J and L be free analytic boundary arcs in ∂G and $\partial\Omega$, respectively. If f is a continuous function on $G \cup J$ that is analytic on G, $f(G) \subseteq \Omega$, and $f(J) \subseteq L$, then for any compact set K contained in J, f has an analytic continuation to an open set containing $G \cup K$.*

Proof. Let $z \in J$ and put $\omega = f(z)$; so $\omega \in L$. By definition there is a neighborhood Δ_ω of ω and a conformal equivalence $h_\omega : \mathbb{D} \to \Delta_\omega$ such that $h_\omega(0) = \omega$, $h_\omega(-1,1) = \Delta_\omega \cap L$, and $h_\omega(\mathbb{D}_+) = \Omega \cap \Delta_\omega$. By continuity, there is a neighborhood U_z about z such that $f\,(U_z \cap \text{cl } G) = f\,(U_z \cap (G \cup J)) \subseteq \Delta_\omega \cap \text{cl } \Omega = \Delta_\omega \cap (\Omega \cup L)$. Since J is a free analytic boundary arc, the preceding lemma implies this neighborhood can be chosen so that there is a conformal equivalence $k_z : \mathbb{D} \to U_z$ with $k_z(0) = z$, $k_z(-1,1) = U_z \cap J$, and $k_z(\mathbb{D}_+) = U_z \cap G$.

Thus $g_z \equiv h_\omega^{-1} \circ f \circ k_z$ is a continuous function on $\mathbb{D}_+ \cup (-1,1)$ that is analytic on $\mathbb{D}_+$ and real valued on $(-1,1)$. In fact, $g_z(-1,1) \subseteq (-1,1)$ and $g_z(\mathbb{D}_+) \subseteq \mathbb{D}_+$. According to Proposition 4.3, g_z has an analytic continuation $g_z^\#$ to $\mathbb{D}$. From the formula for $g_z^\#$ we have that $g_z^\#(\mathbb{D}) \subseteq \mathbb{D}$. Thus $f_z^\# \equiv h_\omega \circ g_z^\# \circ k_z^{-1}$ is a well defined analytic function on U_z that extends $f\,|(U_z \cap G)$. Extend f to a function $\tilde{f}$ on $G \cup U_z$ by letting $\tilde{f} = f$ on G and $\tilde{f} = f_z^\#$ on U_z. It is easy to see that these two definitions of $\tilde{f}$ agree on the overlap so that $\tilde{f}$ is an analytic function on $G \cup U_z$.

Now consider the compact subset K of J and from the open cover $\{U_z : z \in K\}$ extract a finite subcover $\{U_j : 1 \le j \le n\}$ with corresponding analytic functions $f_j : G \cup U_j \to \mathbb{C}$ such that f_j extends f. Write K as the union $K_1 \cup \cdots \cup K_n$, where each K_j is a compact subset of U_j. (The easiest way to do this is to consider a partition of unity $\{\phi_j\}$ on K subordinate to $\{U_j\}$ (see Proposition 18.2.4 below) and put $K_j = \{z \in K : \phi_j(z) \ge 1/n\}$.) Note that if it occurs that $U_i \cap U_j \neq \emptyset$ but $U_i \cap U_j \cap G = \emptyset$, then $K_i \cap K_j = \emptyset$. Indeed, if there is a point z in $K_i \cap K_j$, then z belongs to the open set $U_i \cap U_j$ and so $U_i \cap U_j \cap G \neq \emptyset$. Thus replacing U_i and U_j by smaller open sets that still contain the corresponding compact sets K_i and K_j, we may assume that whenever $U_i \cap U_j \neq \emptyset$ we have that $U_i \cap U_j \cap G \neq \emptyset$.

So if $U_i \cap U_j \neq \emptyset$, f_i and f_j agree on $U_i \cap U_j \cap G$ with f; thus the two extensions must agree on $U_i \cap U_j$. Thus we can obtain an extension $f^\#$ of f to $G \cup \bigcup_{j=1}^n U_j$, which is an open set containing $G \cup K$. $\square$

We close this section with a reflection principle for harmonic functions. First we attack the disk.

4.9 Lemma. *Let u be a continuous real-valued function on $\text{cl}\,\mathbb{D}$ that is* *harmonic on $\mathbb{D}$. If there is an open arc J in $\partial\mathbb{D}$ such that u is constant on J, then there is a region W containing $\mathbb{D} \cup J$ and a harmonic function u_1 on W such that $u_1 = u$ on $\mathbb{D} \cup J$.*

Proof. It suffices to assume that $u \equiv 0$ on J. Suppose $J = \{e^{it} : \alpha < t < \beta\}$, where $-\pi \leq \alpha < \beta \leq \pi$. By using the Poisson kernel we know that

$$\begin{aligned} u(z) &= \frac{1}{2\pi} \int_{-\pi}^{\pi} P_r(\theta - t) u(e^{it}) dt \\ &= \frac{1}{2\pi} \left\{ \int_{-\pi}^{\alpha} + \int_{\beta}^{\pi} \right\} P_r(\theta - t) u(e^{it}) dt \end{aligned}$$

for $z = re^{i\theta}$ in $\mathbb{D}$ (10.2.9). Moreover on $\mathbb{D}$, $u = \operatorname{Re} f$ where f is the analytic function on $W \equiv \mathbb{C} \setminus [\partial\,\mathbb{D} \setminus J]$ defined by

$$f(z) = \frac{1}{2\pi} \left\{ \int_{-\pi}^{\alpha} + \int_{\beta}^{\pi} \right\} \frac{e^{\mathrm{it}} + z}{e^{\mathrm{it}} - z} u(e^{\mathrm{it}}) dt.$$

Thus $u_1 = \operatorname{Re} f$ is the sought for harmonic extension. □

4.10 Theorem. *Suppose G is a region and J is a free analytic boundary arc of G. If $u : G \cup J \to \mathbb{R}$ is a continuous function that is harmonic in G and constant on J, then for any compact subset K of J, u has a harmonic extension u_1 on a region W that contains $G \cup K$.*

Proof. The proof is similar to that of Theorem 4.8; the details are left to the reader. □

Here is a special type of finitely connected region.

4.11 Definition. A region G is a *Jordan region* or *Jordan domain* if it is bounded and the boundary of G consists of a finite number of pairwise disjoint closed Jordan curves. If there are $n + 1$ curves $\gamma_0, \gamma_1, \ldots, \gamma_n$ that make up the boundary of G, then G is called an *n-Jordan region.*

Since G is assumed connected, it follows that one of these curves forms the boundary of the polynomial convex hull of cl G; denote this curve by γ_0 and refer to it as the *outer boundary* of G. It then follows that the insides of the remaining curves are pairwise disjoint. Thus the curves can be suitably oriented so that $\Gamma = \{\gamma_0, \ldots, \gamma_n\}$ is a positive Jordan system.

4.12 Definition. Say that a Jordan curve γ is an *analytic curve* if there is a function f analytic in a neighborhood of $\partial\mathbb{D}$ such that $\gamma = f(\partial\mathbb{D})$. Say that a Jordan region is an *analytic Jordan region* if each of the curves forming the boundary of G is an analytic curve.

It is easy to see that for an analytic Jordan region every arc in its boundary is a free analytic boundary arc. An application of Theorem 4.10 (and

Proposition 1.8) proves the following two results. The details are left to the reader.

4.13 Corollary. *Let G be an analytic Jordan region with boundary curves $\gamma_0, \gamma_1, \dots, \gamma_n$. If u is a continuous real-valued function on $G \cup \gamma_j$ that is harmonic on G and u is a constant on γ_j, then there is an analytic Jordan region G_1 containing $G \cup \gamma_j$ and a harmonic function u_1 on G_1 such that $u_1 = u$ on G.*

4.14 Corollary. *If G is an analytic Jordan region and $u : cl\, G \to \mathbb{R}$ is a continuous function that is harmonic on G and constant on each component of the boundary of G, then u has a harmonic extension to an analytic Jordan region containing* $\operatorname{cl} G$.

As was pointed out above, if G is an analytic Jordan region and $z \in \partial G$, then there is a neighborhood U of z such that $U \cap \partial G$ is a free analytic boundary arc. The converse is also true. If G is a Jordan region and every point of the boundary has a neighborhood that intersects ∂G in an analytic Jordan boundary arc, then G is an analytic Jordan region. This is another of those results about subsets of the plane that seem obvious but require a surprising amount of work to properly prove. See Minda [1977] and Jenkins [1991].

Exercises

1. Let G, Λ, and Ω be simply connected regions and let $f : G \to \Lambda$ be a conformal equivalence satisfying the following: (a) $G \supseteq \mathbb{D}$ and $G \neq \mathbb{D}$; (b) $\Lambda \supseteq \Omega$ and $\Lambda \neq \Omega$; (c) $f(\mathbb{D}) = \Omega$. If J is any open arc of $G \cap \partial\mathbb{D}$, then $f(J)$ is a free analytic boundary arc of Ω.

2. Prove Theorem 4.10.

3. Give the details of the proof of Corollaries 4.13 and 4.14.

§5 Boundary Values for Bounded Analytic Functions

In this section we will state three theorems about bounded analytic functions on $\mathbb{D}$ whose proofs will be postponed. Both the statements and the proofs of these results involve measure theory, though the statements only require a knowledge of a set of measure 0, which will be explained here.

Let U be a (relatively) open subset of the unit circle, $\partial\mathbb{D}$. Hence U is the union of a countable number of pairwise disjoint open arcs $\{J_k\}$. Let $J_k = \{e^{i\theta} : a_k < \theta < b_k\}$, $0 < b_k - a_k < 2\pi$. Define the length of J_k by

$$\ell(U) = \sum_k \ell(J_k).$$

5.1 Definition. A subset E of $\partial\mathbb{D}$ has *measure zero* if for every $\varepsilon > 0$ there is an open set U containing E with $\ell(U) < \varepsilon$.

There are some exercises at the end of this section designed to help the neophyte feel more comfortable with the concept of a set of measure 0. In particular you are asked to show that countable sets have measure 0. There are, however, some uncountable sets with measure 0. For example, if C is the usual Cantor ternary set in $[0, 1]$ and $E = \{e^{2\pi it} : t \in C\}$, then E is an uncountable closed perfect set having measure 0.

A statement will be said to hold *almost everywhere* on $\partial\mathbb{D}$ if it holds for all a in a subset X of $\partial\mathbb{D}$ and $\partial\mathbb{D} \setminus X$ has measure 0; alternately, it is said that the statement holds for almost every a in $\partial\mathbb{D}$. For example, if $f : \partial\mathbb{D} \to \mathbb{C}$ is some function, then the statement that f is differentiable almost everywhere means that there is a subset X of $\partial\mathbb{D}$ such that $\partial\mathbb{D} \setminus X$ has measure 0 and $f'(a)$ exists for all a in X; alternately, $f'(a)$ exists for almost every a in $\partial\mathbb{D}$. The words "almost everywhere" are abbreviated by a.e..

If $f : \mathbb{D} \to \mathbb{C}$ is any function and $e^{i\theta} \in \partial\mathbb{D}$, then f has a *radial limit* at $e^{i\theta}$ if, as $r \to 1-$, the limit of $f(re^{i\theta})$ exists and is finite. The next three theorems will be proved later in this book. Immediately after the statement of each result the location of the proof will be given.

5.2 Theorem. *If $f : \mathbb{D} \to \mathbb{C}$ is a bounded analytic function, then f has radial limits almost everywhere on $\partial\mathbb{D}$.*

This is a special case of Theorem 19.2.12 below.

If f is a bounded analytic function defined on $\mathbb{D}$, then the values of the radial limits of f, when they exist, will also be denoted by $f(e^{i\theta})$ unless it is felt that it is necessary to make a distinction between the analytic function defined on $\mathbb{D}$ and its radial limits. Notice that f becomes a function defined a.e.on $\partial\mathbb{D}$.

5.3 Theorem. *If $f : \mathbb{D} \to \mathbb{C}$ is a bounded analytic function and the radial limits of f exist and are zero on a set of positive measure, then $f \equiv 0$.*

This result is true for a class of analytic functions that is larger than the bounded ones. This more general result is stated and proved in Corollary 20.2.12.

So, in particular, the preceding theorem says that it is impossible for an non-constant analytic function f defined on $\mathbb{D}$ to have a continuous extension $f : \text{cl } \mathbb{D} \to \mathbb{C}$ such that f vanishes on some arc of $\partial\mathbb{D}$. This special case will be used in some of the proofs preceding §20.2, so it is worth noting that this is a direct consequence of the Schwarz Reflection Principle. It turns out that such a function that is continuous on cl $\mathbb{D}$ and analtyic inside can have more than a countable set of zeros without being constantly 0. That, however, is another story.

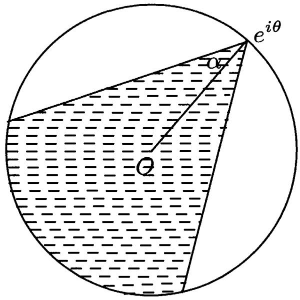

Figure 13.1.

We now consider a more general type of convergence for a function as the variable approaches a boundary point. Fix θ, $0 \leq \theta \leq 2\pi$, and consider the portion of the open unit disk $\mathbb{D}$ contained in an angle with vertex $e^{i\theta} = a$, symmetric about the radius $z = ra$, $0 \leq r \leq 1$, and having opening 2α, where $0 < \alpha < \frac{\pi}{2}$. See Figure 13.1.

Call such a region a *Stolz angle* with vertex a and opening α. The variable z is said to *approach* a *non-tangentially* if $z \to a$ through some Stolz angle. This will be abbreviated $z \to a$ (n.t.). Say that f has a *non-tangential limit* at a if there is a complex number ζ such that $f(z) \to \zeta$ as $z \to a$ through any Stolz angle with vertex a.

5.4 Theorem. *Let $\gamma : [0,1] \to \mathbb{C}$ be an arc with $\gamma([0,1)) \subseteq \mathbb{D}$ and suppose γ ends at the point $\gamma(1) = a$ in $\partial\mathbb{D}$. If $f : \mathbb{D} \to \mathbb{C}$ is a bounded analytic function such that $f(\gamma(t)) \to \alpha$ as $t \to 1-$, then f has non-tangential limit α at a.*

5.5 Corollary. *If a bounded analytic function f has radial limit ζ at a in $\partial\mathbb{D}$, then f has non-tangential limit ζ at a.*

Theorem 5.4 will be proved here, but two results (Exercises 6 and 7) are needed that have not yet been proved. These will be proved later in more generality, but the special cases needed are within the grasp of the reader using the methods of the first volume. For the proof a lemma is needed. In this lemma and the proof of (5.4), the Stolz angle at $z = 1$ of opening 2δ is denoted by S_δ.

5.6 Lemma. *Suppose $0 < r < 1$, $B = B(1;r)$, $\Omega = B \cap \mathbb{D}$, and $I = \{z \in \partial\Omega : \operatorname{Im} z \leq 0 \text{ and } |z| = 1\}$. If ω is the solution of the Dirichlet problem with boundary values χ_I, then for every $\varepsilon > 0$, there is a ρ, $0 < \rho < r$, such that if $|z-1| < \rho$, $0 < \delta < \pi/2$, and $z \in S_\delta$, then $\omega(z) \geq (1/2) - \delta/\pi - \varepsilon$.*

Proof. For w in Ω, let $\phi(w) \in (0,1)$ such that $\pi\phi(w)$ is the angle from the vertical line $\operatorname{Re} z = 1$ counterclockwise to the line passing through 1 and w. It can be verified that $\phi(w) = \pi^{-1} \arg(-i(w-1))$. Thus ϕ is harmonic for $\operatorname{Re} w < 1$ and continuous on $\operatorname{cl}\mathbb{D} \setminus \{1\}$. Let ζ be the end point of the arc I different from 1.

Claim. If we define $(\phi - \omega)(1) = 0$, then $(\phi - \omega) : \operatorname{cl}\Omega \to \mathbb{R}$ is continuous except at ζ.

Since $(\phi - \omega)$ is harmonic on Ω and is the solution of the Dirichlet problem for its boundary values, we need only verify that $(\phi - \omega) : \partial\mathbb{D} \to \mathbb{R}$ is continuous except for the point ζ; by Exercise 6 the only point in doubt here is $w = 1$. Suppose $w \to 1$ with $\operatorname{Im} w < 0$. Here $\phi(w) \to 1$ and $\omega(w)$ is constantly 1. Now suppose $w \to 1$ with $\operatorname{Im} w > 0$. Here $\phi(w) \to 0$ and $\omega(w)$ is constantly 0. Thus the claim.

To finish the proof of the lemma, let $\rho > 0$ such that $|\omega(z) - \phi(z)| < \varepsilon$ for z in $\operatorname{cl}\Omega$ and $|z - 1| < \rho$. If $z \in S_\delta$, then $\phi(z) \geq (1/2) - \delta/\pi$. Thus $\omega(z) = \omega(z) - \phi(z) + \phi(z) \geq (1/2) - \delta/\pi - \varepsilon$. □

Proof of Theorem 5.4. Without loss of generality we may assume that $a = 1$, $\alpha = 0$, and $|f(z)| \leq 1$ for $|z| < 1$. If $0 < r < 1$ there is a number $t_r < 1$ such that $|\gamma(t) - 1| < r$ for $t_r < t < 1$ and $|\gamma(t_r) - 1| = r$. Let γ_r denote the curve γ restricted to $[t_r, 1]$. If $\varepsilon > 0$, then r can be chosen so that $|f(\gamma(t))| < \epsilon$ for $t_r \leq t < 1$. Fix this value of r and let $\Omega = \mathbb{D} \cap B(1; r)$. As in the preceding lemma, let $I_1 = \{z \in \partial\Omega : \operatorname{Im} z \leq 0 \text{ and } |z| = 1\}$ and $I_2 = \{z \in \partial\Omega : \operatorname{Im} z \geq 0 \text{ and } |z| = 1\}$. For $k = 1, 2$ let ω_k be the solution of the Dirichlet problem with boundary values χ_{I_k}; so by Exercise 6, ω_k is continuous on $\operatorname{cl}\Omega$ except at the end points of the arc I_k.

Claim. For z in Ω, $-\log|f(z)| \geq -(\log\varepsilon)\min\{\omega_1(z), \omega_2(z)\}$.

Once this claim is proved, the theorem follows. Indeed, the preceding lemma implies that there is a ρ, $0 < \rho < r$, such that if $|z - 1| < \rho$, $0 < \delta < \pi/2$, and $z \in S_\delta$, then for $k = 1,\ 2$, $\omega_k(z) \geq (1/2) - \delta/\pi - \varepsilon$. (Observe that $\omega_2(w) = \omega_1(\overline{w})$.) Hence $-\log|f(z)| \geq -(\log\varepsilon)[(1/2) - \delta/\pi - \varepsilon]$ for $|z - 1| < \rho$ and $z \in S_\delta$. Therefore for such z, $|f(z)| < \epsilon \exp[(1/2) - \delta/\pi - \varepsilon]$, which can be made arbitrarily small.

To prove the claim, let $v(z) = (\log|f(z)|)/\log\varepsilon$; so v is a superharmonic function on Ω, $v(z) \geq 0$ for all z in Ω, and $v(\gamma(t)) > 1$ for $t_r < t < 1$. So if $z \in \gamma \cap \Omega$, then $v(z) \geq 1 \geq \omega_k(z)$. Suppose that $z \in \Omega \setminus \gamma$ and let U be the component of $\Omega \setminus \gamma$ that contains z. Let ζ_k be the end point of the arc I_k different from 1. Let σ_1 be the path that starts at 1, goes along $\partial\mathbb{D}$ in the positive direction to the point ζ_2, then continues along ∂B until it meets $\gamma(t_r)$. Similarly let σ_2 be the path that starts at $\gamma(t_r)$, goes along ∂B in the positive direction to the point ζ_1, then continues along $\partial\mathbb{D}$ in the positive direction to the point 1. Note that σ_1 and σ_2 together form the entirety of the boundary of Ω. Let $\Gamma_1 = \sigma_1 + \gamma_r$ and $\Gamma_2 = \sigma_2 - \gamma_r$. So $n(\Gamma_1; z) + n(\Gamma_2; z) = n(\partial\Omega; z) = 1$. Thus $n(\gamma_k; z) \neq 0$ for at least one value of $k = 1, 2$.

Suppose $n(\Gamma_1; z) \neq 0$. We now show that $\partial U \subseteq \Gamma_1$. In fact, general topology says that $\partial U \subseteq \partial(\Omega \setminus \gamma_r) = \gamma_r \cup \sigma_1 \cup \sigma_2 = \Gamma_1 \cup \sigma_2$. But if W is the unbounded component of $\mathbb{C} \setminus \Gamma_1$, the assumption that $n(\Gamma_1; z) \neq 0$ implies that $U \cap W = \emptyset$. Also $\sigma_2 \setminus \{1, \gamma(t_r)\} \subseteq W$. Thus $\partial U \subseteq \Gamma_1$.

This enables us to show that $v \geq \omega_1$ on U, and so, in particular, $v(z) \geq \omega_1(z)$. Indeed to show this we need only show that $\limsup_{w \to a} [\omega_1(w) - v(w)] \leq 0$ for all but a finite number of points on ∂U (Exercise 7). Suppose $a \in \partial U$ and $a \neq 1$ or $\gamma(t_r)$. By the preceding paragraph this implies that $a \in \gamma_r$ or $a \in \sigma_1$. If $a \in \gamma_r$ and $a \neq 1$ or $\gamma(t_r)$, then $a \in \mathbb{D}$ and $v(a) \geq 1 \geq \omega_1(a)$. If $a \in \sigma_1$ and $a \neq 1$ or $\gamma(t_r)$, then ω_1 is continuous at a and so $\omega_1(w) \to 0$ as $w \to a$. Since $v(w) \geq 0$, $\limsup_{w \to a} [\omega_1(w) - v(w)] \geq 0$.

In a similar way, if $n(\Gamma_1; z) \neq 0$, then $v(z) \geq \omega_2(z)$. This covers all the cases and so the claim is verified and the theorem is proved. □

Theorem 5.4 is called by some the *Sectorial Limit Theorem*.

Be careful not to think that this last theorem says more than it does. In particular, it does not say that the converse is true. The existence of a radial limit does not imply the existence of the limit along any arc approaching the same point of $\partial\mathbb{D}$. For example, if $f(z) = \exp[(z+1)/(z-1)]$, then f is analytic, $|f(z)| \leq 1$ for all z in $\mathbb{D}$, and $f(t) \to 0$ as $t \to 1-$. So the radial limit of f at $z = 1$ is 0. There are several ways of approaching 1 by a sequence of points (not along an arc) such that the values of f on this sequence approaches any point in cl $\mathbb{D}$.

We wish now to extend this notation of a non-tangential limit to regions other than the disk. To avoid being tedious, in the discussion below most of the details are missing and can be easily provided by the interested reader. For example if $g : \mathbb{D}_+ \to \mathbb{C}$ is a bounded analytic function, it is clear what is meant by non-tangential limits at points in $(-1, 1)$; and that the results about the disk given earlier can be generalized to conclude that g has non-tangential limit a.e. on $(-1, 1)$ and that if these limits are zero a.e. on a proper interval in $(-1, 1)$, then $g \equiv 0$ on $\mathbb{D}_+$.

If J is a free analyticity boundary arc of G and $f : G \to \mathbb{C}$ is a bounded analytic function, it is possible to discuss the non-tangential limits of $f(z)$ as z approaches a point of J. Indeed, it is possible to do this under less stringent requirements than analytically for J, but this is all we require and the discussion becomes somewhat simplified with this restriction. Recall (4.5) that if $a \in J$, there is a neighborhood U of a and a conformal equivalence $h : \mathbb{D} \to U$ such that $h(0) = 0$, $h(-1, 1) = U \cap J$, and $h(\mathbb{D}_+) = G \cap U$. For $0 < \alpha < \pi/2$ and t in $(-1, 1)$, let C be the partial cone $\{z \in \mathbb{D}_+ : \pi 2 < \arg(z - r) < \pi 2 \ + \alpha\}$ with vertex t. Since analytic functions preserve angles, $h(C)$ is a subset of U bounded by two arcs that approach $h(t)$ on the arc J at an angle with the tangent to J at $h(t)$. Say that $z \to h(t)$ *non-tangentially* if z converges to $h(t)$ while remaining in $h(C)$ for some angle α.

Since the Möbius transformation $(z-a)(1-\bar{a}z)$ maps $\mathbb{D}$ to $\mathbb{D}$, $\mathbb{D}_+$ to $\mathbb{D}_+$ and a to 0, it is not hard to see that the definition of non-tangential convergence to a point on J is independent of the choice of neighborhood U and conformal equivalence h. The details are left to the reader.

When we talk about subsets of the arc J having measure zero and the corresponding notion of almost everywhere occurrence, this refers to the arc length measure on J.

5.7 Theorem. *Let G be a region with J a free analytic boundary arc of G and let $f : G \to \mathbb{C}$ be a bounded analytic function.*

(a) *The function f has a non-tangential limit at a.e. point of J.*

(b) *If the non-tangential limit of f is 0 a.e. on a subarc of J, then $f \equiv 0$ on G.*

Proof. Let U and h be as in the discussion of the definition of non-tangential convergence above. Thus $f \circ h$ is a bounded analytic function on $\mathbb{D}_+$ and thus has non-tangential limits a.e. on $(-1,1)$. Clearly this implies that f has non-tangential limits a.e. on $J \cap U$. By covering U with a countable number of such neighborhoods, we have a proof of part (a). The proof of (b) is similar. □

5.8 Corollary. *If G is a analytic Jordan region and $f : G \to \mathbb{C}$ is a bounded analytic function, then f has non-tangential limits a.e. on ∂G.*

Exercises

1. If E is a closed subset of $\partial\mathbb{D}$ having measure 0, then $\partial\mathbb{D} \setminus E$ is an open set having length 2π.

2. If $\{E_k\}$ is a countable number of subsets of $\partial\mathbb{D}$ having measure 0, then $\bigcup_k E_k$ has measure 0.

3. Every countable subset of $\partial\mathbb{D}$ has measure 0.

4. Let $f : \mathbb{D} \to \mathbb{C}$ be defined by $f(z) = \exp\left[(z+1)/(z-1)\right]$. Show that f is analytic, $|f(z)| \le 1$ for all z in $\mathbb{D}$, and $f(t) \to 0$ as $t \to 1-$. If $|\zeta| \le 1$, find a sequence $\{z_n\}$ in $\mathbb{D}$ such that $z_n \to 1$ and $f(z_n) \to \zeta$.

5. Let f_1 and f_2 be bounded analytic functions on $\mathbb{D}$ and suppose f_j has a radial limit at each point of E_j, where $\partial\mathbb{D} \setminus E_j$ has measure 0. Show (by example) that $f_1 + f_2$ and $f_1 f_2$ may have radial limits at a set of points that properly contains $E_1 \cap E_2$.

6. (See Proposition 19.10.4) Let Γ be a rectifiable Jordan curve and let Ω be its inside. If $u : \Gamma \to \mathbb{R}$ is a bounded function that is continuous

except for a finite number of points, then there is a function $\hat{u} : \mathrm{cl}\,\Omega \to \mathbb{R}$ that is harmonic on Ω and continuous at every point of Γ at which the boundary function u is continuous.

7. (See Theorem 21.5.1.b) Let Γ be a rectifiable Jordan curve and let Ω be its inside. (Maximum Principle) If $u : \Omega \to \mathbb{R}$ is a subharmonic function that is bounded above and M is a constant such that $\limsup_{z \to a} u(z) \leq M$ for all but a finite number of points a in Γ, then $u \leq M$ on Ω.

Chapter 14

Conformal Equivalence for Simply Connected Regions

In this chapter a number of results on conformal equivalence for simply connected regions are presented. The first section discusses elementary information and examples. The next three sections present the basics of the theory of prime ends for the study of the boundary behavior of Riemann maps. This will be used in §5 to show that the Riemann map from the unit disk onto the inside of a Jordan curve can be extended to a homeomorphism on the closure of the disk. The chapter then closes with a discussion of the family of all functions that are one-to-one on a simply connected region.

§1 Elementary Properties and Examples

Recall that a conformal equivalence between two regions G and Ω in the complex plane is a one-to-one analytic function f defined on G with $f(G) = \Omega$. From the first volume we know that this implies that $f'(z) \neq 0$ for all z in G. If $f : G \to \mathbb{C}$ is an analytic function whose derivative never vanishes, then we know that f is not necessarily a conformal equivalence (the exponential function being the prime example). If $f'(z) \neq 0$ on G, it does follow, however, that f is locally one-to-one and f is conformal.

In this section the conformal equivalences of some of the standard regions will be characterized and some particular examples will be examined. A slightly weaker version of the first result appeared as Exercise 12.4.2.

1.1 Proposition. *If f is a conformal equivalence from $\mathbb{C}$ onto a subset of $\mathbb{C}$, then $f(z) = az + b$ with $a \neq 0$. In particular, the only conformal equivalences of $\mathbb{C}$ onto itself are the Möbius transformations of the form $f(z) = az + b$ with $a \neq 0$.*

Proof. Clearly every such Möbius transformation is a conformal equivalence of $\mathbb{C}$ onto itself. So assume that $f : \mathbb{C} \to \mathbb{C}$ is a conformal equivalence onto $f(\mathbb{C})$. Since $f(\mathbb{C})$ is simply connected, $f(\mathbb{C}) = \mathbb{C}$. First it will be shown that $f(z) \to \infty$ as $z \to \infty$. Note that this says that f has a pole at infinity and hence f must be a polynomial (Exercise 5.1.13). Since f is a conformal equivalence, it follows that f has degree 1 and thus has the desired form.

If either $\lim_{z\to\infty} f(z)$ does not exist or if the limit exists and is finite, then there is a sequence $\{z_n\}$ in $\mathbb{C}$ such that $z_n \to \infty$ and $f(z_n) \to \alpha$, an element of $\mathbb{C}$. But $f^{-1} : \mathbb{C} \to \mathbb{C}$ is continuous and so $z_n = f^{-1}(f(z_n)) \to f^{-1}(\alpha) \neq \infty$, a contradiction. □

To say that a function is *analytic in a neighborhood of infinity* means that there is an $R > 0$ such that f is analytic in $\{z : |z| > R\}$. For such a function f, $f(z^{-1})$ has an isolated singularity at 0. Thus the nature of the singularity of f at ∞ can be discussed in terms of the nature of the singularity of $f(z^{-1})$ at 0. In particular, f has a removable singularity at ∞ if f is bounded near ∞ and $f(\infty) = \lim_{z\to\infty} f(z)$. If f has a removable singularity at infinity, we will say that f is analytic at ∞. Similarly f has a pole at ∞ if $\lim_{z\to\infty} f(z) = \infty$. In the case of a pole we might say that $f(\infty) = \infty$ and think of f as a mapping of a neighborhood of ∞ in the extended plane to a neighborhood of ∞. The order of a pole at ∞ is the same as the order of the pole of $f(z^{-1})$ at 0.

1.2 Corollary. *If $f : \mathbb{C}_\infty \to \mathbb{C}_\infty$ is a homeomorphism that is analytic on $\mathbb{C}_\infty \setminus \{f^{-1}(\infty)\}$, then f is a Möbius transformation.*

Proof. If $f(\infty) = \infty$, then this result is immediate from the preceding proposition. If $f(\infty) = \alpha \neq \infty$, then $g(z) = (f(z) - \alpha)^{-1}$ is a homeomorphism of $\mathbb{C}_\infty$ onto itself and $g(\infty) = \infty$. Thus the corollary follows. □

1.3 Example. If $\Omega = \mathbb{C} \setminus (-\infty, -r]$ for some $r > 0$, then

$$f(z) = \frac{4rz}{(1-z)^2}$$

is a conformal equivalence of $\mathbb{D}$ onto Ω, $f(0) = 0$, and $f'(0) = 4r$. Thus f is the unique conformal equivalence having these properties.

The uniqueness is, of course, a consequence of the uniqueness statement in the Riemann Mapping Theorem. To show that f has the stated mapping properties, let's go through the process of finding the Riemann map.

Note that the Möbius transformation $f_1(z) = (1+z)(1-z)^{-1}$ maps $\mathbb{D}$ onto $\Omega_1 \equiv \{z : \operatorname{Re} z > 0\}$, $f_1(0) = 1$, $f_1(1) = \infty$, and $f_1(-1) = 0$. Now $f_2(z) = z^2$ maps Ω_1 onto $\Omega_2 \equiv \mathbb{C} \setminus (-\infty, 0]$; $f_3(z) = r(z-1)$ maps Ω_2 onto Ω. The map f above is the composition of these three maps.

Note that the function f in Example 1.3 has a pole of order 2 at $z = 1$ and f has a removable singularity at infinity. In fact $f(\infty) = 0$. Moreover, $f'(\infty) = \lim_{z\to\infty} zf(z) = 4r > 0$. Since $f'(z) = 0$ if and only if $z = -1$, we see that f is conformal on $\mathbb{C} \setminus \{\pm 1\}$.

The next example is more than that.

1.4 Example. For $|\alpha| = 1$ define

$$f(z) = \frac{z}{(1-\alpha z)^2}$$

for z in $\mathbb{D}$. To facilitate the discussion, denote this map by f_α to emphasize its dependence on the parameter α. The function f_1 is a special case of the preceding example and thus maps $\mathbb{D}$ onto $\mathbb{C} \setminus (-\infty, -1/4]$. For an arbitrary α, f_α is the composition of the rotation of the disk by α, followed by f_1, followed by a rotation of $\mathbb{C}$ by $\overline{\alpha}$. Explicitly, $f_\alpha(z) = \overline{\alpha} f_1(\alpha z)$. Thus $f_\alpha(\mathbb{D}) = \mathbb{C} \setminus \{-t\overline{\alpha}/4 : 0 \le t < \infty\}$.

The power series representation of this function is given by

$$\frac{z}{(1-\alpha z)^2} = z + 2\alpha z^2 + 3\alpha^2 z^3 + \cdots = \sum_{n=1}^{\infty} n\alpha^{n-1} z^n.$$

This will be of significance when we discuss the Bieberbach Conjecture in Chapter 17. Also see (7.5) in this chapter.

The function in Example 1.4 for $\alpha = 1$ is called the *Koebe function* and the other functions for arbitrary α are called the rotations of the Koebe function.

The next example was first seen as Exercise 3.3.13. The details are left to the reader.

1.5 Example. If $\Omega = \mathbb{C}_\infty \setminus [-2, 2]$,

$$f(z) = z + \frac{1}{z}$$

is a conformal equivalence of $\mathbb{D}$ onto Ω with $f(0) = \infty$. If g is any other such mapping, then $g(z) = f(e^{i\theta} z)$ for some real constant θ.

Also note that $f(z) = f(z^{-1})$ so that f maps the exterior (in $\mathbb{C}_\infty$) of the closed disk onto Ω.

The next collection of concepts and results applies to arbitrary regions, not just those that are simply connected. They are gathered here because they will be used in this chapter, but they will resurface in later chapters as well.

1.6 Definition. If G is an open subset of $\mathbb{C}$ and $f : G \to \mathbb{C}$ is any function, then for every point a in $\partial_\infty G$ the *cluster* set of f at a is defined by

$$\text{Clu}(f; a) \equiv \cap \{\text{cl}_\infty [f(B(a;\varepsilon) \cap G)] : \varepsilon > 0\}.$$

1.7 Proposition. *For every function f, $\text{Clu}(f; a)$ is a non-empty compact subset of $\mathbb{C}_\infty$. If f is a bounded function, $\text{Clu}(f; a)$ is a compact subset of $\mathbb{C}$.*

Proof. In fact, the sets $\text{cl}_{\infty}[f(B(a;\varepsilon)\cap G)]$ form a decreasing collection of compact subsets of $\mathbb{C}_\infty$ and must have non-empty intersection. The statement about bounded functions is clear. □

1.8 Proposition. *If $a \in \partial_\infty G$ such that there is a $\rho > 0$ for which $G \cap B(a;r)$ is connected for all $r < \rho$ and f is continuous, then* $\text{Clu}(f;a)$ *is a compact connected subset of* $\mathbb{C}_\infty$.

Proof. In this situation the sets $\text{cl}_\infty[f(B(b;\varepsilon)\cap G)]$ form a decreasing collection of compact connected subsets of $\mathbb{C}$ when $\varepsilon < \rho$. The result is now immediate from an elementary result of point set topology. □

The proof of the next proposition is left to the reader (Exercise 5).

1.9 Proposition. *If $a \in \partial_\infty G$, then $\zeta \in$* $\text{Clu}(f;a)$ *if and only if there is a sequence $\{a_n\}$ in G such that $a_n \to a$ and $f(a_n) \to \zeta$.*

1.10 Corollary. *If $a \in \partial_\infty G$, then the limit of $f(z)$ exists as $z \to a$ with z in G if and only if* $\text{Clu}(f;a)$ *is a single point.*

1.11 Proposition. *If $f : G \to \Omega$ is a homeomorphism and $a \in \partial_\infty G$, then* $\text{Clu}(f;a) \subseteq \partial_\infty \Omega$.

Proof. If $\zeta \in \text{Clu}(f;a)$, let $\{a_n\}$ be a sequence in G such that $a_n \to a$ and $f(a_n) \to \zeta$. Clearly $\zeta \in \text{cl}_\infty\Omega$ and if $\zeta \in \Omega$, then the fact that f^{-1} is continuous at ζ implies that $a = \lim_n a_n = \lim_n f^{-1}(f(a_n)) = f^{-1}(\zeta) \in G$, a contradiction. □

We end this section with some widely used terminology.

1.12 Definition. A function on an open set is *univalent* if it is analytic and one-to-one.

Exercises

1. In Example 1.3, what is $f(\mathbb{C}\setminus\mathbb{D})$? $f(\mathbb{C}_\infty\setminus\{\pm 1\})$?
2. Discuss the image of $\partial\mathbb{D}$ under the map $f(z) = z + z^{-1}$.
3. Find a conformal equivalence of $\mathbb{C}_\infty \setminus [-2,2]$ onto $\mathbb{D}$.
4. Give the details of the proof of Proposition 1.8.
5. Prove Proposition 1.9.
6. What is the cluster set of $f(z) = \exp\{(z+1)/(z-1)\}$ at 1?

7. Characterize the conformal equivalences of the upper half plane $\mathbb{H} = \{z : \text{Im } z > 0\}$ onto itself.

8. Characterize the conformal equivalences of the punctured disk onto itself.

9. Characterize the conformal equivalences of the punctured plane onto itself.

§2 Crosscuts

With this section we begin the study of the boundary behavior of a conformal equivalence $\tau : \mathbb{D} \to \Omega$. Much of the discussion here is based on the book of Pommerenke [1975], which has additional material.

We will limit ourselves to the case of a bound region (or bounded Riemann map) as this facilitates the proofs. The reader can consult the literature for the general case.

2.1 Definition. If G is a bounded simply connected region in $\mathbb{C}$, and C_1 is a closed Jordan arc whose end points lie on ∂G and such that $C = C_1$ with its end points deleted lies in G, then C is called a *crosscut* of G.

Usually no distinction will be made between a crosscut C as a curve or its trace. In other words, C may be considered as a set of points or as a parameterized curve $C : (0, 1) \to G$. Recall that C is a Jordan arc if $C(s) \neq C(t)$ for $0 < s < t < 1$. It is possible, however, that $C_1(0) = C_1(1)$ so that C_1 is a Jordan curve. At the risk of confusing the reader, we will not make a distinction between a crosscut C and the corresponding closed Jordan arc C_1. This will have some notational advantages that the reader may notice in the exposition.

Note that if C is a crosscut of G and $f : G \to \mathbb{C}$ is a continuous function, then $\text{cl}_\infty [f(C)] \setminus f(C) \subseteq \text{Clu}(f; a_1) \cup \text{Clu}(f; a_2)$, where a_1 and a_2 are the end points of C.

2.2 Lemma. *If G is a bounded simply connected region in $\mathbb{C}$ and C is a crosscut of G, then $G \setminus C$ has two components and the portion of the boundary of each of these components that lies in G is C.*

Proof. If $\phi : G \to \mathbb{D}$ is a Riemann map, then $f(z) = \phi(z)/(1 - |\phi(z)|)$ is a homeomorphism of G onto $\mathbb{C}$. Hence $f(C)$ is a Jordan arc in $\mathbb{C}$. By Proposition 1.11, $\text{Clu}(f; a) \subseteq \partial_\infty \mathbb{C} = \{\infty\}$ for every point a in ∂G. Hence, by the remark preceding this proposition, $\text{cl}_\infty f(C)$ is a Jordan curve in $\mathbb{C}_\infty$ passing through ∞. If Ω_1 and Ω_2 are the components of $\mathbb{C} \setminus \text{cl}_\infty f(C) = \mathbb{C} \setminus f(C)$, then $f^{-1}(\Omega_1)$ and $f^{-1}(\Omega_2)$ are the components of $G \setminus C$. □

It is now necessary to choose a distinguished point in G. In the following definitions and results, this distinguished point is lurking in the background as part of the scenery and we must forever be aware of its existence. Indeed, the definitions depend on the choice of some distinguished point. It seems wise, however, not to make this point part of the foreground by including it in the notation. We do this by always assuming that $0 \in G$. The assumption that G is bounded will also cease to be made explicit.

The preceding lemma justifies the next definition.

2.3 Definition. If G is a simply connected region containing 0, then for any crosscut C of G that does not pass through 0, let out C denote the component of $G \setminus C$ that contains 0 and let ins C denote the other component. Call out C the *outside* of C and ins C the *inside* of C.

This definition and notation is, of course, in conflict with previous concepts concerning Jordan curves (§13.1). We'll try to maintain peace here by reserving small Greek letters, like σ and γ, for Jordan curves, and capital Roman letters, like X and C, for crosscuts.

From now on we will only consider crosscuts of G that do not pass through the distinguished point 0.

2.4 Definition. A *zero-chain* (or 0-chain) of G is a sequence of crosscuts of G, $\{C_n\}$, having the following properties:

(a) $\text{ins } C_{n+1} \subseteq \text{ins } C_n$;

(b) $\text{cl } C_n \cap \text{cl } C_m = \emptyset$ for $n \neq m$;

(c) $\text{diam } C_n \to 0$ as $n \to \infty$.

Note that the condition in the definition of a 0-chain $\{C_n\}$ that $\text{cl } C_n \cap \text{cl } C_{n+1} = \emptyset$ and $\text{ins } C_{n+1} \subseteq \text{ins } C_n$ precludes the possibility that $\text{cl } C_n \cap \partial G$ is a single point. It is not hard to construct a zero-chain $\{C_n\}$ such that $\inf_n[\text{diam}(\text{ins } C_n)] > 0$. See the examples below.

Why make this definition? Let Ω be a bounded simply connected region and let $\tau : \mathbb{D} \to \Omega$ be a conformal equivalence with $\tau(0) = 0$. We are interested in studying the behavior of $\tau(z)$ as z approaches a point of $\partial\mathbb{D}$. Let $a \in \partial\mathbb{D}$ and construct a 0-chain $\{X_n\}$ in $\mathbb{D}$ such that $\bigcap_n \text{cl } (\text{ins } X_n) = \{a\}$. Clearly $\tau(X_n)$ is an open Jordan arc in Ω. By Proposition 1.11 $[\text{cl } \tau(X_n)] \setminus X_n \subseteq \partial\Omega$. Unfortunately it is not necessarily true that $C_n = \tau(X_n)$ is a crosscut in Ω since $\text{cl } C_n \setminus C_n$ may be an infinite set. We can and will, however, choose the 0-chain $\{X_n\}$ in $\mathbb{D}$ so that not only is each C_n a crosscut in Ω, but $\{C_n\}$ is actually a 0-chain in Ω.

In this way we associate with each point a of $\partial\mathbb{D}$ a 0-chain $\{C_n\}$ in Ω. In fact, we will see in the next section that after we introduce an equivalence relation on the set of 0-chains, there is a way of topologizing $\hat{\Omega}$, the set Ω

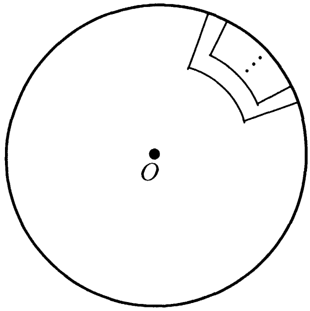

Figure 14.1

together with these equivalence classes, so that τ extends to a homeomorphism $\hat{\tau}$ of cl $\mathbb{D}$ onto $\hat{\Omega}$. This will pave the way for us to study the boundary behavior of τ in future sections when more stringent restrictions are placed on $\partial\Omega$.

The following are some examples of 0-chains. Figure 14.1 has $\Omega = \mathbb{D}$ and shows an example of a 0-chain. Some special 0-chains of this type will be constructed below (see Proposition 2.9).

In Figure 14.2, Ω is a slit disk and the sequence of crosscuts is not a 0-chain since it fails to satisfy property (a) of the definition. Again for Ω a slit disk, Figure 14.3 illustrates a 0-chain.

In Figures 14.4, 14.5, and 14.6, Ω is an open rectangle less an infinite sequence of vertical slits of the same height that converge to the segment $[2, 2+i]$. The sequence of crosscuts $\{C_n\}$ in Figure 14.4 is not a 0-chain since it violates part (a) of the definition; the crosscuts in Figure 14.5 do not form a 0-chain since their diameters do not converge to 0. The crosscuts in Figure 14.6 do form a 0-chain.

Figures 14.7, 14.8, and 14.9 illustrate examples of 0-chains.

We begin with a result on 0-chains of $\mathbb{D}$ that may seem intuitively obvious, but which requires proof. It might be pointed out that the sequence of crosscuts in Figure 14.6 does not satisfy the conclusion of the next proposition.

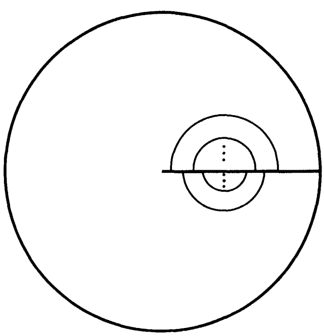

Figure 14.2

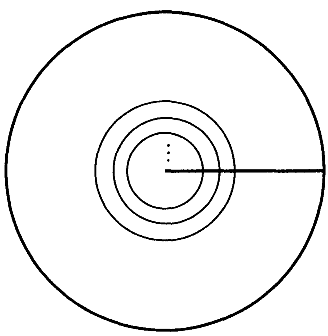

Figure 14.3

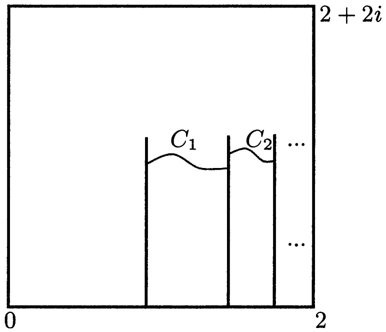

Figure 14.4

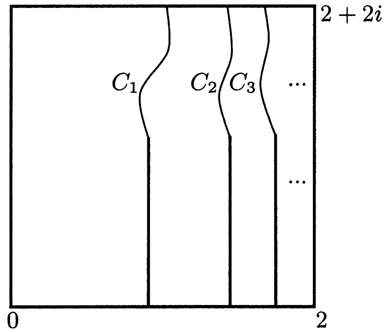

Figure 14.5

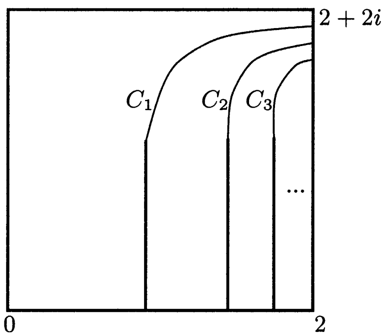

Figure 14.6

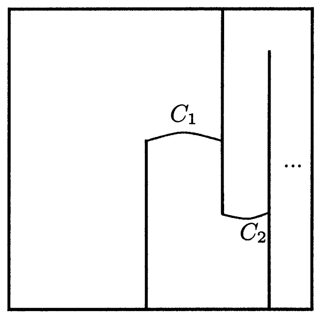

Figure 14.7

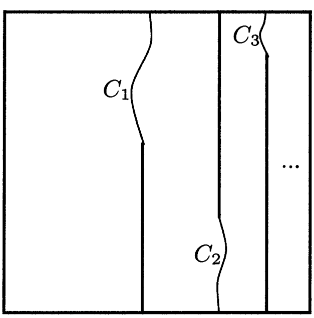

Figure 14.8

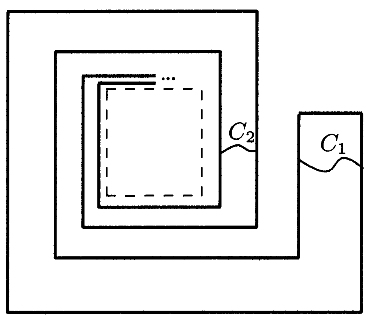

Figure 14.9

2.5 Proposition. *If* $\{X_n\}$ *is a sequence of crosscuts of* $\mathbb{D}$ *with* $\operatorname{ins} X_{n+1} \subseteq \operatorname{ins} X_n$ *and* $\operatorname{diam} X_n \to 0$, *then* $\operatorname{diam}\,(\operatorname{ins} X_n) \to 0$.

Proof. Since cl [ins X_{n+1}] $\subseteq$ cl [ins X_n] for every n, $K \equiv \bigcap_n$ cl [ins X_n] is a non-empty compact connected subset of cl $\mathbb{D}$. Since X_n is a crosscut,

$\partial[\text{ins } X_n] = X_n \cup \gamma_n$ for the closed arc in $\partial\mathbb{D}$, $\gamma_n = \text{cl }[\text{ins } X_n] \cap \partial\mathbb{D}$. In fact, $\gamma_{n+1} = \text{cl }[\text{ins } X_{n+1}] \cap \partial\mathbb{D} \subseteq \gamma_n$ and so $\gamma \equiv \bigcap_n \gamma_n$ is a closed arc in $\partial\mathbb{D}$. It is easy to see that $\gamma = K \cap \partial\mathbb{D}$. Also, since X_n and γ_n have the same end points, diam $\gamma_n \to 0$. Therefore, γ, and hence K, is a single point z_0 in $\partial\mathbb{D}$. A straightforward argument now finishes the proof. □

The remainder of this section is devoted to the construction of 0-chains $\{X_n\}$ in $\mathbb{D}$ such that $\{\tau(X_n)\}$ is a 0-chain in Ω. The process involves the proof of a sequence of lemmas. The first of these necessitates a return to the notion of a set of measure 0. The proof will not be given of the complete statement, but only of the statement that can be obtained by the deletion of any reference to a set of measure 0. The proof of the general statement can be easily obtained from this proof and is left to the reader.

2.6 Lemma. *Let τ be a bounded univalent function defined on $\mathbb{D}$ with $|\tau| \le M$ on $\mathbb{D}$. If E is a subset of $\partial\mathbb{D}$ having measure 0, $1/2 < \rho < 1$, and $0 \le \alpha < \beta \le 2\pi$, then there is a θ with $\alpha < \theta < \beta$ such that $e^{i\theta} \notin E$ and*

$$\int_\rho^1 \left|\tau'(re^{i\theta})\right| dr \le M\sqrt{2\pi}\sqrt{\frac{1-\rho}{\beta-\alpha}}.$$

Proof. In fact an application of the Cauchy-Schwartz Inequality shows that

$$\begin{aligned}\int_\alpha^\beta \left[\int_\rho^1 \left|\tau'(re^{i\theta})\right| dr\right]^2 &\le \int_\alpha^\beta \left[\int_\rho^1 dr\right]\left[\int_\rho^1 \left|\tau'(re^{i\theta})\right|^2 dr\right] d\theta \\ &\le \int_0^{2\pi} (1-\rho)\left[\int_\rho^1 \left|\tau'(re^{i\theta})\right|^2 dr\right] d\theta \\ &< 2(1-\rho)\int_0^{2\pi}\int_{1/2}^1 \left|\tau'(re^{i\theta})\right|^2 r\,dr\,d\theta\end{aligned}$$

since, with $\rho > 1/2$, we have $1 < 2r$ for $\rho \le r$. But by Theorem 13.2.12 this last integral equals Area$(\tau(\{z \in \mathbb{D} : 1/2 < |z| < 1\})) \le \pi M^2$. But

$$(\beta-\alpha)\inf\left\{\left[\int_\rho^1 \left|\tau'(re^{i\theta})\right| dr\right]^2 : \alpha < \theta < \beta\right\} \le \int_\alpha^\beta \left[\int_\rho^1 \left|\tau'(re^{i\theta})\right| dr\right]^2 d\theta.$$

Hence there is at least one value of θ with

$$\left[\int_\rho^1 \left|\tau'(re^{i\theta})\right| dr\right]^2 \le \frac{2(1-\rho)\pi M^2}{\beta-\alpha},$$

whence the lemma. □

Note that the preceding lemma gives a value θ such that $r \to \tau(re^{i\theta})$, $\rho \le r < 1$, is a half open rectifiable Jordan arc (and also gives an estimate on the length of this curve). But a rectifiable open arc cannot wiggle too much, and so the next result is quite intuitive.

2.7 Proposition. *If $\gamma : [0,1) \to G$ is a half open rectifiable arc and $f : G \to \mathbb{C}$ is an analytic function such that $f \circ \gamma$ is also a rectifiable arc, then $\lim_{t\to 1-} f(\gamma(t))$ exists and is finite.*

Proof. Let $L =$ the length of $f \circ \gamma$. Since the shortest distance between two points is a straight line, it follows that $\{f \circ \gamma\} \subseteq \overline{B}\left(f\left(\gamma(1/2)\right); L\right)$. In particular, there is a constant M such that $|f\left(\gamma(t)\right)| \leq M$ for $0 < t < 1$. If $\lim_{t\to 1-} f\left(\gamma(t)\right)$ does not exist, then there are sequences $\{r_n\}$ and $\{s_n\}$ in $(0,1)$ such that $r_n < s_n < r_{n+1}$, $r_n \to 1$ and $s_n \to 1$, $f\left(\gamma(r_n)\right) \to \rho$ and $f\left(\gamma(s_n)\right) \to \sigma$, and $\rho \neq \sigma$. If $0 < 3\delta < |\rho - \sigma|$, then there is an n_0 such that $|f\left(\gamma(r_n)\right) - f\left(\gamma(s_n)\right)| > \delta$ for all $n \geq n_0$. But the length of the path $\{f\left(\gamma(t)\right) : r_n \leq t \leq s_n\}$ is greater than or equal to $|f\left(\gamma(r_n)\right) - f\left(\gamma(s_n)\right)| \geq \delta$. This contradicts the rectifiability of $f \circ \gamma$. □

A combination of the last two results can be used to give a proof that the set of points in $\partial\mathbb{D}$ at which a bounded conformal equivalence τ has radial limits is dense in $\partial\mathbb{D}$. Unfortunately this will not suffice for our purposes and we need more.

2.8 Lemma. *If $\tau : \mathbb{D} \to \mathbb{C}$ is a bounded univalent function, then $(1 - r)\max\{|\tau'(z)| : |z| = r\} \to 0$ as $r \to 1$. Hence $(1 - |z|)\,|\tau'(z)| \to 0$ as $|z| \to 1$.*

Proof. Let $\tau(z) = \sum_n a_n z^n$; hence, using the fact that $|x+y|^2 \leq 2\left(|x|^2 + |y|^2\right)$, we get that

$$(1-|z|)^2\,|\tau'(z)|^2 = (1-|z|)^2\left|\sum_{n=1}^{m-1} n a_n a^{n-1} + \sum_{n=m}^{\infty} n a_n z^{n-1}\right|^2$$

$$\leq 2\,(1-|z|)^2\left|\sum_{n=1}^{m-1} n a_n z^{n-1}\right|^2 + 2\,(1-|z|)^2\left|\sum_{n=m}^{\infty} n a_n z^{n-1}\right|^2.$$

Applying the Cauchy-Schwartz Inequality to the second sum gives

$$\begin{aligned}\left|\sum_{n=m}^{\infty} n a_n z^{n-1}\right|^2 &= \left|\sum_{n=m}^{\infty}\left(\sqrt{n}a_n\right)\left(\sqrt{n}z^{n-1}\right)\right|^2 \\ &\leq \left[\sum_{n=m}^{\infty} n\,|a_n|^2\right]\left[\sum_{n=m}^{\infty} n\,|z|^{2n-2}\right].\end{aligned}$$

But

$$\sum_{n=m}^{\infty} n\,|z|^{2n-2} \leq \sum_{n=1}^{\infty} n\,|z|^{2n-2}$$

$$= \left(1-|z|^2\right)^{-2}.$$

Combining these inequalities gives

$$\begin{aligned}
(1-|z|)^2\left|\tau'(z)\right|^2 &\leq 2(1-|z|)^2\left|\sum_{n=1}^{m-1} n\left|a_n\right||z|^{n-1}\right|^2 \\
&\quad +2(1-|z|)^2\left[\sum_{n=m}^{\infty} n\left|a_n\right|^2\right]\frac{1}{\left(1-|z|^2\right)^2} \\
&= 2(1-|z|)^2\left|\sum_{n=1}^{m-1} n\left|a_n\right||z|^{n-1}\right|^2 \\
&\quad +\frac{2}{(1+|z|)^2}\left[\sum_{n=m}^{\infty} n\left|a_n\right|^2\right].
\end{aligned}$$

But if $|z| \geq 1/2$, then

$$\begin{aligned}
(1-|z|)^2\left|\tau'(z)\right|^2 &\leq 2(1-|z|)^2\left|\sum_{n=1}^{m-1} n\left|a_n\right||z|^{n-1}\right|^2 \\
&\quad +2\left[\sum_{n=m}^{\infty} n\left|a_n\right|^2\right].
\end{aligned}$$

By Corollary 13.2.13, m can be chosen such that the last summand is smaller than $\varepsilon^2/2$. Thus for $1-|z|$ sufficiently small, $(1-|z|)^2\left|\tau'(z)\right|^2$ can be made arbitrarily small. □

2.9 Proposition. *Let Ω be a bounded simply connected region and let $\tau : \mathbb{D} \to \Omega$ be the Riemann map with $\tau(0)=0$ and $\tau'(0)>0$. If $0 \leq \theta \leq 2\pi$ and $\{r_n\}$ is a sequence of positive numbers that converges monotonically to 1, then for every n there are α_n and β_n with $\alpha_n < \theta < \beta_n$, $\alpha_n \to \theta$, and $\beta_n \to \theta$ such that if $Y_n =$ the crosscut of $\mathbb{D}$ defined by*

$$Y_n = \left(e^{i\alpha_n}, r_n e^{i\alpha_n}\right] \cup \left\{r_n e^{it} : \alpha_n \leq t \leq \beta_n\right\} \cup \left[r_n e^{i\beta_n}, e^{i\beta_n}\right),$$

then $\{\tau(Y_n)\}$ is a 0-chain in Ω.

Proof. Let M be an upper bound for $|\tau|$. By Lemma 2.8, positive numbers $\{\varepsilon_n\}$ can be chosen such that $\varepsilon_n \to 0$ and

2.10 $$(1-r_n)\max_t\left|\tau'\left(r_n e^{it}\right)\right| < \varepsilon_n^3.$$

Now apply Lemma 2.6 with $\alpha = \theta - 2(1-r_n)/\varepsilon_n^2$ and $\beta = \theta(1-r_n)/\varepsilon_n^2$ to obtain an α_n with $\theta - 2(1-r_n)/\varepsilon_n^2 < \alpha_n < \theta - (1-r_n)/\varepsilon_n^2$ and

$$\int_{r_n}^{1}\left|\tau'\left(re^{i\alpha_n}\right)\right| dr < M\sqrt{2\pi}\sqrt{\frac{1-r_n}{(1-r_n)/\varepsilon_n^2}} = \varepsilon_n M\sqrt{2\pi}.$$

Similarly, there is a β_n with $\theta < \beta_n$ and

$$\int_{r_n}^{1} \left|\tau'\left(re^{i\beta_n}\right)\right| dr < \varepsilon_n M\sqrt{2\pi}.$$

Actually we want to modify the choice of the points α_n and β_n so that the values of $\tilde{\tau}\left(e^{i\alpha_n}\right)$ and $\tilde{\tau}\left(e^{i\beta}\right)$, the radial limits of τ at the designated points, are all different. This is done as follows. Suppose $\alpha_1, \ldots, \alpha_{n-1}, \beta_1, \ldots, \beta_{n-1}$ have been chosen and let $E = \{\alpha : \theta - 2(1-r_n)/\varepsilon_n^2 < \alpha < \theta - (1-r_n)/\varepsilon_n^2$ and $\tilde{\tau}\left(e^{i\alpha}\right)$ exists and equals one of $\tilde{\tau}\left(e^{i\alpha_1}\right)$, ..., $\tilde{\tau}\left(e^{i\ alpha_{n-1}}\right)$, $\tilde{\tau}\left(e^{i\beta_1}\right)$, ..., $\tilde{\tau}\left(e^{i\beta_{n-1}}\right)\}$. By Theorem 13.5.3, E has measure 0. By Lemma 2.6, α_n can be chosen in the prescribed interval with $\alpha_n \notin E$. By Proposition 2.7, $\tilde{\tau}\left(e^{i\alpha_n}\right)$ exists and differs from $\tilde{\tau}\left(e^{i\alpha_1}\right), \ldots, \tilde{\tau}\left(e^{i\alpha_{n-1}}\right)$, $\tilde{\tau}\left(e^{i\beta_1}\right)$, ..., $\tilde{\tau}\left(e^{i\beta_{n-1}}\right)$. Similarly choose β_n.

Define Y_n as in the statement of the proposition. Clearly $\{Y_n\}$ is a 0-chain in $\mathbb{D}$. Let $C_n = \tau(Y_n)$. Since C_n is rectifiable, each C_n is a crosscut of Ω. Since ins $Y_{n+1} \subseteq$ ins Y_n for every n, ins $C_{n+1} \subseteq$ ins C_n for every n. Since the values $\tilde{\tau}\left(e^{i\alpha_1}\right)$, $\tilde{\tau}\left(e^{i\alpha_2}\right)$, ...; $\tilde{\tau}\left(e^{i\beta_1}\right)$, $\tilde{\tau}\left(e^{i\beta_2}\right)$, ... are all distinct and cl $Y_n \cap$ cl $Y_{n+1} = \emptyset$, cl $C_n \cap$ cl $C_{n+1} = \emptyset$. It remains to prove that diam $C_n \to 0$.

Now (2.10) implies that $\int_{\alpha_n}^{\beta_n} \left|\tau'\left(r_n e^{it}\right)\right| dt \leq (\beta_n - \alpha_n) \frac{\varepsilon_n^2}{1-r_n}$ and $\beta_n - \alpha_n \leq 4(1-r_n)/\varepsilon_n^2$. Hence $\int_{\alpha_n}^{\beta_n} \left|\tau'\left(r_n e^{it}\right)\right| dt \leq 4\varepsilon_n$. This, combined with the preceding estimates, implies that the length of $\tau(Y_n) \leq 4\varepsilon_n + 2\varepsilon_n M (2\pi)^{\frac{1}{2}}$ and thus converges to 0. It is left as an exercise to show that α_n and $\beta_n \to \theta$ as $n \to \infty$. □

Exercise

1. If $\{X_n\}$ is a 0-chain in $\mathbb{D}$, show that $\bigcap_n$ cl (ins X_n) is a single point in $\partial\mathbb{D}$.

§3 Prime Ends

Maintain the notion of the preceding section. Let Ω be a bounded simply connected region and let $\tau : \mathbb{D} \to \Omega$ be the conformal equivalence with $\tau(0) = 0$ and $\tau'(0) > 0$.

3.1 Definition. If $\{C_n\}$ and $\{C_n'\}$ are two zero-chains in Ω, say that they are *equivalent* if for every n there is an m such that ins $C_m \subseteq$ ins C_n' and, conversely, for every i there is a j with ins $C_j' \subseteq$ ins C_i.

It is easy to see that this concept of equivalence for zero-chains in Ω

is indeed an equivalence relation. A *prime end* is an equivalence class of zero-chains.

An examination of the 0-chain in Figure 14.1 will easily produce other 0-chains that are equivalent to the one given. In Figure 14.2 if the crosscuts that are above the slit constitute one 0-chain and the crosscuts below the slit constitute another, then these two 0-chains belong to different prime ends. It can also be seen that the 0-chains appearing in Figures 14.7 and 14.8 are equivalent. The reader is invited to examine the 0-chains appearing in the figures in §2 and to find equivalent ones.

Let $\hat{\Omega}$ denote Ω together with the collection of prime ends. We now want to put a topology on $\hat{\Omega}$. (Apologies to the reader for this notation, which is rather standard but opens up the possibility of confusion with the polynomially convex hull.)

3.2 Definition. Say that a subset U of $\hat{\Omega}$ is *open* if $U \cap \Omega$ is open in Ω and for every p in $U \setminus \Omega$ there exists a zero-chain $\{C_n\}$ in p such that there is an integer n with ins $C_n \subseteq U \cap \Omega$.

Note that from the definition of equivalence and the definition of a 0-chain, if U is an open subset of $\hat{\Omega}$ and $p \in U$, then for every $\{C_n\}$ in p, ins $C_n \subseteq U \cap \Omega$ for all sufficiently large n.

The proof of the next proposition is an exercise.

3.3 Proposition. *The collection of open subsets of $\hat{\Omega}$ is a topology.*

The main result of the section is the following.

3.4 Theorem. *If Ω is a bounded simply connected region in $\mathbb{C}$ and $\tau : \mathbb{D} \to \Omega$ is a conformal equivalence, then τ extends to a homeomorphism of cl $\mathbb{D}$ onto $\hat{\Omega}$.*

Actually, we will want to make specific the definition of $\hat{\tau}(z)$ for every z in $\partial\,\mathbb{D}$ as well as spell out the meaning of the statement that $\hat{\tau}$ is a homeomorphism. If $z \in \partial\,\mathbb{D}$, then Proposition 2.9 implies there is a 0-chain $\{Y_n\}$ in $\mathbb{D}$ such that $\bigcap_n \text{cl }(\text{ins } Y_n) = \{z\}$ and $\{\tau\,(Y_n)\}$ is a 0-chain in Ω. We will define $\hat{\tau}(z)$ to be the equivalence class of $\{\tau\,(Y_n)\}$. We must show that $\hat{\tau}$ is well defined. Thus if $\{X_n\}$ is a second 0-chain in $\mathbb{D}$ with $\bigcap_n \text{cl }(\text{ins } X_n) = \{z\}$ and $\{\tau\,(X_n)\}$ a 0-chain in Ω, we must show that $\{\tau\,(X_n)\}$ and $\{\tau\,(Y_n)\}$ are equivalent. This is not difficult. Fix n; we want to show that ins $\tau\,(Y_m) \subseteq$ ins $\tau\,(X_n)$ for some m. But just examine ins X_n: $\partial\,(\text{ins } X_n) = X_n \cup \gamma_n$, where γ_n is an arc $\partial\,\mathbb{D}$ with z as an interior point. Thus there is a $\delta > 0$ such that $\mathbb{D} \cap B(z;\delta) \subseteq$ ins X_n. Since diam (ins Y_m) $\to 0$ (why?), there is an m with ins $Y_m \subseteq \mathbb{D} \cap B\,(z;\delta) \subseteq$ ins X_n; thus ins $\tau\,(Y_m) = \tau\,(\text{ins } Y_m) \subseteq \tau\,(\text{ins } X_n) =$ ins $\tau\,(X_n)$. Similarly, for every j there is an i with ins $\tau\,(X_i) \subseteq$ ins $\tau\,(Y_j)$.

The proof that $\hat{\tau}$ is well defined also reveals a little something about the disk. Namely, the prime ends of $\mathbb{D}$ are in one-to-one correspondence with the points of $\partial\mathbb{D}$ (something that better be true if the theorem is true). In fact, if $\{X_n\}$ is a 0-chain in $\mathbb{D}$, then Proposition 2.5 implies that $\operatorname{diam}(\operatorname{cl}\ (\operatorname{ins} X_n)) \to 0$. By Cantor's Theorem, $\bigcap_n \operatorname{cl}\ (\operatorname{ins} X_n) = \{z_0\}$ for some point z_0 in cl $\mathbb{D}$. It follows that $z_0 \in \partial\mathbb{D}$ (why?). The preceding paragraph shows that whenever $\{X_n\}$ and $\{Y_n\}$ are equivalent 0-chains in $\mathbb{D}$, $\bigcap_n \operatorname{cl}\ (\operatorname{ins} X_n) = \bigcap_n \operatorname{cl}\ (\operatorname{ins} Y_n)$.

We now proceed to the proof that $\hat{\tau}$ is a homeomorphism. Some preparatory work is required.

3.5 Lemma. *If $\gamma : [0,1) \to \mathbb{D}$ is an arc such that $|\gamma(t)| \to 1$ as $t \to 1$, then the set $Z = \{z :$ there exists $t_k \to 1$ with $\gamma(t_k) \to z\}$ is a closed arc in $\partial\mathbb{D}$. If $f : \mathbb{D} \to \mathbb{C}$ is a bounded analytic function and $\lim_{t\to 1} f(\gamma(t))$ exists, then either Z is a single point or f is constant.*

Proof. Observe that the set $Z = Clu(\gamma; 1)$ (1.6). Thus Z is a closed connected subset of $\partial\mathbb{D}$ (1.7 and 1.8); that is, Z is a closed arc.

Now assume that $f : \mathbb{D} \to \mathbb{C}$ is a bounded analytic function such that $\lim_{t\to 1} f(\gamma(t)) = \omega$ exists and Z is not a single point. It will be shown that f must be constant. In fact, let z be an interior point of Z such that the radial limit of f exists at z. It is easy to see (draw a picture) that the radial segment $[0, z)$ must meet the curve γ infinitely often. Hence there is a sequence $\{t_k\}$ in $[0,1)$ such that $t_k \to 1$, $\gamma(t_k) \to z$, and $\arg(\gamma(t_k)) = \arg z$ for all k. Thus $\lim_{r\to 1} f(rz) = \lim_{k\to\infty} f(\gamma(t_k)) = \omega$. By Theorem 13.5.3, $f \equiv \omega$. $\square$

3.6 Lemma. *Let $\tau : \mathbb{D} \to \Omega$ be a conformal equivalence with $\tau(0) = 0$. If C is a crosscut of Ω, then $X = \tau^{-1}(C)$ is a crosscut of $\mathbb{D}$.*

Proof. Let $C : (0,1) \to \Omega$ be a parameterization of C and, for $q = 0$ or 1, let $a_q = \lim_{t\to q} C(t)$. So $a_q \in \partial\Omega$. Clearly $X(t) = \tau^{-1}(C(t))$ is an open Jordan arc and $|X(t)| \to 1$ as $t \to 0$ or 1. For $q = 0$, 1, let $Z_q = \{z$: there exists $t_k \to q$ with $X(t_k) \to z\}$. But $\lim_{t\to q} \tau(X(t)) = a_q$ and τ is not constant. By Lemma 3.5, Z_q is a single point and so X is a crosscut. $\square$

Now suppose $\{C_n\}$ is a 0-chain in Ω and let $X_n = \tau^{-1}(C_n)$. So each X_n is a crosscut of $\mathbb{D}$ by the preceding proposition. We will see that it is almost true that $\{X_n\}$ is a 0-chain in $\mathbb{D}$. The part of the definition of a 0-chain that will not be fulfilled is that $\operatorname{cl}X_n \cap \operatorname{cl}X_{n+1}$ need not be empty.

Begin by noting that $\tau(\operatorname{ins} X_n) = \operatorname{ins} C_n$; hence $\operatorname{ins} X_{n+1} \subseteq \operatorname{ins} X_n$.

3.7 Proposition. *If $\tau : \mathbb{D} \to \Omega$ is a conformal equivalence with $\tau(0) = 0$ and $\{X_n\}$ is a sequence of crosscuts of $\mathbb{D}$ such that $C_n \equiv \tau(X_n)$ defines a 0-chain of crosscuts in Ω, then* diam $X_n \to 0$.

Proof. First assume that there is an r, $0 < r < 1$, such that $X_n \cap \{z : |z| = r\} \neq \emptyset$ for an infinite number of values of n. Let $z_k \in X_{n_k}$ with $|z_k| = r$ such that $\tau(z_k) \to \zeta_0$; so $\zeta_0 \in \tau(\{z : |z| = r\}) \subseteq \Omega$. But if $\delta > 0$

such that $\overline{B}(\zeta_0;\delta) \subseteq \Omega$, there is a k_0 such that $|\tau(z_k) - \zeta_0| < \delta$ for $k \geq k_0$. But $\tau(z_k) \in C_{n_k} \cap B(\zeta_0;\delta)$ and diam $C_n \to 0$. Hence there is a $k_1 \geq k_0$ such that $C_{n_k} \subseteq B(\zeta_0;\delta)$ for $k \geq k_1$. This implies that cl $C_{n_k} \cap \partial\Omega = \emptyset$. Since C_{n_k} is a crosscut, this is a contradiction. Thus for every $r < 1$, $X_n \cap \{z : |z| = r\} = \emptyset$ for all but a finite number of indices n.

Let $r_n = \inf\{|z| : z \in X_n\}$; by the preceding paragraph, $r_n \to 1$. Since X_n is a crosscut, $\partial[\text{ins } X_n] = X_n \cup \gamma_n$ for some closed arc γ_n of $\partial\,\mathbb{D}$. It follows that $K \equiv \bigcap_n \text{cl }[\text{ins } X_n]$ is a non-empty closed connected subset of $\partial\,\mathbb{D}$ and hence is a closed arc in $\partial\,\mathbb{D}$. Moreover, $K = \bigcap_n \gamma_n$ (why?). It suffices to show that K is a single point.

Suppose K is a proper closed arc in $\partial\,\mathbb{D}$. Then by Theorem 13.5.2 there are distinct interior points z and w of the arc K such that the radial limits of τ exist at both z and w; denote these radial limits by $\tilde{\tau}(z)$ and $\tilde{\tau}(w)$. Since 0 belongs to the outside of each X_n, for each n there are points z_n and w_n on X_n that lie on the rays through z and w, respectively. Thus $\tau(z_n) \to \tilde{\tau}(z)$ and $\tau(w_n) \to \tilde{\tau}(w)$. But $\tau(z_n)$ and $\tau(w_n) \in \tau(X_n) = C_n$ and diam $C_n \to 0$. Hence $\tilde{\tau}(z) = \tilde{\tau}(w)$. Since z and w were arbitrary interior points of K, Theorem 13.5.3 implies that τ is constant, a contradiction. Therefore it must be that K is a single point and so diam $X_n \to 0$. $\square$

3.8 Lemma. *If $\tau : \mathbb{D} \to \Omega$ is a conformal equivalence with $\tau(0) = 0$, $\{C_n\}$ is a 0-chain in Ω, and $X_n =$ the crosscut $\tau^{-1}(C_n)$, then there is a point z_0 on $\partial\mathbb{D}$ and there are positive numbers δ_n and ε_n with $0 < \delta_n < \varepsilon_n$ and $\varepsilon_n \to 0$ such that*

3.9 $$\bigcap_n \text{cl}\,(\text{ins } X_n) = \{z_0\}$$

and

3.10 $$\mathbb{D} \cap B(z_0;\delta_n) \subseteq \text{ins } X_n \subseteq \mathbb{D} \cap B(z_0;\varepsilon_n).$$

Proof. By Propositions 3.7 and 2.5, diam[cl (ins X_n)] $\to 0$. Therefore there is a z_0 in cl $\mathbb{D}$ such that (3.9) holds. By Proposition 1.11, $z_0 \in \partial\,\mathbb{D}$.

It is clear that since diam $X_n \to 0$, the number ε_n can be found. Suppose the number δ_n cannot be found. That is, suppose there is an n (which will remain fixed) such that for every $\delta > 0$, $\mathbb{D} \cap B(z_0;\delta)$ is not contained in ins X_n. Thus for every $\delta > 0$ there are points in $\mathbb{D} \cap B(z_0;\delta)$ that belong to both ins X_n and out X_n; by connectedness, this implies that $X_n \cap B(z_0;\delta) \neq \emptyset$ for every $\delta > 0$. Hence $z_0 \in$ cl X_n. Since ins $X_m \subseteq$ ins X_n for $m \geq n$, the same argument implies that $z_0 \in$ cl X_m for $m \geq n$.

Now construct crosscuts $\{Y_j\}$ as in Proposition 2.9 so that $\bigcap_n \text{cl (ins } Y_j) = \{z_0\}$, diam $Y_j \to 0$, and $\{\tau(Y_j)\}$ is a 0-chain of Ω. It is claimed that $X_n \cap Y_j \neq \emptyset$ for all sufficiently large values of j. In fact, if this were not the case, then, by connectedness and the fact that $z_0 \in$ cl X_n, $X_n \subseteq$ ins Y_j

for all j. But diam[ins Y_j] $\to 0$ and so this implies that X_n is a singleton, a contradiction. Hence there must be a j_0 such that $X_n \cap Y_j \neq \emptyset$ for $j \geq j_0$. Similarly, j_0 can be chosen so that we also have $X_{n+1} \cap Y_j \neq \emptyset$ for $j \geq j_0$. But this implies that $C_n \cap \tau(Y_j) \neq \emptyset \neq C_{n+1} \cap \tau(Y_j)$ for $j \geq j_0$. Therefore, $\text{dist}(C_n, C_{n+1}) \leq \text{diam } \tau(Y_j)$, and this converges to 0. Thus cl $C_n \cap$ cl $C_{n+1} \neq \emptyset$, contradicting the definition of a 0-chain. This implies that (3.10) holds. □

Proof of Theorem 3.4. Define $\hat{\tau} : \text{cl } \mathbb{D} \to \hat{\Omega}$ by letting $\hat{\tau}(z) = \tau(z)$ for $|z| < 1$ and $\hat{\tau}(z) =$ the prime end of Ω corresponding to the 0-chain $\{\tau(Y_j)\}$, where $\{Y_j\}$ is any 0-chain in $\mathbb{D}$ such that $\bigcap_j \text{cl (ins } Y_j) = \{z\}$ and $\{\tau(Y_j)\}$ is a 0-chain in Ω. We have already seen that $\hat{\tau}$ is well defined.

To show that $\hat{\tau}$ is subjective, let $p \in \hat{\Omega} \setminus \Omega$ and let $\{C_n\}$ be a 0-chain in p. If $X_n = \tau^{-1}(C_n)$, then (3.10) and Proposition 2.9 imply we can construct a 0-chain $\{Y_j\}$ in $\mathbb{D}$ with $\bigcap_j \text{cl (ins } Y_j) = \{z\}$, $\{\tau(Y_j)\}$ a 0-chain in Ω, and ins $Y_n \subseteq$ ins X_n for every n. Moreover, for each n, the form of Y_n, (3.9), and the fact that diam[ins X_m] $\to 0$ imply that ins $X_m \subseteq Y_n$ for sufficiently large m. This implies that $\{\tau(Y_j)\}$ and $\{C_n\}$ are equivalent 0-chains in Ω and so $\hat{\tau}(z) = p$.

The proof that $\hat{\tau}$ is one-to-one is left to the reader (Exercise 2).

It remains to show that $\hat{\tau}$ is a homeomorphism. Let U be an open subset of $\hat{\Omega}$; it must be shown that $\hat{\tau}^{-1}(U)$ is (relatively) open in cl $\mathbb{D}$. Clearly $\hat{\tau}^{-1}(U) \cap \mathbb{D} = \tau^{-1}(U \cap \Omega)$, and so this set is open. If $z_0 \in \hat{\tau}^{-1}(U) \cap \partial\mathbb{D}$, it must be shown that there is a $\delta > 0$ with $\mathbb{D} \cap B(z_0; \delta) \subseteq \hat{\tau}^{-1}(U) \cap \mathbb{D}$. Put $p = \hat{\tau}(z_0)$; so $p \in U \setminus \Omega$. Let $\{C_n\} \in p$; by definition, there is an integer n such that ins $C_n \subseteq U \cap \Omega$. If $X_n = \tau^{-1}(C_n)$, then X_n is a crosscut and ins $X_n = \tau^{-1}(\text{ins } C_n) \subseteq \hat{\tau}^{-1}(U) \cap \mathbb{D}$. By (3.10), there is a $\delta > 0$ with $\mathbb{D} \cap B(z_0; \delta) \subseteq$ ins X_n, and so $\hat{\tau}$ is continuous.

Finally, to show that $\hat{\tau}$ is an open map it suffices to fix a z_0 in $\partial\mathbb{D}$ and a $\delta > 0$ and show that $\hat{\tau}(\text{cl } \mathbb{D} \cap B(z_0; \delta))$ contains an open neighborhood of $p = \hat{\tau}(z_0)$. Construct a 0-chain $\{Y_n\}$ as in Proposition 2.9 with $\bigcap_n \text{cl [ins } Y_n] = \{z_0\}$ and cl [ins Y_n] $\subseteq B(z_0; \delta)$ for all n. Thus $\{\tau(Y_n)\} \in p$ and ins $\tau(Y_n) = \tau(\text{ins } Y_n) \subseteq \Omega \cap \hat{\tau}(\text{cl } \mathbb{D} \cap B(z_0; \delta))$. By definition, $\hat{\tau}(\text{cl } \mathbb{D} \cap B(z_0; \delta))$ is a neighborhood of p. □

Some additional material on prime ends will appear in the following two sections. Additional results can be found in Collingwood and Lohwater [1966] and Ohtsuka [1967].

Exercises

1. Prove that the collection of open sets in $\hat{\Omega}$ forms a topology on $\hat{\Omega}$.

2. Supply the details of the proof that the map $\hat{\tau}$ is one-to-one.

3. Can you give a direct proof (that is, without using Theorem 3.4) that $\hat{\Omega}$ is compact?

4. If Ω is the slit disk, describe the topology on $\hat{\Omega}$.

§4 Impressions of a Prime End

We already have seen in §1 the definition of the cluster set of a function $f : G \to \mathbb{C}$ at a point a in ∂G. Here we specialize to a bounded function $f : \mathbb{D} \to \mathbb{C}$ and define the radial cluster set of f at a point a in $\partial\mathbb{D}$. The preliminary results as well as their proofs are similar to the analogous results about the cluster set of a function.

4.1 Definition. If $f : \mathbb{D} \to \mathbb{C}$ is a bounded function and $a \in \partial\mathbb{D}$, the *radial cluster set* of f at a is the set

$$\operatorname{Clu}_r(f;a) \equiv \bigcap_{\varepsilon>0} \operatorname{cl}\ \{f(ra) : 1-\varepsilon \leq r < 1\}.$$

The following results are clear.

4.2 Proposition.

(a) *If $f : \mathbb{D} \to \mathbb{C}$ is a bounded function and $a \in \partial\mathbb{D}$, then $\xi \in \operatorname{Clu}_r(f;a)$ if and only if there is a sequence $\{r_n\}$ increasing to 1 such that $f(r_n a) \to \zeta$.*

(b) *If f is continuous, then $\operatorname{Clu}_r(f;a)$ is a non-empty compact connected set.*

(c) *If f is a homeomorphism of $\mathbb{D}$ onto its image, then $\operatorname{Clu}_r(f;a)$ is a subset of $\partial f(\mathbb{D})$.*

4.3 Proposition. *If $f : \mathbb{D} \to \mathbb{C}$ is a bounded function and $a \in \partial\,\mathbb{D}$, then f has a radial limit at a equal to ζ if and only if $\operatorname{Clu}_r(f;a) = \{\zeta\}$.*

Now let's introduce another pair of sets associated with a prime end of a bounded simply connected region Ω containing 0. The connection with the cluster sets will be discussed shortly.

4.4 Definition. If p is a prime end of a bounded simply connected region Ω, the *impression* of p is the set

$$I(p) \equiv \bigcap_{n=1}^{\infty} \operatorname{cl}\ [\operatorname{ins} C_n]$$

where $\{C_n\} \in p$.

It is routine to show that the definition of the impression does not depend on the choice of the 0-chain $\{C_n\}$ in p so that $I(p)$ is well defined.

4.5 Proposition. *For each prime end p of Ω, the impression $I(p)$ is a non-empty compact connected subset of $\partial\Omega$.*

4.6 Definition. If p is a prime end of Ω, a complex number ζ is called a *principal point* of p if there is a $\{C_n\}$ in p such that $C_n \to \zeta$ in the sense that for every $\varepsilon > 0$ there is an integer n_0 such that $\text{dist}(\zeta, C_n) < \varepsilon$ for all $n \geq n_0$. Let $\Pi(p)$ denote the set of principal points of p.

It might be expected that at this point it would be demonstrated that $\Pi(p)$ is a non-empty compact and possibly even connected subset of $\partial\Omega$. This will in fact follow from the next theorem, so we content ourselves with the observation that $\Pi(p) \subseteq I(p)$.

4.7 Theorem. *If $\tau : \mathbb{D} \to \Omega$ is the Riemann map with $\tau(0) = 0$ and $\tau'(0) > 0$, $a \in \partial\,\mathbb{D}$, and p is the prime end for Ω corresponding to a (that is, $p = \hat{\tau}(a)$), then*

$$\text{Clu}(\tau; a) = I(p) \quad \textit{and} \quad \text{Clu}_r(\tau; a) = \Pi(p).$$

Proof. Let $\zeta \in \text{Clu}(\tau; a)$ and let $\{z_k\}$ be a sequence in $\mathbb{D}$ such that $\tau(z_k) \to \zeta$. Let $\{C_n\} \in p$. By Lemma 3.8 there are positive numbers ε_n and δ_n such that $\mathbb{D} \cap B(a; \delta_n) \subseteq \text{ins } \tau^{-1}(C_n) \subseteq \mathbb{D} \cap B(a; \varepsilon_n)$ for all n. This implies that for every $n \geq 1$ there is an integer k_n such that $z_k \in \text{ins } \tau^{-1}(C_n)$ for $k \geq k_n$. Thus $\tau(z_k) \in \text{ins } C_n$ for $k \geq k_n$ and so $\zeta \in \text{cl (ins } C_n)$. Therefore $\zeta \in I(p)$.

Now assume that $\zeta \in I(p)$. If $\{C_n\} \in p$, then $\zeta \in \text{cl (ins } C_n)$ for all $n \geq 1$. Hence for each $n \geq 1$ there is a point z_n in ins $\tau^{-1}(C_n)$ with $|\tau(z_n) - \zeta| < 1/n$. But an application of Lemma 3.8 shows that $z_n \to a$ and so $\zeta \in \text{Clu}(\tau; a)$.

Let $\zeta \in \text{Clu}_r(\tau; a)$ and let $r_n \uparrow 1$ such that $\tau(r_n a) \to \zeta$; define the crosscuts $\{Y_n\}$ as in Proposition 2.9 so that $\{\tau(Y_n)\}$ is a 0-chain in Ω and $\tau(r_n a) \in \tau(Y_n)$ for each n. Note that of necessity $\{\tau(Y_n)\} \in p$. Thus $\zeta \in \Pi(p)$.

Finally assume that $\zeta \in \Pi(p)$ and let $\{C_n\}$ be a 0-chain in p such that $C_n \to \zeta$. An application of Lemma 3.8 implies that $\zeta \in \text{Clu}_r(\tau; a)$. The details are left to the reader. $\square$

An immediate corollary of the preceding theorem can be obtained by assuming that the two cluster sets are singletons. Before stating this explicitly, an additional type of prime end is introduced that is equivalent to such an assumption. Say that a prime end p of Ω is *accessible* if there is a Jordan arc $\gamma : [0, 1] \to \text{cl } \Omega$ with $\gamma(t)$ in Ω for $0 \leq t < 1$ and $\gamma(1)$ in $\partial\Omega$ such that for some $\{C_n\}$ in p, $\gamma \cap C_n \neq \emptyset$ for all sufficiently large n. Note that if p is an accessible prime end, then for every $\{C_n\}$ in p, $\gamma \cap C_n \neq \emptyset$ for sufficiently large n.

4.8 Corollary. *Let $\tau : \mathbb{D} \to \Omega$ be the Riemann map with $\tau(0) = 0$ and $\tau'(0) > 0$ and let $a \in \partial\,\mathbb{D}$ with $p = \hat{\tau}(a)$.*

(a) $\lim_{z \to a} \tau(z)$ *exists if and only if $I(p)$ is a singleton.*

(b) *The following statements are equivalent.*

 (i) $\lim_{r \to 1} \tau(ra)$ *exists.*

 (ii) $\Pi(p)$ *is a singleton.*

 (iii) *p is an accessible prime end.*

Proof. The proof of (a) is clear in light of the theorem and Proposition 4.3. (b) The equivalence of (i) and (ii) is also immediate from the theorem and Proposition 4.3. Assume (i) and let $\zeta = \lim_{r \to 1} \tau(ra)$. So $\zeta \in \partial\Omega$ and $\gamma(r) = \tau(ra)$ is the requisite arc to demonstrate that p is an accessible prime end. This proves (iii).

Now assume that (iii) holds and let $\gamma : [0,1] \to \mathbb{C}$ be the Jordan arc as in the definition of an accessible prime end; let $\zeta = \gamma(1)$. Thus $\sigma(t) = \tau^{-1}(\gamma(t))$ for $0 \leq t < 1$ is a Jordan arc in $\mathbb{D}$. Let $\{C_n\} \in p$ and put $X_n = \tau^{-1}(C_n)$. According to Lemma 3.8 there are sequences of positive numbers $\{\varepsilon_n\}$ and $\{\delta_n\}$ that converge to 0 such that for every $n \geq 1$, $\mathbb{D} \cap B(a; \delta_n) \subseteq \text{ins } X_n \subseteq \mathbb{D} \cap B(a; \varepsilon_n)$, where

$$\{a\} = \bigcap_{n=1}^{\infty} \text{cl } [\text{ins } X_n].$$

If $\varepsilon > 0$ is arbitrary, choose n_0 such that $\varepsilon_n < \varepsilon$ and $\gamma \cap C_n \neq \emptyset$ for $n \geq n_0$. Fix $n \geq n_0$ and let t_0 be such that $\gamma(t) \in \text{ins } C_n$ for $t_0 < t < 1$. Thus $\sigma(t) \in \text{ins } X_n$ and hence $|\sigma(t) - a| < \varepsilon$ when $t_0 < t$. This says that $\sigma(t) \to a$ as $t \to 1$; define $\sigma(1) = a$. By Theorem 13.5.4, τ has a radial limit at a. □

Exercises

1. Prove that the definition of $I(p)$ (4.4) does not depend on the choice of 0-chain $\{C_n\}$.

2. Let K be a non-empty compact connected subset of $\mathbb{C}$ such that K has no interior and $\mathbb{C} \setminus K$ is connected. Show that there is a simply connected region Ω for which $K = I(p)$ for some prime end p of Ω. (The converse of this is not true as the next exercise shows.)

3. Let $\gamma(t) = e^{-t^{-1} + it}$ for $0 \leq t < \infty$ and put $\Omega = \mathbb{D} \setminus \{\gamma\}$. Show that Ω has a prime end p such that $I(p) = \partial\,\mathbb{D}$.

§5 Boundary Values of Riemann Maps

In this section we address the problem of continuously extending a Riemann map τ from $\mathbb{D}$ onto a simply connected region Ω to a continuous map from the closure of $\mathbb{D}$ to the closure of Ω. First note that Proposition 1.11 implies that if $\tau : \mathbb{D} \to \Omega$ has a continuous extension, $\tau : \text{cl } \mathbb{D} \to \text{cl } \Omega$, then $\tau(\partial \mathbb{D}) \subseteq \partial\Omega$. Thus τ maps cl $\mathbb{D}$ onto cl Ω and so $\tau(\partial \mathbb{D}) = \partial\Omega$. This also shows that $\partial\Omega$ is a curve. If τ extends to a homeomorphism, then $\partial\Omega$ is a Jordan curve. The principal results of this section state that the converse of these observations is also true. If $\partial\Omega$ is a curve, τ has a continuous extension to cl $\mathbb{D}$; if $\partial\Omega$ is a Jordan curve, τ extends to a homeomorphism of cl $\mathbb{D}$ onto cl Ω.

This is a remarkable result and makes heavy use of the fact that τ is analytic. It is not difficult to show that $\phi(z) = z \exp(i/(1-|z|))$ is a homeomorphism of $\mathbb{D}$ onto itself. However each radial segment in $\mathbb{D}$ is mapped onto a spiral and so $\text{Clu}(\phi; a) = \partial \mathbb{D}$ for every a in $\partial \mathbb{D}$. So ϕ cannot be continuously extended to any point of the circle.

Why suspect that conformal equivalences behave differently from homeomorphisms? Of course we have seen that the conformal equivalences of $\mathbb{D}$ onto itself have homeomorphic extensions to the closure. Also imagine the curves $\tau_r(\theta) = \tau(re^{i\theta})$, $r > 0$; the images of the circles of radius r. τ maps the radial segments onto Jordan arcs that are orthogonal to this family of curves. If $\partial\,\Omega$ is a Jordan curve, then the curves $\{\tau_r : r > 0\}$ approach $\partial\,\Omega$ in some sense. You might be led to believe that τ has a nice radial limit at each point of $\partial\,\mathbb{D}$.

We begin with some topological considerations.

5.1 Definition. A compact metric space X is *locally connected* if for every $\varepsilon > 0$ there is a $\delta > 0$ such that whenever x and y are points in X with $|x - y| < \delta$, there is a connected subset A of X containing x and y and satisfying diam $A < \varepsilon$.

The proofs of the following topology facts concerning local connectedness are left to the reader. An alternative is to consult Hocking and Young [1961].

5.2 Proposition. *If X is a compact metric space, the following statements are equivalent.*

(a) *X is locally connected.*

(b) *For every $\varepsilon > 0$ there are compact connected sets $A_1, \ldots, A_m$ with* diam $A_j < \varepsilon$ *for $1 \leq j \leq m$ and such that $X = A_1 \cup \ldots \cup A_m$.*

(c) *For every $\varepsilon > 0$ and for every x in X, there is a connected open set U such that $x \in U \subseteq B(x; \varepsilon)$.*

Part (c) of the preceding proposition is the usual definition of local connectedness. Indeed it can be easily extended to a definition that can be made for arbitrary topological spaces. Definition 5.1 was chosen for the definition here because it is the property that will be used most often in subsequent proofs.

The next topological fact is easier to prove.

5.3 Proposition. *If X and Y are compact metric spaces, $f : X \to Y$ is a continuous surjection, and X is locally connected, then Y is locally connected.*

Note that as a result of this proposition it follows that every path is locally connected. We need one final topological lemma that will be used in the proof of the main results of this section.

5.4 Lemma. *If A_1 and A_2 are compact connected subsets of $\mathbb{C}$ with $A_1 \cap A_2$ connected and non-empty and x and y are points such that neither A_1 nor A_2 separates x from y, then $A_1 \cup A_2$ does not separate x and y.*

Proof. Without loss of generality it can be assumed that $x = 0$ and $y = \infty$. For $j = 1, 2$ let $\gamma_j : [0,1] \to \mathbb{C}_\infty \setminus A_j$ be a path with $\gamma_j(0) = 0$ and $\gamma_j(1) = \infty$. Since $A_1 \cap A_2$ is a connected subset of $\mathbb{C} \setminus (\gamma_1 \cup \gamma_2)$ there is a component G of $\mathbb{C} \setminus (\gamma_1 \cup \gamma_2)$ containing $A_1 \cap A_2$. Thus $A_1 \setminus G$ and $A_2 \setminus G$ are disjoint compact subsets of $\mathbb{C}\setminus\gamma_1$ and $\mathbb{C}\setminus\gamma_2$; therefore there are disjoint open sets V_1 and V_2 such that for $j = 1, 2$, $A_j \setminus G \subseteq V_j \subseteq \mathbb{C} \setminus \gamma_j$.

Let $U = G \cup V_1 \cup V_2$ so that $A_1 \cup A_2 \subseteq U$. Proposition 13.1.7 implies there is a smooth Jordan curve σ in U that separates $A_1 \cup A_2$ from ∞; thus $A_1 \cup A_2 \subseteq \text{ins}\ \sigma$. It will be shown that 0 is in the outside of σ so that $A_1 \cup A_2$ does not separate 0 from ∞.

Note that each component $\mathbb{C}\setminus\gamma_j$ is simply connected and does not contain 0. Thus there is a branch of the logarithm $f : \mathbb{C} \setminus \gamma_j \to \mathbb{C}$. Moreover these functions can be chosen so that $f_1(z) = f_2(z)$ on G. Therefore

$$f(z) = \begin{cases} f_1(z) & \text{if } z \in V_1 \\ f_2(z) & \text{if } z \in V_2 \\ f_1(z) = f_2(z) & \text{if } z \in G \end{cases}$$

is a well defined branch of the logarithm on U. Since $f'(z) = z^{-1}$ on U, the winding number of σ about 0 is 0. Therefore 0 is in the outside of σ. $\square$

Pommerenke [1975] calls the preceding lemma Janiszewski's Theorem. Now for one of the main theorems in this section.

5.5 Theorem. *Let Ω be a bounded simply connected region and let $\tau : \mathbb{D} \to \Omega$ be the Riemann map with $\tau(0) = 0$ and $\tau'(0) > 0$. The following statements are equivalent.*

(a) *τ has a continuous extension to the closure of* $\mathbb{D}$.

(b) $\partial\Omega$ *is a continuous path.*

(c) $\partial\Omega$ *is locally connected.*

(d) $\mathbb{C}_\infty \setminus \Omega$ *is locally connected.*

Proof. It has already been pointed out that (a) implies (b) implies (c). Assume that (c) holds. To prove (d), let $\varepsilon > 0$ and choose $\delta > 0$ so that for x and y in $\partial\Omega$ and $|x-y| < \delta$ there is a connected subset B of $\partial\Omega$ that contains x and y and satisfies diam $B < \varepsilon/3$. Choose δ so that $\delta < \varepsilon/3$. It suffices to show that if z and $w \in X \equiv \mathbb{C} \setminus \Omega$ such that $|z-w| < \delta$, then there is a connected subset A of $\mathbb{C} \setminus \Omega$ that contains z and w and satisfies diam $Z < \varepsilon$. (Why?) Examine $[z,w] \cap \partial\Omega$ and let x and y be the points in this set that are nearest z and w, respectively; thus $|x-y| < \delta$. Let B be the subset of $\partial\Omega$ as above and put $A = [z,x] \cup B \cup [y,w]$. So diam $A < \varepsilon$ and z, $w \in A$.

Now assume that (d) holds. To prove (a) it suffices to show that for every prime end p of Ω the impression $I(p)$ is a singleton. Fix the prime end p and let $\{C_n\} \in p$. Let $0 < \varepsilon < \operatorname{dist}(0, \partial\Omega)$ and let $\delta > 0$ be chosen as in the definition of locally connected; also choose $\delta < \varepsilon$. Find an integer n_0 such that diam $C_n < \delta$ for $n \geq n_0$. Thus if a_n and b_n are the end points of C_n, $|a_n - b_n| < \delta$. Since $X \equiv \mathbb{C} \setminus \Omega$ is locally connected there is a compact connected subset B_n of X that contains a_n and b_n and satisfies diam $B_n < \varepsilon$. Observe that $B_n \cup C_n$ is a connected subset of $B(a_n;\varepsilon)$.

Thus if $|\zeta - a_n| > \varepsilon$, then 0 and ζ are not separated by $B_n \cup C_n$. If in addition $\zeta \in \Omega$, then 0 and ζ are not separated by X. But $(B_n \cup C_n) \cap X = B_n$ is connected. Thus the preceding lemma implies that for ζ in Ω with $|\zeta - a_n| > \varepsilon$, 0 and ζ are not separated by $(B_n \cup C_n) \cup X = C_n \cup X$. That is, both 0 and ζ belong to the same component of $\mathbb{C} \setminus (C_n \cup X) = \Omega \setminus C_n$. Hence $\zeta \in \operatorname{out} C_n$ if $\zeta \in G$ and $|\zeta - a_n| > \varepsilon$. But this says that $\operatorname{ins} C_n \subseteq B(a_n;\varepsilon)$ and so diam $C_n < \varepsilon$ for $n \geq n_0$. Thus $I(p)$ is a singleton. $\square$

It is now a rather easy matter to characterize those Riemann maps that extend to a homeomorphism on cl $\mathbb{D}$,

5.6 Theorem. *If Ω is a bounded simply connected region and $\tau : \mathbb{D} \to \Omega$ is the Riemann map with $\tau(0) = 0$ and $\tau'(0) > 0$, then τ extends to be a homeomorphism of* cl $\mathbb{D}$ *onto* cl Ω *if and only if $\partial\Omega$ is a Jordan curve.*

Proof. If τ extends to a homeomorphism of cl $\mathbb{D}$ onto cl Ω, then, as mentioned before, $\tau(\partial\,\mathbb{D}) = \partial\Omega$ and so $\partial\Omega$ is a Jordan curve. Conversely, assume that $\partial\Omega$ is a Jordan curve. By Theorem 5.5, τ has a continuous extension to $\tau :$ cl $\mathbb{D} \to$ cl Ω. It remains to prove that τ is one-to-one on $\partial\Omega$.

Suppose $w_1, w_2 \in \partial\,\mathbb{D}$ and $\tau(w_1) = \tau(w_2)$; let $\sigma_i = \{\tau(rw_i) : 0 \leq r \leq 1\}$. So σ_1 and σ_2 are two Jordan arcs with end points $\tau(0) = 0$ and $\omega_0 =$

$\tau(w_1) = \tau(w_2)$ that lie inside Ω except for the final point. Taken together, these arcs form a (closed) Jordan curve σ; let $\Lambda = \text{ins}\ \sigma$. (In fact, σ is a crosscut.) By Corollary 13.1.11, $\Lambda \subseteq \Omega$. Put $\lambda_1 = (\text{out}\ \sigma) \cap \Omega; \Lambda_1$ is connected (verify). Also $\Lambda \cup \Lambda_1 \cup (\sigma \cap \Omega) = \Omega$. Observe that $(\text{cl}\ \Lambda) \cap \partial\Omega$ is the singleton $\{\omega_0\}$.

Let V and V_1 be the two components of $\mathbb{D} \setminus \{rw_i : 0 \le r < 1,\ i = 1, 2\}$. Since $\tau(V \cup V_1) = \Lambda \cup \Lambda_1$, a connectedness argument shows that either $\tau(V) = \Lambda$ or $\tau(V) = \Lambda_1$. Assume that $\tau(V) = \Lambda$; hence $\tau(V_1) = \Lambda_1$.

The proof now proceeds to show that if w belongs to the arc $\partial\,\mathbb{D} \cap \partial V$, then $\tau(w) = \omega_0$. Fix w in this arc; so $\lambda = \{\tau(rw) : 0 \le r \le 1\}$ is a path in Λ except for the points 0 and $\tau(w)$. But $\tau(w) \in (\text{cl}\ \Lambda) \cap \partial\Omega = \{\omega_0\}$. Since w was an arbitrary point in the arc $\partial\,\mathbb{D} \cap \partial V$, this shows that the bounded analytic function τ is constant along an arc of $\partial\,\mathbb{D}$. By Theorem 13.5.3, τ is a constant function, a contradiction. □

A *Jordan region* is a simply connected region whose boundary is a Jordan curve.

5.7 Corollary. *If G and Ω are two Jordan regions and $f : G \to \Omega$ is a conformal equivalence, then f has an extension to a homeomorphism of* cl G *onto* cl Ω.

Recall that a curve $\gamma : \partial\,\mathbb{D} \to \mathbb{C}$ is rectifiable if $\theta \to \gamma(e^{i\theta})$ is a function of bounded variation and the length of the curve γ is given by $\int d|\gamma|(e^{i\theta}) = V(\gamma)$, the total variation of γ. If the boundary of a simply connected region Ω is a rectifiable curve, Theorem 5.5 can be refined.

In Chapter 20 the class of analytic functions H^1 will be investigated. Here this class will be used only as a notational device though one result from the future will have to be used. In fact H^1 consists of those analytic functions f on $\mathbb{D}$ such that

$$\sup\left\{\int_0^{2\pi} \left|f(re^{i\theta})\right| d\theta : 0 < r < 1\right\} < \infty.$$

Note that if f is an analytic function on $\mathbb{D}$ and $0 < r < 1$, then $\gamma_r(\theta) = f(re^{i\theta})$ defines a rectifiable curve. The length of this curve is given by

$$r\int_0^{2\pi} \left|f'(re^{i\theta})\right| d\theta.$$

Thus the condition that $f' \in H^1$ is precisely the condition that the curves $\{\gamma_r\}$ have uniformly bounded lengths. This leads to the next result.

5.8 Theorem. *Assume that Ω is a Jordan region and let $\tau : \mathbb{D} \to \Omega$ be the Riemann map with $\tau(0) = 0$ and $\tau'(0) > 0$. The following statements are equivalent.*

(a) *$\partial\Omega$ is a rectifiable Jordan curve.*

(b) $\tau' \in H^1$.

(c) *The function* $\theta \to \tau(e^{i\theta})$ *is a function of bounded variation.*

(d) *The function* $\theta \to \tau(e^{i\theta})$ *is absolutely continuous.*

Proof. Using Theorem 5.6, extend τ to a homeomorphism of cl $\mathbb{D}$ onto cl Ω. Assume that γ is a rectifiable parameterization of $\partial\Omega$ and let $\alpha(\theta) = \tau(e^{i\theta})$. Since both α and γ are one-to-one, there is a homeomorphism $\sigma : [0, 2\pi] \to [0, 2\pi]$ such that $\alpha(\theta) = \gamma(\sigma(\theta))$ for all θ. So σ is either increasing or decreasing. If $0 = \theta_0 < \theta_1 < \cdots < \theta_n = 2\pi$, then $\sum |\alpha(\theta_k) - \alpha(\theta_{k-1})| = \sum |\gamma(\sigma(\theta_k)) - \gamma(\sigma(\theta_{k-1}))| \leq V(\gamma)$ since $\{\sigma(\theta_0), \ldots, \sigma(\theta_n)\}$ is also a partition of $[0, 2\pi]$. Thus α is a bounded variation. This shows that (a) implies (c). Clearly (c) implies (a).

Now let's show that (b) implies (c). Assume that $\tau' \in H^1$ and let $0 < r < 1$. If $0 = \theta_0 < \cdots < \theta_n = 2\pi$, then

$$\begin{aligned}
\sum_{j=1}^{n} \left|\tau(r^{i\theta_j}) - \tau(re^{i\theta_{j-1}})\right| &= \sum_{j=1}^{n} \left| \int_{\theta_{j-1}}^{\theta_j} r\tau'(re^{i\theta}) ie^{i\theta} d\theta \right| \\
&\leq \int_0^{2\pi} \left|\tau'(re^{i\theta})\right| d\theta \\
&\leq C,
\end{aligned}$$

where C is the constant whose existence is guaranteed by the assumption that τ' belongs to H^1. Letting $r \to 1$ we get that

$$\sum_{j=1}^{n} \left|\tau(e^{i\theta_j} - \tau(e^{i\theta j-1})\right| \leq C$$

and so τ is a function of bounded variation of $\partial\mathbb{D}$.

The fact that (c) and (d) are equivalent and imply (b) will be shown in Theorem 20.4.8 below. □

Now for an application of Theorem 5.6 to a characterization of the simply connected regions whose boundaries are Jordan curves.

5.9 Definition. For any region Ω a boundary point ω is a *simple boundary point* if whenever $\{\omega_n\}$ is a sequence in Ω converging to ω there is a path $\alpha : [0, 1] \to \mathbb{C}$ having the following properties:

(a) $\alpha(t) \in \Omega$ for $0 \leq t < 1$;

(b) $\alpha(1) = \omega$;

(c) there is a sequence $\{t_n\}$ in $[0, 1)$ such that $t_n \to 1$ and $\alpha(t_n) = \omega_n$ for all $n \geq 1$.

It is not hard to see that each point of $\partial\,\mathbb{D}$ is a simple boundary point (Exercise 4). The region in Figure 14.4 furnishes examples of boundary points of a simply connected region that are not simple boundary points. Exercise 4 also states that not every point in the boundary of a slit disk is a simple boundary point. Here is one way of getting some examples and a precursor of the main result to come.

5.10 Proposition. *If Ω is a simply connected region, $g : \Omega \to \mathbb{D}$ is a conformal equivalence, and $\omega \in \partial\Omega$ such that g has an extension to a continuous map from $\Omega \cup \{\omega\}$ onto $\mathbb{D} \cup \{a\}$ for some a in $\partial\,\mathbb{D}$, then ω is a simple boundary point of Ω.*

Proof. Exercise. □

5.11 Corollary. *If Ω is a Jordan region, then every point of $\partial\Omega$ is a simple boundary point.*

The preceding corollary is a geometric fact that was derived from a theorem of analysis (5.6). Giving a purely geometric proof seems quite hard.

5.12 Theorem.

(a) *Let Ω be a bounded simply connected region and let $g : \Omega \to \mathbb{D}$ be a conformal equivalence. If ω is a simple boundary point of Ω, then g has a continuous extension to $\Omega \cup \{\omega\}$.*

(b) *If R is the collection of simple boundary points of Ω, then g has a continuous one-to-one extension to $\Omega \cup R$.*

Proof. (a) If g does not have such an extension, then there is a sequence $\{\omega_n\}$ in Ω that converges to ω such that $g(\omega_{2n}) \to w_2$, $g(\omega_{2n+1}) \to w_1$, and $w_1 \neq w_2$. It is easy to see that w_1 and w_2 belong to $\partial\,\mathbb{D}$ (1.11). Let $\alpha : [0,1] \to \Omega \cup \{\omega\}$ be a path such that $\alpha(1) = \omega$, $\alpha(t_n) = \omega_n$, and $t_n \to 1$. Put $p(t) = g\left(\alpha(t)\right)$. It follows that $|p(t)| \to 1$ as $t \to 1$ (why?). Let J_1 and J_2 be the two open arcs in $\partial\,\mathbb{D}$ with end points w_1 and w_2. By drawing a picture it can be seen that one of these arcs, say J_1, has the property that for every w on J_1 and for $0 < s < 1$, there is a t with $s < t < 1$ and $p(t)$ lying on the radius $[0, w]$ (exercise). If $\tau = g^{-1} : \mathbb{D} \to \Omega$, then τ is a bounded analytic function since Ω is bounded. So for almost every w in J_1, $\lim_{r\to 1} \tau(rw)$ exists; temporarily fix such a w. But the property of J_1 just discussed implies there is a sequence $\{s_n\}$ in $(0,1)$ such that $s_n \to 1$ and $p(s_n) \to w$ radially. Thus $\tau\left(p(s_n)\right) \to \tau(w)$. But $\tau\left(p(s_n)\right) = \alpha(s_n) \to \omega$; so $\tau(w) = \omega$ for every point of J_1 at which τ has a radial limit. By Theorem 13.5.3, τ is constant, a contradiction.

(b) Let g denote its own extension to $\Omega \cup R$. Suppose ω_1 and ω_2 are distinct points in R and $g(\omega_1) = g(\omega_2)$; we may assume that $g(\omega_1) = g(\omega_2) = -1$. Since ω_1 and ω_2 are simple boundary points, for $j = 1, 2$ there is a path $\alpha_j : [0,1] \to \Omega \cup \{\omega_j\}$ such that $\alpha_j(1) = \omega_j$ and $\alpha_j(t) \in \Omega$ for

$t < 1$. Put $p_j(t) = g(\alpha_j(t))$; so $p_j([0,1)) \subseteq \mathbb{D}$ and $p_j(1) = -1$. Let $t_0 < 1$ such that for $t_0 \le s,\ t \le 1$

$$\textbf{5.13} \qquad |\alpha_1(s) - \alpha_2(t)| > \frac{1}{2}|\omega_1 - \omega_2|.$$

Choose $\delta > 0$ sufficiently small that

$$p_j([0,t_0]) \cap B(-1;\delta) = \emptyset$$

for $j = 1, 2$ and put $A_\delta = \mathbb{D} \cap B(-1;\delta)$. Since each of the curves p_j terminates at -1, whenever $0 < r < \delta$, there is a $t_j > t_0$ so that $p_j(t_j) = w_j$ satisfies $|1 + w_j| = r$. Again letting $\tau = g^{-1}$ we have that (5.13) implies that

$$\begin{aligned} \frac{1}{2}|\omega_1 - \omega_2| &< |\tau(w_1) - \tau(w_2)| \\ &= \left| \int_{\theta_1}^{\theta_2} \tau'(-1 + re^{i\theta}) rie^{i\theta} d\theta \right|. \end{aligned}$$

For each value of r, let θ_r be the angle less than $\pi/2$ such that $1+re^{i\theta} \in \mathbb{D}$ for $|\theta| < \theta_r$. The above inequalities remain valid if the integral is taken from $-\theta_r$ to θ_r. Do this and then apply the Cauchy-Schwartz Inequality for integrals to get

$$\begin{aligned} \frac{|\omega_1 - \omega_2|^2}{4} &\le \left[\int_{-\theta_r}^{\theta_r} \left|\tau'(-1 + re^{i\theta})\right|^2 d\theta \right] \left[\int_{-\theta_r}^{\theta_r} r^2 d\theta \right] \\ &\le \pi r^2 \int_{-\theta_r}^{\theta_r} \left|\tau'(-1 + re^{i\theta})\right|^2 d\theta. \end{aligned}$$

Thus, performing the necessary algebraic manipulations and integrating with respect to r from 0 to δ, we get

$$\begin{aligned} \frac{|\omega_1 - \omega_2|^2}{4\pi} \int_0^\delta \frac{1}{r} dr &\le \int_0^\delta r \int_{-\theta_r}^{\theta_r} \left|\tau'(-1 + re^{i\theta}\right|^2 d\theta dr \\ &= \text{Area } (\tau(A_\delta)) \\ &\le \text{Area } \Omega. \end{aligned}$$

Since Ω is bounded, the right hand side of this inequality is finite. The only way the left hand side can be finite is if $\omega_1 = \omega_2$, contradicting the assumption that they are distinct.

The proof that g is continuous on $\Omega \cup R$ is left to the reader. □

5.14 Corollary. *If Ω is a bounded simply connected region in the plane and every boundary point is a simple boundary point, then $\partial\Omega$ is a Jordan curve.*

Finally, the results of this section can be combined with the results on reflection across an analytic arc.

5.15 Proposition. *Let Ω be a simply connected region and let $g : \Omega \to \mathbb{D}$ be a conformal equivalence. If L is a free analytic boundary arc of Ω and K is a compact subset of L, then g has an analytic continuation to a region Λ containing $\Omega \cup K$.*

Proof. Use Exercise 7 and Theorem 13.4.8. □

Note that even though the function g in Proposition 5.15 is one-to-one, its extension need not be.

5.16 Example. Let $\Omega = \mathbb{D}_+$ and define $g : \Omega \to \mathbb{D}$ as the composition of the function $h(z) = z + z^{-1}$ and the Möbius transformation $T(z) = (z-i)(z+i)^{-1}$:

$$g(z) = T \circ h(z) = \frac{z^2 - iz + 1}{z^2 + iz + 1}.$$

The function h maps $\mathbb{D}_+$ onto the upper half plane and T maps this half plane onto $\mathbb{D}$. The upper half circle L is a free analytic boundary arc of Ω so that g has an analytic continuation across L. In fact it is easy to see that $h^{\#}(z) = h(z)$ on Ω and so $g^{\#}(z) = g(z)$. Thus even though g is univalent on Ω, $g^{\#}$ is not.

The following question arises. If Ω is a Jordan region and the curve that forms the boundary of Ω has additional smoothness properties, does the boundary function of the Riemann map $\tau : \mathbb{D} \to \Omega$ have similar smoothness properties? If $\partial\Omega$ is an analytic curve, we have that τ has an analytic continuation to a neighborhood of cl $\mathbb{D}$ by the Schwarz Reflection Principle. But what if $\partial\Omega$ is just C^∞; or C^1? A discussion of this question is in Bell and Krantz [1987]. In particular, they show that if $\partial\Omega$ is C^∞, then so is the boundary function of τ.

Exercises

1. Prove Proposition 5.2.

2. This exercise will obviate the need for Theorem 13.5.3 in the proof of Theorem 5.6. Let τ be a bounded analytic function on $\mathbb{D}$ and let J be an open arc of $\partial\mathbb{D}$. Show, without using Theorem 13.5.3, that if τ has a radial limit at each point of J and this limit is 0, then $\tau \equiv 0$. (Hint: For a judicious choice of $w_1, \ldots, w_m$ in $\partial\mathbb{D}$, consider the function $h(z) = \tau(w_1 z)\tau(w_2 z)\ldots\tau(w_m z)$.)

3. Let G be a region and suppose that $\zeta_0 \in \partial G$ such that there is a $\delta > 0$ with the property that $B(a;\delta) \cap G$ is simply connected and

$B(a;\delta) \cap \partial G$ is a Jordan arc γ. Let Ω be a finitely connected region whose boundary consists of pairwise disjoint Jordan curves. Show that if $f : G \to \Omega$ is a conformal equivalence, then f has a continuous one-to-one extension to $G \cup \gamma$.

4. Show that every point of $\partial\,\mathbb{D}$ is a simple boundary point. If Ω is the slit disk $\mathbb{D} \setminus (-1, 0]$, show that points of the interval $(-1, 0)$ are not simple boundary points while all the remaining points are.

5. Show that if ω is a simple boundary point of Ω, then there is a $\delta > 0$ such that $B(\omega;\delta) \cap \Omega$ is connected.

6. Show that the conclusion of Theorem 5.12 remains valid if Ω is not assumed to be bounded but $\mathbb{C} \setminus \text{cl}\ \Omega$ has interior. Is the conclusion always valid?

7. Show that if J is a free analytic boundary arc of Ω, then every point of J is a simple boundary point.

8. (a) If $g : \text{cl}\ \mathbb{D} \to \mathbb{C}$ is a continuous function that is analytic in $\mathbb{D}$, show that there is a sequence of polynomials $\{p_n\}$ that converges uniformly on cl $\mathbb{D}$ to g. (Hint: For $0 < r < 1$, consider the function $g_r : \text{cl}\ \mathbb{D} \to \mathbb{C}$ defined by $g_r(z) = g(rz)$.) (b) If γ is a Jordan curve, Ω is the inside of γ, and $f : \text{cl}\ \Omega \to \mathbb{C}$ is a continuous function that is analytic on Ω, show that there is a sequence of polynomials $\{p_n\}$ that converges uniformly on cl Ω to f.

9. If Ω is any bounded region in the plane and $f : \text{cl}\ \Omega \to \mathbb{C}$ is a continuous function that is analytic on Ω and if there is a sequence of polynomials $\{p_n\}$ that converges uniformly on cl Ω to f, show that f has an analytic continuation to int $[\hat{\Omega}]$, where $\hat{\Omega}$ is the polynimally convex hull of Ω.

10. Suppose that G and Ω are simply connected Jordan regions and f is a continuous function on cl G such that f is analytic on G and $f(G) \subseteq \Omega$. Show that if f maps $\partial\,G$ homeomorphically onto $\partial\,\Omega$, then f is univalent on G and $f(G) = \Omega$.

§6 The Area Theorem

If f is analytic near infinity, then it is analytic on a set of the form $G = \{z : |z| > R\} = \text{ann}(0; R, \infty)$, and thus f has a Laurent expansion in G

$$f(z) = \sum_{n=-\infty}^{\infty} \alpha_n z^n;$$

this series converges absolutely and uniformly on compact subsets of G. With this notion, f has a pole at ∞ of order p if $\alpha_n = 0$ for $n > p$. Note that this is the opposite of the discussion of poles at finite points. The residue of f at ∞ is the coefficient α_1 and f has a removable singularity at ∞ if this expansion has the form

$$f(z) = \alpha_0 + \frac{\alpha_1}{z} + \frac{\alpha_2}{z^2} + \cdots.$$

Here $\alpha_0 = f(\infty)$, $\alpha_1 = f'(\infty) = \lim_{z\to\infty} z\left(f(z) - \alpha_0\right), \ldots$.

Consider the collection $\mathcal{U}$ of functions f that are univalent in $\mathbb{D}^* \equiv \{z : |z| > 1\}$ and have the form

6.1
$$f(z) = z + \alpha_0 + \frac{\alpha_1}{z} + \cdots.$$

In other words, $\mathcal{U}$ consists of all univalent functions on $\mathbb{D}^*$ with a simple pole at ∞ and residue 1. This class of functions can be characterized without reference to the Laurent expansion. The easy proof is left to the reader.

6.2 Proposition. *A function f belongs to the class $\mathcal{U}$ if and only if f is a univalent analytic function on $\mathbb{D}^*$ such that $f(\infty) = \infty$ and $f - z$ has a removable singularity at ∞.*

6.3 Area Theorem. *If $f \in \mathcal{U}$ and f has the expansion (6.1), then*

$$\sum_{n=1}^{\infty} n|\alpha_n|^2 \leq 1.$$

Proof. For $r > 1$, let Γ_r be the curve that is the image under f of the circle $|z| = r$. Because f is univalent, Γ_r is a smooth Jordan curve; let Ω_r be the inside of Γ_r. Applying Green's Theorem to the function $u = \bar{z}$ we get that

$$\begin{aligned} \text{Area}(\Omega_r) &= \iint_{\Omega_r} \bar{\partial}(\bar{z}) \\ &= \frac{1}{2i}\int_{\Gamma_r} \bar{z} \\ &= \frac{1}{2i}\int_0^{2\pi} \overline{\Gamma_r(t)}\Gamma'(t)dt. \end{aligned}$$

Since $\Gamma_r(t) = f(re^{it})$, this means that

$$\text{Area}(\Omega_r) = \frac{r}{2}\int_0^{2\pi} \overline{f(re^{it})}f'(re^{it})e^{it}dt.$$

Using (6.1) we can calculate that

$$\begin{aligned}
\overline{f(re^{it})} &= re^{-it} + \overline{\alpha}_0 + \sum_{n=1}^{\infty} \frac{\overline{\alpha_n}}{r_n} e^{int}, \\
f'(re^{it}) &= 1 - \sum_{m=1}^{\infty} m \frac{\alpha_m}{r^{m+1}} e^{-i(m+1)t}.
\end{aligned}$$

Using the fact that $\int_0^{2\pi} e^{int} dt = 0$ unless $n = 0$, in which case the integral is 2π, the uniform convergence of the above series implies that

$$\begin{aligned}
0 \leq \text{Area}(\Omega_r) &= \frac{r}{2}\left[2\pi r - \sum_{n=1}^{\infty} \frac{n|\alpha_n|^2}{r^{2n+1}}(2\pi)\right] \\
&= \pi r^2 - \pi \sum_{n=1}^{\infty} \frac{n|\alpha_n|^2}{r^{2n}}. \qquad \textbf{6.4}
\end{aligned}$$

Therefore

$$1 \geq \sum_{n=1}^{\infty} \frac{n|\alpha_n|^2}{r^{2n+2}}$$

for all $r > 1$. If the inequality is not valid for $r = 1$, then there is an integer N such that $1 < \sum_1^N n|\alpha_n|^2$. But since $r > 1$ this also gives that $1 < \sum_1^N n|\alpha_n|^2/r^{2n+2}$, a contradiction. □

Part of the proof of the preceding theorem, namely Equation 6.4, indicates why this result has its name. What happens to (6.4) when r is allowed to approach 1? Technically we must appeal to measure theory but the result is intuitively clear.

If $X_r = f(\{z : |z| \geq r\})$ for $r > 1$, then $\Omega_r = \mathbb{C} \setminus X_r$. Thus $\bigcap_r \Omega_r = \mathbb{C} \setminus \bigcup_r X_r = \mathbb{C} \setminus (\{z|z| > 1\}) = E$, a closed set. As $r \to 1$, $\text{Area}(\Omega_r) \to \text{Area}(E)$. Thus the following corollary.

6.5 Corollary. *If $f \in \mathcal{U}$, f has the Laurent expansion (6.1), and $E = \mathbb{C} \setminus f(\mathbb{D}^*)$, then*

$$\text{Area}(E) = \pi - \pi \sum_{n=1}^{\infty} n|\alpha_n|^2.$$

Thus $\text{Area}(E) = 0$ *if and only if equality occurs in the Area Theorem.*

The next proposition provides a uniqueness statement about the mappings in the class $\mathcal{U}$. Note that if $f \in \mathcal{U}$ and f is considered as a mapping on the extended plane $\mathbb{C}_\infty$, then $f(\infty) = \infty$.

6.6 Proposition. *If $f \in \mathcal{U}$ and $f(\mathbb{D}^*) = \mathbb{D}^*$, then $f(z) = z$ for all z.*

Proof. If f is as in the statement of the proposition, then Corollary 6.5 implies that $\pi = \pi - \pi \sum_n n|\alpha_n|^2$, so that $\alpha_n = 0$ for all $n \geq 1$. Thus

$f(z) = z+\alpha_0$. On the other hand, the hypothesis on the mapping properties of f also implies that $|f(z)| \to 1$ as $|z| \to 1$. Letting $z \to 1$, this implies that $|f(z)|^2 = |z+\alpha_0|^2 = |z|^2+2\text{Re}\,(\alpha_0\bar{z})+|\alpha_0|^2 \to 1+2\text{Re}\,(\alpha_0)+|\alpha_0|^2 = 1$. Thus $\text{Re}\,(\alpha_0)+|\alpha_0|^2 = 0$. Similarly, letting $z \to -1$ show that $-\text{Re}\,(\alpha_0)+|\alpha_0|^2 = 0$. We now conclude that $\alpha_0 = 0$ and so $f(z) = z$. $\square$

The preceding proposition can also be proved using Schwarz's Lemma (Exercise 3).

6.7 Proposition. *If $f \in \mathcal{U}$ and f has the expansion (6.1), then $|\alpha_1| \leq 1$. Moreover $|\alpha_1| = 1$ if and only if the set $E = \mathbb{C} \setminus f(\mathbb{D}^*)$ is a straight line segment of length 4. In this case $f(z) = z + \alpha_0 + \alpha_1 z^{-1}$ and $E = [-2\lambda + \alpha_0,\ 2\lambda + \alpha_0]$, where $\lambda^2 = \alpha_1$.*

Proof. Since $|\alpha_1|$ is one of the terms in the sum appearing in the Area Theorem, it is clear that $|\alpha_1| \leq 1$. If $|\alpha_1| = 1$, then $\alpha_n = 0$ for $n \geq 2$. Thus $f(z) = z + \alpha_0 + \alpha_1 z^{-1}$. It can be seen by using Exercise 2 that in this case $E = [-2\lambda + \alpha_0, 2\lambda + \alpha_0]$, where $\lambda^2 = \alpha_1$. In particular, E is a straight line segment of length 4.

Conversely assume that E is a straight line segment of length 4; so E has the form $E = [-2\mu + \beta_0, 2\mu + \beta_0]$, where β_0 and μ are complex numbers and $|\mu| = 1$. If $g(z) = z+\beta_0+\mu^2 z^{-1}$, then $g \in \mathcal{U}$ and $g(\mathbb{D}^*) = \mathbb{C}\setminus E = f(\mathbb{D}^*)$. Therefore $f \circ g^{-1} \in \mathcal{U}$ and maps $\mathbb{D}^*$ onto itself. By Proposition 6.6, $f = g$ and $\alpha_1 = \mu^2$, so that $|\alpha_1| = 1$. $\square$

The next proposition is a useful estimate of the derivative of a function in $\mathcal{U}$.

6.8 Proposition. *If $f \in \mathcal{U}$, then*

$$|f'(z)| \leq \frac{|z|^2}{|z|^2 - 1}$$

whenever $|z| > 1$. Equality occurs at some number a with $|a| > 1$ if and only if f is given by the formula

$$f(z) = z + \alpha_0 - \frac{|a|^2 - 1}{\bar{a}(\bar{a}z - 1)}.$$

Proof. Since $f'(z) = 1 - \alpha_1 z^{-2} - \alpha_2 z^{-3} - \cdots = 1 - \sum_n n\alpha_n z^{-n-1}$, an application of the Cauchy-Schwarz Inequality as well as the Area Theorem shows that

$$\begin{aligned} |f'(z) - 1| &= \left| \sum_{n=1}^{\infty} \left(\sqrt{n}\alpha_n\right) \left(\sqrt{n}z^{-n-1}\right) \right| \\ &\leq \left[\sum_{n=1}^{\infty} n|\alpha_n|^2 \right]^{\frac{1}{2}} \left[\sum_{n=1}^{\infty} n|z|^{-2n-2} \right]^{\frac{1}{2}} \end{aligned}$$

$$
\begin{aligned}
&\leq \left[\sum_{n=1}^{\infty} n|z|^{-2n-2}\right]^{\frac{1}{2}} \\
&= |z|^{-2}\left[\sum_{n=1}^{\infty} n\left(|z|^{-2}\right)^{n-1}\right]^{\frac{1}{2}} \\
&= |z|^{-2}\frac{1}{1-\frac{1}{|z|^2}} \\
&= \frac{1}{|z|^2-1}.
\end{aligned}
$$

It now follows that $|f'(z)| \leq |f'(z)-1|+1 \leq (|z|^2-1)^{-1}+1$ $= |z|^2(|z|^2-1)^{-1}$.

Now suppose that there is a complex number a, $|a|>1$, such that the inequality becomes an equality when $z=a$. Thus $|f'(a)| \leq |f'(a)-1|+1 \leq |a|^2(|a|^2-1)^{-1} = |f'(a)|$. This implies that the two inequalities in the above display become equalities when $z=a$. The fact that the first of these becomes an equality means there is equality in the Cauchy-Schwarz Inequality. Therefore there is a complex number b such that $\alpha_n = b\overline{a}^{-n-1}$ for all $n \geq 1$. The fact that the second inequality becomes an equality means that

$$
\begin{aligned}
1 &= \sum_{n=1}^{\infty} n|\alpha_n|^2 \\
&= |b|^2\sum_{n=1}^{\infty} n|a|^{-2n-2} \\
&= |b|^2\frac{1}{(|a|^2-1)^2}.
\end{aligned}
$$

Thus $|b| = |a|^2-1$. Substituting these relations in the Laurent expansion for f gives

$$
\begin{aligned}
f(z) &= z+\alpha_0+\sum_{n=1}^{\infty}\frac{b\overline{a}^{n-1}}{z^n} \\
&= z+\alpha_0+\frac{b}{\overline{a}}\sum_{n=1}^{\infty}\left(\frac{1}{\overline{a}z}\right)^n \\
&= z+\alpha_0+\frac{b}{\overline{a}}\left[\left(1-\frac{1}{\overline{a}z}\right)^{-1}-1\right] \\
&= z+\alpha_0-\frac{b}{\overline{a}}\left[\frac{1}{\overline{a}z-1}\right].
\end{aligned}
$$

Now use this formula for $f(z)$ to compute $f'(a)$:

$$
f'(a) = 1-\frac{b}{\overline{a}}(\overline{a}a-1)^{-2}\overline{a}
$$

$$
\begin{aligned}
&= 1 - b(|a|^2 - 1)^{-2} \\
&= \frac{(|a|^2 - 1)^2 - b}{(|a|^2 - a)^2} \\
&= \frac{\bar{b}b - b}{\bar{b}b} \\
&= \frac{\bar{b} - 1}{\bar{b}}.
\end{aligned}
$$

By assumption

$$
\begin{aligned}
|f'(a)|^2 &= \left(\frac{|a|^2}{|a|^2 - 1}\right)^2 \\
&= \frac{(|b| - 1)^2}{|b|^2}.
\end{aligned}
$$

Equating the two expressions for $|f'(a)|^2$ we get that $\operatorname{Re} b = \operatorname{Re} \bar{b} = -|b|$. It follows that $b = -|b| = 1 - |a|^2$ and so f has the desired form.

If f is given by the stated formula it is routine to check that equality occurs when $z = a$. □

Exercises

1. Show that for r and β any complex numbers, $f(z) = 4r\beta z(\beta - z)^{-2}$ is the composition $f = f_3 \circ f_2 \circ f_1$, where $f_1(z) = (\beta + z)/(\beta - z)$, $f_2(z) = z^2$, and $f_3(z) = r(z-1)$. Use this to show that f is a conformal equivalence of $\{z : |z| < |\beta|\}$ as well as $\{z : |z| > |\beta|\}$ onto the split plane $\mathbb{C} \setminus \{z = -rt : t \geq 1\}$.

2. For a complex number λ, show that $f(z) = z + \lambda^2 z^{-1}$ is the composition $f_2 \circ f_1$, where $f_1(z) = z(\lambda - z)^{-2}$ and $f_2(z) = (1 + 2\lambda z)/z$. Use this to show that f is a conformal equivalence of both $\{z : |z| < |\lambda|\}$ and $\{z : |z| > |\lambda|\}$ onto $\mathbb{C} \setminus [-2\lambda, 2\lambda]$.

3. Prove Proposition 6.6 using Schwarz's Lemma.

§7 Disk Mappings: The Class $\mathcal{S}$

In this section attention is focused on a class of univalent functions on the open unit disk, $\mathbb{D}$. Since each simply connected region is the image of $\mathbb{D}$ under a conformal equivalence, the study of univalent functions on

$\mathbb{D}$ is equivalent to the study of univalent functions on arbitrary simply connected regions. If ∞ is adjoined to the region $\mathbb{D}^*$, the resulting region (also denoted by $\mathbb{D}^*$) is also simply connected, so that it is equivalent to consider univalent functions on $\mathbb{D}^*$. After suitable normalization this amounts to a consideration of functions in the class $\mathcal{U}$. The class of univalent functions $\mathcal{S}$ on $\mathbb{D}$ defined below is in one-to-one correspondence with a subset of the class $\mathcal{U}$. The study of $\mathcal{S}$ is classical and whether to study $\mathcal{S}$ or $\mathcal{U}$ depends on your perspective, though one class sometimes offers certain technical advantages over the other.

7.1 Definition. The class $\mathcal{S}$ consists of all univalent functions f on $\mathbb{D}$ such that $f(0) = 0$ and $f'(0) = 1$.

The reason for the use of the letter $\mathcal{S}$ to denote this class of functions is that they are called *Schlicht functions.*

If h is any univalent function on $\mathbb{D}$, then $f = [h - h(0)]/h'(0)$ belongs to $\mathcal{S}$, so that information about the functions in $\mathcal{S}$ gives information about all univalent functions on $\mathbb{D}$. If $f \in \mathcal{S}$, then the power series expansion of f about zero has the form

7.2 $$f(z) = z + a_2 z^2 + a_3 z^3 + \cdots.$$

As mentioned the class $\mathcal{S}$ and the class $\mathcal{U}$ from the preceding section are related. This relation is given in the next proposition.

7.3 Proposition.

(a) *If $g \in \mathcal{U}$ and g never vanishes, then $f(z) = \left[g(z^{-1})\right]^{-1} \in \mathcal{S}$ and, conversely, if $f \in \mathcal{S}$, then $g(z) = \left[f(z^{-1})\right]^{-1} \in \mathcal{U}$ and g never vanishes.*

(b) *If $f \in \mathcal{S}$ with power series given by (7.2) and $\left[f(w^{-1})\right]^{-1} = g(w) = w^{-1} + \sum_0^\infty \alpha_n z^{-n}$ for w in $\mathbb{D}^*$, then $\alpha_0 = -a_2$.*

Proof. (a) Suppose $g \in \mathcal{U}$ and $f(z) = \left[g(z^{-1})\right]^{-1}$ for z in $\mathbb{D}$. Since $g(\infty) = \infty$, it is clear that f is univalent on $\mathbb{D}$ and $f(0) = 0$. Moreover $g(z)/z \to 1$ as $z \to \infty$ and so it follows that $f'(0) = \lim_{z\to 0} f(z)/z = 1$ and $f \in \mathcal{S}$. The proof of the converse is similar.

(b) Just use the fact that for $|z| < 1$, $g(z^{-1})f(z) = 1$, perform the required multiplication of the corresponding series, and set equal to 0 all the coefficients of the non-constant terms. □

7.4 Proposition.

(a) *If $f \in \mathcal{S}$ and n is any positive integer, then there is a unique function g in $\mathcal{S}$ such that $g(z)^n = f(z^n)$. For such a function g, $g(\omega z) = \omega g(z)$ for any n-th root of unity ω and all z in $\mathbb{D}$. Conversely, if $g \in \mathcal{S}$ and $g(\omega z) = \omega g(z)$ for any n-th root of unity ω and all z in $\mathbb{D}$, then there is a function f in $\mathcal{S}$ such that $g(z)^n = f(z^n)$.*

(b) *Similarly, if $f \in \mathcal{U}$, then there is a unique function g in $\mathcal{U}$ such that $g(z)^n = f(z^n)$. For such a function g, $g(\omega z) = \omega g(z)$ for any n-th root of unity ω and all z in $\mathbb{D}$. Conversely, if $g \in \mathcal{U}$ such that $g(\omega z) = \omega g(z)$ for any n-th root of unity ω and all z in $\mathbb{D}$, then there is a function f in $\mathcal{U}$ such that $g(z)^n = f(z^n)$.*

Proof. (a) Assume that $f \in \mathcal{S}$ and let $h(z) = f(z^n)$ for $|z| < 1$. The only zero of h in $\mathbb{D}$ is the one at $z = 0$ and this has order n. Thus $h(z) = z^n h_1(z)$ and h_1 is analytic on $\mathbb{D}$ and does not vanish. Moreover the fact that $f'(0) = 1$ implies that $h_1(0) = 1$. Thus there is a unique analytic function g_1 on $\mathbb{D}$ such that $g_1^n = h_1$ and $g_1(0) = 1$. Put $g(z) = zg_1(z)$; clearly $g(z)^n = f(z^n)$, $g(0) = 0$, and $g'(0) = \lim_{z\to 0} g(z)/z = g_1(0) = 1$. Notice that these properties uniquely determine g. Indeed, if k is any analytic function on $\mathbb{D}$ such that $k(z)^n = f(z^n)$ and $k'(0) = 1$, then $[g/k]^n = 1$ and g/k is analytic, whence the conclusion that $k = g$.

If the power series of f is given by (7.2), then a calculation shows that $h_1(z) = 1 + a_2 z^n + a_3 z^{2n} + \cdots$, so that $h(\omega z) = h(z)$ whenever $\omega^n = 1$. Thus for an n-th root of unity ω, $k(z) = zg_1(\omega z)$ has the property that $k(z)^n = f(z^n)$ and $k'(0) = 1$. By the uniqueness statement above, $k = g$. Thus $g_1(\omega z) = g_1(z)$. From here it follows that $g(\omega z) = \omega g(z)$ whenever $\omega^n = 1$.

To complete the proof that $g \in \mathcal{S}$ it remains to show that g is univalent. If $g(z) = g(w)$, then $f(z^n) = f(w^n)$ and so $z^n = w^n$; thus there is an n-th root of unity such that $w = \omega z$. So $g(z) = g(\omega z) = \omega g(z)$. Clearly we can assume that $z \neq 0$ so that $g(z) \neq 0$ and hence $\omega = 1$; that is, $w = z$.

For the converse, if $g \in \mathcal{S}$ and $g(\omega z) = \omega g(z)$ for any n-th root of unity ω and all z in $\mathbb{D}$, then g has a power series representation of the form

$$g(z) = z + b_{n+1} z^{n+1} + b_{2n+1} z^{2n+1} + \cdots.$$

Thus

$$g(z)^n = z^n + c_{2n} z^{2n} + \cdots.$$

Let

$$f(z) = z + c_{2n} z^2 + \cdots.$$

The radius of convergence of this power series is at least 1, $f(0) = 0$, and $f'(0) = 1$. If z and $w \in \mathbb{D}$ and $f(z) = f(w)$, let z_1 and w_1 be points in $\mathbb{D}$ with $z_1^n = z$ and $w_1^n = w$. So $g(z_1)^n = g(w_1)^n$. It is left to the reader to show that this implies there is an n-th root of unity ω such that $z_1 = \omega w_1$. Hence $z = w$ and so f is univalent. That is, $f \in \mathcal{S}$.

(b) This proof is similar. □

The celebrated Bieberbach Conjecture concerns the class $\mathcal{S}$. Precisely, this says that if $f \in \mathcal{S}$ and its power series is given by (7.2), then

7.5 $$|a_n| \leq n$$

for all $n \geq 2$. Moreover, equality occurs if and only if f is the Koebe function or one of its rotations (1.4).

We will prove the Bieberbach conjecture now for $n = 2$. The proof of the general case is due to L deBranges [1985]. This material is presented in Chapter 17. We start with a corresponding inequality for the class $\mathcal{U}$, which is stated separately.

7.6 Theorem. *If $g \in \mathcal{U}$ with Laurent series $g(z) = z+\alpha_0+\alpha_1 z^{-1}+\cdots$, then $|\alpha_0| \leq 2$. Equality occurs if and only if $g(z) = z+2\lambda+\lambda^2 z^{-1} = z^{-1}(z+\lambda)^2$, where $|\lambda| = 1$. In this case g maps $\mathbb{D}^*$ onto $\mathbb{C} \setminus [0, 4\lambda]$.*

Proof. Let $h \in \mathcal{U}$ such that $h(z)^2 = g(z^2)$ for z in $\mathbb{D}^*$ and let the expansion of h be given by $h(z) = z + \beta_0 + \beta_1 z^{-1} + \cdots$. Thus

$$\begin{aligned} h(z)^2 &= z^2 + 2\beta_0 z + (\beta_0^2 + 2\beta_1) + \cdots \\ &= g(z^2) \\ &= z^2 + \alpha_0 + \alpha_1 z^{-2} + \cdots. \end{aligned}$$

Hence $\beta_0 = 0$ and $\alpha_0 = 2\beta_1$. But according to Proposition 6.7, $|\beta_1| \leq 1$ so $|\alpha_0| \leq 2$. The equality $|\alpha_0| = 2$ holds if and only if $|\beta_1| = 1$, in which case $h(z) = z + \lambda z^{-1}$, where $\lambda^2 = \beta_1$ (so $|\lambda| = 1$). But in this case, $g(z^2) = h(z)^2 = (z+\lambda z^{-1})^2 = z^2+2\lambda+\lambda^2 z^{-2}$, so that $g(z) = z+2\lambda+\lambda^2 z^{-1}$. The mapping properties of this function are left for the reader to verify. (See Exercise 6.2) □

7.7 Theorem. *If $f \in \mathcal{S}$ with power series given by (7.2), then $|a_2| \leq 2$. Equality occurs if and only if f is a rotation of the Koebe function.*

Proof. Let g be the corresponding function in the class $\mathcal{U}$: $g(z) = \left[f(z^{-1})\right]^{-}$ for z in $\mathbb{D}^*$. It follows that g has the Laurent series

$$g(z) = z - a_2 + (a_2^2 - a_3)z^{-1} + \cdots.$$

The fact that $|a_2| \leq 2$ now follows the preceding theorem. Moreover equality occurs if and only if there is a λ, $|\lambda| = 1$, such that $g(z) = z^{-1}(z+\lambda)^2$. This is equivalent to having f be a rotation of the Koebe function. □

As an application this theorem is used to demonstrate the Koebe "1/4-theorem."

7.8 Theorem. *If $f \in \mathcal{S}$, then $f(\mathbb{D}) \supseteq \{\zeta : |\zeta| < 1/4\}$.*

Proof. Fix f in $\mathcal{S}$ and let ζ_0 be a complex number that does not belong to $f(\mathbb{D})$; it must be shown that $|\zeta| \geq 1/4$. Since $0 \in f(\mathbb{D})$, $\zeta_0 \neq 0$ and so $g(z) = f(z)\left[1 - \zeta_0^{-1} f(z)\right]^{-1}$ is an analytic function on $\mathbb{D}$. In fact $g \in \mathcal{S}$. To see this first observe that $g(0) = 0$ and $g'(0) = \lim_{z\to 0}[g(z)/z] = f'(0) = 1$. Finally g is the composition of f and a Möbius transformation and hence must be univalent.

Since $f(0) = 0$, there is a small value of r such that $|f(z)| < |\zeta_0|$ for $|z| < r$. In this neighborhood of 0 we get

$$\left[1 - \zeta_0^{-1} f(z)\right]^{-1} = 1 + \zeta_0^{-1} f(z) + \zeta_0^{-2} f(z)^2 + \cdots.$$

Substituting the power series expansion (7.2) of f and collecting terms we get that for $|z| < r$

$$g(z) = z + \left(\zeta_0^{-1} + a_2\right) z^2 + \cdots.$$

(In fact this power series converges throughout the unit disk.) By Theorem 7.7 this implies that $|\zeta_0^{-1} + a_2| \leq 2$. But $|a_2| \leq 2$ so that $|\zeta_0^{-1}| \leq 4$, or $|\zeta_0| \geq 1/4$. □

Consideration of the Koebe function shows that the constant $1/4$ is sharp. The next result is often called the *Koebe Distortion Theorem.*

7.9 Theorem. *If $f \in \mathcal{S}$ and $|z| < 1$, then:*

(a)
$$\frac{1 - |z|}{(1 + |z|)^3} \leq |f'(z)| \leq \frac{1 + |z|}{(1 - |z|)^3};$$

(b)
$$\frac{|z|}{(1 + |z|)^2} \leq |f(z)| \leq \frac{|z|}{(1 - |z|)^2}.$$

Equality holds for one of these four inequalities at some point $z \neq 0$ if and only if f is a rotation of the Koebe function.

Proof. For each complex number a in $\mathbb{D}$ define the function

$$f_a(z) = \frac{f\left(\frac{z+a}{1+\bar{a}z}\right) - f(a)}{(1 - |a|^2) f'(a)}.$$

It is easy to see that f_a is univalent on $\mathbb{D}$ since it is the composition of univalent functions. Also $f_a(0) = 0$ and a routine calculation shows that $f_a'(0) = 1$. Therefore $f_a \in \mathcal{S}$. Let $f_a(z) = z + b_2 z^2 + \cdots$ in $\mathbb{D}$.

Another computation reveals that

$$f_a''(0) = \left[\left(1 - |a|^2\right) \frac{f''(a)}{f'(a)} - 2\bar{a}\right].$$

Since $b_2 = f_a''(0)/2$ and $|b_2| \leq 2$ by Theorem 7.7, this shows that

$$\left|\left(1 - |a|^2\right) \frac{f''(a)}{f'(a)} - 2\bar{a}\right| \leq 4.$$

Thus

$$\left|\frac{f''(a)}{f'(a)} - \frac{2\bar{a}}{1 - |a|^2}\right| \leq \frac{4}{1 - |a|^2}.$$

Multiply both sides of this inequality by $|a|$ and substitute $z = a$ to get

7.10
$$\left| z\frac{f''(z)}{f'(z)} - \frac{2|z|^2}{1-|z|^2} \right| \leq \frac{4|z|}{1-|z|^2}.$$

Now f' does not vanish on $\mathbb{D}$, so there is an analytic branch of $\log f'(z)$ with $\log f'(0) = 0$. Using the chain rule,

$$\begin{aligned} \frac{\partial}{\partial r}\left[\log f'(re^{i\theta})\right] &= \frac{\partial}{\partial r}\left[\log f'(z)\right]\frac{\partial}{\partial r}[z] + \frac{\partial}{\partial \bar{z}}\left[\log f'(z)\right]\frac{\partial}{\partial r}[\bar{z}] \\ &= \frac{f''(re^{i\theta})}{f'(re^{i\theta})}e^{i\theta}. \end{aligned}$$

Now for any function g, $\text{Re}[\partial g/\partial r] = \partial\,\text{Re}[g]/\partial r$, so

7.11
$$r\frac{\partial}{\partial r}\left[\log|f'(re^{i\theta})|\right] = \text{Re}\left[z\frac{f''(z)}{f'(z)}\right].$$

Thus (7.10) implies

$$\left| r\frac{\partial}{\partial r}\left[\log|f'(re^{i\theta})|\right] - \frac{2r^2}{1-r^2} \right| \leq \frac{4r}{1-r^2}.$$

Dividing by r and performing some algebraic manipulation gives that

$$\frac{2r-4}{1-r^2} \leq \frac{\partial}{\partial r}\left[\log|f'(re^{i\theta})|\right] \leq \frac{2r+4}{1-r^2}$$

for all $re^{i\theta}$ in $\mathbb{D}$. Thus, for $\rho < 1$,

7.12
$$\int_0^\rho \frac{2r-4}{1-r^2}dr \leq \int_0^\rho \frac{\partial}{\partial r}\left[\log|f'(re^{i\theta})|\right]dr \leq \int_0^\rho \frac{2r+4}{1-r^2}dr$$

or

$$\log\left[\frac{1-\rho}{(1+\rho)^3}\right] \leq \log|f'(\rho e^{i\theta})| \leq \log\left[\frac{1+\rho}{(1-\rho)^3}\right].$$

Now take the exponential of both sides of these inequalities to obtain the inequality in (a) for $z = \rho e^{i\theta}$.

Suppose for some $z = \rho e^{i\theta}$ one of the inequalities in (a) is an equality; for specificity, assume that equality occurs in the lower bound. It follows that the first inequality in (7.12) is an equality. Thus the integrands are equal for $0 \leq r \leq \rho$. Using (7.11) for $0 \leq r \leq \rho$ and letting $r \to 0$ we get that $-4 = \text{Re}[e^{i\theta}f''(0)]$, so that $|f''(0)| \geq 4$. By Theorem 7.7, $|f''(0)| = 4$ and f is a rotation of the Koebe function. The proof of the case for equality in the upper bound is similar.

To prove part (b) note that $|f(z)| = \left|\int_{[0,z]} f'(\zeta)d\zeta\right| \leq \int_{[0,z]} |f'(\zeta)||d\zeta|$. Parameterize the line segment by $\zeta = tz$, $0 \leq t \leq 1$, and use part (a) to get an upper estimate for $|f'(\zeta)|$. After performing the required calculations this shows that

$$|f(z)| \leq \frac{|z|}{(1-|z|)^2},$$

the right hand side of part (b).

To get the left hand side of part (b), first note that an elementary argument using calculus shows that $t(1+t)^2 \leq 1/4$ for $0 \leq t \leq 1$; so it suffices to establish the inequality under the assumption that $|f(z)| < 1/4$. But here Koebe's 1/4-theorem implies that $\{\zeta : |\zeta| > 1/4\} \subseteq f(\mathbb{D})$. So fix z in $\mathbb{D}$ with $|f(z)| < 1/4$ and let γ be the path in $\mathbb{D}$ from 0 to z such that $f \circ \gamma$ is the straight line segment $[0, f(z)]$. That is, $f(\gamma(t)) = tf(z)$ for $0 \leq t \leq 1$. Thus $|f(z)| = |\int_\gamma f'(w)dw| = |\int_0^1 f'(\gamma(t))\gamma'(t)dt|$. Now $f'(\gamma(t))\gamma'(t) = [tf(z)]' = f(z)$ for all t. Thus $|f(z)| = \int_\gamma |f'(w)||dw|$. Using the appropriate part of (a) we get that $|f'(w)| \geq (1-|w|)(1+|w|)^{-3}$. On the other hand if we take $0 \leq s < t \leq 1$, $|\gamma(t) - \gamma(s)| \leq ||\gamma(t)| - |\gamma(s)||$ and so (symbolically), $|dw| \leq d|w|$. Combining these inequalities gives that

$$\begin{aligned} |f(z)| &\geq \int_0^{|z|} \frac{1-r}{(1+r)^3}dr \\ &= \frac{|z|}{(1+|z|)^2}. \end{aligned}$$

This proves (b). It is left to the reader to show that f is a rotation of the Koebe function if one of these two inequalities is an equality. □

Before giving an important corollary of this theorem, here is a lemma that appeared as Exercise 7.2.10.

7.13 Lemma. *If $\{f_n\}$ is a sequence of univalent functions on a region G and $f_n \to f$ in $H(G)$, then either f is univalent or f is constant.*

7.14 Corollary. *The set $\mathcal{S}$ of univalent functions is compact in $H(\mathbb{D})$.*

Proof. By Montel's Theorem (7.2.9) and Theorem 7.9, $\mathcal{S}$ is a normal family. It remains to show that $\mathcal{S}$ is closed (8.1.15). But if $\{f_n\} \subseteq \mathcal{S}$ and $f_n \to f$ in $H(\mathbb{D})$, then the preceding lemma implies that either f is univalent or f is constant. But $f_n'(0) = 1$ for all n so that $f'(0) = 1$ and f is not constant. Clearly $f(0) = 0$ and so $f \in \mathcal{S}$. □

The next result is almost a corollary of the preceding corollary, but it requires a little more proof than one usually associates with such an appelation.

7.15 Proposition. *If G is a region, $a \in G$, and b is any complex number, then $\mathcal{S}(G, a, b) \equiv \{f \in H(G) : f$ is univalent, $f(a) = b$, and $f'(a) = 1\}$ is compact in $H(G)$.*

Proof. By a simple translation argument it may be assumed that $a = b = 0$; let $\mathcal{S}(G) = \mathcal{S}(G, 0, 0)$. Let $R > 0$ such that $\overline{B}(0; R) \subseteq G$. If $f \in \mathcal{S}(G)$ and $f_R(z) = R^{-1} f(Rz)$ for z in $\mathbb{D}$, then $f_R \in \mathcal{S} = \mathcal{S}(\mathbb{D})$. Thus $f_R(\mathbb{D}) \supseteq \{\zeta : |\zeta| < 1/4\}$ and so $|f_R(e^{i\theta})| \geq 1/4$ for all θ (7.8). Hence $|f(Re^{i\theta})| \geq R/4$ for all θ and $\overline{B}(0; R/4) \subseteq f(\overline{B}(0; R))$. So f maps $G \setminus \overline{B}(0; R)$ into $f(G) \setminus \overline{B}(0; R/4)$; that is, $|f| \geq R/4$ for z in $G \setminus \overline{B}(0; R)$. Therefore $\phi_f(z) = z/f(z)$ is an analytic function on G, $|\phi_f(z)| \leq 4$ for $|z| \leq R$ and $|\phi_f(z)| \leq 4|z|R^{-1}$ for z in $G \setminus \overline{B}(0; R)$. Thus $\Phi \equiv \{\phi_f : f \in \mathcal{S}(G)\}$ is a locally bounded family of analytic functions on G and hence must be normal.

By an argument similar to that used to prove Corollary 7.14, $\mathcal{S}(G)$ is closed. So to prove the proposition, it suffices to show that $\mathcal{S}(G)$ is a normal family. Let $\{f_n\}$ be a sequence in $\mathcal{S}(G)$ and let $\{\phi_n\}$ be the corresponding sequence in Φ. By passing to a subsequence if necessary, it may be assumed that $\phi_n \to \phi$ for some analytic function ϕ on G. Clearly the functions ϕ_n have no zero in G so either $\phi \equiv 0$ or ϕ does not vanish in G (7.2.6). Also for each n, $\phi_n(0) = f_n'(0) = 1$ and so $\phi(0) = 1$ and hence ϕ has no zeros on G. Let $f(z) = z/\phi(z)$. Now $f'(0) = 1$, so f is not constant. Clearly $f_n(z) \to f(z)$ for all z in G. If K is a compact subset of G, let $\varepsilon > 0$ such that $|\phi| \geq 2\varepsilon$ on K. It follows that $|\phi_n| \geq \varepsilon$ on K for all n sufficiently large (see Exercise 4). This implies that $\{f_n\}$ is locally bounded on G and, hence, a normal family. □

We close with an extension of the Distortion Theorem; you might call this the Generalized Distortion Theorem.

7.16 Theorem. *If K is a compact subset of the region G, then there is a constant M (dependent on K) such that for every univalent function f on G and every pair of points z and w in K,*

$$\frac{1}{M} \leq \frac{|f'(z)|}{|f'(w)|} \leq M.$$

Proof. By interchanging the roles of z and w, it suffices to prove the second of these inequalities. Let $0 < 2d < \operatorname{dist}(K, \partial G)$ and cover K by a finite collection $\mathcal{B}$ of open disks of radius $d/8$. Suppose B_1 and B_2 are two of the disks from $\mathcal{B}$ such that $B_1 \cap B_2 \neq \emptyset$. Let $z_i \in B_i$, $i = 1, 2$. So $|z_1 - z_2| < d/2$ and $\overline{B}(z_i; d) \subseteq G$. Consider the function

$$g(z) = \frac{f(z_1 + dz) - f(z_1)}{df'(z_1)}.$$

This function belongs to the class $\mathcal{S}$. According to Theorem 7.9,

$$|g'(z)| \leq \frac{1 + |z|}{(1 - |z|)^3}$$

for $|z| < 1$. Making the appropriate substitutions we get

$$\left|\frac{f'(z_1 + dz)}{f'(z_1)}\right| \leq \frac{1 + |z|}{(1 - |z|)^3}.$$

Take $z = (z_1 - z_2)/d$ so that $|z| < 1/2$ and we get that

$$\left|\frac{f'(z_2)}{f'(z_1)}\right| \leq \frac{1 + \frac{1}{2}}{\left(1 - \frac{1}{2}\right)^3} = M_0.$$

If z and w are arbitrary points of K, then there are points $z_1 = z, z_2, \ldots, z_n = w$ such that each consecutive pair of points is in a disk from $\mathcal{B}$, $n \leq N$, the total number of disks in $\mathcal{B}$, and these disks are pairwise intersecting. Therefore

$$\left|\frac{f'(z)}{f'(w)}\right| = \left|\frac{f'(z_1)}{f'(z_2)}\right| \left|\frac{f'(z_2)}{f'(z_3)}\right| \cdots \left|\frac{f'(z_{n-1})}{f'(z_n)}\right| \leq M_0^{n-1} \leq M_0^{N-1} \equiv M.$$

□

Exercises

1. Let f and g be as in part (b) of Proposition 7.3 and show that $\alpha_1 = a_2^2 - a_3$. Show that $\alpha_2 = -a_4 + 2a_2a_3 - a_2^3$.

2. Let f be the function given in (7.5) and show that $f \in \mathcal{S}$ and if the power series of f is given by (7.2), then $|a_n| = n$ for all $n \geq 2$. Show that the image of $\mathbb{D}$ under f is the plane minus the radial slit from $\lambda/4$ to ∞ that does not pass through the origin.

3. Let $f_n(z) = z + nz^2$ and show that even though $f_n(0) = 0$ and $f_n'(0) = 1$ for all $n \geq 1$, $\{f_n\}$ is not a normal sequence.

4. Let $\{g_n\}$ be a normal sequence of analytic functions on a region G such that each g_n has no zeros in G and $g_n \to g$, where g is not identically 0. Show that if K is a compact subset of G, then there is an $\varepsilon > 0$ such that $|g_n| \geq \varepsilon$ on K for all $n \geq 1$.

5. Let G be a region and fix a point a in G. For a choice of positive constants C, m, and M show that $\mathcal{F} = \{f \in H(G) : f$ is univalent, $|f(a)| \leq C$, and $m \leq |f'(a)| \leq M\}$ is a compact subset of $H(G)$.

6. For the set $\mathcal{U}$ of univalent functions on $\mathbb{D}^* = \{\infty\} \cup \{z : |z| > 1\}$, show that $\mathcal{U} \cup \{\infty\}$ is compact in the space $C(\mathbb{D}^*, \mathbb{C}_\infty)$.

7. Show that for each integer $n \geq 2$ there is a function f in $\mathcal{S}$ such that if f has the power series expansion (7.2), then $a_n \geq g^{(n)}(0)/n!$ for all g in $\mathcal{S}$.

8. Show that if K is a compact subset of a region G, then there is a constant M such that for every pair of points z and w in K and for every univalent function f on G, $|f(z)| \leq |f(w)| + M|f'(w)|$.

9. If G is a simply connected region, $a \in G$, and $f : G \to \mathbb{D}$ is the conformal equivalence with $f(0) = a$ and $f'(0) > 0$, show that $[4\,\text{dist}(a, \partial G)]^{-1} \leq f'(a) \leq [\text{dist}(a, \partial G)]^{-1}$.

10. Show that if τ is a bounded univalent function on $\mathbb{D}$ with $G = \tau(\mathbb{D})$, then $|\tau'(z)| \leq 4(1 - |z|^2)^{-1}\text{dist}(\tau(z), \partial G)$. (Hint: Use Exercise 9 and Schwarz's Lemma.)

Chapter 15

Conformal Equivalence for Finitely Connected Regions

In this chapter it will be shown that each finitely connected region is conformally equivalent to a variety of canonical regions. Subject to certain normalizations, such conformal equivalences are unique. We begin with some basic facts about complex analysis on finitely connected regions.

§1 Analysis on a Finitely Connected Region

Say that a region G in $\mathbb{C}$ is *n-connected* if $\mathbb{C}_\infty \setminus G$ has $n+1$ components. Thus a 0-connected region is simply connected. Say that G is a *non-degenerate n-connected* region if it is an an n-connected region and no component of its complement in $\mathbb{C}_\infty$ is a singleton. A region G is *finitely connected* if it is n-connected for some non-negative integer n. Note that if G is any region in $\mathbb{C}_\infty$ and K is any component of $\mathbb{C}_\infty \setminus G$ that does not contain ∞, then K must be a compact subset of $\mathbb{C}$; in fact, such components of $\mathbb{C}_\infty \setminus G$ are precisely the bounded components of $\mathbb{C} \setminus G$.

Throughout this section the following notation will be fixed: G is an n-connected region in $\mathbb{C}$ and $K_1, \ldots, K_n$ are the bounded components of $\mathbb{C} \setminus G$; K_0 will be the component of $\mathbb{C}_\infty \setminus G$ that contains ∞. Note that for any j, $1 \leq j \leq n$, $G \cup K_j$ is an $(n-1)$-connected region.

If E is any compact subset of G, then by Proposition 13.1.5 there is a positive Jordan system $\Gamma = \{\gamma_0, \gamma_1, \ldots, \gamma_n\}$ in G having the following properties:

1.1
$$\begin{array}{ll} \text{(i)} & E \subseteq \text{ins}\,\Gamma \subseteq \text{G}; \\ \text{(ii)} & \text{for } 1 \leq j \leq n, \ K_j \subseteq \text{ins}\,\gamma_\text{j}; \\ \text{(iii)} & \text{cl}(\text{ins}\,\gamma_\text{j}) \cap \text{cl}(\text{ins}\,\gamma_\text{k}) = \emptyset \text{ for j} \neq \text{k}. \end{array}$$

The idea here is that for $1 \leq j \leq n$ and a in K_j, $n(\gamma_j; a) = -1$ while $n(\gamma_0; a) = 1$.

A positive Jordan system Γ satisfying (1.1) with $E = \emptyset$ will be called a *curve generating system* for G. In fact, the curves $\gamma_1, \ldots, \gamma_n$ are a set of generators for the first homology group of G as well as the fundamental group of G.

In the case of a finitely connected region the condition that a harmonic function has a conjugate can be greatly simplified. In fact, the infinite number of conditions in part (c) of Theorem 13.3.4 can be replaced by a finite number of conditions.

1.2 Theorem. *If G is an n-connected region and $\Gamma = \{\gamma_0, \gamma_1, \ldots, \gamma_n\}$ is a curve generating system for G, then a harmonic function $u : G \to \mathbb{R}$ has a harmonic conjugate if and only if $\int_{\gamma_j} (u_x\, dy - u_y\, dx) = 0$ for $1 \le j \le n$.*

Proof. From Theorem 13.3.4 it is easy to see that it suffices to assume that $\int_{\gamma_j} f = 0$ for $1 \le j \le n$, where $f = \partial u$, and prove that f has a primitive. Fix a point a_j in K_j for $1 \le j \le n$. If γ is any closed rectifiable curve in G, put $m_j = n(\gamma; a_j)$. It follows that the system of curves $\{\gamma, -m_1\gamma_1, \ldots, -m_n\gamma_n\}$ is homologous to 0 in G. By Cauchy's Theorem,

$$0 = \int_\gamma f - \sum m_j \int_{\gamma_j} f.$$

But this implies that $\int_\gamma f = 0$ by assumption. Hence f has a primitive. □

What is going on in the preceding theorem is that the first homology group of G is a free abelian group on n generators and the curves $\{\gamma_j\}$ form a system of generators for this group. Morever, if Γ is an element of the first homology group, then Γ corresponds to a system of closed curves in G and the map $\Gamma \to \int_\Gamma f$ is a homomorphism of this group into the additive group $\mathbb{C}$. Thus the condition of the preceding theorem is that this homomorphism vanishes on the generators, and hence vanishes identically.

Let G be an n-connected region and let $\Gamma = \{\gamma_0, \gamma_1, \ldots, \gamma_n\}$ be a curve generating system in G. For a harmonic function $u : G \to \mathbb{R}$ with conjugate differential ${}^*du$, the numbers

$$c_j = \frac{1}{2\pi} \int_{\gamma_j} {}^*du,$$

$1 \le j \le n$, are called the *periods* of u. Note that the periods of u are real numbers since u is real-valued. So a rephrasing of Theorem 1.2 is that a real-valued harmonic function u on G has a harmonic conjugate if and only if all its periods are 0.

Theorem 1.2 can be used to describe exactly how a harmonic function differs from one that has a harmonic conjugate.

1.3 Theorem. *Let G be an n-connected region with $K_1, \ldots, K_n$ the bounded components of its complement; for $1 \le j \le n$, let $a_j \in K_j$. If u is a real-valued harmonic function on G and $c_1, \ldots, c_n$ are its periods, then there is an analytic function h on G such that*

$$u = \operatorname{Re} h + \sum_{j=1}^{n} c_j \log|z - a_j|.$$

Proof. Consider the harmonic function

$$U = u - \sum_j c_j \log|z - a_j|.$$

Now $\ell(z) = \log|z|$ is harmonic on the punctured plane $\mathbb{C}_0 \equiv \mathbb{C} \setminus \{0\}$ and an elementary computation shows that ℓ_x - $i\ell_y = \bar{z}/|z|^2 = z^{-1}$. Thus for any closed rectifiable curve γ not passing through 0,

$$\int_\gamma (\ell_x \, dy - \ell_y \, dx) = -i \int_\gamma z^{-1} dz = 2\pi \; n(\gamma; 0).$$

So if $L_k(z) = \log|z - a_k|$, $\int_{\gamma_j} (L_{k\,x} dy - L_{k\,y} dx) = 2\pi$ if $k = j$ and 0 otherwise. It is straightforward to see that the choice of c_j gives that $\int_{\gamma_j} (U_x \, dy - U_y \, dx) = 0$ for $1 \leq j \leq n$. By Theorem 1.2, there is an analytic function h on G such that $U = \operatorname{Re} h$. $\square$

This theorem has several interesting consequences. Here are a few.

1.4 Corollary. *If u is a real-valued harmonic function in the punctured disk $B_0(a; R) \equiv B(a; R) \setminus \{a\}$, then there are real constants b and c such that*

$$\int_{-\pi}^{\pi} u\left(a + re^{i\theta}\right) d\theta = b \log r + c.$$

Proof. By Theorem 1.3 there is a real constant b and an analytic function h on $B_0(a; R)$ such that $u = \operatorname{Re} h + b \log|z - a|$. Thus,

$$\begin{aligned} \int_{-\pi}^{\pi} u(a + re^{i\theta}) d\theta &= b \int_{-\pi}^{\pi} \log|re^{i\theta}| \, d\theta + \int_{-\pi}^{\pi} \operatorname{Re} h(a + re^{i\theta}) d\theta \\ &= 2\pi b \log r + \int_{-\pi}^{\pi} \operatorname{Re} h(a + re^{i\theta}) d\theta \end{aligned}$$

But it is easy to see that $\int_{-\pi}^{\pi} \operatorname{Re} h(a + re^{i\theta}) d\theta$ is the imaginary part of the integral

$$\int_\gamma h(z)(z - a)^{-1} dz,$$

where $\gamma(\theta) = a + re^{i\theta}$, and hence is a constant. $\square$

The next result might have been expected.

1.5 Corollary. *If u is a bounded harmonic function in the punctured disk $B_0(a; R)$, then u has a harmonic extension to $B(a; R)$.*

Proof. It suffices to assume that u is real-valued. Suppose $|u| \leq M$ on $B_0(a; R)$. According to Theorem 1.3, there is a real constant b and an analytic function h on $B_0(a;R)$ such that $u = b\log|z-a| + \text{Re}\ h(z)$. Also there is a real constant c such that for $r < R$

$$\int_{-\pi}^{\pi} u\left(a + \mathrm{r}e^{\mathrm{i}\theta}\right) d\theta = 2\pi b\ \log r + c.$$

On the other hand, $|u| \le M$ so $|2\pi\, b\ \log r + c| \le 2\pi M$ for $r < R$. But this is impossible if $b \ne 0$ so it must be that $b = 0$.

Hence $u = \text{Re}\ h$ in $B_0(a;R)$. This implies that h is an analytic function on $B_0(a;R)$ whose real part is bounded. Consider $g = \exp(h)$; g is bounded in $B_0(a;R)$ since $|g| = \exp(\text{Re h})$. Therefore a is a removable singularity for g. This implies that a is a removable singularily for $g' = h\ g$ and thus also for $h = g'/g$. This provides the required harmonic extension of u. □

Most of the preceding material is taken from Axler [1986].

1.6 Definition. If none of the components $K_0, K_1, \ldots, K_n$ of $\mathbb{C}_\infty \setminus G$ is a point, then the *harmonic basis* for G is the collection of continuous functions $\omega_1, \omega_2, \ldots, \omega_n$ on $cl_\infty G$ that are harmonic on G and satisfy $\omega_j|\partial K_j \equiv 1$ and $\omega_j|\partial K_i \equiv 0$ for $i \ne j$.

Note that these functions exist since the hypothesis guarantees that we can solve the Dirichlet problem for G (10.4.17). Also the function ω_0 that is 1 on ∂K_0 and 0 on the boundary of the bounded components is not included here since this function can be obtained from the others by means of the formula $\omega_0 = 1 - \sum_1^n \omega_j$.

In the literature these functions in the harmonic basis are often called the *harmonic measures* for G. This terminology originated before the full blossoming of measure theory. Later a harmonic measure for G will be introduced that is indeed a measure. In order to avoid confusion, the classical terminology has been abandoned.

The next lemma will be used in the proof of some conformal mapping results for finitely connected regions.

1.7 Proposition. *If G is a non-degenerate n-connected region, $\Gamma = \{\gamma_0, \ldots, \gamma_n\}$ is a curve generating system for G, and for $1 \le j,\ k \le n$*

$$c_{jk} = \frac{1}{2\pi}\int_{\gamma_j}{}^*d\omega_k,$$

where $\{\omega_1, \ldots, \omega_n\}$ is the harmonic basis for G, then the $n \times n$ matrix $[c_{jk}]$ is invertible.

Proof. It suffices to show that this matrix is an injective linear transformation of $\mathbb{R}^n$ into $\mathbb{R}^n$. So suppose there are real scalars $\lambda_1, \ldots, \lambda_n$ such that

$$\begin{cases} c_{11}\lambda_1 + c_{12}\lambda_2 + \ldots + c_{1n}\lambda_n = 0 \\ \vdots \\ c_{n1}\lambda_1 + c_{n2}\lambda_2 + \ldots + c_{nn}\lambda_n = 0\,. \end{cases}$$

We want to show that $\lambda_1 = \cdots = \lambda_n = 0$.

Put $u = \sum_1^n \lambda_k \omega_k$. So u is continuous on $cl_\infty G$ and harmonic on G. Also $u \equiv \lambda_k$ on ∂K_k for $1 \le k \le n$ and $u \equiv 0$ on ∂K_0. Moreover for $1 \le j \le n$,

$$\int_{\gamma_j} {}^*du = \sum_{k=1}^n \lambda_k \int_{\gamma_j} {}^*d\omega_k = \sum_{k=1}^n c_{jk}\lambda_k = 0.$$

Therefore Theorem 1.2 implies there is an analytic function f on G such that $u = \text{Re } f$.

Put $\lambda_0 = 0$ and let $\zeta_0 \in \mathbb{C}$ such that Re $\zeta_0 \ne \lambda_0, \lambda_1, \ldots, \lambda_\mathrm{n}$. For $0 \le k \le n$, pick constants $\varepsilon_k > 0$ such that if $\Omega_k = \{\zeta \in \mathbb{C} : |\text{Re}\,\zeta - \lambda_k| < \varepsilon_k\}$, then $\zeta_0 \notin \bigcup_{k=0}^n \text{cl}\,\Omega_k$. Now $u = \text{Re} f$ is continuous on $\text{cl}_\infty\, G$ so $\{z \in \text{cl}_\infty G : |u(z) - \lambda_k| < \varepsilon_k\}$ is a relatively open subset of $\text{cl}_\infty G$ and contains ∂K_k. Therefore if U_k equals the union of K_k and the component of $\{z \in \text{cl}_\infty G : |u(z) - \lambda_k| < \varepsilon_k\}$ that contains ∂K_k, then U_k is open in $\mathbb{C}$ and contains K_k. (The fact that some of the constants λ_k may coincide forces some of this awkward language. If it were the case that $\lambda_k \ne \lambda_j$ for $k \ne j$, then the ε_k could have been chosen so that the sets Ω_k have pairwise disjoint closures and the language would be simpler.) By Proposition 13.1.5, for $1 \le k \le n$ there is a Jordan curve σ_k contained in U_k such that $K_k \subseteq \text{ins}\,\sigma_\mathrm{k}$ and $n(\sigma_k; z) = -1$ for all z in K_k. Similarly, there is a Jordan curve σ_0 in U_0 such that $K_0 \subseteq \text{out}\;\sigma_0$ and $n(\sigma_0; z) = +1$ for all z in $K_1 \cup \cdots \cup K_n \cup \sigma_1 \cup \cdots \cup \sigma_n$. Hence $\Sigma = \{\sigma_0, \sigma_1, \ldots, \sigma_n\}$ is a Jordan system in G; in fact, Σ is a curve generating system for G.

Now

$$\sum_{k=0}^n \frac{1}{2\pi i} \int_{\sigma_k} \frac{f'}{f - \zeta_0} = \sum_{k=0}^n n(f \circ \sigma_k; \zeta_0).$$

But $f \circ \sigma_k$ is a rectifiable closed curve lying inside the vertical strip Ω_k and ζ_0 lies outside this strip. Thus $n(f \circ \sigma_k; \zeta_0) = 0$ for $0 \le k \le n$. By the Argument Priniciple this implies that the equation $f(z) = \zeta_0$ has no solutions in ins Σ. But $f(U_k) \subseteq \Omega_k$ and so $f(z) = \zeta_0$ has no solution in $\cup_{k=0}^n U_k$. Therefore $\zeta_0 \notin f(G)$. Since ζ_0 was an arbitrary point with $\text{Re}\,\zeta_0 \ne \lambda_0, \lambda_1, \ldots, \lambda_n$ we get that $f(G) \subseteq$ the union of the lines $\{\zeta : \text{Re}\,\zeta = \lambda_k\}$ for $k = 0, \ldots, n$. By the Open Mapping Theorem, f, and hence u, must be constant. But $u = 0$ on ∂K_0 so $u \equiv 0$. In particular $\lambda_1 = \cdots = \lambda_n = 0$. $\square$

Exercises

1. If G is an $(n+1)$-connected region in $\mathbb{C}$, $\text{Har}(G)$ is the vector space of harmonic functions on G, and $\text{Har}_c(G)$ is the subspace of $\text{Har}(G)$ consisting of all those harmonic functions with a harmonic conjugate, then the quotient space $\text{Har}(G)/\text{Har}_c(G)$ is an n-dimensional vector space over $\mathbb{R}$. Find a basis for this space.

2. Find the harmonic basis for the annulus.

3. (The $n+1$ Constants Theorem) Let G be an n-Jordan region with boundary curves $\gamma_0, \gamma_1, \ldots, \gamma_n$ and let $\lambda_0, \lambda_1, \ldots, \lambda_n$ be real constants. Show that if u is a subharmonic function on G such that for a in γ_j, $0 \leq j \leq n$,
$$\limsup_{z \to a} u(z) \leq \lambda_j,$$
then
$$u(z) \leq \sum_{j=0}^{n} \lambda_j \omega_j(z),$$
where $\{\omega_1, \ldots, \omega_n\}$ is the harmonic basis for G and $\omega_0 = 1 - (\omega_1 + \cdots + \omega_m)$.

4. Let G be an annulus and $f : G \to \mathbb{C}$ an analytic function. Use the preceeding exercise with $u = |f|$ to deduce Hadamard's Three Circles Theorem (6.3.13).

§2 Conformal Equivalence with an Analytic Jordan Region

Recall the definition of an analytic Jordan region (13.4.11). The main result of this section is the following.

2.1 Theorem. *If G is a non-degenerate n-connected region, then G is conformally equivalent to an analytic n-Jordan region Ω. Moreover, Ω can be chosen so that its outer boundary is $\partial\mathbb{D}$ and $0 \notin \Omega$.*

Proof. The proof consists of an iterative application of the Riemann Mapping Theorem. Let $K_0, K_1, \ldots, K_n$ be the components of $\mathbb{C}_\infty \setminus G$ with K_0 containing ∞. Consider $G_0 = \mathbb{C} \setminus K_0$ and note that G_0 is simply connected since its complement in $\mathbb{C}_\infty$, K_0, is connected. Let $\phi_0 : G_0 \to \mathbb{D}$ be a Riemann map and put $\Omega_0 = \phi_0(G)$.

So Ω_0 is a finitely connected region and $\mathbb{C}_\infty \setminus \Omega_0 = (\mathbb{C}_\infty \setminus \mathbb{D}) \cup \phi_0(K_1) \cup \cdots \cup \phi_0(K_n) = K_{00} \cup K_{01} \cup \cdots \cup K_{0n}$. Now let $G_1 = \mathbb{C}_\infty \setminus K_{01}$; again G_1 is a simply connected region in $\mathbb{C}_\infty$ containing ∞ and the region Ω_0. Let $\phi_1 : G_1 \to \mathbb{D}$ be the Riemann map with $\phi_1(\infty) = 0$ and put $\Omega_1 = \phi_1(\Omega_0)$. Again Ω_1

is an n-connected region. The components of its complement in $\mathbb{C}_\infty$ are $K_{11} \equiv \mathbb{C}_\infty \setminus \mathbb{D}$, $K_{10} \equiv \phi_1(K_{00})$, $K_{12} \equiv \phi_1(K_{02}), \ldots,$ $K_{1n} \equiv \phi_1(K_{0n})$. Note that the components of the boundary of Ω_1 are $\partial\,\mathbb{D}$, $\phi_1(\partial\,\mathbb{D})$, $\phi_1(\partial\,\mathbb{D}) = \phi_1(\partial K_{00})$, $\partial K_{12}, \ldots, \partial K_{1n}$.

So at this stage we have that G is conformally equivalent to a region Ω_1 contained in $\mathbb{D}$, two of whose boundary components are analytic Jordan curves, one of them being the unit circle. Continue the process to get that G is conformal equivalent to an n-Jordan region Ω_{n-1} contained in $\mathbb{D}$ and having $\partial\,\mathbb{D}$ as its outer boundary. Now pick a in $\mathbb{D} \setminus \mathrm{cl}\ \Omega_{n-1}$, let ϕ_a be the Schwarz map $(z-a)(1-\bar{a}z)^{-1}$, and put $\Omega = \phi_a(\Omega_{n-1})$. □

The preceding proposition has value for problems that involve properties that are invariant under conformal equivalence, but its value diminishes when this is not the case.

The next result is the analogue of the classical Green Identities using the conjugate differential and the ∂-derivative. For the remainder of the section it is assumed that G is an analytic Jordan region with oriented boundary $\Gamma = \{\gamma_0, \ldots, \gamma_n\}$, with γ_0 as the outer boundary.

2.2 Proposition. *Let G be an analytic Jordan region with oriented boundary Γ.* (a) *If u and v are functions that are C^1 on $\mathrm{cl}\, G$, then*

$$\begin{aligned}\int_\Gamma (v\,du - u\,dv) &= 4i \int\!\!\int_G (\bar{\partial} v\,\partial u - \bar{\partial} u\,\partial v) \\ &= 2\int\!\!\int_G (u_y\,v_x - u_x v_y).\end{aligned}$$

(b) *If u and v are functions that are C^1 on $\mathrm{cl}\, G$ and C^2 on G, then*

$$\int_\Gamma (v\,{}^*du - u\,{}^*dv) = 4\int\!\!\int_G (v\partial\bar{\partial}u - u\partial\bar{\partial}v).$$

Proof. (a) Using the definition of du and dv, applying Green's Theorem, and then simplifying we get

$$\begin{aligned}\int_\Gamma (v\,du - u\,dv) &= \int_\Gamma [(v\,u_x - u\,v_x)dx + (v\,u_y - u\,v_y)dy] \\ &= 2\int\!\!\int_G (u_y v_x - u_x v_y).\end{aligned}$$

Again using the definitions of the expressions and simplifying, we also have

$$\int\!\!\int_G (\bar{\partial} v\ \partial u - \bar{\partial} u\ \partial v) = -\frac{i}{2}\int\!\!\int_G (u_y\ v_x - u_x v_y),$$

proving part (a).

(b) Using Exercise 13.3.2 we have

$$\int_\Gamma v\,{}^*du = -2i\int_\Gamma v\,\partial u + i\int_\Gamma v\,du,$$

$$\int_\Gamma u\,{}^*dv = -2i\int_\Gamma u\,\partial v + i\int_\Gamma u\,dv.$$

Performing the required algebra yields

$$\int_\Gamma (v\,{}^*du - u\,{}^*dv) = -2i\int_\Gamma (v\,\partial u - u\,\partial v) + i\int_\Gamma (v\,du - u\,dv).$$

Now apply Green's Theorem to the first integral and part (a) to the second in order to get

$$\begin{aligned}\int_\Gamma (v\,{}^*du - u\,{}^*dv) &= -2i\left[2i\int\!\!\int_G [\bar\partial(v\,\partial u - u\,\partial v)]\right] \\ &\quad + i(4i)\int\!\!\int_G (\bar\partial v\,\partial u - \bar\partial u\,\partial v) \\ &= 4\int\!\!\int_G [\bar\partial v\,\partial u + v\,\bar\partial\partial u - \bar\partial u\,\partial v - u\,\bar\partial\partial v] \\ &\quad -4\int\!\!\int_G [\bar\partial(v\,\partial u - u\,\partial v)] \\ &= 4\int\!\!\int_G (v\,\partial\bar\partial u - u\,\partial\bar\partial v).\end{aligned}$$

□

Recall (13.2.3) that the Laplacian of a function u is $4\partial\bar\partial u$.

2.3 Corollary. *If u and v are functions that are C^1 on $\text{cl}\,G$ and harmonic on G, then $\int_\Gamma v\,{}^*du = \int_\Gamma u\,{}^*dv$.*

Recall (§10.5) that there is a Green function $g(z,a)$ for G. If a is fixed, $g_a(z) = -\log|z-a| + R_a(z)$, where R_a is harmonic on G. Now g_a is harmonic on $G\setminus\{a\}$ and identically 0 on Γ. By Corollary 13.4.12, g_a can be extended to a harmonic function defined in a neighborhood of Γ. We will always assume that g_a has been so extended. Since $\log|z-a|$ is also harmonic in a neighborhood of Γ, it follows that the same holds for R_a. In particular it is legitimate to discuss the integrals of the functions g_a and R_a as well as their derivatives and conjugate differentials on Γ.

The first application of these notions is a formula for the solution of the Dirichlet Problem, but first a lemma.

2.4 Lemma. *Let $r > 0$ such that $\overline{B}(a;r) \subseteq G$ and put $B_r = B(a;r)$. If u is a C^1 function on cl G that is harmonic inside G, then*

$$\int_{\partial B_r} (g_a\,{}^*du - u\,{}^*dg_a) = 2\pi\, u(a) + \eta(r),$$

where $\eta(r) \to 0$ as $r \to 0$.

Proof. Put $\ell(\zeta) = \log|\zeta - a|$. Using the definition of ${}^*d\ell$ and parametrizing the circle ∂B_r by $\gamma(\theta) = a + re^{i\theta}$, we get that ${}^*d\ell = d\theta$ on ∂B_r. Thus

$$\int_{\partial B_r} u\,{}^*d\ell = \int_0^{2\pi} u(a + re^{i\theta})d\theta = 2\pi\, u(a)$$

by the Mean Value Property for harmonic functions. Also there is a constant M such that $\left|\int_{\partial B_r} u\,{}^*dR_a\right| \le Mr$. Hence $\int_{\partial B_r} u\,{}^*dg_a \to -2\pi\, u(a)$. On the other hand, there are constants C_1 and C_2 such that for all r,

$$\begin{aligned}\left|\int_{\partial B_r} g_a\,{}^*du\right| &= \left|-\int_{\partial B_r} \ell\,{}^*du + \int R_a\,{}^*du\right| \\ &\le C_1 r\,\log r + C_2\, r.\end{aligned}$$

This completes the proof of the lemma. □

2.5 Theorem. *If G is an analytic Jordan region, Γ is the positively oriented boundary of G, h is a continuous function on Γ, and $\hat{h}$ is the solution of the Dirichlet Problem with boundary values h, then*

$$\hat{h}(z) = -\frac{1}{2\pi}\int_\Gamma h\,{}^*dg_z,$$

where g_z is Green's function for G with singularity at z.

Proof. Both sides of the above equation behave properly if a sequence of functions $\{h_n\}$ converges uniformly to a function h on Γ. Thus, it suffices to prove this under the extra assumption that h is a smooth function on Γ. Let $r > 0$ such that $\overline{B}(a;\, r) \subseteq G$. Put $G_r = G \setminus \overline{B}(a;\, r)$ and $\Gamma_r = \partial G_r$ with positive orientation. Let $B_r = B(a;\, r)$ and always consider ∂B_r as having positive orientation. Put $u = \hat{h}$. According to Corollary 2.3

$$\int_{\Gamma_r} g_a\,{}^*du = \int_{\Gamma_r} u\,{}^*dg_a.$$

Now g_a is identically 0 on Γ so that

$$\begin{aligned}\int_\Gamma u\,{}^*dg_a &= \int_{\partial B_r} (u\,{}^*dg_a - g_a\,{}^*du) \\ &= -2\pi\, u(a) + \eta(r)\end{aligned}$$

by the preceding lemma. Therefore taking the limit as $r \to 0$ we get that

$$\int_\Gamma h\, {}^*dg_a = -2\pi \hat{h}(a).$$

This proves the theorem. □

The reader might note that the converse of Theorem 2.5 is valid. If h is a continuous function on Γ, then the formula in (2.5) gives the solution of the Dirichlet problem. Thus (2.5) is a generalization of the Poisson Formula for $\mathbb{D}$. See Exercise 3. We will see more of these matters in Chapters 19 and 21.

Using the definition of normal derivative, the formula for the solution of the Dirichlet Problem in the preceding theorem can be rewritten as

$$\hat{h}(z) = \int_\Gamma h(w)\, \frac{\partial g_z}{\partial n}(w)\, |dw|.$$

2.6 Corollary. *If $\{\omega_1, \ldots, \omega_n\}$ is the harmonic basis for G, then for $0 \leq j \leq n$*

$$\frac{1}{2\pi}\int_{\gamma_j} {}^*dg_a = -\omega_j(a).$$

Proof. Take h in the preceding theorem to be the characteristic function of γ_j. □

Exercises

1. If G is an m-connected region and n of the components of its complement in $\mathbb{C}_\infty$ are not trivial, then G is conformally equivalent to an analytic n-Jordan region with $m - n$ points removed.

2. Show that the matrix $[c_{jk}]$ from Proposition 1.7 is positive definite.

3. Prove that with the hypothesis of Theorem 2.5, if $h : \Gamma \to \mathbb{R}$ is a continuous function, then the solution of the Dirichlet Problem is given by the formula there. If $G = \mathbb{D}$, show that the formula in (2.5) is precisely the Poisson Integral Formula.

4. Show that the matrix $[c_{jk}]$ from Proposition 1.7 is a conformal invariant for analytic Jordan regions.

§3 Boundary Values for a Conformal Equivalence Between Finitely Connected Jordan Regions

The results from §14.5 on boundary values for a conformal equivalence between two simply connected Jordan regions can be extended to a conformal equivalence between two finitely connected Jordan regions. To do this it is not necessary to start from the beginning, but rather the results from the simply connected case can be used to carry out the extension.

We begin by showing that if G and Ω are finitely connected regions and $\phi : G \to \Omega$ is a conformal equivalence, then ϕ defines a correspondence between the components of the boundaries of the two sets. Recall (13.1.2) that the map $F \to F \cap \partial_\infty G$ defines a bijection between the components of $\mathbb{C}_\infty \setminus G$ and the components of $\partial_\infty G$. For our discussion fix an n-connected region G and let $K_0, \ldots, K_n$ be the components of $\mathbb{C}_\infty \setminus G$ with ∞ in K_0 and let $C_0, \ldots, C_n$ be the corresponding boundary components. Let Ω be another n-connected region and assume there is a conformal equivalence $\phi : G \to \Omega$. For $0 \le j \le n$ and $\varepsilon > 0$, let $\Phi_j(\varepsilon) = \text{cl}[\phi(\{z \in G : \text{dist}(z, K_j) < \varepsilon\})] = \text{cl}[\phi(\{z \in G : \text{dist}(z, C_j) < \varepsilon\})]$ and put $\Phi_j = \bigcap_\varepsilon \Phi_j(\varepsilon)$. Here the distance involved is the metric of the extended plane, though the usual metric for the plane can be employed if Ω is bounded. It will be shown that the sets $\Phi_0, \ldots, \Phi_n$ are precisely the components of $\partial_\infty \Omega, \gamma_0, \ldots, \gamma_n$.

3.1 Lemma.

(a) $\Phi_j(\varepsilon_1) \subseteq \Phi_j(\varepsilon_2)$ *if* $\varepsilon_1 < \varepsilon_2$.

(b) *If* $\{\varepsilon_k\}$ *monotonically decreases to* 0, *then* $\Phi_j = \cap_k \Phi_j(\varepsilon_k)$.

(c) *If* U *is an open set containing* Φ_j, *then there is an* $\varepsilon > 0$ *such that* $\Phi_j(\varepsilon) \subseteq U$.

(d) Φ_j *is a connected subset of* $\partial_\infty \Omega$.

Proof. The proofs of parts (a),(b), and (c) are left to the reader. It is also left to the reader to show that $\Phi_j \subseteq \partial_\infty \Omega$. Suppose that U and V are disjoint open sets and $\Phi_j \subseteq U \cup V$. By part (c) there is an $\varepsilon > 0$ such that $\Phi_j(\varepsilon) \subseteq U \cup V$. But $\Phi_j(\varepsilon)$ is clearly a connected set as it is the closure of a connected set. This contradiction establishes part (d). □

3.2 Proposition. *The sets* $\{\Phi_0, \Phi_1, \ldots, \Phi_n\}$ *are precisely the components of* $\partial_\infty \Omega$.

Proof. By part (d) of the preceding lemma, each set Φ_j is contained in one of the components γ_i of $\partial_\infty \Omega$. On the other hand, similar arguments involving ϕ^{-1} show that each point of $\partial_\infty \Omega$ must belong to one of the sets Φ_j. Thus a simple counting argument shows that $\{\Phi_0, \Phi_1, \ldots, \Phi_n\}$ are the components $\{\gamma_0, \gamma_1, \ldots, \gamma_n\}$. □

If $\Phi_j = \gamma_i$, then we will say that ϕ *associates* C_j *with* γ_i. To maintain some flexibility we might also say in this situation that ϕ associates K_j with γ_i. The idea here is that even though ϕ may not extend as a function to the boundary of G, it is possible to think of ϕ as mapping the components of $\partial_\infty G$ onto the components of $\partial_\infty \Omega$.

The next result is just a restatement of Theorem 2.1 with the added piece of information that any component of the complement of G can be made to be associated with $\partial\, \mathbb{D}$.

3.3 Theorem. *If G is a non-degenerate n-connected region and C is any component of its extended boundary, then there is a conformal equivalence ϕ of G onto an analytic n-Jordan region Ω such that the outer boundary of Ω is $\partial\, \mathbb{D}, 0 \notin \Omega$, and ϕ associates C with $\partial\, \mathbb{D}$.*

Proof. Refer to the proof of Theorem 2.1. Using the notation there it follows that the map ϕ constructed there associates ∂K_n with $\partial\, \mathbb{D}$. If $C = \partial K_j$ for $1 \leq j \leq n$, then a simple relabelling proves the present theorem. If $C = \partial K_0$, then look at the image G_1 of G under the Möbius transformation $(3-a)^{-1}$ for an appropriate choice of a. Here C corresponds to the boundary of a component of the complement of G_1 that does not contain ∞ and the previous argument applies. $\square$

3.4 Theorem. *If G and Ω are two finitely connected Jordan regions and $\phi : G \to \Omega$ is a conformal equivalence, then ϕ extends to a homeomorphism of* $\operatorname{cl} G$ *onto* $\operatorname{cl} \Omega$.

Proof. Using the above notation and letting $C_j = \partial K_j$, we can assume that ϕ associates the boundary curve C_j of G with the boundary curve γ_j of Ω, $0 \leq j \leq n$. It suffices to assume that G is an analytic Jordan region. In fact if this case is done, then for the general case apply Theorem 2.1 to find two analytic Jordan regions G_1 and Ω_1 and conformal equivalences $f : G \to G_1$ and $\tau : \Omega \to \Omega_1$. Let $\phi_1 : G_1 \to \Omega_1$ be defined by $\phi_1 = \tau \circ \phi \circ f^{-1}$. Now observe that if the theorem is established for the special case where the domain is an analytic Jordan region, then by taking inverses the theorem also holds when the range is an analytic Jordan region. Thus each of the maps f, τ, and ϕ_1 as well as their inverses extends to a homeomorphism. Hence ϕ extends to a homeomorphism.

So we assume that each of the curves $C_0, \ldots, C_n$ is an analytic Jordan curve. Let $0 < r < \operatorname{dist}(\gamma_j, \gamma_k)$ for $j \neq k$. Let $0 < \varepsilon < \operatorname{dist}(C_j, C_k)$ for $j \neq k$; by Proposition 3.2 and Lemma 3.1, ε can be chosen such that $\phi(\{z \in G : \operatorname{dist}(z, C_j) < \varepsilon\}) \subseteq \{\zeta \in \Omega : \operatorname{dist}(\zeta, \gamma_j) < r\}$. By Corollary 13.5.7, ϕ has non-tangential limits a.e. on each C_j.

Fix j, $0 \leq j \leq n$, and let a_1 and a_2 be any two points on C_j at which ϕ has a non-tangential limit. Thus there are Jordan arcs $\eta_i : [0, 1) \to G$, $i = 1, 2$, such that $\eta_i(t) \to a_i$ as $t \to 1$, $\eta_1(0) \neq \eta_2(0)$, and $\lim_{t \to 1} \phi(\eta_i(t)) = \alpha_i$, a point in γ_i. Note that $\alpha_1 \neq \alpha_2$ because if they were equal, then the non-tangential limit of ϕ would be this common vlaue at a.e. point on one of the two subarcs of C_j that joins a_1 and a_2. (See the proof of Theorem 14.5.6.) Let η_0 be one of the two arcs in C_j that joins a_1 and a_2 and let η_3 be a Jordan arc in G joining $\eta_2(0)$ to $\eta_1(0)$. Thus $C = \eta_0\eta_1\eta_3\eta_2^{-1}$ is a Jordan curve and ins $C \subseteq G$. Let $\lambda_i = \phi \circ \eta_i$ for $i = 1, 2, 3$; so each λ_i is a Jordan arc in Ω and $\lambda_1\lambda_3\lambda_2^{-1}$ is a Jordan arc joining α_2 to α_1. If λ_0 is either of the two subarcs of γ_j joining α_1 to α_2, then $\gamma = \lambda_1\lambda_3\lambda_2^{-1}\lambda_0$ is a Jordan curve that is disjoint from $\phi(\text{ins } C)$; choose the arc λ_0 such that $\phi(\text{ins } C) = \text{ins } \gamma$. Since ins $\gamma \subseteq \Omega$ and ϕ is surjective, it follows that $\phi(\text{ins } C) = \text{ins } \gamma$. By Corollary 14.5.7, ϕ extends to a homeomorphism of $\text{cl}(\text{ins}\, C)$ onto $\text{cl}(\text{ins}\, \gamma)$. Thus ϕ maps η_0 homeomorphically onto λ_0.

By examining the other subarc of C_j that joins a_1 to a_2, we see that ϕ extends to a homeomorphism of $G \cup C_j$ onto $\Omega \cup \gamma_j$. The details of this argument as well as the remainder of the proof are left to the reader. □

The proofs of the next results are similar to the preceding proof and will not be given. The following extends Theorem 14.5.5.

3.5 Theorem. *If G is a finitely connected Jordan region and $\phi : G \to \Omega$ is a conformal equivalence, then the following are equivalent.*

(a) *ϕ has a continuous extension to the closure of G.*

(b) *Each component of $\partial\Omega$ is a continuous path.*

(c) *$\partial\Omega$ is locally connected.*

(d) *$\mathbb{C}_\infty \setminus \Omega$ is locally connected.*

Recall the definition of a simple boundary point (14.5.9). The next theorem extends Theorem 14.5.12.

3.6 Theorem. *Let Ω be a bounded finitely connected region, G a finitely connected Jordan region, and let $g : \Omega \to G$ be a conformal equivalence.*

(a) *If ω is a simple boundary point of Ω, then g has a continuous extension to $\Omega \cup \{\omega\}$.*

(b) *If R is the collection of simple boundary points of Ω, then g has a continuous one-to-one extension to $\Omega \cup R$.*

These results on conformal equivalences between finitely connected regions can be used to extend some of the results of §14.5 to unbounded simply connected regions. Rather than listing all the possibilities, we examine a couple that will be of use later.

3.7 Proposition. *Let $\gamma : [0, \infty) \to \mathbb{C}$ be a Jordan arc such that $\gamma(t) \to \infty$ as $t \to \infty$ and let $\Omega = \mathbb{C} \setminus \gamma$. If $\tau : \mathbb{D} \to \Omega$ is the Riemann map with*

$\tau(0) = \alpha$ and $\tau'(0) > 0$, then τ extends to a continuous map of $\operatorname{cl}\mathbb{D}$ *onto* $\mathbb{C}_\infty$.

Proof. Consider the Möbius transformation $T(z) = (z-a)^{-1}$ and put $\sigma = T \circ \tau$. So σ is a conformal equivalence of $\mathbb{D}$ onto $\Lambda \equiv$ the complement in the extended plane of the arc $\lambda = T \circ \gamma$ and $\sigma(0) = \infty$. Thus the cluster values of σ at points of $\partial\mathbb{D}$ all lie on the arc λ. Put $A = \{z : 1/2 < |z| < 1\}$; thus $\sigma(A)$ is the region bounded by the arc λ and the Jordan curve $\sigma(\{|z| = 1/2\})$. According to Theorem 3.5, σ has a continuous extension to $\operatorname{cl} A$. From here it easily follows that $\tau = (1 - \alpha\sigma)/\sigma$ has a continuous extension to $\operatorname{cl}\mathbb{D}$. □

Such regions as Ω in the preceding proposition are called *slit domains* and will play an important role in Chapter 17 below. Another fact about mappings between a slit domain and the disk that will be used later is the following.

3.8 Proposition. *If Ω is a slit domain as in the preceding proposition and $g : \Omega \to \mathbb{D}$ is a conformal equivalence, then g can be continuously extended to $\Omega \cup \{\gamma(0)\}$.*

Proof. Of course $\omega_0 = \gamma(0)$ is a simple boundary point of Ω, but Ω is not bounded so that Theorem 3.6 is not immediately available. However if $\omega = g^{-1}(0)$, $T(\zeta) = (\zeta - \omega)^{-1}$, and $\Lambda = T(\Omega) \subseteq \mathbb{C}_\infty$, then $h = g \circ T^{-1}$ is a conformal equivalence of Λ onto $\mathbb{D}$ with $h(\infty) = 0$. If $A = \{z : 1/2 < |z| < 1\}$ and $\Lambda_1 = h^{-1}(A)$, then Λ_1 is a bounded region and $T(\omega_0)$ is a simple boundary point. It is left to the reader to apply Theorem 3.6 to Λ_1 and h and then unravel the regions and maps to conclude the proof. □

Exercises

1. Give the details in the proof of Theorem 3.5.

2. In Proposition 3.7 show that the extension of τ to $\operatorname{cl}\mathbb{D}$ has the property that there are unique points a and b on $\partial\mathbb{D}$ that map to $\gamma(0)$ and ∞ and that every other point of γ has exactly two points in its preimage.

3. A continuous map $f : G \to \Omega$ is *proper* if for every compact set K contained in Ω, $f^{-1}(K)$ is compact in G. (a) Show that a continuous function f is proper if and only if for each a in ∂G, $\operatorname{Clu}(f, a) \subseteq \partial\Omega$. (b) If $f : G \to \Omega$ is analytic, show that f is proper if and only if there is an integer n such that for each ζ in Ω, the equation $f(z) = \zeta$ has exactly n solutions, counting multiplicites.

4. If in Exercise 3, $G = \Omega = \mathbb{D}$ and $f : \mathbb{D} \to \mathbb{D}$ is a proper analytic map, show that f is a finite Blaschke product. (See Exercise 7.5.4 for the definition.)

§4 Convergence of Univalent Functions

It is desirable to extend the concept of convergence of analytic functions as discussed in Chapter 7. (In this section the regions will be assumed to be arbitrary; it is not assumed that they are finitely connected.) To begin, assume that for every positive integer n there is a region G_n that contains the origin and an analytic function $f_n : G_n \to C$. How can we give meaning to the statement that $\{f_n\}$ converges to a function $f : G \to \mathbb{C}$? Without some restriction on the behavior of the regions G_n there is no hope of a meaningful concept.

4.1 Definition. If $\{G_n\}$ is a sequence of regions each of which contains 0, define the *kernel* of $\{G_n\}$ (with respect to 0) to be the component of the set

$$\{z : \text{there is an } \ r > 0 \text{ such that } \ \overline{B}(z;r) \subseteq G_n$$

$$\text{for all but a finite number of integers } \ n\}$$

that contains 0, provided that this set is non-empty. If the above set is empty, then $\{G_n\}$ does not have a kernel. When $\{G_n\}$ has a kernel it is denoted by $\ker\{G_n\}$. Say that $\{G_n\}$ converges to G if G is the kernel of every subsequence of $\{G_n\}$; this is denoted by $G_n \to G$.

Note that if $\{G_n\}$ is an increasing sequence of regions and G is their union, then $G_n \to G$. Also if $\{G_n\}$ is defined by letting $G_n = \mathbb{D}$ when n is even and $G_n =$ the unit square with vertices $\pm 1 \pm i$ when n is odd, then $\mathbb{D} = \ker\{G_n\}$ but $\{G_n\}$ does not converge to $\mathbb{D}$. Also notice that there is nothing special about 0. If there is any point common to all the regions G_n, it is possible to define the kernel of $\{G_n\}$ with respect to this point. This quasigenerality will not be pursued here.

We also note that the kernel of $\{G_n\}$ is the largest region G containing 0 such that if K is a compact subset of G, then there is an n_0 such that $K \subseteq G_n$ for $n \geq n_0$.

With the notion of a kernel, the extended idea of convergence of functions can be defined.

4.2 Definition. Suppose that $\{G_n\}$ is a sequence of regions each of which contains 0 and such that $G = \ker\{G_n\}$ exists. If $f_n : G_n \to \mathbb{C}$ is a function for all $n \geq 1$, say that $\{f_n\}$ *converges uniformly on compacta* to $f : G \to \mathbb{C}$ if for every compact subset K of G and for every $\varepsilon > 0$ there is an n_0

such that $|f_n(z) - f(z)| < \varepsilon$ for all z in K and all $n \geq n_0$. This will be abbreviated to $f_n \to f\ (uc)$.

When the notation "$f_n \to f\ (uc)$" is used, it will be assumed that all the notation preceding the definition is in force. The following notation will be used often in this section.

4.3 For each $n \geq 1$, G_n is a region containing 0 and $f_n : G_n \to \mathbb{C}$ is an analytic function with $f_n(0) = 0$, $f_n'(0) > 0$, and

$$\Omega_n = f_n(G_n).$$

The reader is invited to revisit Chapter 7 and verify that most of the results about convergence of analytic functions there carry over to the present setting. In particular, if each f_n is analytic and $f_n \to f\ (uc)$, then f is analytic and $f_n' \to f'(uc)$. We also have the following.

4.4 Proposition. *Assume (4.3). If each $f_n : G_n \to \mathbb{C}$ is a univalent function and $f_n \to f\ (uc)$, then either f is univalent or f is identically 0.*

4.5 Lemma. *Assume (4.3). If each f_n is a univalent function, $f_n \to f\ (uc)$, and f is not constant, then the sequence of regions $\{\Omega_n\}$ has a kernel that contains $f(G)$.*

Proof. Let $\Omega_n = f_n(G_n)$ and $\Omega = f(G)$; so Ω is a region containing 0. Let K be a compact subset of Ω. Now G can be written as the union of the open sets $\{H_k\}$, where $\operatorname{cl} H_k$ is compact and contained in H_{k+1}. Thus $\Omega = \cup_k f(H_k)$ and so there is a $k \geq 1$ such that $K \subseteq f(H_k)$. By the definition of a kernel, there is an n_0 such that $\operatorname{cl} H_k \subseteq G_n$ for $n \geq n_0$. Thus $K \subseteq \Omega_n$ for all $n \geq n_0$. $\square$

The reader might want to compare the next result with Proposition 14.7.15, whose proof is similar.

4.6 Lemma. *Assume (4.3). If $G = \ker\{G_n\}$ exists, each f_n is univalent, and $f_n'(0) = 1$ for all n, then there is a subsequence $\{f_{n_k}\}$ such that $G = \ker\{G_{n_k}\}$ exists and $\{f_{n_k}\}$ converges (uc) to a univalent function $f : G \to \mathbb{C}$.*

Proof. Let $R > 0$ such that $\overline{B}(0; R) \subseteq G$; let N_0 be an integer such that $\overline{B}(0; R) \subseteq G_n$ for all $n \geq N_0$. Put $g_n(z) = z/f_n(z)$. As in the proof of Proposition 14.7.15, $|g_n(z)| \leq 4$ for $|z| \leq R$ and $|g_n(z)| \leq 4|z|R^{-1}$ for z in $G_n \setminus \overline{B}(0; R)$.

Write $G = \bigcup_{j=1}^{\infty} D_j$, where each D_j is a region containing $\overline{B}(0; R)$ and $\operatorname{cl} D_j$ is a compact set that is included in D_{j+1}. Let $N_0 < N_1 < \cdots$ such that $\operatorname{cl} D_j \subseteq G_n$ for $n \geq N_j$. From the preceding observations, $\{g_n : n \geq N_j\}$ is uniformly bounded on D_j. Let $\{g_n : n \in A_1\}$ be a subsequence that converges uniformly on compact subsets of D_1 to an analytic function

$h_1 : D_1 \to \mathbb{C}$. There is a subsequence $\{g_n : n \in A_2\}$ of $\{g_n : n \in A_1\}$ that converges in $H(D_2)$ to a function h_2. Continue to obtain infinite sets of integers $\{A_j\}$ with $A_{j+1} \subseteq A_j$, $n \geq N_j$ for all n in A_j, and such that $\{g_n : n \in A_j\}$ converges in $H(D_j)$ to an analytic function h_j defined on D_j. From the nature of subsequences it is clear that $h_j = h_{j+1}$ for all j. Hence there is an analytic function $g : G \to \mathbb{C}$ such that $g|D_j = h_j$ for all j.

Let n_j be the j-th integer in A_j. Since $D_k \subseteq G_{n_j}$ for $j \geq k, G = \ker\{G_{n_j}\}$. Also $\{g_{n_j}\}$ is a subsequence of each $\{g_n : n \in A_k\}$ and so $g_{n_j} \to g$ (uc). Since g_n never vanishes on G_n and $g_n(0) = 1$ for all n, $g(0) = 1$ and hence g does not vanish on G (why?). It is easy to check (as in the proof of Proposition 14.7.15) that $f_{n_j} \to f = z/g$ (uc). □

The next result can be considered the principal result of this section.

4.7 Theorem. *Assume (4.3). If $G_n \to G$, each f_n is univalent, and $f_n'(0) = 1$ for all n, then there is a univalent function f on G such that $f_n \to f$ (uc) if and only if $\Omega_n \to \Omega$ for some region Ω. When this happens, $\Omega = f(G)$ and $\phi_n = f_n^{-1} \to f^{-1}(uc)$.*

Proof. Let us first assume that $f_n \to f$ (uc) for a univalent function f defined on G; put $\Omega = f(G)$. According to Lemma 4.5, $\Lambda \equiv \ker\{\Omega_n\}$ exists and $\Omega \subseteq \Lambda$. Let $\phi_n = f_n^{-1} : \Omega_n \to G_n$. According to Lemma 4.6 there is a subsequence $\{\phi_{n_k}\}$ such that $\Lambda = \ker\{\Omega_{n_k}\}$ and a univalent function $\phi : \Lambda \to \mathbb{C}$ such that $\phi_{n_k} \to \phi$ (uc).

Fix $\varepsilon > 0$. Since $\phi(0) = 0$, there is a $\rho > 0$ such that $\overline{B}(0;\rho) \subseteq \Omega$ and $|\phi(\zeta) - \phi(\zeta')| < \varepsilon/2$ whenever ζ and ζ' are in $B(0;\rho)$. Let k_1 be chosen so that $\overline{B}(0;\rho) \subseteq \Omega_{n_k}$ when $k \geq k_1$. Pick $r > 0$ such that $f(\overline{B}(0;r)) \subseteq B(0;\rho)$ and choose $k_2 > k_1$ such that $f_{n_k}(B(0;r)) \subseteq B(0;\rho)$ for $k \geq k_2$. Finally pick $k_3 > k_2$ such that $|\phi_{n_k}(\zeta) - \phi(\zeta)| < \varepsilon/2$ for $k \geq k_3$ and $|\zeta| < \rho$. Thus for $k \geq k_3$ and $|z| < r$ we have

$$\begin{aligned} |\phi_{n_k}(f_{n_k}(z)) - \phi(f(z))| &\leq |\phi_{n_k}(f_{n_k}(z)) - \phi(f_{n_k}(z))| \\ &+ |\phi(f_{n_k}(z)) - \phi(f(z))| \\ &< \varepsilon. \end{aligned}$$

But for each k, $\phi_{n_k}(f_{n_k}(z)) = z$ and so $\phi(f(z)) = z$ for all z in $B(0;r)$. But $f(G) = \Omega \subseteq \Lambda$ so we get that $\phi(f(z)) = z$ for all z in G.

Now Lemma 4.5 applied to the sequence $\{\phi_{n_k}\}$ implies that $G \subseteq \phi(\Lambda) \subseteq \ker\{\phi_{n_k}(\Omega_{n_k})\} = \ker\{G_{n_k}\}$, which equals G since $G_n \to G$. From here it follows that $\Lambda = \Omega$ and $\phi = f^{-1}$.

Note that the preceding argument can be applied to any subsequence of $\{f_n\}$. That is, for any subsequence $\{f_{m_k}\}$ there is a further subsequence $\{f_{m_{k_j}}\}$ such that $\Omega = \ker\{\Omega_{m_{k_j}}\}$ and $\phi_{m_{k_j}} \to \phi = f^{-1}$.

Now to prove that $\Omega_n \to \Omega$. If not, then there is a subsequence $\{\Omega_{m_k}\}$ that either has no kernel or does not have Ω as its kernel. In either case there is a compact subset K of Ω such that $K \setminus \Omega_n \neq \emptyset$ for infinitely many n. Thus there is a subsequence $\{\Omega_{m_k}\}$ such that $K \setminus \Omega_{m_k} \neq \emptyset$ for all k. But then the reasoning of the preceding paragraph is applied, and we get a subsequence of $\{\Omega_{m_k}\}$ that has Ω as its kernel, giving a contradiction.

Now for the converse. Assume that $\Omega_n \to \Omega$ and put $\phi_n = f_n^{-1}$. So $\phi_n(0) = 0$ and $\phi_n'(0) = 1$. If $\{f_n\}$ is not convergent, then there is an $\varepsilon > 0$, a compact subset K of G, and a subsequence $\{f_{n_k}\}$ such that $\sup\{|f_{n_k}(z) - f_{n_j}(z)| \geq \varepsilon : z \in K\}$ for all $n_k \neq n_j$. Once again Lemma 4.6 implies there is a subsequence $\{\phi_{n_{k_j}}\}$ of $\{\phi_{n_k}\}$ such that $\Omega = \ker\{\Omega_{n_{k_j}}\}$ and $\phi_{n_{k_j}} \to \phi$ (uc), a univalent function on Ω. But we already know that $\Omega_{n_{k_j}} \to \Omega$. Now we can apply the first half of the proof to this sequence to obtain the fact that $G_{n_{k_j}} \to G = \phi(\Omega)$ and $f_{n_{k_j}} \to f = \phi^{-1}(uc)$. This contradicts the fact that $\sup\{|f_{n_k}(z) - f_{n_j}(z)| : z \in K\} \geq \varepsilon$ for all $n_k \neq n_j$. Thus there is a function f on G such that $f_n \to f$ (uc). Since $f'(0) = 1$, it must be that f is univalent. $\square$

Before proving an extension of the preceding theorem, here is a result that will be useful in this proof and holds interest in itself. This proposition is true for all regions (as it is stated), but the proof given here will only be valid for a smaller class of regions. The complete proof will have to await the proof of the Uniformization Theorem; see Corollary 16.5.6 below.

4.8 Proposition. *Let G be a region in $\mathbb{C}$ that contains zero and is not equal to $\mathbb{C}$. If f is a conformal equivalence of G onto itself with $f(0) = 0$ and $f'(0) > 0$, then $f(z) = z$ for all z in G.*

Proof. Let C be a non-trivial component of the complement of G in $\mathbb{C}_\infty$. Put $\Omega = \mathbb{C}_\infty \setminus C$; so Ω is a simply connected region containing G. Let $\phi : \Omega \to \mathbb{D}$ be the Riemann map with $\phi(0) = 0$ and $\phi'(0) > 0$. Put $G_1 = \phi(G)$, $\phi_1 = \phi|G$, and $f_1 = \phi_1 \circ f \circ \phi_1^{-1}$. So f_1 is a conformal equivalence of G_1 onto itself with $f_1(0) = 0$ and $f_1'(0) > 0$. If it is shown that f_1 is the identity map, it follows that f is the identity. Thus it can be assumed that the region G is bounded.

Let M be a constant with $|z| \leq M$ for all z in G and pick $R > 0$ such that $\overline{B}(0; R) \subseteq G$. For $n \geq 1$, put $f_n = f \circ f \circ \cdots \circ f$, the composition of f with itself n times. So $f_n(0) = 0$ and $f_n'(0) = [f'(0)]^n$. Using Cauchy's Estimate it follows that $0 < [f'(0)]^n = f_n'(0) \leq M/R$. Thus $f'(0) \leq 1$. But applying the same reasoning to the inverse of f implies that $[f'(0)]^{-1} \leq 1$. Hence $f'(0) = 1$.

Let a_m be the first non-zero coefficient in the power series expansion of f about 0 with $m \geq 2$. So $f(z) = z + a_m z^m + \cdots$. By an induction argument, $f_n(z) = z + n a_m z^m + \cdots$. Once again Cauchy's estimate gives that $n|a_m| = n|f_n^{(m)}(0)|/n! \leq M/R^m$. But this implies that $a_m = 0$, a contradiction. Hence the only non-zero coefficient in the power series expansion for f is the first. Thus $f(z) = z$. $\square$

4.9 Corollary. *If G is a region not equal to $\mathbb{C}$, and if f and g are conformal equivalences of G onto a region Ω such that for some point a in G, $f(a) = g(a)$, $f'(a) > 0$, and $g'(a) > 0$, then $f = g$.*

The next result is a variation on Theorem 4.7.

4.10 Theorem. *Assume (4.3) and suppose that $G \neq \mathbb{C}$. If $G_n \to G$, each f_n is univalent, and $f_n'(0) > 0$ for each n, then there is a univalent function f on G such that $f_n \to f$ (uc) if and only if $\Omega_n \to \Omega$ for some region Ω. When this happens, $f(G) = \Omega$.*

Proof. First assume that $f_n \to f$ (uc) for some univalent function f. Thus $f'(0) > 0$ and $f_n'(0) \to f'(0)$. Thus if $g_n = f_n/f_n'(0)$ and $g = f/f'(0), g_n \to g$ (uc) on G.

According to the preceding theorem, $g_n(G_n) \to g(G)$. But $g_n(G_n) = [f_n(0)]^{-1}\Omega_n$, which converges to $[f'(0)]^{-1}\Omega$ (see Exercise 2). Thus $\Omega_n \to \Omega$.

Now assume that $\Omega_n \to \Omega$ for some region Ω containing 0. To avoid multiple subscripts, observe that anything demonstrated for the sequence $\{f_n\}$ applies as well to any of its subsequences. Put $g_n = f_n/f_n'(0)$ and assume that $f_n'(0) \to 0$. By Lemma 4.6 there is a subsequence $\{g_{n_k}\}$ that converges (uc) to a univalent function g on G. By Theorem 4.7, $[f_{n_k}'(0)]^{-1}\Omega_{n_k} = g_{n_k}(G_{n_k}) \to g(G)$. According to Exercise 2, $g(G) = \mathbb{C}$. Since g is univalent, this implies that $G = \mathbb{C}$, a contradiction. Now assume that $f_n'(0) \to \infty$. The same argument shows that there is a subsequence and a univalent function g on G such that $[f_{n_k}'(0)]^{-1}\Omega_{n_k} = g_{n_k}(G_{n_k}) \to g(G)$. Again Exercise 2 applies and we conclude that $\{g_{n_k}(G_{n_k})\}$ can have no kernel, a contradiction.

Thus it follows that there are constants c and C such that $c \leq f_n'(0) \leq C$. Suppose that $f_n'(0) \to \alpha$ for some non-zero scalar α. Maintaining the notation of the preceding paragraph, there is a subsequence $\{g_{n_k}\}$ such that $g_{n_k} \to g$ (uc) for some univalent function g on G. Thus $f_{n_k} = f_{n_k}'(0)g_{n_k} \to \alpha g = f$. Since $\alpha \neq 0$, f is univalent and, by Theorem 4.7, f maps G onto Ω. Note that $f'(0) = \alpha$.

Now suppose that the sequence of scalars $\{f_n'(0)\}$ has two distinct limit points, α and β. The preceding paragraph implies there are conformal equivalences f and h mapping G onto Ω with $f(0) = h(0) = 0$, $f'(0) = \alpha$, and $h'(0) = \beta$. By Corollary 4.9, $f = h$ and $\alpha = \beta$. It therefore follows that the sequence $\{f_n'(0)\}$ has a unique limit point α and so $f_n'(0) \to \alpha$. As above, this implies that $\{f_n\}$ converges (uc) to a univalent function f on G that has $f(G) = \Omega$. $\square$

This concept of the kernel of regions was introduced by Carathéodory, who proved the following.

4.11 Corollary. (The Cathéodory Kernel Theorem) *If for each $n \geq 1$, f_n is a univalent function on $\mathbb{D}$ with $f_n(0) = 0$, $f_n'(0) > 0$, and $f_n(\mathbb{D}) =$*

Ω_n, then $\{f_n\}$ converges uniformly on compact subsets of $\mathbb{D}$ to a univalent function if and only if $\{\Omega_n\}$ has a kernel $\Omega \neq \mathbb{C}$ and $\Omega_n \to \Omega$.

Theorem 4.10 is false when $G = \mathbb{C}$. Indeed, if $G_n = \mathbb{C} = \Omega_n$ for all n and $f_n(z) = n^{-1}z$, then $G_n \to \mathbb{C}$, $\Omega_n \to \mathbb{C}$, but $\{f_n\}$ does not converge to a univalent function.

Note that the general purpose of the main results of this section is to provide a geometrically equivalent formulation of the convergence of a sequence of univalent functions.

Much of this section is based on §V.5 of Goluzin [1969].

Exercises

1. If $G = \ker\{G_n\}$ and T is a Möbius transformation, then $T(G) = \ker\{T(G_n)\}$; similarly, if $G_n \to G$, then $T(G_n) \to T(G)$. Give conditions on an analytic function f so that $f(G) = \ker\{f(G_n)\}$ whenever $G = \ker\{G_n\}$.

2. Assume (4.3) and let $\{c_n\}$ be a sequence of complex scalars. (a) If $G_n \to G$ and $c_n \to c$, then $c_nG_n \to cG$. (b) If $\{G_n\}$ has a kernel and $c_n \to \infty$, then $c_nG_n \to \mathbb{C}$. (c) If $\{G_n\}$ has a kernel and $c_n \to 0$, then $\{c_nG_n\}$ has no kernel.

3. Give the details in the proof of Corollary 4.11.

4. Let G be the region obtained by deleting a finite number of non-zero points from $\mathbb{C}$ and show that the conclusion of Proposition 4.8 holds for G.

5. Assume (4.3). If $\{G_n\}$ has a kernel G and $G_n \to G$, then for every a in ∂G there is a sequence $\{a_n\}$ with $a_n \in G_n$ such that $a_n \to a$.

6. Give an example of a sequence of regions $\{G_n\}$ such that for each $n \geq 1$, $\mathbb{C} \setminus G_n$ has an infinite number of components and $G_n \to \mathbb{D}$.

7. Assume that G is a finitely connected Jordan region and $\varphi : G \to G$ is a conformal equivalence; so φ extends to $\mathrm{cl}G$. If G is not simply connected and there is a point a in ∂G such that $\varphi(a) = a$, then φ is the identity.

§5 Conformal Equivalence with a Circularly Slit Annulus

This section begins the presentation of some results concerning regions that are conformally equivalent to a finitely connected region. The reader might consider these as extensions of the Riemann Mapping Theorem. We know that each simply connected region is conformally equivalent to either the unit disk or the whole plane, this latter case only occurring when the region in question is the plane itself. The picture for finitely connected regions is more complicated but still manageable.

For our discussion fix an n-connected region G and let $K_0, \ldots, K_n$ be the components of $\mathbb{C}_\infty \setminus G$ with ∞ in K_0. Let us agree to call a region Ω a *circularly slit annulus* if it has the following form:

$$\Omega = \{\zeta \in \mathbb{C} : r_1 < |\zeta| < 1 \text{ and } \zeta \notin C_j \text{ for } 2 \leq j \leq n\},$$

where C_j is a closed proper arc of the circle $|\zeta| = r_j, r_1 < r_j < 1$, and these arcs are pairwise disjoint. For notational convenience set $r_0 = 1$. Note that this slit annulus in an n-connected region. Of course there is nothing in what follows that is insistent on having the outer radius of the associated annulus equal to 1; this is done for normalization purposes (see Exercise 1). For notational convenience C_0 will be the unit circle $\partial\,\mathbb{D}$ and C_1 will be the circle $\{z : |z| = r_1\}$, and these will be referred to as the *outer circle* and the *inner circle* of Ω. The main result of this section is the following.

5.1 Theorem. *If G is a non-degenerate n-connected region in $\mathbb{C}$, A and B are two components of $\partial_\infty G$, and $a \in G$, then there is a unique circularly slit annulus Ω and a conformal equivalence $\phi : G \to \Omega$ such that ϕ associates A with the outer circle and B with the inner circle and $\phi'(a) > 0$.*

The uniqueness statement will follow from the next lemma. Also the proof of this lemma will provide some motivation for the proof of existence.

Note that if Ω and Λ are two circularly slit annuli and $f : \Omega \to \Lambda$ is a conformal equivalence such that f associates the outer circle of Ω with the outer circle of Λ, then f extends continuously to the outer circle of Ω and maps this onto the outer circle of Λ. (In fact, f extends analytically across the outer circle.) Similar statements apply to the inner circle.

5.2 Lemma. *If Ω and Λ are two circularly slit annuli and $f : \Omega \to \Lambda$ is a conformal equivalence such that f associates the outer and inner circles of Ω with the outer and inner circles of Λ, respectively, then there is a complex number α with $|\alpha| = 1$ such that $\Lambda = \alpha\Omega$ and $f(z) = \alpha z$ for all z in Ω.*

Proof. Let G be an analytic n-Jordan region such that there is a conformal equivalence $\phi : G \to \Omega$. We can assume that $G \subseteq \mathbb{D}$, the outer boundary of G is $\partial\,\mathbb{D}$, and that ϕ associates $\partial\,\mathbb{D}$ with $\partial\,\mathbb{D}$. Let $\gamma_1, \ldots, \gamma_n$ be the remaining Jordan curves in the boundary of G and put $\gamma_0 = \partial\,\mathbb{D}$. Number these curves so that ϕ associates γ_1 with the inner circle of Ω. Let $\omega_1, \ldots, \omega_n$ be the harmonic basis for G. Note that each ω_k extends to a harmonic function in a neighborhood of $\operatorname{cl} G$.

Adopt the notation in the paragraph before the statement of the theorem and let $\Lambda = \{\zeta : \rho_1 < |\zeta| < 1 \text{ and } \zeta \notin \bigcup_{j=2}^n D_j\}$, where D_j is a closed proper arc in the circle $\{\zeta : |\zeta| = \rho_j\}$. Let D_1 and D_0 be the inner and outer circles of Λ. Let $\psi : G \to \Lambda$ be the conformal equivalence $\psi = f \circ \phi$. It follows that ϕ and ψ extend continuously to $\operatorname{cl} G$ (3.5).

Since ϕ does not vanish on G, $u = \log|\phi|$ is a continuous function on $\operatorname{cl} G$ that is harmonic on G. Because of the behavior of ϕ on ∂G, u is constant on each component of ∂G; let $u \equiv \lambda_j$ on γ_j. Thus $u - \sum_1^n \lambda_k \omega_k$ is harmonic and vanishes on $\gamma_1, \dots, \gamma_n$. But $u \equiv 0$ on $\gamma_0 = \partial\mathbb{D}$ and so $u - \sum_1^n \lambda_k \omega_k \equiv 0$ on ∂G. Therefore $u - \sum_1^n \lambda_k \omega_k \equiv 0$. Defining c_{jk} as in Proposition 1.7 we get

$$\begin{aligned}
\sum_{k=1}^n c_{jk}\lambda_k &= \sum_{k=1}^n \frac{1}{2\pi}\lambda_k \int_{\gamma_j} {}^*d\omega_k \\
&= \frac{1}{2\pi}\int_{\gamma_j} {}^*du \\
&= \frac{1}{\pi i}\int_{\gamma_j} \partial u \\
&= \frac{1}{2\pi i}\int_{\gamma_j} \frac{\phi'}{\phi} \\
&= n(\phi(\gamma_j); 0) \\
&= \begin{cases} -1 & \text{if} \quad j = 1 \\ 0 & \text{if} \quad 2 \le j \le n. \end{cases}
\end{aligned}$$

Now carry out the analogous argument with ψ. So $\log|\psi| = \sum_1^n \eta_k \omega_k$ and $\sum_{k=1}^n c_{jk}\eta_k = \sum_{k=1}^n c_{jk}\lambda_k$ for $1 \le j \le n$. But Proposition 1.7 says that the matrix $[c_{jk}]$ is invertible and so $\eta_j = \lambda_j$ for $1 \le j \le n$. Thus $\log|\phi| = \log|\psi|$. Hence $|\phi| = |\psi|$ on G and this implies the existence of a constant α with $|\alpha| = 1$ such that $\psi = \alpha\phi$. Thus $\Lambda = \alpha\Omega$ and $f(\phi(z)) = \alpha\,\phi(z)$ for all z in G. $\square$

It is easy to construct an example of a circularly slit annulus Ω for which no rotation takes Ω onto itself. So in this case if f is a conformal equivalence of Ω onto itself that maps the outer and inner circles onto themselves, the conclusion of the lemma is that f is the identity function.

Proof of Theorem 5.1. According to Theorem 3.3 we may assume that G is an analytic n-Jordan region such that the component A of ∂G is $\gamma_0 = \partial\mathbb{D}$, the outer boundary of G. The component B of ∂G is another curve. Let b be a point inside this curve; so $b \notin G$. If T is a Möbius transformation that maps $\mathbb{D}$ onto $\mathbb{D}$ and $T(b) = 0$, then replacing G by $T(G)$ we may assume that 0 belongs to the inside of B. Denote this boundary curve B by γ_1 and let $\gamma_2, \dots, \gamma_n$ be the remaining boundary curves. So $n(\gamma_1; 0) = -1$, $n(\gamma_0, 0) = 1$, and $n(\gamma_j; 0) = 0$ for $2 \le j \le n$. For $1 \le j, k \le n$ let

$$c_{jk} = \frac{1}{2\pi}\int_{\gamma_j} {}^*d\omega_k.$$

According to Proposition 1.7 the matrix $[c_{jk}]$ is invertible.

Corollary 13.4.14 implies there is an anlytic Jordan region W containing $\operatorname{cl} G$ such that each harmonic function ω_k, $1 \leq k \leq n$, has an extension to a function harmonic in W; also denote this extension by ω_k. Let $\alpha_1 = 0$ and pick points $\alpha_2, \ldots, \alpha_n$ in the inside of $\gamma_2, \ldots, \gamma_n$ so that they lie in the complement of $\operatorname{cl} W$.

Since $[c_{jk}]$ is invertible, there are (unique) real numbers $\lambda_1, \ldots, \lambda_n$ such that

5.3
$$\begin{cases} \displaystyle\sum_{k=1}^{n} c_{1k}\lambda_k = -1 \\ \displaystyle\sum_{k=1}^{n} c_{jk}\lambda_k = 0 \quad \text{for } 2 \leq j \leq n. \end{cases}$$

Let u be the harmonic function on W given by

$$u = \sum_{k=1}^{n} \lambda_k \omega_k.$$

By Theorem 1.3 there is an analytic function h on W such that

$$u(z) = \operatorname{Re} h(z) + \sum_{k=1}^{n} c_k \log|z - \alpha_k|,$$

where $c_1, \ldots, c_n$ are the periods of u. Let's calculate these periods.

For $1 \leq j \leq n$,

$$\begin{aligned} c_j &= \frac{1}{2\pi}\int_{\gamma_j} {}^*du \\ &= \sum_{k=1}^{n} \lambda_k \frac{1}{2\pi}\int_{\gamma_j} {}^*d\omega_k \\ &= \sum_{k=1}^{n} \lambda_k\, c_{jk} \\ &= \begin{cases} -1 & j = 1 \\ 0 & 2 \leq j \leq n. \end{cases} \end{aligned}$$

Thus $u = \operatorname{Re} h - \log|z|$.

Put $\phi = z^{-1}e^h$. So ϕ is an analytic function on W that does not vanish there. Note that $|\phi| = |z|^{-1}e^{\operatorname{Re} h} = \exp u$. Thus $|\phi|$ is constant on each of the boundary curves of G, $\gamma_1, \ldots, \gamma_n$. In fact on γ_j, $|\phi| \equiv r_j \equiv e^{\lambda_j}$. It is claimed that ϕ is the desired conformal equivalence. To establish this, many things must be checked.

For any complex number ζ, let $N(\zeta)$ be the number of solutions, counting multiplicities, of the equation $\phi(z) = \zeta$ that lie in the region G. From the Argument Principle

5.4 $$N(\zeta) = \sum_{j=0}^{n} \frac{1}{2\pi i} \int_{\gamma_j} \frac{\phi'}{\phi - \zeta} = \sum_{j=0}^{n} n(\phi(\gamma_j); \zeta).$$

Now $\phi(\gamma_j)$ is a closed curve (possibly not a Jordan curve) and, since $|\phi| = r_j$ on γ_j, this closed curve must be contained in the circle $A_j = \{\zeta : |\zeta| = r_j\}$. Thus $n(\phi(\gamma_j); \zeta) = 0$ for $|\zeta| > r_j$ and $n(\phi(\gamma_j); \zeta) = n(\phi(\gamma_j); 0)$ for $|\zeta| < r_j$. Using Proposition 13.3.5 we get that $0 \le j \le n$ and $|\zeta| \ne r_j$,

$$\begin{aligned} n(\phi(\gamma_j); 0) &= \frac{1}{2\pi i} \int_{\gamma_j} \frac{\phi'}{\phi} \\ &= \frac{1}{2\pi i} \int_{\gamma_j} \left(\frac{1}{z} + h' \right) \\ &= \frac{1}{\pi i} \int_{\gamma_j} \partial u \\ &= \frac{1}{2\pi} \int_{\gamma_j} {}^*du. \end{aligned}$$

For $1 \le j \le n$ this last integral is c_j, which was calculated previously. For $j = 0$ first observe that ∂u is an analytic function on W so that

$$\begin{aligned} n(\phi(\gamma_0); 0) &= \frac{1}{\pi i} \int_{\gamma_0} \partial u \\ &= -\sum_{j=1}^{n} \frac{1}{\pi i} \int_{\gamma_j} \partial u \\ &= -\sum_{j=1}^{n} c_j \\ &= 1. \end{aligned}$$

Therefore we obtain

5.5 $$n(\phi(\gamma_j); 0) = \begin{cases} 1 & \text{if } \ j = 0 \\ -1 & \text{if } \ j = 1 \\ 0 & \text{if } \ 2 \le j \le n. \end{cases}$$

Substituting in (5.4) this implies that if $|\zeta| \ne r_0 (= 1), r_1, \ldots, r_n$, then $N(\zeta) = n(\phi(\gamma_0); \zeta) + n(\phi(\gamma_1); \zeta)$ and so a consideration of all the possible cases (save one) gives that

5.6
$$N(\zeta)=\begin{cases} 0 & \text{if } |\zeta|>1 \text{ and } |\zeta|>r_1 \\ 0 & \text{if } |\zeta|<1 \text{ and } |\zeta|<r_1 \\ 1 & \text{if } r_1<|\zeta|<1. \end{cases}$$

The one possibility that is left out is to have $N(\zeta)=-1$ for $1<|\zeta|<r_1$. But this is nonsense; the equation $\phi(z)=\zeta$ cannot have -1 solutions in G. Thus $r_1\le 1$. But $\phi(G)$ is open so that $\phi(\zeta)=\zeta$ must have some solutions. Thus we have that
$$r_1<1.$$

Equation 5.6 also shows that ϕ is a one-to-one map of G onto its image and $\phi(G)\subseteq R\equiv\{\zeta : r_1<|\zeta|<1\}$. Let $C_j=\phi(\gamma_j)\subseteq A_j$. Again (5.6) shows that $N(\zeta)=1$ for ζ in R and $|\zeta|\ne r_j(2\le j\le n)$. So $\phi(G)\supseteq\{\zeta\in R : |\zeta|\ne r_2,\ldots,r_n\}$. Because ϕ is a homeomorphism of G onto $\phi(G), \partial\phi(G)=\phi(\partial G)=\cup_{j=2}^n C_j\cup\{\zeta : |\zeta|=1 \text{ or } r_1\}$. This implies two things. First it must be that $r_1<r_j<1$ for $2\le j\le n$. Second
$$\phi(G)=\Omega\equiv R\setminus\bigcup_{j=2}^n C_j.$$

Now $\phi(G)$ and hence Ω, must be a connected set and so each C_j is a proper closed arc in the circle A_j. That is, Ω is a slit annulus.

What about $\phi'(0)$? It may be that $\phi'(0)$ is not positive. However by replacing ϕ by $e^{i\theta}\phi$ for a suitable θ and replacing Ω by $e^{i\theta}\Omega$, this property is insured.

The proof of uniqueness is an easy consequence of Lemma 5.2 and is left to the reader. □

Consider the annulus $G=\{z : R<|z|<1\}$. The map $\phi(z)=\lambda R/z$ for a scalar λ with $|\lambda|=1$ is a conformal equivalence of G onto itself. Thus the uniqueness of the conformal equivalence obtained in Theorem 5.1 is dependent on the assignment of two of the boundary components. The use of a rotation shows that in addition to assigning two boundary components it is also necessary to specify the sign of the derivative at a point.

What happens if some of the components of $\mathbb{C}_\infty\setminus G$ are singletons? Suppose that $\mathbb{C}_\infty\setminus G$ has $n+1$ non-trivial components $K_0,\ldots,K_n$ and m components $a_1,\ldots,a_m$ that are singletons. By the application of a simple Möbius transformation, it can be assumed that $\infty\in K_0$. Let $H=G\cup\{a_1,\ldots,a_m\}$. According to Theorem 5.1 there is a conformal equivalence $\phi: H\to\Omega$ for some circularly slit annulus Ω. Let $\alpha_i=\phi(a_i)$. This leads to the following.

5.7 Theorem. *If $n \geq 1$ and G is an $(n+m)$-connected region with only $(n+1)$ of the components of its complement in $\mathbb{C}_\infty$ non-trivial, then G is conformally equivalent to a circularly slit annulus with m points removed.*

The above thereoms have an appealing form in the case that $n = 1$ and it is worth stating this separately.

5.8 Theorem. *If G is a 1-connected region in $\mathbb{C}$, then the following statements hold.*

(a) *If each component of $\mathbb{C}_\infty \setminus G$ is a point, then G is conformally equivalent to the punctured plane $\mathbb{C}_0$.*

(b) *If one component of $\mathbb{C}_\infty \setminus G$ is a point and the other is not, then G is conformally equivalent to $\{z : 1 < |z| < \infty\}$.*

(c) *If neither component of $\mathbb{C}_\infty \setminus G$ is a point, then there is a finite number r such that G is conformally equivalent to $\{z : 1 < |z| < r\}$.*

If $A_r \equiv \{z : 1 < |z| < r\}$ for $1 < r \leq \infty$, then A_{r_1} and A_{r_2} are conformally equivalent if and only if $r_1 = r_2$.

Proof. The proofs of (a), (b), and (c) are straightforward. The proof that A_{r_1} and A_{r_2} are conformally equivalent if and only if $r_1 = r_2$ follows from the uniqueness part of Theorem 5.1. □

Exercises

1. Show that every proper annulus is conformally equivalent to one of the form $\{z : r < |z| < 1\}$.

2. Assume G is an analytic Jordan region and Ω is a circularly slit annulus as in the proof of Theorem 5.1; adopt the notation of that proof. If $\phi : \operatorname{cl} G \to \operatorname{cl} \Omega$ is the continuous extension of the conformal equivalence of G onto Ω, show that for $2 \leq j \leq n$, ϕ is two-to-one on γ_j. (Hint: Let σ_ε be a Jordan curve in G that contains γ_j in its inside and has the remaining boundary curves of G in its outside. Note that $\phi(\sigma_\varepsilon)$ is a Jordan curve in Ω that contains precisely one boundary arc of Ω in its inside. What happens as $\varepsilon \to 0$?)

3. If $0 \leq r_j < R_j \leq \infty, j = 1, 2$, show that $\operatorname{ann}(0; r_1, R_1)$ and $\operatorname{ann}(0; r_2, R$ are conformally equivalent if and only if $R_1/r_1 = R_2/r_2$.

4. Let $A = \operatorname{ann}(0; r, R), 0 \leq r < R \leq \infty$, and characterize all the analytic functions $f : A \to A$ that are bijective.

5. (a) Let G be a non-degenerate n-connected region with boundary components $C_0, C_1, \ldots, C_n$ and let $\varphi : G \to G$ be a conformal equivalence. Show that if $n \geq 2$ and for three values of k, φ associates C_k with itself in the sense of §3, then φ is the identity mapping. (b) If $n \geq 2$, show that the group of all conformal equivalences of G onto itself is finite. Give a bound on the order of the group. (See Heins [1946].)

§6 Conformal Equivalence with a Circularly Slit Disk

In this section we will see another collection of canonical n-connected regions that completely model the set of all n-connected regions.

6.1 Definition. A *circularly slit disk* is a region Ω of the form

$$\Omega = \mathbb{D} \setminus \bigcup_{j=1}^{n} C_j,$$

where for $1 \leq j \leq n$, C_j is a proper closed arc in the circle $|z| = r_j$, $0 < r_j < 1$.

Note that as defined a circularly slit disk contains 0. The point 0 will be used to give the uniqueness statement in Theorem 6.2 below. The main result of this section is the following.

6.2 Theorem. *If G is a non-degenerate n-connected region, $a \in G$, and A is any component of ∂G, then there is a unique circularly slit disk Ω and a unique conformal equivalence $\phi : G \to \Omega$ such that ϕ associates A with $\partial \mathbb{D}$, $\phi(a) = 0$, and $\phi'(a) > 0$.*

As in the preceding section, we will prove a lemma that will imply the uniqueness part of the theorem and also motivate the existence proof.

6.3 Lemma. *If Ω and Λ are slit disks and $f : G \to \Lambda$ is a conformal equivalence such that $f(0) = 0$ and $f(\partial \mathbb{D}) = \partial \mathbb{D}$, then there is a complex number α with $|\alpha| = 1$ such that $\Lambda = \alpha\Omega$ and $f(z) = \alpha z$ for all z in Ω.*

Proof. Some details will be omitted as this proof is similar to that of Lemma 5.2. Let G be an analytic Jordan region with outer boundary $\gamma_0 = \partial \mathbb{D}$ such that there is a conformal equivalence $\phi : G \to \Omega$ and $\phi(\partial \mathbb{D}) = \partial \mathbb{D}$. Let $a \in G$ such that $\phi(a) = 0$. Denote the remaining boundary curves of G by $\gamma_1, \ldots, \gamma_n$. Adopt the notation in Definition 6.1.

Let $u = \log |\phi|$. So u is a negative harmonic function on $G \setminus \{a\}$ and, for $0 \leq j \leq n$, $u \equiv \rho_j$, where $\exp(\rho_j) = r_j$. Now $\phi = (z - a)h$, where h is an analytic function that never vanishes on G. So $u = \log |z - a| + \log |h|$ and

$\log|h|$ is harmonic on G. If $\omega_1, \ldots, \omega_n$ is the harmonic basis for G, $[u - \sum_j \rho_j \omega_j] - \log|z-a|$ is harmonic on G and $u - \sum_j \rho_j \omega_j$ vanishes on ∂G. Thus

$$u(z) = -g(z,a) + \sum_{j=1}^{n} \rho_j \omega_j(z),$$

where $g(z,a) = g_a(z)$ is the Green function for G with singularity at a. Thus

$$\begin{aligned}
\frac{1}{2\pi}\int_{\gamma_j} {}^*du &= -\frac{1}{2\pi}\int_{\gamma_j} {}^*dg_a + \sum_{k=1}^{n} \frac{\rho_k}{2\pi}\int_{\gamma_j} {}^*d\omega_k \\
&= \omega_j(a) + \sum_{k=1}^{n} c_{jk}\rho_k,
\end{aligned}$$

where c_{jk} is one of the periods of ω_k. On the other hand,

$$\begin{aligned}
\frac{1}{2\pi}\int_{\gamma_j} {}^*du &= \frac{1}{\pi i}\int_{\gamma_j} \partial u \\
&= \frac{1}{2\pi i}\int_{\gamma_j} \frac{\phi'}{\phi} \\
&= 0
\end{aligned}$$

for $1 \le j \le n$. Therefore $\rho_1, \ldots, \rho_n$ are the unique solutions of the equations

6.4
$$\sum_{k=1}^{n} c_{jk}\,\rho_k = -\omega_j(a).$$

Arguing as in the proof of Lemma 5.2, if $\psi = f \circ \phi$, then $|\phi| = |\psi|$ and so $\psi = \alpha\phi$ for some scalar α with $|\alpha| = 1$. □

Proof of Theorem 6.2. Without loss of generality we may assume that G is an analytic Jordan region with outer boundary $A = \gamma_0 = \partial\mathbb{D}$; let $\gamma_1, \ldots, \gamma_n$ be the remaining boundary curves. Let $g_a = -\log|z-a| + R_a$ be the Green function for G and let $\omega_1, \ldots, \omega_n$ be the harmonic basis for G. If $\{c_{jk}\}$ are the periods for the harmonic basis, let $\rho_1, \ldots, \rho_n$ be the unique scalars such that (6.4) is satisfied. Put $u = -g_a + \sum_k \rho_k\omega_k$ and put $v = u - \log|z-a|$. So v is harmonic on G and a computation shows that for $1 \le j \le n$

$$\begin{aligned}
\frac{1}{2\pi}\int_{\gamma_j} {}^*dv &= -\frac{1}{2\pi}\int_{\gamma_j} {}^*d\,\log|z-a| - \frac{1}{2\pi}\int_{\gamma_j} {}^*dg_a + \sum_{k=1}^{n} c_{jk}\rho_k \\
&= -\frac{1}{2\pi}\int_{\gamma_j} {}^*d\,\log|z-a| \\
&= n(\gamma_j, a) \\
&= 0.
\end{aligned}$$

Therefore there is an analytic function h on G such that $v = \operatorname{Re} h$. Moreover h can be chosen so that $h(a)$ is real. So $u = \log|z-a| + \operatorname{Re} h$.

Let $\phi = (z-a)e^h$ so that $\log|\phi| = u$. It follows that $|\phi| = r_0 \equiv 1$ on γ_0 and $|\phi| = r_j \equiv e^{\rho_j}$ on γ_j for $1 \le j \le n$. For any complex number ζ let $N(\zeta)$ be the number of solutions of the equation $\phi(z) = \zeta$, counting multiplicities. As in the proof of Theorem 5.1,

$$N(\zeta) = \sum_{j=0}^{n} n(\phi(\gamma_j);\zeta).$$

Now $\phi(\gamma_j) \subseteq \{\zeta : |\zeta| = r_j\}$ so that to calculate $N(\zeta)$ it suffices to calculate $n(\phi(\gamma_j);0)$. But

$$\begin{aligned} n(\phi(\gamma_j);0) &= \frac{1}{2\pi i}\int_{\gamma_j} \frac{\phi'}{\phi} \\ &= \frac{1}{2\pi i}\int_{\gamma_j} \left(\frac{1}{z-a} + h'\right) \\ &= n(\gamma_j;a) \\ &= \begin{cases} 1 & \text{if } j = 0 \\ 0 & \text{if } 1 \le j \le n. \end{cases} \end{aligned}$$

Collating the various pieces of information we get that if $|\zeta| \ne r_j$ for $0 \le j \le n$, then

$$N(\zeta) = \begin{cases} 0 & \text{if } |\zeta| > 1 \\ 1 & \text{if } |\zeta| < 1. \end{cases}$$

From here it follows that $\Omega = \phi(G)$ is a slit disk and ϕ is a conformal equivalence of G onto Ω with $\phi(a) = 0$. Since $h(a)$ is a real number, a calculation shows that $\phi'(a) > 0$.

The uniqueness follows from Lemma 6.3 and is left to the reader. $\square$

Exercises

1. Denote the conformal equivalence obtained in Theorem 6.2 by $\phi(z,a)$; so $\phi(a,a) = 0$ and $\partial_1\phi(a,a) > 0$. Show that if $[d_{jk}]$ is the inverse of the matrix $[c_{jk}]$, then

$$\log|\phi(z,a)| = -g(z,a) + \sum_{j,k=1}^{n} d_{jk}\,\omega_j(z)\,\omega_k(a).$$

Thus $\phi(z,a) = \phi(a,z)$.

2. Show that if G is an n-connected region and $a, b \in G$, then there is a conformal equivalence ψ on G that maps G onto the extended plane with circular slits and $\psi(a) = 0$ and $\psi(b) = \infty$. (Hint: Let $\psi(z) = \phi(z,a)/\phi(z,b)$ (notation from Exercise 1); use the Argument Principle to show that ψ is one-to-one.)

3. How must Theorem 6.2 be changed if some of the components of the complement of G are trivial?

§7 Conformal Equivalence with a Circular Region

A region Ω is a *circular region* if its boundary consists of a finite number of disjoint non-degenerate circles. In this section it will be shown that every n-connected region is conformally equivalent to a circular region bounded by $n+1$ circles. This proof will be accomplished by the use of Brouwer's Invariance of Domain Theorem combined with previously proved conformal mapping results. But first the uniqueness question for such regions will be addressed. Recall that the *oscillation* of a function f on a set E is $\operatorname{osc}(f; E) = \sup\{|f(x) - f(y)| : x, y \in E\}$. For a curve γ, $\ell(\gamma)$ denotes its length.

7.1 Lemma. *Let K be a compact subset of the region G and let f be a bounded analytic function on $G \setminus K$. If, for every $\varepsilon > 0$ and every open set U containing K and contained in G, there are smooth Jordan curves $\{\gamma_1, \dots, \gamma_n\}$ in U that contain K in the union of their insides such that*

$$\sum_{j=1}^{n} \ell(\gamma_j)^2 < \varepsilon \qquad \text{and} \qquad \sum_{j=1}^{n} [\operatorname{osc}(f, \gamma_j)]^2 < \varepsilon,$$

then f has an analytic continuation to G.

Proof. Let γ_0 be a smooth positively oriented Jordan curve in $G \setminus K$ that contains K in its inside. Fix a point z in $G \setminus K$ and inside γ_0. For $\varepsilon > 0$ let $\gamma_1, \dots, \gamma_n$ be as in the statement of the lemma arranged so that they lie inside γ_0 and the point z is outside each of them. Give the curves $\gamma_1, \dots, \gamma_n$ negative orientation. So $\Gamma = \{\gamma_0, \gamma_1, \dots, \gamma_n\}$ is a positively oriented Jordan system in G. Let $d = \operatorname{dist}(z, \Gamma)$. Hence

$$f(z) = \frac{1}{2\pi i} \int_{\gamma_0} \frac{f(w)}{w - z} dw + \sum_{j=1}^{n} \frac{1}{2\pi i} \int_{\gamma_j} \frac{f(w)}{w - z} dw.$$

Fix a point w_j on γ_j. Since z is in the outside of γ_j,

$$\frac{1}{2\pi i} \int_{\gamma_j} \frac{f(w)}{w - z} dw \quad = \quad -\frac{f(w_j)}{2\pi i} \int_{\gamma_j} \frac{dw}{w - z} + \frac{1}{2\pi i} \int_{\gamma_j} \frac{f(w)}{w - z} dw$$

$$= \frac{1}{2\pi i}\int_{\gamma_j}\frac{f(w)-f(w_j)}{w-z}dw.$$

Hence if $\delta_j = \operatorname{osc}(f,\gamma_j)$,

$$\begin{aligned}\left|\sum_{j=1}^{n}\frac{1}{2\pi i}\int_{\gamma_j}\frac{f(w)}{w-z}dw\right| &\leq \frac{1}{2\pi}\sum_{j=1}^{n}\int_{\gamma_j}\left|\frac{f(w)-f(w_j)}{w-z}\right||dw| \\ &\leq \sum_{j=1}^{n}\frac{\ell(\gamma_j)\delta_j}{2\pi\, d} \\ &\leq \frac{1}{2\pi\, d}\left[\sum_{j=1}^{n}\ell(\gamma_j)^2\right]^{1/2}\left[\sum_{j=1}^{n}\delta_j^2\right]^{1/2} \\ &\leq \frac{\varepsilon}{2\pi\, d}.\end{aligned}$$

Since ε was arbitrary,

$$f(z) = \frac{1}{2\pi i}\int_{\gamma_0}\frac{f(w)}{w-z}dw$$

for all z inside γ_0 and lying in $G\setminus K$. Thus this formula gives a means of defining f on the set K that furnishes the required analytic continuation. □

For the following discussion, let's fix some notation. Let Ω be a circular region whose outer boundary is $\gamma_0 = \partial\,\mathbb{D}$. Let $\gamma_1,\ldots,\gamma_n$ be the remaining circles that form the boundary of Ω; put $\gamma_j = \partial B(a_j;r_j)$. Now look at the region that is the reflection of Ω across the circle γ_j. Recall (§3.3) that the reflection of a point z across the circle γ_j is the point w given by

7.2 $$w = a_j + \frac{r_j^2}{\bar{z}-\bar{a}j}.$$

Note that this formula is the conjugate of a Möbius transformation. Thus the image of Ω under this transformation is another circular region; call it Ω_{1j} for the moment. Note that the outer boundary of Ω_{1j} is the circle γ_j. Thus $\Omega(1) \equiv \Omega\cup\Omega_{11}\cup\cdots\cup\Omega_{1n}\cup\gamma_1\cup\cdots\cup\gamma_n$ is also a circular region, though its complement has more components (how many more?). Now for each Ω_{1j} and for each of its boundary circles γ, look at the image of Ω_{1j} under the reflection across γ. Denote the resulting circular regions by $\{\Omega_{2j} : 1\leq j\leq N_2\}$. Note that each of these is the image of Ω under two successive reflections and hence is the image of Ω under the composition of two transformations of the type given in (7.2). It is easy to check that the composition of two such transformations is a Möbius transformation. Let

$\Omega(2)$ be the union of $\Omega(1)$, the regions $\{\Omega_{2j} : 1 \le j \le N_2\}$, together with the circles forming the inner boundary of $\Omega(1)$.

Continue. This produces for each integer k a collection of circular regions $\{\Omega_{kj} : 1 \le j \le N_k\}$ and an increasing sequence of circular regions $\{\Omega(k)\}$, where $\Omega(k)$ is the union of $\Omega(k-1)$, the regions $\{\Omega_{kj} : 1 \le j \le N_k\}$, together with the circles that form the inner boundary of $\Omega(k-1)$. Put $\Omega(\infty) = \cup_k \Omega(k)$. So

$$\Omega(\infty) = \Omega \cup \bigcup \{\text{cl}\,\Omega_{kj} : k \ge 1 \,\text{and}\, 1 \le j \le N_k\}.$$

For $k \ge 1$ and $1 \le j \le N_k$, $\Omega_{kj} = T_{kj}(\Omega)$, where T_{kj} is a Möbius transformation if k is even and the conjugate of a Möbius transformation if k is odd. Let γ_{kji}, $1 \le i \le n$, be the circles that are the components of $\partial\Omega_{kj}$ exclusive of its outer boundary. So the components of $\partial\Omega(k)$ are the unit circle together with the circles $\{\gamma_{kji} : 1 \le j \le N_k, 1 \le i \le n\}$. Put $r_{kji} =$ the radius of γ_{kji}.

7.3 Lemma. *With the preceding notation,*

$$\sum_{k=1}^{\infty} \sum \{r_{kji}^2 : 1 \le j \le N_k,\ 1 \le i \le n\} < \infty.$$

Proof. First note that the regions $\{\Omega_{kj} : k \ge 1,\ 1 \le j \le N_k\}$ are pairwise disjoint. Let D_{kj} be the derivative of T_{kj} when k is even and the derivative of the conjugate of T_{kj} when k is odd. Let $B = B(a; r)$ be a disk contained in Ω and consider the disks $T_{kj}(B)$. By Koebe's 1/4-Theorem, $T_{kj}(B)$ contains a disk of radius $r\,|D_{kj}(a)|/4$. Thus $\text{Area}(\Omega_{kj}) \ge \pi\, r^2\, |D_{kj}(a)|^2/16$. Thus

7.4
$$\sum_{k=1}^{\infty} \sum \{|D_{kj}(a)|^2 : 1 \le j \le N_k\} < \infty.$$

According to the Distortion Theorem (14.7.14), for $1 \le i \le n$ there is a constant M_i such that for $k \ge 1$ and $1 \le j \le N_k$,

$$\sup\{|D_{kj}(z)| : z \in \gamma_i\} \le M_i\, |D_{kj}(a)|\,.$$

Now $\gamma_{kji} = T_{kj}(\gamma_i)$ for $1 \le i \le n$. Thus

$$2\pi\, r_{kji} = \int_{\gamma_i} |D_{kj}(z)|\ |dz| \le M_i\, |D_{kj}(a)|\, 2\pi\, r_i.$$

Combining this with (7.4) gives the proof of the lemma. □

7.5 Proposition. *If Ω and Λ are circular regions and $f : \Omega \to \Lambda$ is a conformal equivalence, then f is a Möbius transformation.*

Proof. By Theorem 3.4 it follows that f maps each circle in the boundary of Ω homeomorphically onto a circle in the boundary of Λ. By using appropriate Möbius transformations, it suffices to consider the case that the outer boundaries of Ω and Λ are the unit circle and f maps $\partial\,\mathbb{D}$ onto itself. If $\eta_1, \ldots, \eta_n$ are the other boundary circles of Λ, the numbering can be arranged so that $f(\gamma_i) = \eta_i$ for $1 \le i \le n$.

Adopt the notation of Lemma 7.3 and the analogous notation for the circular region Λ. By using the Reflection Principle there is for each $k \ge 1$ a conformal equivalence $f_k : \Omega(k) \to \Lambda(k)$ that continues f. Hence we get a conformal equivalence $f_\infty : \Omega(\infty) \to \Lambda(\infty)$. Note that $K = \mathbb{D} \setminus \Omega(\infty)$ is compact.

Now apply Lemma 7.1 to show that f_∞ has a continuation to $\mathbb{D}$. Once this is done the proof will be complete. Indeed, if $g = f^{-1} : \Lambda \to \Omega$, the same argument shows that g has a continuation to $\mathbb{D}$ that is a conformal equivalence on $\Lambda(\infty)$. In fact $g_\infty = f_\infty^{-1}$. Since $z = g_\infty \circ f_\infty = f_\infty \circ g_\infty$, it must be that the continuation of f to $\mathbb{D}$ is a conformal equivalence whose inverse is the continuation of g to $\mathbb{D}$. Thus f is a conformal equivalence of $\mathbb{D}$ onto itself. According to Theorem 6.2.5, f is a Möbius transformation.

To see that Lemma 7.1 is applicable to f_∞, let U be an open subset of $\mathbb{D}$ that contains K and let $\varepsilon > 0$. By Lemma 7.3 and an easy topological argument, there is an integer m such that for $k \ge m$ each of the circles γ_{kji} is contained in U and

$$\sum_{k=m}^{n} \sum \{r_{kji}^2 : 1 \le j \le N_k,\ 1 \le i \le n\} < \varepsilon.$$

If ρ_{kji} is the radius of η_{kji}, m can also be chosen so that

$$\sum_{k=m}^{\infty} \sum \{\rho_{kji}^2 : 1 \le j \le N_k,\ 1 \le i \le n\} < \varepsilon.$$

But f_k maps the circle γ_{kji} onto the circle η_{kji}. So if $\delta_{kji} = \operatorname{osc}(f, \gamma_{kji})$, this last inequality implies that

$$\sum_{k=m}^{\infty} \sum \{\delta_{kji}^2 : 1 \le j \le N_k,\ 1 \le i \le n\} < \varepsilon.$$

Thus the curves $\{\gamma_{mji} : 1 \le i \le n,\ 1 \le j \le N_m\}$ are those required by Lemma 7.1. $\square$

The topological lemma that follows will be used in the existence proof.

7.6 Lemma. (a) *Let $\{\Omega_k\}$ be circularly slit disks such that $\partial\,\mathbb{D}$ is the outer boundary of each and each Ω_k is n-connected. If Ω is a non-degenerate n-connected region with outer boundary $\partial\,\mathbb{D}$ and $\Omega_k \to \Omega$ in the sense of (4.1), then Ω is a circularly slit disk.*

(b) *Let $\{G_k\}$ be circular regions such that $\partial\,\mathbb{D}$ is the outer boundary of each and each G_k is n-connected. If G is a non-degenerate n-connected region with outer boundary $\partial\,\mathbb{D}$ and $G_k \to G$ in the sense of (4.1), then G is a circular region.*

Proof. (a) Let $\gamma_1, \ldots, \gamma_n$ be the bounded components of the complement of Ω; put $\gamma_0 = \partial\,\mathbb{D}$. Let $0 < \delta < \text{dist}(\gamma_j, \gamma_i)$ for $i \neq j$ and $0 \leq i, j \leq n$. Choose an integer k_1 such that for $k \geq k_1$, $K \equiv \{z \in \mathbb{D}; \text{dist}(z, \partial\Omega) \geq \delta\} \subseteq \Omega_k$. If $0 \leq j \leq n$ and $a_j \in \gamma_j$, then Exercise 4.5 implies there is an integer $k_2 > k_1$ such that for $k \geq k_2$, $\text{dist}(a_j, \partial\Omega_k) < \delta/2$. Fix $k \geq k_2$ for the moment and let $a_{jk} \in \partial\Omega_k$ with $|a_{jk} - a_j| < \delta/2$. If γ_{jk} is the component of $\partial\Omega_k$ that contains a_{jk}, then it must be that $\gamma_{jk} \subseteq \{z : \text{dist}(z, \gamma_j) < \delta\}$. Indeed the fact that $K \subseteq \Omega_k$ and $\Omega \subseteq \mathbb{D}$ implies that $\gamma_{jk} \subseteq \mathbb{D} \setminus K$. Since γ_{jk} is connected, the choice of δ implies that $\gamma_{jk} \subseteq \{z : \text{dist}(z, \gamma_j) < \delta\}$.

Thus we have that for $k \geq k_2$, the proper arcs that form the bounded components of the complement of Ω_k can be numbered $\gamma_{1k}, \ldots, \gamma_{nk}$ so as to satisfy

7.7 $$\gamma_{jk} \subseteq (\gamma_j)_\delta \equiv \{z : \text{dist}(z, \gamma_j) < \delta\}.$$

Now fix j, $1 \leq j \leq n$. For each ζ in γ_j, Exercise 4.5 implies there is a sequence $\{\zeta_k\}$ with ζ_k in $\partial\Omega_k$ such that $\zeta_k \to \zeta$. By (7.7), $\zeta_k \in \gamma_{jk}$ for $k \geq k_2$. If γ_{jk} is contained in the circle $\{\zeta : |\zeta| = \rho_{jk}\}$, this implies that $\rho_{jk} = |\zeta_{jk}| \to |\zeta|$. Since ζ was an arbitrary point of γ_j, this shows that γ_j is contained in the circle $\{\gamma : |\gamma| = \rho_j\}$, where $\rho_{jk} \to \rho_j$ as $k \to \infty$. But Ω is connected and no component of the complement of Ω is trivial. Thus γ_j is a proper closed arc in this circle and hence Ω is a circularly slit disk.

The proof of part (b) is similar. □

7.8 Lemma.
(a) *Let $\{G_k\}$ and G be circular regions such that $\partial\,\mathbb{D}$ is the outer boundary of each and each is n-connected; for each $k \geq 1$ let $f_k : G_k \to \Omega_k$ be a conformal equivalence onto a circularly slit disk Ω_k with outer boundary $\partial\,\mathbb{D}$ such that $f_k(0) = 0$, $f_k'(0) > 0$, and $f_k(\partial\,\mathbb{D}) = \partial\,\mathbb{D}$. If $G_k \to G$ in the sense of (4.1), then $f_k \to f$ (uc), where f is a conformal equivalence of G onto a circularly slit region Ω with outer boundary $\partial\,\mathbb{D}$, and $\Omega_k \to \Omega$.*
(b) *Let $\{\Omega_k\}$ and Ω be circularly slit disks such that $\partial\,\mathbb{D}$ is the outer boundary of each and each is n-connected; for each $k \geq 1$ let $\phi_k : \Omega_k \to G_k$ be a conformal equivalence onto a circular region G_k with outer boundary $\partial\,\mathbb{D}$ such that $\phi_k(0) = 0$, $\phi_k'(0) > 0$, and $\phi_k(\partial\,\mathbb{D}) = \partial\,\mathbb{D}$. If $\Omega_k \to \Omega$ in the sense of (4.1), then $\phi_k \to \phi$ (uc), where ϕ is a conformal equivalence of Ω onto a circularly slit region G with outer boundary $\partial\,\mathbb{D}$, and $G_k \to G$.*

Proof. As with the preceding lemma, the proofs of (a) and (b) are similar, so only the proof of (a) will be presented.

First we show that $\{f_k'(0)\}$ is bounded away from 0. Let $C_1, \ldots, C_n$ be the circles in the boundary of G that are different from $\partial\,\mathbb{D}$ and choose $\varepsilon > 0$

such that the closure of $V = \{z : \operatorname{dist}(z, \partial\mathbb{D}) < \varepsilon\}$ is disjoint from each C_j. Since $G_k \to G$, there is a k_1 such that for $k \ge k_1, \operatorname{cl} V \cap \partial G_k = \partial\mathbb{D}$. Thus each f_k with $k \ge k_1$ admits a univalent analytic continuation to $G_k \cup V$, which will also be denoted by f_k (Proposition 13.4.4). It is straightforward to check that $G_k \cup V \to G \cup V$.

We now want to construct certain paths from 0 to points on the unit circle. Let $z \in \partial\mathbb{D}$ and consider the radius $[0, z]$. This radius may meet some of the circles C_j. Enlarge these circles to circles $D_1, \ldots, D_n$ so that they remain pairwise disjoint, do not meet $\partial\mathbb{D}$, and do not surround 0. Whenever $[0, z]$ meets D_j, replace the segment of $[0, z]$ by half of this circle. Each circle D_j has radius less than 1 so that we arrive at a path from 0 to z that lies entirely in G, stays well away from the circles $C_1, \ldots, C_n$, and has length less than $1 + n\pi$. Thus we can find an open subset U of G such that $\partial\mathbb{D} \subseteq \operatorname{cl} U \subseteq G \cup V$, $0 \in U$, and for each z in $\partial\mathbb{D}$ there is a path in U from 0 to z having length $\le 1 + n\pi$. Let $k_2 > k_1$ such that $\operatorname{cl} U \subseteq G_k \cup V$ for $k \ge k_2$. According to Theorem 14.7.14 there is a constant M such that $|f_k'(z)| \le M\,|f_k'(0)|$ for all $k \ge k_2$ and z in $\operatorname{cl} U$. If $|z| = 1$, let γ be a path in U for 0 to z with length $(\gamma) \le 1 + n\pi$. Thus $1 = |f_k(z)| \le \int_\gamma |f_k'|\,|dw| \le M\,|f_k'(0)|(1 + n\pi)$. Hence $\{f_k'(0)\}$ is bounded below.

Now we will show that $\{f_k'(0)\}$ is bounded above. Let $c_k = [f_k'(0)]^{-1}$ and put $g_k = c_k f_k$. If $\{f_k'(0)\}$ is unbounded, there is a subsequence $\{c_{k_j}\}$ that converges to 0. But according to Lemma 4.6, by passing to a further subsequence if necessary, there is a univalent function $g : G \to \mathbb{C}$ such that $g_{k_j} \to g\,(uc)$. Theorem 4.7 implies that $c_{k_j}\Omega_{k_j} = g_{k_j}(G_{k_j}) \to g(G)$. But the sets Ω_{k_j} are all contained in $\mathbb{D}$ and since $c_{k_j} \to 0, \{c_{k_j}\Omega_{k_j}\}$ can have no kernel, a contradiction. Thus $\{f_k'(0)\}$ must be bounded.

Remember that anything proved about the sequences $\{f_k\}$ or $\{f_k'(0)\}$ is also true about any of their subsequences. Suppose that $f_k'(0) \to a$, a non-zero scalar. Using the notation of the preceding paragraph, there is a subsequence $\{g_{k_j}\}$ and a univalent function g on G such that $g_{k_j} \to g\,(uc)$. Thus $\Omega_{k_j} = f_{k_j}(G_{k_j}) \to a\,g(G)$.

Thus $f \equiv a\,g$ is a conformal equivalence of G onto a region Ω and $\Omega_{k_j} \to \Omega$. Since the outer boundary of Ω is $\partial\mathbb{D}$, Lemma 7.6 implies that Ω is a circularly slit disk.

Now suppose that $\{f_{k_j}\}$ and $\{f_{m_j}\}$ are two subsequences of $\{f_k\}$ such that $f_{k_j} \to f$ and $f_{m_j} \to h$, where f and h are conformal equivalences of G onto circularly slit disks Ω and Λ, respectively, with $f(0) = h(0) = 0$, $f'(0) > 0$, and $h'(0) > 0$. Thus $\phi = f \circ h^{-1}$ is a conformal equivalence of Λ onto Ω with $\phi(0) = 0, \phi'(0) > 0$, and $\phi(\partial\mathbb{D}) = \partial\mathbb{D}$. By Lemma 6.3, $\Lambda = \Omega$ and $\phi(\zeta) = \zeta$ for all ζ in Λ. Thus $h = f$.

To recapitulate, each subsequence of $\{f_k\}$ has a subsequence that converges to a conformal equivalence of G onto a circularly slit disk, and each convergent subsequence of $\{f_k\}$ has the same limit point. This implies that $\{f_k\}$ converges to a conformal equivalence f of G onto the circularly slit disk Ω. Hence $\Omega_k \to \Omega$ by Theorem 4.10. $\square$

7.9 Theorem. *If G is a non-degenerate finitely connected region, A is a component of the extended boundary of G, and $a \in G$, then there is a unique circular region Ω and a unique conformal equivalence $f : G \to \Omega$ such that f associates A with $\partial\mathbb{D}, f(a) = 0$, and $f'(a) > 0$.*

Proof. Once existence is established, uniqueness follows from Proposition 7.5 as follows. Suppose that for $j = 1, 2$, $f_j : G \to \Omega_j$ is a conformal equivalence that associates A with $\partial\mathbb{D}$ such that $f_j(a) = 0$ and $f_j'(a) > 0$. Then $g = f_2 \circ f_1^{-1}$ is a conformal equivalence of Ω_1 onto Ω_2, $g(\partial\mathbb{D}) = \partial\mathbb{D}$, $g(0) = 0$, and $g'(0) > 0$. It follows that $g(\mathbb{D}) = \mathbb{D}$ and so $g(z) = \lambda(z-a)(1-\overline{a}z)^{-1}$. The remaining information about g shows that $g(z) = z$ for all z.

Now for the proof of existence. Let $\mathcal{G}$ be the collection of all circular regions G that are n-connected such that $0 \in G$ and $\partial\mathbb{D}$ is the outer boundary of G. Let $\mathcal{H}$ be the collection of all circularly slit disks Ω that are n-connected such that $\partial\mathbb{D}$ is the outer boundary of Ω. According to Theorem 6.2, for every G in $\mathcal{G}$ there is a unique Ω in $\mathcal{H}$ and a conformal equivalence $f : G \to \Omega$ such that $f(0) = 0, f'(0) > 0$, and f associates $\partial\mathbb{D}$ with $\partial\mathbb{D}$. This defines a map $F : \mathcal{G} \to \mathcal{H}$ by $F(G) = \Omega$ whenever G and Ω are conformally equivalent. To prove the theorem it suffices to show that F is surjective.

We now topologize $\mathcal{G}$ and $\mathcal{H}$. If $G \in \mathcal{G}$, let $C_1, \ldots C_n$ be the circles that form the boundaries of the bounded components of the complement of G. Each circle C_j is determined by its center $z_j = a_j + ib_j$ and its radius r_j. Thus G can be identified with the point in $\mathbb{R}^{3n}$ with coordinates $(a_1, b_1, r_1, \ldots, a_n, b_n, r_n)$; let $\mathcal{G}'$ be the set of such points in $\mathbb{R}^{3n}$ that are so obtained. Note that $\mathcal{G}'$ is a subset of

$$\{(a_1, b_1, r_1, \ldots, a_n, b_n, r_n) : 0 < a_j^2 + b_j^2 < 1 \text{ and } 0 < r_j < 1\}.$$

$\mathcal{G}'$ is a proper subset of this set since we must have that the circles comprising the boundary of G do not intersect. If $G \in \mathcal{G}$, let G' be the corresponding point in $\mathcal{G}'$.

If $\Omega \in \mathcal{H}$, let $\gamma_1, \ldots, \gamma_n$ be the closed arcs that constitute the bounded components of the complement of Ω. Each γ_j is determined by its beginning point $\zeta_j = \alpha_j + i\beta_j$ and its length θ_j as measured in a counterclockwise direction. Thus we also have that each Ω in $\mathcal{H}$ can be identified with a point Ω' in $\mathbb{R}^{3n}$; let $\mathcal{H}' = \{\Omega' : \Omega \in \mathcal{H}\}$. The function $F : \mathcal{G} \to \mathcal{H}$ gives rise to a function $F' : \mathcal{G}' \to \mathcal{H}'$.

Give $\mathcal{G}'$ and $\mathcal{H}'$ their relative topologies from $\mathbb{R}^{3n}$. It is left to the reader to show that a sequence $\{G_k\}$ in $\mathcal{G}$ converges to G in $\mathcal{G}$ (in the sense of Definition 4.1) if and only if $G_k' \to G'$ in $\mathcal{G}'$. Similarly for convergent sequences in $\mathcal{H}$ and $\mathcal{H}'$. We will show that F is surjective by showing that F' is a homeomorphism.

7.10 Claim. Both $\mathcal{G}'$ and $\mathcal{H}'$ are connected open subsets of $\mathbb{R}^{3n}$.

The proof of this claim is left to the readers so that they might more thoroughly familiarize themselves with the notation and the identifications. At this point it seems safe to abandon the distinction between the regions and the corresponding points in $\mathbb{R}^{3n}$ and we do so.

Note that F is injective by the uniqueness statement of Theorem 6.2. Let $G_k \to G$ in $\mathcal{G}$ and put $\Omega_k = F(G_k)$ and $\Omega = F(G)$. According to Lemma 7.8, $\Omega_k \to \Omega$ and so F is continuous. By the Invariance of Domain Theorem, $F : \mathcal{G} \to \mathcal{H}$ is an open map. Suppose F is not surjective; let $\Omega_1 \in \mathcal{H} \setminus F(\mathcal{G})$ and let $\Omega_0 = F(G_0) \in F(\mathcal{G})$. Since $\mathcal{H}$ is an open connected subset of $\mathbb{R}^{3n}$, there is a path $\Omega : [0,1] \to \mathcal{H}$ with $\Omega(0) = \Omega_0$ and $\Omega(1) = \Omega_1$. Since $F(\mathcal{G})$ is open, there is a τ with $0 < \tau \leq 1$ such that $\Omega(\tau) \notin F(\mathcal{G})$ and $\Omega(t) = F(G(t)) \in F(\mathcal{G})$ for $0 \leq t < \tau$. Let $0 < t_k < \tau$ such that $t_k \to \tau$; so $\Omega(t_k) \to \Omega(\tau)$. If $G_k = F^{-1}[\Omega(t_k)]$, then Lemma 7.8 implies that $G_k \to G$, a circular region, and it must be that $F(G) = \Omega(\tau)$, contradicting the fact that $\Omega(\tau) \notin F(\mathcal{G})$. Thus F is surjective, proving the thereom. $\square$

Exercises

1. Give an example of a sequence of circularly slit disks that converges to $\mathbb{D}$ in the sense of (4.1).

2. What happens in Theorem 7.9 if some of the components of the complement of G are trivial?

3. Refer to Exercise 3.3 for the definitions of a proper map. Let G be a non-degenerate n-connected region and let $\text{Aut}(G)$ be the group of all conformal equivalence of G onto itself. Show that if $f : G \to G$ is a proper map, then $f \in \text{Aut}(G)$ (Radó [1922]) (Hint: Take G to be a circular region with outer boundary $\partial\mathbb{D}$. Use the hypothesis that f is proper to show that f extends to $\text{cl}G$. Now extend f by the Schwarz Reflection Principle. Now f defines a permutation of the boundary circles $\gamma_1, \ldots, \gamma_n$ of G. Show that for some integer $m \geq 1$, the m-th iterate of f, f^m, defines the identity permutation. Thus, without loss of generality, we may assume that f defines the identity permutation of the boundary circles. Now use the method of the proof of Proposition 7.5 to extend f to a proper map of $\mathbb{D}$ onto $\mathbb{D}$ and use Exercise 3.4. Also show that f fixes the points of $\mathbb{D} \setminus G(\infty)$.)

4. If G is a non-degenerate n-connected region and $f : G \to G$ is a conformal equivalence such that $f(z) = z$ for three distinct points z in G, then f is the identity.

Chapter 16

Analytic Covering Maps

In this chapter it will be shown that for every region Ω in the plane such that $\mathbb{C} \setminus \Omega$ has at least two points, there is an analytic covering map $\tau : \mathbb{D} \to \Omega$. This is the essential part of what is called the Uniformization Theorem. The reader might want to review §9.7 before going much further. The reader will be assumed to be familiar with some basic topological notions such as the fundamental group and its properties. Some topological facts will be proved (especially in the first section) even though they may seem elementary and assumable to many.

§1 Results for Abstract Covering Spaces

Recall that if Ω is a topological space, a *covering space* of Ω is a pair (G, τ) where G is also a connected topological space and $\tau : G \to \Omega$ is a surjective continuous function with the property that for every ζ in Ω there is a neighborhood Δ of ζ such that each component of $\tau^{-1}(\Delta)$ is mapped by τ homeomorphically onto Δ. Such a neighborhood Δ of ζ is called a *fundamental neighborhood* of ζ.

We will be concerned in this book with covering spaces (G, τ) of regions Ω in $\mathbb{C}$ where G is also a region in $\mathbb{C}$ and τ is an analytic function. Such a covering space will be called an *analytic covering space*. It is not difficult to check that $(\mathbb{C}, \exp)$ and $(\mathbb{C} \setminus \{0\},\ z^n)$, for n a non-zero integer, are both analytic covering spaces of the punctured plane. Of course any homeomorphism yields a covering space and a conformal equivalence gives rise to an analytic covering space. In fact our main concern will be analytic covering spaces (G, τ) of regions Ω in $\mathbb{C}$ where $G = \mathbb{D}$. But for the moment in this section we remain in the abstract situation where G and Ω are metric spaces. In fact the following assumption will remain in force until it is supplanted with an even more restrictive one.

Assumption. *Both G and Ω are arcwise connected and locally arcwise connected metric spaces, (G, τ) is a covering space of Ω, $a_0 \in G$, and $\alpha_0 = \tau(a_0) \in \Omega$.*

Recall that a topological space is said to be locally arcwise connected if for each point in the space and each neighborhood of the point there is a smaller neighborhood that is arcwise connected. A good reference for the general theory of covering spaces is Massey [1967]. Because certain notions will be so frequently used and to fix the notation, we recall a few facts from §9.7.

If $\gamma : [0,1] \to \Omega$ is a path with $\gamma(0) = \alpha_0$, then there is a unique path $\tilde{\gamma} : [0,1] \to G$ with $\tilde{\gamma}(0) = a_0$ and $\tau \circ \tilde{\gamma} = \gamma$ (9.7.5). Such a path $\tilde{\gamma}$ is called an *a_0-lifting* (or *a_0-lift*) of γ. Moreover if $\sigma : [0,1] \to \Omega$ is another path with initial point α_0 and $\tilde{\sigma}$ is its a_0-lifting,then $\tilde{\gamma}$ and $\tilde{\sigma}$ have the same final point in G if γ and σ are fixed end point (FEP) homotopic in Ω (9.7.6). Indeed $\tilde{\gamma}$ and $\tilde{\sigma}$ are FEP homotopic under this hypothesis. A *loop* in Ω is a closed path. If $\gamma : [0,1] \to \Omega$ is a loop with $\gamma(0) = \gamma(1) = \alpha_0$, say that γ is a loop with base point α_0.

We begin with some basic results about "liftable" continuous functions.

1.1 Lemma. *Suppose (G,τ) is a covering space of Ω, X is a locally connected space, $f : X \to \Omega$ is a continuous function, and $T : X \to G$ is a continuous function such that $\tau \circ T = f$. If $x \in X$, Δ is a fundamental neighborhood of $\xi = f(x)$, U is the component of $\tau^{-1}(\Delta)$ containing $z = T(x)$, and W is a connected neighborhood of x such that $f(W) \subseteq \Delta$, then $T|W = (\tau|U)^{-1} \circ (f|W)$.*

Proof. Using the above notation, $T(W)$ is connected and contained in $\tau^{-1}(\Delta)$, and $z \in T(w)$; therefore $T(W) \subseteq U$. Since $f(w) = \tau(T(w))$ for all w in W, the lemma follows. □

1.2 Proposition. *Suppose (G,τ) is a covering space of Ω, X is a connected locally connected space, $f : X \to \Omega$ is a continuous function, and S and T are continuous functions from X into G such that $f = \tau \circ T = \tau \circ S$. If there is a point x_0 in X for which $T(x_0) = S(x_0)$, then $T = S$.*

Proof. Set $Y = \{x \in X : T(x) = S(x)\}$. By hypothesis $Y \neq \emptyset$ and clearly Y is closed. It suffices to show that Y is also open. If $x \in Y$, let $z = T(x) = S(x)$, $\xi = f(x)$, Δ a fundamental neighborhood of ξ, and let U be the component of $\tau^{-1}(\Delta)$ that contains z. If W is a neighborhood of x such that $f(W) \subseteq \Delta$, then the preceding lemma implies that $T|W = (\tau|U)^{-1} \circ (f|W)$ and also $S|W = (\tau|U)^{-1} \circ (f|W)$. Thus $W \subseteq Y$ and Y is open. □

1.3 Theorem. *Suppose (G,τ) is a covering space of Ω, X is a connected locally connected space, and $f : X \to \Omega$ is a continuous function with $f(x_0) = \alpha_0 = \tau(a_0)$. If X is simply connected, then there is a unique continuous function $T : X \to G$ such that $f = \tau \circ T$ and $T(x_0) = a_0$.*

Proof. If $x \in X$, let σ be a path in X from x_0 to x. So $f \circ \sigma$ is a path in Ω from α_0 to $f(x)$. Let $\tilde{\gamma}$ bet the a_0-lift to G. Define $T(x) = \tilde{\gamma}(1)$; it must be shown that $T(x)$ is well defined. So suppose that σ_1 is another path in X from x_0 to x. Since X is simply connected, $\sigma_1 \sim \sigma$ (FEP) in X. Thus $f \circ \sigma_1 \sim f \circ \sigma$ (FEP) in Ω. By the Abstract Monodromy Theorem, the

a_0-lift of $f \circ \sigma_1$ has the same final point as $\tilde{\gamma}$. Therefore the definition of $T(x)$ does not depend on the choice of the curve σ and is well defined.

To prove continuity let $x \in X$ and let $z = T(x)$. Let σ be a path in X from x_0 to x, $\gamma = f \circ \sigma$, and let $\tilde{\gamma}$ be the a_0-lift to G. So $z = \tilde{\gamma}(1)$. Let Δ be a fundamental neighborhood of $\zeta = f(x)$ and let U be the component of $\tau^{-1}(\Delta)$ that contains z. Choose Δ so that it is arcwise connected. Let W be an arcwise connected neighborhood of x in X such that $f(W) \subseteq \Delta$. If w is any point in W, let λ be a path in W from x to w. So $f \circ \lambda$ is a path in Δ from $\zeta = f(x)$ to $f(w)$. Thus the z-lift of $f \circ \lambda$ is $\tilde{\lambda} = (\tau|U)^{-1} \circ f \circ \lambda$. But $\lambda\sigma$ is a path in X from x_0 to w and this leads to the fact that $T(w) = \tilde{\lambda}(1) = (\tau|U)^{-1}(f(w))$. Thus $T|W = (\tau|U)^{-1} \circ f|W$ and T is continuous.

It is easy to check from the definition that $\tau \circ T = f$ and $T(x_0) = a_0$. Uniqueness is a consequence of Proposition 1.2. □

1.4 Definition. If (G_1, τ_1) and (G_2, τ_2) are covering spaces of Ω, a *homomorphism* from G_1 to G_2 is a continuous map $T : G_1 \to G_2$ such that $\tau_2 \circ T = \tau_1$. If T is a homeomorphism as well, then T is called an *isomorphism* between the covering spaces.

If (G, τ) is a covering space of Ω, an *automorphism* of the covering space is a covering space isomorphism of G onto itself. Let $\text{Aut}(G, \tau)$ denote the collection of all automorphisms of (G, τ).

Note that the inverse of an isomorphism is an isomorphism and $\text{Aut}(G, \tau)$ is a group under composition. The next result collects some facts about homomorphisms of covering spaces that are direct consequences of the preceding results.

1.5 Proposition. *Suppose that (G_1, τ_1) and (G_2, τ_2) are covering spaces of Ω with $\tau(a_1) = \tau(a_2) = \alpha_0$.*

(a) *If $T : G_1 \to G_2$ is a homomorphism of the covering spaces, then T is a surjective local homeomorphism.*

(b) *If G_1 is simply connected, there is a unique homomorphism $T : G_1 \to G_2$ with $T(a_1) = a_2$.*

(c) *Any two covering spaces of Ω that are simply connected are isomorphic.*

Proof. (a) Without loss of generality we may asume that $T(a_1) = a_2$. To see that T is surjective, let z_2 be an arbitrary point in G_2 and let $\tilde{\gamma}_2$ be a path in G_2 from a_2 to z_2. Let $\gamma = \tau_2 \circ \tilde{\gamma}_2$ and let $\tilde{\gamma}_1$ be the a_1-lift of γ. Now $T \circ \tilde{\gamma}_1$ is a path in G_2 with initial point a_2 and $\tau_2 \circ T \circ \tilde{\gamma}_1 = \gamma$. By the uniqueness of path lifts, $T \circ \tilde{\gamma}_1 = \tilde{\gamma}_2$. Thus $z_2 = \tilde{\gamma}_2(1) = T(\tilde{\gamma}_1(1))$ and T is surjective. Since τ_1 is a local homeomorphism, the fact that T is a local homeomorphism is immediate from Lemma 1.1.

(b) This follows from Theorem 1.3.

(c) Let $T : G_1 \to G_2$ be a homomorphism with $T(a_1) = a_2$ and let $S : G_2 \to G_1$ be a homomorphism with $S(a_2) = a_1$. Thus $S \circ T$ is a homorphism of G_1 into itself that fixes the point a_1. From Proposition 1.2 it follows that $S \circ T$ is the identity homorphism of G_1 and so T must be a isomorphism (with S as its inverse). □

Thus we say that a simply connected covering space of Ω is the *universal covering space* of Ω. The reason for the word "universal" here is contained in (b) of the preceding proposition. The reason for the use of the word "the" is contained in (c). Of course this uniqueness statement does not imply existence. The existence of a universal covering space for subsets of the plane will be established before the end of this chapter. Existence results for more general spaces can be found in any standard reference.

1.6 Corollary. *If (G, τ) is the universal covering space of Ω and a_1 and a_2 are two points in G with $\tau(a_1) = \tau(a_2)$, then there is a unique T in $\mathrm{Aut}(G, \tau)$ with $T(a_1) = a_2$.*

Proof. Apply Theorem 1.3 and, as in the proof of the preceding proposition, show that the resulting homomorphism is an automorphism. □

1.7 Corollary. *If (G, τ) is the universal covering space of Ω and $z \in G$, then $\tau^{-1}(\tau(z)) = \{T(z) : T \in \mathrm{Aut}(G, \tau)\}$.*

1.8 Theorem. *If (G, τ) is the universal covering space of Ω, then $\mathrm{Aut}(G, \tau)$ is isomorphic to the fundamental group of Ω, $\pi(\Omega)$.*

Proof. Let $T \in \mathrm{Aut}(G, \tau)$ and let $\tilde{\gamma}$ be any path in G from a_0 to $T(a_0)$. Since $T \in \mathrm{Aut}(G, \tau)$, $\gamma = \tau \circ \tilde{\gamma}$ is a loop in Ω with base point α_0. If $\tilde{\sigma}$ is another path in G from a_0 to $T(a_0)$, then the simple connectedness of G implies that $\tilde{\sigma}$ and $\tilde{\gamma}$ are FEP homotopic in G. Thus $\tau \circ \tilde{\sigma}$ and γ are homotopic in Ω. This says that for each T in $\mathrm{Aut}(G, \tau)$ there is a well defined element γ_T in $\pi(\Omega)$. It will be shown that the map $T \to \gamma_T$ is an anti-isomorphism of $\mathrm{Aut}(G, \tau)$ onto $\pi(\Omega)$. (The prefix "anti" is used to denote that the order of multiplication is reversed.) This implies that $T \to \gamma_T^{-1}$ defines an isomorphism between the two groups, proving the theorem.

Let S and T be two automorphisms of the covering space; it will be shown that $\gamma_{ST} = \gamma_T \gamma_S$. To this end let $\tilde{\gamma}$ and $\tilde{\sigma}$ be paths in G from a_0 to $T(a_0)$ and $S(a_0)$, respectively. So $\gamma_T = \tau \circ \tilde{\gamma}$ and $\gamma_S = \tau \circ \tilde{\sigma}$. Now $S \circ \tilde{\gamma}$ is a path in G from $S(a_0)$ to $S(T(a_0))$ so that $(S \circ \tilde{\gamma})\tilde{\sigma}$ is a path in G from a_0 to $S(T(a_0))$. Therefore $\gamma_{ST} = \tau \circ [(S \circ \tilde{\gamma})\tilde{\sigma}] = [\tau \circ (S \circ \tilde{\gamma})]\ [\tau \circ \tilde{\sigma}] = \gamma_T\ \gamma_S$.

To show that the map is surjective, let $\gamma \in \pi(\Omega, \alpha_0)$ and let $\tilde{\gamma}$ be the a_0-lift of γ. According to Corollary 1.6 there is a unique automorphism T such that $T(a_0) = \tilde{\gamma}(1)$. It follows from the definition that $\gamma_T = \gamma$.

Finally let's show that the map is injective. Suppose $T \in \mathrm{Aut}(G, \tau)$ and

$\gamma = \gamma_T \sim 0$ in Ω. Let $\tilde{\gamma}$ be the a_0-lift of $\tilde{\gamma}$; so $T(a_0) = \tilde{\gamma}(1)$. Now $\gamma \sim 0$ implies that γ is homotopic to the constant path α_0. But the Abstract Monodromy Theorem implies that the a_0-lifts of α_0 and γ have the same end point. Since the a_0-lift of the constant path α_0 is the constant path a_0, this says that $T(a_0) = a_0$. By Proposition 1.2 this implies that T is the identity automorphism. □

Exercises

1. Suppose (G, τ) is a covering space of Ω and Δ is a subset of Ω that is both arcwise connected and locally arcwise connected. Show that if H is a component of $\tau^{-1}(\Delta)$, then (H, τ) is a covering space of Δ.
2. If $\tau(t) = e^t$, show that $(\mathbb{R}, \tau)$ is a covering space of $\partial\mathbb{D}$. Find $\text{Aut}(\mathbb{R}, \tau)$.
3. For the covering space $(\mathbb{C}, \exp)$ of $\mathbb{C} \setminus \{0\}$, find $\text{Aut}(\mathbb{C}, \exp)$.
4. For the covering space $(\mathbb{C} \setminus \{0\},\ z^n)$ of $\mathbb{C} \setminus \{0\}$, find $\text{Aut}(\mathbb{C} \setminus \{0\},\ z^n)$.
5. For any z in G, show that $\{T(z) : T \in \text{Aut}(G, \tau)\}$ is a closed discrete subset of G.
6. If $G = \{z : 0 < \text{Re } z < r\}$ and $\tau(z) = e^z$, show that (G, τ) is a covering space of $\text{ann}(0; 1, e^r)$ and find $\text{Aut}(G, \tau)$.
7. If $G = \{z : 0 < \text{Re } z\}$ and $\tau(z) = e^z$, show that (G, τ) is a covering space of $\{z : 1 < |z| < \infty\}$ and find $\text{Aut}(G, \tau)$.
8. Prove that for the universal covering space, the cardinality of $\{T(z) : T \in \text{Aut}(G, \tau)\}$ is independent of the choice of z.

§2 Analytic Covering Spaces

In this section we restrict our attention to analytic covering spaces and derive a few results that are pertinent to this situation. Assume that Ω and G are regions in the plane and $\tau : G \to \Omega$ is an analytic function that is also a covering map.

2.1 Proposition. *If (G, τ) is an analytic covering space of Ω, H is an open subset of the plane, $f : H \to \Omega$ is an analytic function, and $T : H \to G$ is a continuous function such that $\tau \circ T = f$, then T is analytic.*

Proof. This is an immediate consequence of Lemma 1.1. □

2.2 Corollary. *If (G, τ) and (G_2, τ_2) are analytic covering spaces of Ω and $T : G_1 \to G_2$ is a homomorphism, then T is analytic.*

2.3 Corollary. *If (G, τ) is an analytic covering space of Ω, then every function in* $\mathrm{Aut}(G, \tau)$ *is a conformal equivalence of G.*

2.4 Corollary. *If (G_1, τ_1) and (G_2, τ_2) are analytic covering spaces of Ω and $T : G_1 \to G_2$ is an isomorphism, then T is a conformal equivalence.*

2.5 Corollary. *If $\tau : \mathbb{D} \to \Omega$ is an analytic covering map, $\tau(0) = \alpha_0$, and $\tau'(0) > 0$, then τ is unique.*

Proof. Suppose that $\mu : \mathbb{D} \to \Omega$ is another such map. By Proposition 1.5 there is an isomorphism $f : (\mathbb{D}, \tau) \to (\mathbb{D}, \mu)$ of the covering spaces with $f(0) = 0$. By the preceding corollary, f is a conformal equivalence of the disk onto itself and thus must be a Möbius transformation. But $f(0) = 0$ and $f'(0) > 0$; therefore $f(z) = z$ for all z. □

2.6 Corollary. *Suppose (G, τ) is an analytic covering space of Ω, X is a region in the plane, and $f : X \to \Omega$ is an analytic function with $f(x_0) = \alpha_0 = \tau(a_0)$. If X is simply connected, then there is a unique analytic function $T : X \to G$ such that $f = \tau \circ T$ and $T(x_0) = a_0$.*

Proof. Just combine the preceding proposition with Theorem 1.3. □

It will be shown later (4.1) that if Ω is any region in the plane such that its complement in $\mathbb{C}$ has at least two points, then there is an analytic covering map from the unit disk $\mathbb{D}$ onto Ω. The next result establishes that for this to be the case it must be that the complement has at least two points. Recall that $\mathbb{C}_0$ denotes the punctured plane.

2.7 Proposition. *The pair (G, τ) is a universal analytic covering space of $\mathbb{C}_0$ if and only if $G = \mathbb{C}$ and $\tau(z) = \exp(az + b)$ for some pair of complex numbers a and b with $a \neq 0$.*

Proof. If $a, b \in \mathbb{C}$ with $a \neq 0$, then $az + b$ is a Möbius transformation of $\mathbb{C}$ onto itself. It is easy to see that if $\tau(z) = \exp(az + b)$, then $(\mathbb{C}, \tau)$ is a covering map of $\mathbb{C}_0$.

For the converse assume that G is a simply connected region in $\mathbb{C}$ and (G, τ) is a covering space of $\mathbb{C}_0$. We have already seen that $(\mathbb{C}, \exp)$ is a covering space for $\mathbb{C}_0$ so Proposition 1.5 and Corollary 2.4 imply there is a conformal equivalence $h : \mathbb{C} \to G$ such that $\exp z = \tau(h(z))$ for all z in $\mathbb{C}$. But the only way such a conformal equivalence can exist is if $G = \mathbb{C}$. But then Proposition 14.1.1 implies that $h(z) = az + b$ for complex numbers a and b with $a \neq 0$. □

2.8 Example. Let Ω be the annulus $\{z : 1 < |z| < \rho\}$, where $\rho = e^{\pi}$. If

$$\tau(z) = \exp\left\{ i \log\left(\frac{1+z}{1-z}\right) + \frac{\pi}{2} \right\},$$

then $(\mathbb{D}, \tau)$ is the universal analytic covering space for Ω. The maps in the group $\text{Aut}(\mathbb{D}, \tau)$ are the Möbius transformations

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

where

$$\beta_n = \frac{e^{2\pi n} - e^{-2\pi n}}{e^{2\pi n} + e^{-2\pi n}} = \tanh(2\pi n).$$

To see this first observe that the map τ above can be expressed as a composition of two maps: the first is a conformal equivalence of $\mathbb{D}$ onto a vertical strip; the second is the exponential map, which wraps the strip around the annulus an infinite number of times. To show that the automorphisms of this covering space have the requisite form uses some algebra and the following observation. (Another verification of the statements in this example can be obtained by using Exercise 1.6 and the form of the conformal equivalence of $\mathbb{D}$ onto the relevant vertical strip.)

If $(\mathbb{D}, \tau)$ is the universal analytic covering space for Ω, then every T in $\text{Aut}(\mathbb{D}, \tau)$ is necessarily a conformal equivalence of $\mathbb{D}$ onto itself. Hence T is a Schwarz map,

$$T(z) = e^{i\theta} \frac{z - \beta}{1 - \overline{\beta} z},$$

for some choice of θ and β, $|\beta| < 1$. For convenience, whenever $(\mathbb{D}, \tau)$ is the universal analytic covering space for a region Ω in $\mathbb{C}$, we will let

$$\mathcal{G}_\tau = \text{Aut}(\mathbb{D}, \tau).$$

By Theorem 1.8, $\mathcal{G}_\tau \approx \pi(\Omega)$. It is known that for regions Ω in $\mathbb{C}$, $\pi(\Omega)$ is a free group (see Exercise 3). If $\mathbb{C}_\infty \setminus \Omega$ has $n+1$ components, then $\pi(\Omega)$ is the free group with n generators. If $\mathbb{C}_\infty \setminus \Omega$ has an infinite number of components, then $\pi(\Omega)$ is the free group on a countable number of generators.

Exercises

1. For $|z| < 1$, define $\tau(z) = \exp[z(1 - |z|)^{-1}]$ and show that $(\mathbb{D}, \tau)$ is a covering space of $\mathbb{C} \setminus \{0\}$.

2. Show that for an n-connected region Ω, $\pi(\Omega)$ is a free group on n generators.

3. Let K be a compact subset of $\mathbb{C}$ and put $\Omega = \mathbb{C} \setminus K$. If $\gamma_1, \ldots, \gamma_n$ are paths in Ω, let $\delta > 0$ such that $\text{dist}(\gamma_j, K) > \delta$ for $1 \le j \le n$. Let $\Omega(\delta) = \{\omega \in \Omega : \text{dist}(\omega, K) > \delta\}$. Show that $\Omega(\delta)$ is finitely connected and contains the paths $\gamma_1, \ldots, \gamma_n$. Show that the fundamental group of Ω is a countably generated free group.

4. Let $\Omega = \{z : 1 < |z|\}$ and let $\tau(z) = \exp\left[\frac{z+1}{z-1}\right]$. Explicitly determine $\text{Aut}(\mathbb{D}, \tau)$.

§3 The Modular Function

Here we examine a special analytic function called the modular function. This is a special analytic covering map from the upper half plane onto the plane with the points 0 and 1 deleted.

3.1 Definition. A *modular transformation* is a Möbius transformation

3.2
$$M(z) = \frac{az+b}{cz+d}$$

such that the coefficients a, b, c, d are integers and $ad - bc = 1$. The set of all modular transformations is called the *modular group* and is dentoed by $\mathcal{M}$.

The designation of $\mathcal{M}$ as a group is justified by the following proposition. The proof is left as an exercise.

3.3 Proposition. *$\mathcal{M}$ is a group under composition with identity the identity transformation $I(z) = z$. If $M(z)$ is given by (3.2), then*

$$M^{-1}(z) = \frac{dz-b}{-cz+a}.$$

Let $\mathbb{H}$ denote the upper half plane, $\{z : \text{Im } z > 0\}$.

3.4 Proposition. *If $M \in \mathcal{M}$, then $M(\mathbb{R}_\infty) = \mathbb{R}_\infty$ and $M(\mathbb{H}) = \mathbb{H}$.*

Proof. The first equality follows because the coefficients of M are real numbers. Thus the Orientation Principle implies that $M(\mathbb{H})$ is either the upper or lower half plane. Using the fact that $M \in \mathcal{M}$ and consequently has determinant 1, it follows that

3.5
$$\text{Im } M(z) = \frac{1}{|cz+d|^2}\text{Im } z.$$

This concludes the proof. □

Let $\mathcal{G}$ denote the subgroup of $\mathcal{M}$ generated by the modular transformations

3.6
$$S(z) = \frac{z}{2z+1} \qquad \text{and} \qquad T(z) = z + 2.$$

Let

3.7
$$G = \{z \in \mathbb{H} : -1 \leq \text{Re } z < 1,\ |2z+1| \geq 1,\ |2z-1| > 1\}.$$

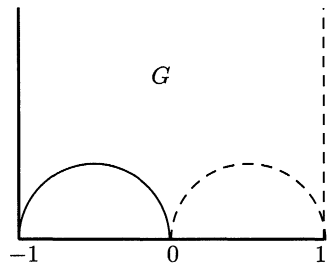

Figure 16.1.

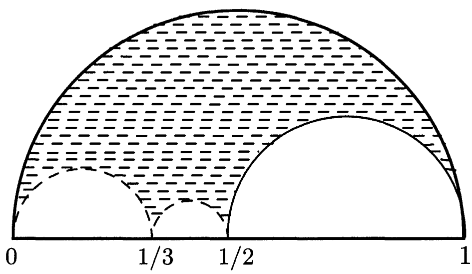

Figure 16.2.

This set G is illustrated in Figure 16.1. Note that $G \cap \mathbb{R} = \emptyset$. The reason for defining G in this way and excluding portions of its boundary while including other parts of ∂G will become clear as we proceed. For now, gentle reader, please accept the definition of G as it is.

3.8 Example. If S is as in (3.6) and G is as above, then $S(G) = \{\zeta : \text{Im } \zeta > 0,\ |2\zeta - 1| \le 1,\ |4\zeta - 3| \ge 1,\ |6\zeta - 1| > 1, \text{and } |12\zeta - 5| > 1\}$. Consequently, $S(G) \cap G = \emptyset$.

The region described as $S(G)$ is depicted in Figure 16.2. To see that the assertation of this example is true, first let L_- and L_+ be the rays $\{z : \text{ Im } z > 0 \text{ and Re } z = \pm 1\}$ and let C_- and C_+ be the half circles $\{z : \text{Im } z > 0 \text{ and } |2z \mp 1| = 1\}$. Observe that $S(L_\pm)$ and $S(C_\pm)$ must be circles that are perpendicular to the real line. Since $S(0) = 0$, $S(-1) = 1$, $S(1) = 1/3$, and $S(\infty) = 1/2$, applications of the Orientation Principle show that $S(G)$ has the desired form.

3.9 Lemma. *Let G and $\mathcal{G}$ be as in (3.6) and (3.7).*

(a) *If M_1 and $M_2 \in \mathcal{G}$ and $M_1 \neq M_2$, then $M_1(G) \cap M_2(G) = \emptyset$.*

(b) $\mathbb{H} = \bigcup\{M(G) : M \in \mathcal{G}\}$.

(c) *$\mathcal{G}$ consists of all modular transformations M having the form (3.2) such that the coefficients a and d are odd and b and c are even integers.*

Proof. Let $\mathcal{G}_1$ be the collection of all modular transformations described in part (c) and note that if S and T are defined as in (3.6), then they belong to $\mathcal{G}_1$. The reader can verify directly that $\mathcal{G}_1$ is a group under composition, and so it follows that $\mathcal{G} \subseteq \mathcal{G}_1$.

3.10 Claim. If $M_1,\ M_2 \in \mathcal{G}_1$ and $M_1 \neq M_2$, then $M_1(G) \cap M_2(G) = \emptyset$.

Because $\mathcal{G}_1$ is a group, to prove this claim it suffices to show that if $M \in \mathcal{G}_1$ and $M \neq I$, then $M(G) \cap G = \emptyset$. This will be done by considering two possible cases. The first case is that $c = 0$ in (3.2); so $M(z) = (az+b)/d$, where a and d are odd integers and b is even. Since $1 = ad - bc = ad$, $a = d = \pm 1$. Thus $M(z) = z + 2n$, with n in $\mathbb{Z}$ and $n \neq 0$ since $M \neq I$. It is now clear that $M(G) \cap G = \emptyset$.

Now assume that M has the form (3.2), $M \in \mathcal{G}_1$, and $c \neq 0$. Notice that the closed disk $\overline{B}(-1/2; 1/2)$ meets G without containing in its interior any of the points -1, 0, or 1, while any other closed disk whose center lies on the real axis and that meets G must have one of these points in its interior. This leads us to conclude that if $M(z) \neq S(z)+b$, where S is the transformation in (3.6) and b is an even integer, then $|cz + d| > 1$ for all z in G. Indeed, if there is a point z in G with $|cz + d| \leq 1$, then $\overline{B}(-d/c;\ 1/|c|) \cap G \neq \emptyset$. For the moment, assume this closed disk is not the disk $\overline{B}(-1/2; 1/2)$. As observed, this implies that 0, $+1$, or $-1 \in B(-d/c;\ 1/|c|)$; let k be this integer. So $|k + d/c| < 1/|c|$ and hence $|kc + d| < 1$. But c is even and d is odd, so that $kc + d$ is odd, and this furnishes a contradiction. Thus

3.11 $$|cz + d| > 1 \text{ for all } z \text{ in } G,$$

provided $-d/c \neq -1/2$ or $1/|c| \neq 1/2$. On the other hand, if $-d/c = -1/2$ and $1/|c| = 1/2$, then $c = \pm 2$ and $d = \pm 1$. All the entries in a Möbius transformation can be multiplied by a constant without changing the transformation, so we can assume that $c = 2$ and $d = 1$. But the condition that the determinant of M is 1 implies that $a - 2b = 1$, so $a = 1 + 2b$. Thus

$$\begin{aligned} M(z) &= \frac{az+b}{2z+1} \\ &= \frac{z+2bz+b}{2z+1} \\ &= S(z) + b. \end{aligned}$$

So (3.11) holds whenever M is not the transformation $S(z) + b$.

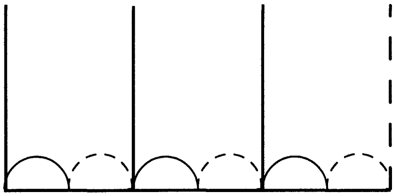

Figure 16.3.

Note that (3.5) implies that if $M \neq S + b$ for S as in (3.6) and b is even, then

$$\text{Im } M(z) < \text{Im } z \text{ for all } z \text{ in } G.$$

The definition of G and (3.5) show that when M equals $S(z) + b$, we still have that

$$\text{Im } M(z) \leq \text{Im } z \text{ for all } z \text{ in } G.$$

Now let M be an arbitrary element of the group $\mathcal{G}_1$. It is left to the reader to show that either M or M^{-1} is not of the form $S + b$, with S as in (3.6) and b an even integer. Thus either Im $M(z) <$ Im z for all z in G or Im $M^{-1}(z) <$ Im z for all z in G; assume for the moment that the former is the case. If there is a z in $G \cap M(G)$, then Im z = Im $M^{-1}M(z) <$ Im $M(z) <$ Im z, a contradiction. Thus $G \cap M(G) = \emptyset$. If $M^{-1}(z) <$ Im z for all z in G, a similar argument also shows that $G \cap M(G) = \emptyset$. This establishes Claim 3.10.

Now let $L = \cup\{M(G) : M \in \mathcal{G}\}$, a subset of $\mathbb{H}$. If T is as in (3.6), then $T^n(z) = z + 2n$. So for each n in $\mathbb{Z}$, L contains $T^n(G)$, which is the translate of G by $2n$. As discovered in Example 3.8, S maps the circle $|2z + 1| = 1$ onto the circle $|2z - 1| = 1$. Combining these last two facts (and looking at Figure 16.3) we get that

3.12 L contains every z in $\mathbb{H}$ that satisfies $|2z - k| \geq 1$ for all odd integers k.

Fix a ζ in $\mathbb{H}$. Because $\{c\zeta + d : \ c, \ d \in \mathbb{Z}\}$ has no limit point in the plane, there is an element of this set having minimum modulus. Thus there is a transformation $M_0(z) = (a_0z + b_0)/(c_0z + d_0)$ in $\mathcal{G}$ such that $|c_0\zeta + d_0| \leq |c\zeta + d|$ for all $M(z) = (az + b)/(cz + d)$ in G. By virtue of (3.5) we get that Im $M_0(\zeta) \geq$ Im $M(\zeta)$ for all M in $\mathcal{G}$. Putting $z = M_0(\zeta)$ and realizing that $MM_0 \in \mathcal{G}$ whenever $M \in \mathcal{G}$, this shows that

3.13 Im $z \geq$ Im $M(z)$ for $z = M_0(\zeta)$ and for all M in $\mathcal{G}$.

Let $n \in \mathbb{Z}$ and continue to have $z = M_0(\zeta)$ for a fixed ζ in $\mathbb{H}$. Applying (3.13) to $M = ST^{-n}$ and using (3.5) as well as a little algebra shows that

$$\text{Im } z \geq \text{Im } ST^{-n}(z) = \frac{\text{Im } z}{|2z - 4n + 1|^2}.$$

Now apply (3.13) to $M = S^{-1}T^{-n}$ and perform similar calculations to get that

$$\text{Im } z \geq \frac{\text{Im } z}{|-2z + 4n + 1|^2}.$$

But Im $z > 0$ so these inequalities become

$$|2z - 4n + 1| \geq 1 \text{ and } |2z - 4n - 1| \geq 1 \text{ for all } n \text{ in } \mathbb{Z}.$$

But $\{4n - 1,\ 4n + 1 : n \in \mathbb{Z}\}$ is the collection of all odd integers, so (3.12) implies that $z = M_0(\zeta) \in L$. Thus $\zeta = M_0^{-1}(z) \in M_0^{-1}(L) = L$. Since ζ was arbitrary, this proves (b).

Remember that we have already proved that $\mathcal{G} \subseteq \mathcal{G}_1$. Let $M_1 \in \mathcal{G}_1$. By part (b) there is a transformation M in $\mathcal{G}$ such that $M(G) \cap M_1(G) \neq \emptyset$. But both M and M_1 are in $\mathcal{G}_1$, so Claim 3.10 implies that $M_1 = M \in \mathcal{G}$. This proves (c). By (3.10), part (a) also holds. □

Now the stage is set for the principal result of this section. It is convenient to let $\mathbb{C}_{0,1} \equiv \mathbb{C} \setminus \{0, 1\}$. The reader might want to carry out Exercise 1 simultaneously with this proof.

3.14 Theorem. *If G and $\mathcal{G}$ are as in (3.6) and (3.7), then there is an analytic function $\lambda : \mathbb{H} \to \mathbb{C}$ having the following properties.*

(a) *$\lambda \circ M = \lambda$ for every M in $\mathcal{G}$.*

(b) *λ is univalent on G.*

(c) *$\lambda(\mathbb{H}) = \mathbb{C}_{0,1}$.*

(d) *λ is not analytic on any region that properly contains $\mathbb{H}$.*

(e) *$(\mathbb{H}, \lambda)$ is a covering space of $\mathbb{C}_{0,1}$.*

Proof. Let

$$G_0 = \{z : \text{Im } z > 0,\ 0 < \text{Re } z < 1, \text{ and } |2z - 1| > 1\};$$

so $G_0 \subseteq G$. Let $f_0 : G_0 \to \mathbb{H}$ be any conformal equivalence and extend f_0 to a homeomorphism of $\text{cl}_\infty G_0$ onto $\text{cl}_\infty \mathbb{H}$ (14.5.7). Let A be a Möbius transformation that maps $f_0(0)$ to $0, f_0(1)$ to 1, $f_0(\infty)$ to ∞, and takes $\mathbb{H}$ onto itself. Hence $f = A \circ f_0$ is a homeomorphism of $\text{cl}_\infty G_0$ onto $\text{cl}_\infty \mathbb{H}$, a conformal equivalence of G_0 onto $\mathbb{H}$, and fixes the points 0, 1, and ∞.

Since f is real-valued on ∂G_0, f can be extended to int G by reflecting across Re $z = 0$. This extended version of f satisfies $f(x+iy) = f(-x+iy)$ for $x+iy$ in G_0. (Note that f is also real-valued on the portions of $|2z-1| = 1$ and $|2z+1| = 1$ that lie in cl G, so that by successively reflecting in these circles and the circles and vertical lines of the various reflected images of G it is possible to extend f to all of $\mathbb{H}$. The argument that follows does just this but with a little more finesse and accuracy by using the action of the group $\mathcal{G}$.)

Define the following sets:

$$\begin{aligned} L_1 &= \{z : \text{Im } z \geq 0, \ \text{Re } z = 0\} \cup \{\infty\}; \\ L_2 &= \{z : \text{Im } z \geq 0, \ |2z - 1| = 1\}; \\ L_3 &= \{z : \text{Im } z \geq 0, \ \text{Re } z = 1\} \cup \{\infty\}. \end{aligned}$$

So $L_1 \cup L_2 \cup L_3 = \partial_\infty G_0$. For $j = 1, 2, 3, L_j$ is connected and so the same holds for $f(L_j)$. By an orientation argument (supply the details)

$$\begin{aligned} f(L_1) &= \{z : z = \text{Re } z \leq 0\} \cup \{\infty\}; \\ f(L_2) &= \{z : z = \text{Re } z \text{ an } 0 \leq z \leq 1\}; \\ f(L_3) &= \{z : z = \text{Re } z \geq 1\} \cup \{\infty\}. \end{aligned}$$

Thus

$$\begin{aligned} f(\text{int } G) &= \mathbb{C} \setminus [0, \infty); \\ f(G) &= \mathbb{C}_{0,1} \equiv \Omega. \end{aligned}$$

Extend f to a function $\lambda : \mathbb{H} \to \mathbb{C}$ by letting

3.15 $$\lambda(z) = f(M^{-1}(z))$$

whenever $M \in \mathcal{G}$ and $z \in M(G)$. According to Lemma 3.9 this function λ is well defined.

Why is λ analytic? Observe that if S and T are defined as in (3.6), then λ is analytic on $V \equiv \text{int}[G \cup T^{-1}(G) \cup S^{-1}(G)]$ and V is an open set that contains G. Thus for every M in $\mathcal{G}$, λ is analytic on a neighborhood of $M(G)$. Thus λ is analytic on all of $\mathbb{H}$.

Clearly condition (a) of the theorem holds because of the definition of λ. Because f is defined on G by reflecting a conformal equivalence, it is one-to-one on G, so (b) holds. Since $f(G) = \Omega$, (c) is also true. Part (d) is a consequence of the following.

3.16 Claim. $\{M(0) : M \in \mathcal{G}\}$ is dense in $\mathbb{R}$.

In fact if this claim is established and λ has an analytic continuation to a region Λ that contains $\mathbb{H}$, then Λ must contain a non-trivial open interval (a, b) of $\mathbb{R}$. But (3.16) implies that for every α in (a, b) there is a sequence $\{M_k\}$ in $\mathcal{G}$ such that $M_k(0) \to \alpha$. Let $y \to 0$ through positive values. Then $\lambda(M_k(0)) = \lim_{y \to 0} \lambda(M_k(iy)) = \lim_{y \to 0} \lambda(iy) = \lambda(0) = 0$. Thus α is an accumulation point of zeros of λ, a contradiction.

To prove Claim 3.16 let $M \in \mathcal{G}$ and suppose M is given by (3.2). So $M(0) = b/d$. Now b and c are even integers, a and d are odd, and $ad - bc = 1$. It suffices to show that for every even integer b and odd integer d such that b and d have no common divisor, there is an odd integer d such that b and d have no common divisor and there is an odd integer a and an even integer c with $ad - bc = 1$. Equivalently, show that given integers m and n such that $2m + 1$ and $2n$ have no common divisor, there are integers p and q such that

$$\begin{aligned} 1 &= (2p+1)(2m+1) - (2n)(2q) \\ &= 4pm + 2m + 2p + 1 - 4nq. \end{aligned}$$

This happens if and only if p and q can be found so that $-m = p(2m + 1) - (2n)q$. But $\{p(2m+1) - (2n)q : p,\ q \in \mathbb{Z}\}$ is the ideal in the ring $\mathbb{Z}$ generated by $2m + 1$ and $2n$. Since these two integers have no common divisors, this ideal is all of $\mathbb{Z}$.

It remains to prove (e). If $\zeta \in \mathbb{C} \setminus [0, \infty)$ and $\delta > 0$ is chosen sufficiently small that $B = B(\zeta; \delta) \subseteq \mathbb{C} \setminus [0, \infty)$, let $U_0 = f^{-1}(B) \subseteq \operatorname{int} G$. It is easy to verify that $\lambda^{-1}(B) = \cup\{M(U_0) : M \in \mathcal{G}\}$ and so $\{M(U_0) : M \in \mathcal{G}\}$ are the components of $\lambda^{-1}(B)$. Now assume that $\zeta = t \in (0, 1)$ and choose $\delta > 0$ sufficiently small that $B = B(t; \delta) \subseteq \mathbb{C}_{0,1}$. An examination of the definition of f as the reflection of a homeomorphism of $\mathrm{cl}_\infty G_0$ onto $\mathrm{cl}_\infty \mathbb{H}$, shows that $f^{-1}(t) = \{z_+, z_-\} \subseteq \partial G$, where $|2z_\pm \mp 1| = 1$. Also $f^{-1}(B)$ consists of two components, U_+ and U_-, where $z_\pm \in U_\pm$. Thus $f(U_\pm) = B \cap \mathrm{cl}\,(\pm\mathbb{H})$. If S is the Möbius transformation defined in (3.6), then it can be verified that S maps the circle $|2z + 1| = 1$ onto the circle $|2z - 1| = 1$. (The reader has probably already done this in Example 3.8.) Since $t = f(z_\pm) = \lambda(z_\pm) = \lambda(S(z_\pm))$, it follows that $S(z_-) = z_+$. Hence $U_0 \equiv U_+ \cup S(U_-)$ is a neighborhood of z_+ and $\lambda(U_0) = f(U_+) \cup \lambda(S(U_-)) = f(U_+) \cup f(U_-) = B$. Therefore $\{M(U_0) : M \in \mathcal{G}\}$ are the components of $\lambda^{-1}(B)$ and clearly λ maps each of these homeomorphically onto B. The final case for consideration, where $\zeta = t \in (1, \infty)$, is similar to the preceding one and is left to the reader. □

3.17 Example. If $\tau(z) = \lambda(i(1 - z)/(1 + z))$, then $(\mathbb{D}, \tau)$ is an analytic covering space of $\mathbb{C}_{0,1}$.

3.18 Definition. The function λ obtained in Theorem 3.14 is called the *modular function.*

Calling λ the modular function is somewhat misleading and the reader should be aware of this when perusing the literature. First λ is not unique, as can be seen from the proof, since its definition is based on taking a conformal equivalence f_0 of G_0 onto $\mathbb{H}$. Given f_0, λ is unique. It is possible to so construct the function λ that $\lambda(0) = 0$, $\lambda(1) = \infty$, and $\lambda(\infty) = 0$.

Having done this the function is unique and is the classical modular function of complex analysis.

Remarks.

1. The material at the beginning of this section has connections with the study of the group $SL_2(\mathbb{R})$, modular forms, and number theory. The interested reader should see Lang [1985]. Also see Ford [1972] for the classical theory.

2. A nice reference for the Uniformization Theorem for Riemann surfaces is Abikoff [1981]. Also see Ahlfors [1973].

Exercises

1. This exercise constructs the exponential function by a process similar to that used to construct the modular function and is meant to help the student feel more comfortable with the proof of Theorem 3.14. (Thanks to David Minda for the suggestion.) Let G_0 be the strip $\{z : 0 < \text{Im } z < \pi\}$ and let ϕ_0 be the conformal equivalence of G_0 onto $\mathbb{H}$ with $\phi_0(-\infty) = 0$, $\phi_0(+\infty) = \infty$, and $\phi_0(0) = 1$. Now extend ϕ_0 to an analytic function ϕ defined on $G = \{z : -\pi < \text{Im } z < \pi\}$ by reflection. If $T_n(z) = z + 2\pi\, in$ for n in $\mathbb{Z}$, extend ϕ to a function E on $\mathbb{C}$ by letting $E(z) = \phi(T_n(z))$ for an appropriate choice of n. Show that E is a well defined entire function that is a covering map of $\mathbb{C}_0$ and prove that E is the exponential function.

2. If λ is the modular function, what is $\text{Aut}(\mathbb{H}, \lambda)$? (By Theorem 1.8, $Aut(\mathbb{H}, \lambda) \approx \pi(\mathbb{C}_{0,1})$, which is a free group on two generators. So the question asked here can be answered by finding the two generators of $\text{Aut}(\mathbb{H}, \lambda)$.)

3. Let $\{a, b\}$ be any two points in $\mathbb{C}$ and find a formula for an analytic covering $(\mathbb{D}, \tau)$ of $\mathbb{C} \setminus \{a, b\}$. If $\alpha_0 \in \mathbb{C} \setminus \{a, b\}$, show that τ can be chosen with $\tau(0) = \alpha_0$ and $\tau'(0) > 0$.

§4 Applications of the Modular Function

In this section the Picard theorems are proved as applications of the modular function and the material on analytic covering spaces. Proofs of these results have already been seen in (12.2.3) and (12.4.2), where the proofs were "elementary."

4.1 Little Picard Theorem. *If f is an entire function that omits two values, then f is constant.*

Proof. Suppose f is an entire function and there are two complex numbers a and b such that $a, b \notin f(\mathbb{C})$. By composing with a Möbius transformation, we may assume that $f(\mathbb{C}) \subseteq \mathbb{C}_{0,1}$. Let $\tau : \mathbb{D} \to \mathbb{C}_{0,1}$ be a universal covering map. According to Proposition 2.1, there is an analytic function $T : \mathbb{C} \to \mathbb{D}$ such that $\tau \circ T = f$. But the T is a bounded entire function and hence constant. Thus f is constant. $\square$

As we know from §14.4, the key to the proof of the Great Picard Theorem is to prove the Montel-Carathéodory Theorem.

4.2 Montel-Carathéodory Theorem. *If X is any region in the plane and $\mathcal{F} = \{f : f$ is an analytic function on X with $f(X) \subseteq \mathbb{C}_{0,1}\}$, then $\mathcal{F}$ is a normal family in $C(X, \mathbb{C}_\infty)$.*

Proof. To prove that $\mathcal{F}$ is normal, it suffices to show that for every disk $B = B(z_0; R)$ contained in X, $\mathcal{F}|\mathcal{B}$ is normal in $C(B, \mathbb{C}_\infty)$. To prove this, it suffices to show that for any sequence $\{f_n\}$ in $\mathcal{F}$, there is either a subsequence that is uniformly bounded on compact subsets of B or a subsequence that converges to ∞ uniformly on compact subsets of B (why?). So fix such a sequence $\{f_n\}$. By passing to a subsequence if necessary, we may assume that there is a point α in $\mathbb{C}_\infty$ such that $f_n(z_0) \to \alpha$. We consider some cases.

Claim. If $\alpha \in \mathbb{C}$ and $\alpha \neq 0, 1$, then $\{f_n\}$ has a subsequence that is uniformly bounded on compact subsets of B.

Let $\tau : \mathbb{D} \to \mathbb{C}_{0,1}$ be a universal analytic covering map and let $\Delta = B(\alpha; \rho)$ be a fundamental neighborhood of α. Fix a component U of $\tau^{-1}(\Delta)$. Since $f_n(z_0) \to \alpha$, we may assume that $f_n(z_0) \in \Delta$ for all $n \geq 1$. According to Theorem 1.3 and Proposition 2.1, for each $n \geq 1$ there is an analytic function $T_n : B \to \mathbb{D}$ such that $T_n(z_0) \in U$ and $\tau \circ T_n = f_n$ on B. But $\{T_n\}$ is a uniformly bounded sequence of analytic functions and so there is a subsequence $\{T_{n_k}\}$ that converges in $H(B)$ to an analytic function T. Clearly $|T(z)| \leq 1$ for all z in B. If there is a z in B such that $|T(z)| = 1$, then T is constantly equal to the number λ with $|\lambda| = 1$. In particular, $T_{n_k}(z_0) \to \lambda$. But $(\tau|U)^{-1}(\alpha) = \lim_k (\tau|U)^{-1}(f_{n_k}(z_0)) = \lim_k T_{n_k}(z_0) = \lambda$, a contradiction. Thus it must be that $|T(z)| < 1$ for all w in B.

Let K be an arbitrary compact subset of B. From the discussion just concluded, there is a number r such that $M = \max\{|T(z)| : z \in K\} < r < 1$. Let k_0 be an integer such that $|T_{n_k}(z) - T(z)| < r - M$ for all z in K and $k \geq k_0$. Hence $|T_{n_k}(z)| \leq r$ for all z in K and $k \geq k_0$. But τ is bounded on $B(0; r)$. It follows that $\{f_{n_k}\} = \{\tau \circ T_{n_k}\}$ is uniformly bounded on K. Since K was arbitrary, this proves the claim.

Now assume that $\alpha = 1$. Since each function f_n never vanishes, there is an analytic function $g_n : B \to \mathbb{C}$ with $g_n^2 = f_n$. Since $f_n(z_0) \to 1$, we can choose the branch of the square root so that $g_n(z_0) \to -1$. But once again, $g_n(B) \subseteq \mathbb{C}_{0,1}$. So the claim implies there is a subsequence $\{g_{n_k}\}$ that is uniformly bounded on compact subsets of B. Clearly this implies that $\{f_{n_k}\}$ is uniformly bounded on compact subsets of B.

Assume that $\alpha = 0$. Here let $g_n = 1 - f_n$. So $g_n(z_0) \to 1$ and g_n never assumes the values 0 and 1. The preceding case, when applied to $\{g_n\}$, shows that $\{f_n\}$ has a subsequence that is uniformly bounded on compact subsets of B.

Finally assume that $\alpha = \infty$. Let $g_n = 1/f_n$. Again g_n is analytic, $g_n(B) \subseteq \mathbb{C}_{0,1}$, and $g_n(z_0) \to 0$. Therefore the preceding case implies there is a subsequence $\{g_{n_k}\}$ that converges uniformly on compact subsets of B to an analytic function g. But the functions $\{g_{n_k}\}$ have no zeros while $g(z_0) = 0$. By Hurwitz's Theorem (7.2.5), $g \equiv 0$. It now follows that $f_{n_k}(z) \to \infty$ uniformly on compact subsets of B. □

4.3 The Great Picard Theorem. *If f has an essential singularity at $z = a$, then there is a complex number α such that for $\zeta \neq \alpha$ the equation $f(z) = \zeta$ has an infinite number of solutions in any punctured neighborhood of a.*

For a proof the reader can consult Theorem 7.4.2.

Exercises

1. Prove Landau's Theorem: If a_0 and a_1 are complex numbers with $a_0 \neq 0, 1$ and $a_1 \neq 1$, then there is a constant $L(a_0, a_1)$ such that whenever $r > 0$ and there is an analytic function f on $r\mathbb{D}$ with $f(0) = a_0$, $f'(0) = a_1$, and f omits the values 0 and 1,then $r \leq L(a_0, a_1)$. (Hint: Let $\tau : \mathbb{D} \to \mathbb{C}_{0,1}$ with $\tau(0) = a_0$ and $\tau'(0) > 0$. Use Corollary 2.6 to find an analytic function $T : r\mathbb{D} \to \mathbb{D}$ such that $\tau \circ T = f$ and $T(0) = 0$. Now apply Schwarz's Lemma.)

2. Prove Schottky's Theorem: For α in $\mathbb{C}_{0,1}$ and $0 < \beta < 1$, there is a constant $C(\alpha, \beta)$ such that if f is an analytic function on $\mathbb{D}$ that omits the values 0 and 1 and $f(0) = \alpha$, then $|f(z)| \leq C(\alpha, \beta)$ for $|z| \leq \beta$. (See 12.3.3.)

§5 The Existence of the Universal Analytic Covering Map

The purpose of this section is to prove the following theorem.

5.1 Theorem. *If Ω is any region in $\mathbb{C}$ whose complement in $\mathbb{C}$ has at* *least two points and* $\alpha_0 \in \Omega$, *then there is a unique analytic covering map* $\tau : \mathbb{D} \to \Omega$ *with* $\tau(0) = \alpha_0$ *and* $\tau'(0) > 0$.

The proof requires some preliminary material. The first lemma is a restatement of Exercise 3.3.

5.2 Lemma. *If* a, b, *and* $\alpha_0 \in \mathbb{C}$ *with* $\alpha_0 \neq a, b$, *then there is an analytic covering map* $\tau : \mathbb{D} \to \mathbb{C} \setminus \{a, b\}$ *with* $\tau(0) = \alpha_0$ *and* $\tau'(0) > 0$.

The next lemma is technical but it contains the vital construction needed in the proof.

5.3 Lemma. *Suppose* Ω *is a region in* $\mathbb{C}$ *and* (G, τ) *is an analytic covering of* Ω *with* $\tau(a) = \alpha_0$ *for some point* a *in* G. *If* $\Delta_0 = B(\alpha_0; r_0)$ *is a fundamental neighborhood of* α_0 *and* g_0 *is the local inverse of* τ *defined on* Δ_0 *such that* $g_0(\alpha_0) = a$, *then the following statements hold:*

(a) $g_0'(\alpha_0) > 0$ *if* $\tau'(a) > 0$.

(b) *If* $\gamma : [0, 1] \to \Omega$ *is a path with* $\gamma(0) = \alpha_0$, *then there is an analytic continuation* $\{(g_t, \Delta_t) : 0 \le t \le 1\}$ *of* (g_0, Δ_0) *such that* $g_t(\Delta_t) \subseteq G$ *for all* t.

(c) *If* (g_1, Δ_1) *and* (g_2, Δ_2) *are obtained from* (g_0, Δ_0) *by analytic continuation and* $g_1(\omega_1) = g_2(\omega_2)$ *for some points* ω_1 *in* Δ_1 *and* ω_2 *in* Δ_2, *then* $\omega_1 = \omega_2$ *and* $g_1(\omega) = g_2(\omega)$ *for all* ω *in* $\Delta_1 \cap \Delta_2$.

(d) *If* (g, Δ) *is obtained by analytic continuation of* (g_0, Δ_0), *then* $\tau(g(\omega)) = \omega$ *for all* ω *in* Δ.

Proof. Part (a) is trivial. To prove (b), use Lebesgue's Covering Lemma to find a $\delta > 0$ with $\delta < r_0$ and such that for each t, $0 \le t \le 1$, $B(\gamma(t); \delta)$ is a fundamental neighborhood. Let $0 = t_0 < t_1 < \cdots < t_n = 1$ be such that $\gamma(t) \in B(\gamma(t_j); \delta)$ for $t_{j-1} \le t \le t_j$, $1 \le j \le n$.

In particular, $\gamma(0) = \alpha_0 \in \Delta_1 \equiv B(\gamma(t_1); \delta)$; let $g_1 : \Delta_1 \to G$ be the local inverse of τ such that $g_1(\alpha_0) = a$. Clearly $g_1 = g_0$ on $\Delta_0 \cap \Delta_1$. Hence (g_1, Δ_1) is a continuation of (g_0, Δ_0). Similarly, let $\Delta_2 \equiv B(\gamma(t_2); \delta)$ and let $g_2 : \Delta_2 \to G$ be the local inverse of τ such that $g_2(\gamma(t_1)) = g_1(\gamma(t_1))$; so $g_2 = g_1$ on $\Delta_1 \cap \Delta_2$. Continuing this establishes (b).

Let $\{(g_t, \Delta_t) : 0 \le t \le 1\}$ be any continuation of (g_0, Δ_0) along a path γ and put $F = \{t \in [0, 1] : \tau(g_t(z)) = z$ for all z in $\Delta_t\}$. Since $0 \in F$, $F \neq \emptyset$. It is left as an exercise for the reader to show that F is closed and relatively open in [0,1] and so $F = [0, 1]$. This proves (d).

Using the notation from (c), the fact that g_1 and g_2 are local inverses of τ (by (d)) implies that $\omega_1 = \tau(g_1(\omega_1)) = \tau(g_2(\omega_2)) = \omega_2$. The rest of (c) follows by the uniqueness of the local inverse. □

Proof of Theorem 5.1. Let a and b be two distinct points in the complement of Ω and put $\Omega_0 = \mathbb{C} \setminus \{a, b\}$. By Lemma 5.2 there is an analytic covering map $\tau_0 : \mathbb{D} \to \Omega_0$ with $\tau_0(0) = \alpha_0$ and $\tau_0'(0) > 0$. Let Δ_0 be a fundamental neighborhood of α_0 and let $h_0 : \Delta_0 \to \mathbb{D}$ be the local inverse of τ_0 with $h_0(\alpha_0) = 0$ and $h_0'(0) > 0$. Define $\mathcal{F}$ to be the collection of all analytic functions $g : \Delta_0 \to \mathbb{D}$ having the following properties:

5.4
$$\begin{cases} \text{(a)} & g(\alpha_0) \text{ and } g'(\alpha_0) > 0; \\ \text{(b)} & \text{If } \gamma : [0,1] \to \Omega \text{ is a path with } \gamma(0) = \alpha_0, \text{ then} \\ & \text{there is an analytic continuation } \{(g_t, \Delta_t) : 0 \le t \le 1\} \\ & \text{of } (g, \Delta_0) \text{ such that } g_t(\Delta_t) \subseteq \mathbb{D} \text{ for all } t; \\ \text{(c)} & \text{If } (g_1, \Delta_1) \text{ and } (g_2, \Delta_2) \text{ are obtained from } (g, \Delta_0) \\ & \text{by analytic continuation and } g_1(\omega_1) = g_2(\omega_2) \\ & \text{for some points } \omega_1 \text{ in } \Delta_1 \text{ and } \omega_2 \text{ in } \Delta_2, \text{ then} \\ & \omega_1 = \omega_2 \text{ and } g_1(\omega) = g_2(\omega) \text{ for all } \omega \text{ in} \\ & \Delta_1 \cap \Delta_2. \end{cases}$$

Note that conditions (b) and (c) are the conditions (b) and (c) of Lemma 5.3 with G replaced by $\mathbb{D}$. Since $h_0 \in \mathcal{F}$, $\mathcal{F} \neq \emptyset$. Since each function in $\mathcal{F}$ is bounded by 1, $\mathcal{F}$ is a normal family. The strategy here, reminiscent of the proof of the Riemann Mapping Theorem, is to show that the function in $\mathcal{F}$ with the largest derivative at α_0 will lead to the proof of the theorem.

Put $\kappa = \sup\{g'(\alpha_0) : g \in \mathcal{F}\}$; so $0 < \kappa < \infty$. Let $\{h_k\} \subseteq \mathcal{F}$ such that $h_k'(\alpha_0) \to \kappa$. Because $\mathcal{F}$ is a normal family, we can assume that there is an analytic function h on Δ_0 such that $h_k \to h$ uniformly on compact subsets of Δ_0. So $h(\alpha_0) = 0$ and $h'(\alpha_0) = \kappa > 0$. Thus (5.4.a) is satisfied; it will be shown that (b) and (c) of (5.4) also are satisfied so that $h \in \mathcal{F}$.

Let γ be a path in Ω with $\gamma(0) = \alpha_0$ and choose $\delta > 0$ so that $B(\alpha_O; \delta) \subseteq \Delta_0$ and $B(\gamma(t); \delta) \subseteq \Omega$ for $0 \le t \le 1$. Let $0 = t_0 < t_1 < \cdots < t_n = 1$ be such that $\gamma(t) \in \Delta_j \equiv B(\gamma(t_j); \delta)$ for $t_{j-1} \le t \le t_j$. By hypothesis each function h_k admits unrestricted analytic continuation throughout Δ_1, and so the Monodromy Theorem implies there is an analytic function $h_{1k} : \Delta_1 \to \mathbb{D}$ with $h_{1k} = h_k$ on $\Delta_1 \cap \Delta_0$. Continuing, for $1 \le j \le n$ we get an analytic function $h_{jk} : \Delta_j \to \mathbb{D}$ such that $h_{jk} = h_{j-1,k}$ on $\Delta_j \cap \Delta_{j-1}$. Fix j for the moment. So $\{h_{jk} : k \ge 1\}$ is a uniformly bounded sequence of analytic functions on Δ_j and must therefore have a subsequence that converges uniformly on compact subsets of Δ_j to an analytic function $g_j : \Delta_j \to \mathbb{D}$. By successively applying this argument we see that the functions g_j can be obtained so that $g_j = g_{j-1}$ on $\Delta_j \cap \Delta_{j-1}$. Thus $\{(g_j, \Delta_j) : 1 \le j \le n\}$ is an analytic continuation of (h, Δ_0) and (5.4.b) is satisfied.

Now adopt the notation of (5.4.c) with h replacing the function g. Arguing as in the preceding paragraph, there are sequences of analytic functions $\{(h_{1k}, \Delta_1)\}$ and $\{(h_{2k}, \Delta_2)\}$ such that for $j = 1, 2, (h_{jk}, \Delta_j)$ is a continuation of (h_k, Δ_0) and $h_{jk} \to g_j$ uniformly on compact subsets of Δ_j as $k \to \infty$. Now Hurwitz's Theorem (and Exercise 7.2.11) implies that for $j = 1, 2$ and all sufficiently large k there are points ω_{jk} in Δ_j such that $h_{1k}(\omega_{1k}) = g_1(\omega_1) = g_2(\omega_2) = h_{2k}(\omega_{2k})$; these points also can be chosen so that $\omega_{jk} \to \omega_j$ as $k \to \infty$. But $h_k \in \mathcal{F}$ and so (5.4.c) implies that for each k, $\omega_{1k} = \omega_{2k}$ and $h_{1k} = h_{2k}$ on $\Delta_1 \cap \Delta_2$. Thus $\omega_1 = \omega_2$ and $g_1 = g_2$ on $\Delta_1 \cap \Delta_2$. That is, h satisfies (5.4.c). Therefore $h \in \mathcal{F}$.

Fix the notation that h is the function in $\mathcal{F}$ for which $h'(\alpha_0) = \kappa$.

5.5 Claim. For every z in $\mathbb{D}$ there is an analytic continuation (g_1, Δ_1) of (h, Δ_0) such that $z \in g_1(\Delta_1)$.

Suppose the (5.5) is false and there is a complex number c in $\mathbb{D}$ such that no continuation of h ever assumes the value c. Since $h(\alpha_0) = 0$, $0 < |c| < 1$. Let T be the Möbius transformation $(c - z)/(1 - \bar{c}z)$; so T is a conformal equivalence of $\mathbb{D}\backslash\{0\}$ onto $\mathbb{D}\backslash\{c\}$ with $T(0) = c$. Thus if $\mu(z) = T(z^2)$, $((\mathbb{D}\backslash\{0\}),\ \mu)$ is an analytic covering space of $\mathbb{D}\backslash\{c\}$.

Let a be a square root of c; so $\mu(a) = 0$ and $\mu'(a) = -2a/(1 - |c|^2)$. Let g be a local inverse of μ defined in a neighborhood of 0 with $g(0) = a$ and apply Lemma 5.3 to $((\mathbb{D}\backslash\{0\}),\ \mu)$. Thus g satisfies properties (b), (c), and (d) of Lemma 5.3 (with $\alpha_0 = 0$, $G = \mathbb{D}\backslash\{0\}$, $\tau = \mu$, and $\Omega = \mathbb{D}\backslash\{c\}$). Now $g \circ h$ is defined and analytic near α_0 and $g(h(\alpha_0)) = a$. If (h_1, Δ_1) is any analytic continuation of (h, Δ_0) with $h_1(\Delta_1) \subseteq \mathbb{D}$, then $h_1(\Delta_1) \subseteq \mathbb{D}\backslash\{c\}$. Since, by Lemma 5.3.b, g admits unrestricted analytic continuation in $\mathbb{D}\backslash\{c\}$, it follows that $g \circ h$ admits unrestricted analytic continuation in Ω.

We want to have that $g \circ h \in \mathcal{F}$, but two things are lacking: $g(h(\alpha_0)) = a \neq 0$ and $(g \circ h)'(\alpha_0)$ may not be positive. These are easily corrected as follows. Let M be the Möbius transformation $M(z) = b(a - z)/(1 - \bar{a}z)$, where $b = a/|a| = a/\sqrt{|c|}$. Put $f = M \circ g \circ h$. Thus $f(\alpha_0) = 0$ and f satisfies (5.4.b) and (5.4.c) since $g \circ h$ does. Also

$$\begin{aligned} f'(\alpha_0) &= M'(g(h(\alpha_0))g'(h(\alpha_O))\kappa \\ &= M'(a)g'(0)\kappa. \end{aligned}$$

Now $z = \mu(g(z))$, so $1 = \mu'(g(0))g'(0) = -2a(1-|c|^2)^{-1}g'(0)$. Thus $g'(0) = -(1 - |c|^2)/2a$. A computation of $M'(a)$ shows that the above equation becomes

$$\begin{aligned} f'(\alpha_0) &= \frac{b}{1 - |c|}\left(\frac{1 - |c|^2}{2a}\right)\kappa \\ &= \frac{1 + |c|}{2\sqrt{|c|}}\kappa \\ &> 0. \end{aligned}$$

Therefore $f \in \mathcal{F}$. But this also shows that $f'(\alpha_0) > \kappa$, contradicting the definition of κ. Thus Claim 5.5 must hold.

Now to define τ, the covering map of $\mathbb{D}$ onto Ω; this is the easy part. As above, let h be the function in $\mathcal{F}$ with $h'(\alpha_0) = \kappa$. If $z \in \mathbb{D}$, Claim 5.5 implies there is an analytic continuation (h_1, Δ_1) of h with z in $h_1(\Delta_1)$; say $z = h_1(\omega_1)$ for ω_1 in Δ_1. If (h_2, Δ_2) is another continuation of h with $z = h_2(\omega_2)$ for some ω_2 in Δ_2, then (5.4.c) implies that $\omega_1 = \omega_2$ and $h_1 = h_2$ on $\Delta_1 \cap \Delta_2$. Since (5.4.c) also implies that h_1 is univalent on Δ_1, we can safely define τ on $U \equiv h_1(\Delta_1)$ by $\tau = (h_1|\Delta_1)^{-1}$. Since z was arbitrary, τ is defined on $\mathbb{D}$. From the definition it is clear that τ is analytic; it is just as clear that $\tau(\mathbb{D}) = \Omega$. Since τ is locally univalent, $\tau'(z) \neq 0$ for all z in $\mathbb{D}$.

The only remaining part of the existence proof is to show that τ is a covering map. To do this fix a point ω in Ω; we must find a neighborhood Δ of ω such that each component of $\tau^{-1}(\Delta)$ is mapped by τ homeomorphically onto Δ. To find Δ just take any analytic continuation (h_1, Δ) of h, where Δ is a disk about ω. Suppose (h_2, Δ_2) is another continuation of h with ω in Δ_2. According to (5.4.b), h_2 has a continuation to every point of Δ. Thus the Monodromy Theorem implies that h_2 extends to an analytic function on $\Delta_2 \cup \Delta$. So we may assume that h_2 is defined on Δ. That is, we may assume that every continuation (h_2, Δ_2) of (h, Δ_0) with ω in Δ_2 has $\Delta \subseteq \Delta_2$. Let $\mathcal{H}_\omega = \{(h_i, \Delta) : i \in I\}$ be the collection of all the distinct analytic continuations of (h, Δ_0) to Δ. (At this time we do not know that $\mathcal{H}_\omega$ is countable, though this will turn out to be the case.) From the definition of τ,

$$\tau^{-1}(\Delta) = \bigcup_{i \in I} h_i(\Delta).$$

But (5.4.c) implies that $h_i(\Delta) \cap h_j(\Delta) = \emptyset$ for $i \neq j$. Thus $\{h_i(\Delta) : i \in I\}$ are the components of $\tau^{-1}(\Delta)$ (and so I is countable). Clearly τ maps each $h_i(\Delta)$ homeomorphically onto Δ since $\tau\,|h_i(\Delta) = (h_i|\,\Delta)^{-1}$. This completes the proof of existence.

The uniqueness statement follows by Corollary 2.5. □

Theorem 5.1 is the essential part of what is called the Uniformization Theorem. The treatment here is from Pfluger [1969] but with considerable modification. The reader can also see Fisher [1983], Fisher, Hubbard, and Wittner [1988], and Veech [1967] as well as the survey article Abikoff [1981] for different appraches and additional information.

We can now furnish the proof of Proposition 15.4.8 in its entirety. Recall that the proof given there was only valid for the case that the complement of the region G had a component that was not trivial.

5.6 Corollary. *Let G be a proper region in $\mathbb{C}$ that contains zero. If f is a conformal equivalence of G onto itself with $f(0) = 0$ and $f'(0) > 0$, then $f(z) = z$ for all z in G.*

Proof. If G is the entire plane less a single point, the proof is left to the reader. If the complement of G has at least two points, let $\tau : \mathbb{D} \to G$ be the universal analytic covering map with $\tau(0) = 0$ and $\tau'(0) > 0$. It follows that $f \circ \tau : \mathbb{D} \to G$ is also an analytic covering map with $(f \circ \tau)(0) = 0$ and $(f \circ \tau)'(0) > 0$. Thus $f \circ \tau = \tau$. That is, $f(\tau(z)) = \tau(z)$ for all z in $\mathbb{D}$ and the corollary follows. $\square$

Exercises

1. Show that if Ω is simply connected and τ is the covering map obtained in Theorem 5.1, then τ is the Riemann map.

2. Let Ω be a region in $\mathbb{C}$ whose complement has at least two points and let $\tau : \mathbb{D} \to \Omega$ be the analytic covering map with $\tau(0) = \alpha_0$ and $\tau'(0) > 0$. If $f : \Omega \to \mathbb{C}$ is an analytic function, show that $g = f \circ \tau$ is an analytic function on $\mathbb{D}$ such that $g \circ T = g$ for all T in $\mathcal{G}_\tau$. Conversely, if $g : \mathbb{D} \to \mathbb{C}$ is an analytic function with $g \circ T = g$ for all T in $\mathcal{G}_\tau$, then there is an analytic function f on Ω such that $g = f \circ \tau$.

3. Suppose Ω is an n-Jordan region and τ is as in Theorem 5.1. (a) Show that for any point a in $\mathbb{D}$, $\{T(0) : T \in \mathcal{G}_\tau\}$ and $\{T(a) : T \in \mathcal{G}_\tau\}$ have the same set of limit points and these limit points are contained in $\partial\mathbb{D}$. Let K denote this set of limit points. (b) Show that if $w \in \partial\mathbb{D}\backslash K$, then there is a neighborhood U of w such that τ is univalent on $U \cap \mathbb{D}$ and τ has a continuous extension to $\mathbb{D} \cup (\partial\mathbb{D}\backslash K)$.

4. Prove a variation of Corollary 5.6 by assuming that $\mathbb{C} \backslash G$ has at least two points but only requiring F to be a covering map of G onto itself.

Exercises 5 through 8 give a generalization of Schwarz's Lemma. The author thanks David Minda for supplying these.

5. If G is a proper simply connected region, $a \in G$, and f is an analytic function on G such that $f(G) \subseteq G$ and $f(a) = a$, then $|f'(a)| \leq 1$. The equality $|f'(a)| = 1$ occurs if and only if f is a conformal equivalence of G onto itself that fixes a.

6. Let G be a region such that its complement has at least two points, let $a \in G$, and let $\tau : \mathbb{D} \to G$ be the analytic covering map with $\tau(0) = a$ and $\tau'(0) > 0$. If f is an analytic function on G with $f(G) \subseteq G$ and $f(a) = a$, show that there is an analytic function $F : \mathbb{D} \to \mathbb{D}$ with $F(0) = 0$ and $f \circ \tau = \tau \circ F$. Also show that $F(\tau^{-1}(a)) \subseteq \tau^{-1}(a)$.

7. Let f and G be as in Exercise 6. (a) Show that $|f'(a)| \leq 1$. (b) If f is a conformal equivalence of G onto itself, show that $|f'(a)| = 1$ and $f'(a) = 1$ if and only if $f(z) = z$ for all z. (c) Let $\operatorname{Aut}(G, a)$ be all the

conformal equivalences of G onto itself that fix the point a and prove that the map $f \to f'(a)$ is a group monomorphism of $\mathrm{Aut}(G,a)$ into the circle group, $\partial\mathbb{D}$, and hence $\mathrm{Aut}(G,a)$ is abelian. (d) If G is not simply connected, show that G is a finite cyclic group. What happens when G is simply connected?

8. Let G, f, τ, and F be as in Exercise 6. This exercise will show that if $|f'(a)| = 1$, then $f \in \mathrm{Aut}(G,a)$. For this purpose we can assume that G is not simply connected or, equivalently, that $\tau^{-1}(a)$ is infinite. (a) Show that if $|f'(a)| = 1$, then there is a constant c such that $F(z) = cz$ for all z in $\mathbb{D}$. (b) Examine the action of F on $\tau^{-1}(a)$ and conclude that c is an n-th root of unity for some positive integer n. (c) Conclude that $f \in \mathrm{Aut}(G,a)$.

9. Let f and G be as in Exercise 6. For $n \geq 1$ let f_n be the composition of f with itself n times. Prove that if $f \notin \mathrm{Aut}(G,a)$, then $f_n(z) \to a$ uniformly on compact subsets of G.

10. In this exercise let G be a proper region in the plane, let a and b be distinct points in G, and let f be an analytic function on G with $f(G) \subseteq G$, $f(a) = a$, and $f(b) = b$. (a) If G is simply connected and $f(a) = a$ and $f(b) = b$, show that $f(z) = z$ for all z in G. (b) Give an example to show that (a) fails if G is not assumed to be simply connected. (c) If G is not simply connected but the complement of G has at least two points, show that f is a conformal equivalence of G onto itself and there is a positive integer n such that the composition of f with itself n times is the identity. (d) What happens when $G = \mathbb{C}$ or $\mathbb{C}_0$?

11. Suppose f is an analytic function on $\mathbb{D}$ such that $|f(z)| \leq 1$ and $f(0) = 0$. Prove that if z_1 is a point in $\mathbb{D}$ different from 0 and $f(z_1) = 0$, then $|f'(0)| \leq |z_1|$. (Hint: Consider $f(z)/zT(z)$ for an appropriate Möbius transformation T.)

12. *The Aumann-Carathéodory Rigidity Theorem.* Let G be a region that is not simply connected and whose complement has at least two points; let $a \in G$. Show that there is a constant α depending only on G and a with $0 \leq \alpha < 1$ such that if f is an analytic function on G with $f(G) \subseteq G$ and $f(a) = 0$, and if f is not a conformal equivalence of G onto itself, then $|f'(a)| \leq \alpha$. (Hint: Let τ be the analytic covering map of $\mathbb{D}$ onto G with $\tau(0) = a$ and $\tau'(0) > 0$ and let F be as in Exercise 6. Write $\tau^{-1}(a) = \{0, z_1, z_2, \ldots\}$ with $0 < |z_1| \leq |z_2| \leq \cdots$. Observe that $F(\tau^{-1}(a)) \subseteq \tau^{-1}(a)$ and try to apply Exercise 11.)

13. Show that if G is simply connected, no such constant α as that obtained in the preceding exercise can be found.

Chapter 17

De Branges's Proof of the Bieberbach Conjecture

In this chapter we will see a proof of the famous Bieberbach conjecture. Bieberbach formulated his conjecture in 1916 and in 1984 Louis de Branges announced his proof of the conjecture. In March of 1985 a symposium on the Bieberbach conjecture was held at Purdue University and the proceedings were published in Baernstein et al. [1986]. This reference contains a number of research papers as well as some personal accounts of the history surrounding the conjecture and its proof. The history will not be recounted here.

The original ideas of de Branges for the proof come from operator theory considerations. The proof presented here (in §6) is based on the proof given in Fitzgerald and Pommerenke [1985]. This avoids the operator theory, though some additional preliminaries are needed. Another proof can be found in Weinstein [1991]. Though this is shorter in the published form than the Fitzgerald-Pommerenke proof, it actually becomes longer if the same level of detail is provided.

The strategy for the proof, as it evolves in the course of this chapter, is well motivated and clear from both a mathematical and a historical point of view. The tactics for the proof of the final crucial lemma (Lemma 6.8) are far from intuitive. Perhaps this is part of the nature of the problem and connected with the conjecture lying unproved for so many decades. The tools needed in the proof have existed for quite some time: Loewner chains and Loewner's differential equation (1923); Lebediv-Millin inequalities (1965). In fact, de Branges brought a new source of intuition to the problem. This insight was rooted in operator theory and is lost in the shorter proofs of Fitzgerald-Pommerenke and Weinstein. The serious student who wishes to pursue this topic should look at de Branges's paper (de Branges [1985]) and the operator theory that has been done in an effort to more fully explicate his original ideas.

§1 Subordination

In this section we will see an elementary part of function theory that could have been presented at a much earlier stage of the reader's education.

1.1 Definition. If G is a region including the origin and f and g are two analytic functions on G, then f is *subordinate* to g if there is an analytic function $\phi : G \to G$ such that $\phi(0) = 0$ and $f = g \circ \phi$.

It follows from the definition that if f is subordinate to g, then $f(0) = g(0)$. Also note that if f is subordinate to g, then $f(G) \subseteq g(G)$. In Corollary 16.2.6 it was shown that if $g : \mathbb{D} \to \Omega$ is the universal analytic covering map for Ω with $g(0) = \alpha$ and f is any analytic function on $\mathbb{D}$ with $f(0) = \alpha$ and $f(\mathbb{D}) \subseteq \Omega = g(\mathbb{D})$, then f is subordinate to g. In this section we will use this result for the case that Ω is simply connected and g is the Riemann map. The proof for this case is easy and so it is presented here so as to make this section independent of Chapter 16.

The fact that subordination is valid in a more general setting should be a clue for the reader that this is a more comprehensive subject than it will appear from this section. In fact, it has always impressed the author as a favorite topic for true function theorists as it yields interesting estimates and inequalities. For a further discussion of subordination, see Goluzin [1969] and Nehari [1975].

1.2 Proposition. *If g is a univalent function on $\mathbb{D}$, then an analytic function f on $\mathbb{D}$ is subordinate to g if and only if $f(0) = g(0)$ and $f(\mathbb{D}) \subseteq g(\mathbb{D})$. If f is subordinate to g and ϕ is the function such that $f = g \circ \phi$, then ϕ is unique.*

Proof. If $f(\mathbb{D}) \subseteq g(\mathbb{D})$, let$\phi : \mathbb{D} \to \mathbb{D}$ be defined by $\phi = g^{-1} \circ f$; this shows that f is subordinate to g. The converse was observed prior to the statement of this proposition. □

1.3 Proposition. *If f and g are analytic functions on $\mathbb{D}$, f is subordinate to g, and ϕ is the function satisfying $f = g \circ \phi$ and $\phi(0) = 0$, then for each $r > 1$:*

(a) $\{f(z) : |z| < r\} \subseteq \{g(z) : |z| < r\}$;

(b) $\max\{|f(z)| : |z| \leq r\} \leq \max\{|g(z)| : |z| \leq r\}$;

(c) $(1 - |z|^2)|f'(z)| \leq (1 - \phi(z)|^2)|g'(\phi(z))|$;

(d) $|f'(0)| \leq |g'(0)|$ *with equality if and only if there is a constant c with $|c| = 1$ such that $f(z) = g(cz)$ for all z.*

Proof. Let $\phi : \mathbb{D} \to \mathbb{D}$ be the analytic function such that $f = g \circ \phi$. Since $\phi(0) = 0$, Schwarz's Lemma implies that $|\phi(z)| \leq |z|$ on $\mathbb{D}$. This immediately yields (a) and part (b) is a consequence of (a).

To prove part (c), first recall from (6.2.3) that

$$(1 - |z|^2)|\phi'(z)| \leq 1 - |\phi(z)|^2$$

for all z in $\mathbb{D}$. Thus

$$\begin{aligned}(1-|z|^2)|f'(z)| &= (1-|z|^2)|\phi'(z)||g'(\phi(z))| \\ &\le (1-|\phi(z)|^2)|g'(\phi(z))|,\end{aligned} \tag{1.4}$$

establishing (c).

Since $f'(0) = g'(0)\phi'(0)$ and Schwarz's Lemma implies that $|\phi'(0)| \le 1$, we have that $|f'(0)| \le |g'(0)|$. If $|f'(0)| = |g'(0)|$, then $|\phi'(0)| = 1$ and so there is a constant c with $|c| = 1$ such that $\phi(z) = cz$ for all z. □

We now apply these results to a particular class of functions on $\mathbb{D}$. Let $\mathcal{P}$ be the set of all analytic functions p on $\mathbb{D}$ such that $\operatorname{Re} p(z) > 0$ on $\mathbb{D}$ and $p(0) = 1$. The first part of next result is a restatement of Exercise 6.2.3.

1.5 Proposition. *If $p \in \mathcal{P}$ and $z \in \mathbb{D}$, then*

$$\frac{1-|z|}{1+|z|} \le |p(z)| \le \frac{1+|z|}{1-|z|}$$

and

$$|p'(z)| \le \frac{2}{(1-|z|)^2}.$$

These inequalities are sharp for $p(z) = (1+z)/(1-z)$.

Proof. The Möbius transformation $T(z) = (1+z)/(1-z)$ maps to $\mathbb{D}$ onto the right half plane and $T(0) = 1$. Thus Proposition 1.2 implies p is subordinate to T. Let $\phi : \mathbb{D} \to \mathbb{D}$ be the analytic function with $\phi(0) = 0$ and

$$p(z) = T(\phi(z)) = \frac{1+\phi(z)}{1-\phi(z)}$$

for all z in $\mathbb{D}$. But for any ζ in $\mathbb{D}$,

$$\frac{1-|\zeta|}{1+|\zeta|} \le \left|\frac{1+\zeta}{1-\zeta}\right| \le \frac{1+|\zeta|}{1-|\zeta|},$$

so

$$\frac{1-|\phi(z)|}{1+|\phi(z)|} \le \left|\frac{1+\phi(z)}{1-\phi(z)}\right| \le \frac{1+|\phi(z)|}{1-|\phi(z)|}.$$

But $|\phi(z)| \le |z|$, whence the first inequality.

For the second inequality, apply (1.3.c) to obtain that

$$\begin{aligned}|p'(z)| &\le 2\frac{1-|\phi(z)|^2}{|1-\phi(z)|^2}\,\frac{1}{1-|z|^2} \\ &\le 2\frac{1-|\phi(z)|^2}{(1-|\phi(z)|)^2}\,\frac{1}{1-|z|^2} \\ &= 2\frac{1+|\phi(z)|}{1-|\phi(z)|}\,\frac{1}{1-|z|^2}\end{aligned}$$

$$\begin{aligned} &\leq 2\frac{1+|z|}{1-|z|}\,\frac{1}{1-|z|^2} \\ &= \frac{2}{(1-|z|)^2}. \end{aligned}$$

□

1.6 Corollary. *$\mathcal{P}$ is a compact subset of $H(\mathbb{D})$.*

Exercises

1. Show that for p in $\mathcal{P}$ and $|z| < 1$: (a) $\frac{1-|z|}{1+|z|} \leq \text{Re } p(z) \leq \frac{1+|z|}{1-|z|}$; (b) $|\text{Im } p(z)| \leq \frac{2|z|}{1-|z|^2}$; (c) $|\arg p(z)| \leq \sin^{-1}\left[\frac{2|z|}{1+|z|^2}\right]$; (d) $\frac{|p'(z)|}{\text{Re } p(z)} \leq \frac{1}{1-|z|^2}$.

2. (a) Show that if f is an analytic function on $\mathbb{D}$ with $f(0) = 0$ and $f'(0) = 1$, then the convex hull of $f(\mathbb{D})$ contains $B(0;1/2)$. (Hint: Show that if this is not the case, then f is subordinate to a rotation of the map $g(z) = z/(1-z)$.) (b) Show that if $f \in \mathcal{S}$ (14.7.1) and $f(\mathbb{D})$ is convex, then $f(\mathbb{D})$ contains $B(0;1/2)$.

3. (a) Let f_α be the rotation of the Koebe function (Example 14.1.4). Show that if f is an analytic function on $\mathbb{D}$ with $f(0) = 0$ and $f'(0) = 1$, then $f(\mathbb{D})$ meets the ray $\{r\overline{\alpha} : r \geq 1/4\}$ unless $f \neq f_\alpha$. (b) Suppose $f \in \mathcal{S}$ (14.7.1) and $f(\mathbb{D})$ is a star like region; show that $f(\mathbb{D})$ contains $B(0;1/4)$. (This is a special case of the Koebe 1/4-Theorem (14.7.8).)

4. Show that if f is an analytic function on $\mathbb{D}$ with $0 < |f(z)| < 1$ for all z, then $|f'(0)| \leq 2/e$ and this bound is sharp.

§2 Loewner Chains

If $f : \mathbb{D} \times [0,\infty) \to \mathbb{C}$, then $f'(z,t)$ is defined to be the partial derivative of f with respect to the complex variable z, provided this derivative exists. The derivative of f with respect to the real variable t is denoted by $\dot{f}(z,t)$.

2.1 Definition. A *Loewner chain* is a continuous function $f : \mathbb{D} \times [0,\infty) \to \mathbb{C}$ having the following properties:

(a) for all t in $[0,\infty)$, the function $z \to f(z,t)$ is analytic and univalent;

(b) $f(0,t) = 0$ and $f'(0,t) = e^t$;

(c) for $0 \leq s < t < \infty$, $f(\mathbb{D}, s) \subseteq f(\mathbb{D}, t)$.

Let $\mathcal{L}$ denote the collection of all Loewner chains.

The first question is the existence of Loewner chains. If g is the Koebe function (14.1.4) and $f(z,t) = e^t g(z) = e^t z(1-z)^{-2}$, then it is easy to check that f is a Loewner chain with $f(\mathbb{D},t) = \mathbb{C} \setminus (-\infty, -e^t/4]$. If g is any function from the class $\mathcal{S}$, then $f(z,t) = e^t g(z)$ satisfies conditions (a) and (b) of the definition of a Loewner chain, but it may not satisfy condition (c). An amplification of this existence question appears in Theorem 2.16 below.

The first result gives some properties of the parametrized family of simply connected regions $\Omega(t) = f(\mathbb{D},t)$ associated with a Loewner chain that will be used in the construction of the examples just alluded to.

2.2 Proposition. *If f is a Loewner chain and for each $t \geq 0$, $\Omega(t) = f(\mathbb{D},t)$, then:*

2.3
$$\begin{cases} (a) & \textit{for } 0 \leq s < t < \infty,\ \Omega(s) \subseteq \Omega(t) \\ & \textit{and } \Omega(s) \neq \Omega(t); \\ (b) & \textit{if } t_n \to t,\ \textit{then } \Omega(t_n) \to \Omega(t)\ \textit{in the} \\ & \textit{sense of Definition 15.4.1;} \\ (c) & \textit{if } t_n \to \infty,\ \textit{then } \Omega(t_n) \to \mathbb{C}. \end{cases}$$

Proof. The proof that $\Omega(s) \subseteq \Omega(t)$ when $s < t$ is immediate from the definition. Since $f'(0,s) \neq f'(0,t)$, the uniqueness part of the Riemann Mapping Theorem implies that $\Omega(s) \neq \Omega(t)$. Part (b) is clear from Theorem 15.4.10. Note that the Koebe 1/4-Theorem (14.7.8) implies that $f(\mathbb{D},t) \supseteq (e^t/4)\mathbb{D}$ so that $f(\mathbb{D},t) \to \mathbb{C}$ as $t \to \infty$. □

We now prove what is essentially a converse of the preceding proposition.

2.4 Proposition. *Let $\{\Omega(t) : 0 \leq t < \infty\}$ be a family of simply connected regions satisfying (2.3) and for each $t > 0$ let $h_t : \mathbb{D} \to \Omega(t)$ be the Riemann map with $h_t(0) = 0$ and $h_t'(0) = \beta(t) > 0$. If $h(z,t) = h_t(z)$ and $\beta_0 = \beta(0)$, then the following hold.*

(a) *The function β is continuous, strictly increasing, and $\beta(t) \to \infty$ as $t \to \infty$.*

(b) *If $\lambda(t) = \log[\beta(t)/\beta_0]$ and $f(z,t) = \beta_0^{-1} h(z, \lambda^{-1}(t))$, then f defines a Loewner chain with $f(\mathbb{D},t) = \beta_0^{-1}\Omega(\lambda^{-1}(t))$.*

Proof. Let
$$h(z,t) = \beta(t)[z + b_2(t)z^2 + \cdots].$$

By (2.3.b), if $\{t_n\}$ is a sequence in $[0, \infty)$ such that $t_n \to t$, then $\Omega(t_n) \to \Omega(t)$ and so $h_{t_n} \to h_t$ in $H(\mathbb{D})$ (15.4.10). In particular, $h : \mathbb{D} \times [0, \infty) \to \mathbb{C}$ is a continuous function that satisfies conditions (a) and (c) of Definition 2.1 as well as $h(0, t) = 0$ for all t. So h fails to be a Loewner chain only because it lacks the property that $h'(0, t) = e^t$. Indeed, there is no reason to think that $\beta(t) = e^t$. However it is easy enough to reparametrize.

Note that if $t_n \to t$, $h_{t_n} \to h_t$ in $H(\mathbb{D})$ and so $\beta(t_n) \to \beta(t)$. Hence β is continuous.

Fix $s < t$; so $\Omega(s) \subseteq \Omega(t)$. By Proposition 1.2 there is an analytic function $\phi : \mathbb{D} \to \mathbb{D}$ with $\phi(0) = 0$ and such that $h(z, s) = h(\phi(z), t)$ for all z in $\mathbb{D}$. By Schwarz's Lemma $|\phi(z)| < |z|$ and $|\phi'(0)| < 1$, where the strict inequality occurs because $\Omega(s) \neq \Omega(t)$. Thus $\beta(s) = h'(0, s) = h'(0, t)\phi'(0) = \beta(t)\phi'(0)$. But $\phi'(0) > 0$ and we have that β is a strictly increasing function from $[0, \infty)$ into $(0, \infty)$. Moreover $\Omega(t) \to \mathbb{C}$ as $t \to \infty$ so that it must be that $\beta(t) \to \infty$ as $t \to \infty$. This proves (a).

So we have $\beta : [0, \infty) \to [\beta_0, \infty)$ is strictly increasing, continuous, and surjective. Thus $\lambda(t) = \log[\beta(t)/\beta_0]$ is a strictly increasing continuous function from $[0, \infty)$ onto itself. Define $f(z, t)$ as in part (b). So $f : \mathbb{D} \times [0, \infty) \to \mathbb{C}$ is a continuous function that is easily seen to satisfy conditions (a) and (c) in (2.1) and $f(0, t) = 0$. If $\tau = \lambda^{-1}(t)$, then $t = \lambda(\tau)$ so that $e^t = \beta(\tau)/\beta_0$. So $f'(0, t) = \beta_0^{-1}\beta(\tau) = e^t$. Thus f is a Loewner chain and it is clear that $f(\mathbb{D}, t) = \beta_0^{-1}\Omega(\lambda^{-1}(t))$. □

Of course the constant β_0^{-1} must appear in (b) of the preceding result. The function f_0 in a Loewner chain belongs to the class $\mathcal{S}$, and for an arbitrary region $\Omega(0)$ there is no reason to think that the Riemann map of $\mathbb{D}$ onto $\Omega(0)$ comes from the class $\mathcal{S}$.

The following example will prove to be of more value than merely to demonstrate the existence of a Loewner chain.

2.5 Example. Let $\gamma : [0, \infty) \to \mathbb{C}$ be a Jordan arc that does not pass through 0 and is such that $\gamma(t) \to \infty$ as $t \to \infty$ and $\gamma(0) = a_0$. For $0 \leq t < \infty$, let γ_t be the restriction of γ to $[t, \infty)$ and put $\Omega(t) = \mathbb{C} \setminus \gamma_t$. It is easy to see that (2.3) is satisfied. By means of Proposition 2.4 we have an example of a Loewner chain.

For a Loewner chain f, let f_t denote the univalent function on $\mathbb{D}$ defined by $f_t(z) = f(z, t)$. Think of a Loewner chain as a parametrized family of univalent functions on $\mathbb{D}$, $\{f_t\}$, indexed by time, where f_0 is the starting point, and as time approaches ∞ the functions expand to fill out the plane.

2.6 Proposition. *If $f \in \mathcal{L}$ and $0 \leq s \leq t < \infty$, then there is a unique analytic function $z \to \phi(z, s, t)$ defined on $\mathbb{D}$ having the following properties.*

(a) *$\phi(z, s, t) \in \mathbb{D}$ and $f(z, s) = f(\phi(z, s, t), t)$ for all z in $\mathbb{D}$.*

(b) *$z \to \phi(z, s, t)$ is univalent, $\phi(0, s, t) = 0$, $|\phi(z, s, t)| \leq |z|$ for all z in*

$\mathbb{D}$, *and* $\phi'(0,s,t) = e^{s-t}$.

(c) $\phi(z,s,s) = z$ *for all* z *in* $\mathbb{D}$.

(d) *If* $s \le t \le u$, *then* $\phi(z,s,u) = \phi(\phi(z,s,t),t,u)$ *for all* z *in* $\mathbb{D}$.

Proof. Because $f_s(\mathbb{D}) \subseteq f_t(\mathbb{D})$, there is a unique analytic function $\phi(z,s,t)$ defined on $\mathbb{D}$ with values in $\mathbb{D}$ and such that $\phi(0,s,t) = 0$ and $f_t(\phi(z,s,t)) = f_s(z)$ (Proposition 1.2). Since both $f_t(z)$ and $f_s(z)$ are univalent on $\mathbb{D}$, ϕ is also. This shows (a). The fact that $|\phi(z,s,t)| \le |z|$ for all z in $\mathbb{D}$ follows by Schwarz's Lemma. Taking the derivative of both sides of the equation in (a) at $z=0$ gives that $e^s = f'(0,s) = f'(\phi(0.s.t),t)\phi'(0.s.t) = e^t\phi'(0.s.t)$. This proves part (b).

Part (c) follows by the equation in (a) and the fact that the function $\phi(z,s.t)$ is unique. Finally, to show (d) observe that the properties of the functions imply that for $w = \phi(z,s,t)$, $f(\phi(w,t,u),u) = f(w,t) = f(z,s)$, so that (d) follows by the uniqueness of ϕ. □

2.7 Definition. The function $\phi(z,s,t)$ defined for z in $\mathbb{D}$ and $0 \le s \le t < \infty$ and satisfying

2.8 $$f(z,s) = f(\phi(z,s,t),t)$$

for a Loewner chain f is called the *transition function* for the Loewner chain.

Note that the transition function is given by the equation $\phi(z,s,t) = f_t^{-1}(f_s(z))$.

2.9 Lemma. *If* $f \in \mathcal{L}$, *then for all* z *in* $\mathbb{D}$ *and* $0 \le t < \infty$,

2.10 $$e^t\frac{1-|z|}{(1+|z|)^3} \le |f'(z,t)| \le e^t\frac{1+|z|}{(1-|z|)^3}$$

2.11 $$e^t\frac{|z|}{(1+|z|)^2} \le |f(z,t)| \le e^t\frac{|z|}{(1-|z|)^2}.$$

Proof. In fact, the function $e^{-t}f(z,t) \in \mathcal{S}$, the class of univalent functions defined in §14.7. Thus this lemma is an immediate consequence of the Koebe Distortion Theorem (14.7.9). □

The preceding lemma quickly implies that for any $T > 0$, $\{f(z,t) : 0 \le t \le T\}$ is a normal family in $H(\mathbb{D})$. However we will prove much more than this in Proposition 2.15 below.

2.12 Lemma. *If f is a Loewner chain with transition function ϕ, then the function p is defined for z in $\mathbb{D}$ and $0 \le s \le t < \infty$ by*

$$\begin{aligned} p(z,s,t) &= \frac{1+e^{s-t}}{1-e^{s-t}}\left[\frac{1-z^{-1}\phi(z,s,t)}{1+z^{-1}\phi(z,s,t)}\right] \\ &= \frac{1+e^{s-t}}{1-e^{s-t}}\left[\frac{z-\phi(z,s,t)}{z+\phi(z,s,t)}\right] \end{aligned}$$

belongs to the class $\mathcal{P}$ and $p(0,s,t) = 1$.

Proof. Let $\phi(z) = \phi(z,s,t)$ for s and t fixed. The fact that p belongs to $\mathcal{P}$ is a consequence of the fact that $|\phi(z)| \le |z|$ and hence belongs to $\mathbb{D}$ for all z in $\mathbb{D}$. □

2.13 Lemma. *If $f \in \mathcal{L}$, $|z| < 1$, and $0 \le s \le t \le u, \infty$, then the following inequality holds:*

$$|f(z,s) - f(z,t)| \le \frac{8|z|}{(1-|z|)^4}(e^t - e^s).$$

Proof. For the moment, fix s and t, $s \le t$, and put $\phi(z) = \phi(z,s,t)$. According to (2.10), if $|\zeta| \le |z| < 1$, then $|f'(\zeta,t)| \le 2e^t(1-|\zeta|)^{-3} \le 2e^t(1-|z|)^{-3}$. Since $|\phi(z)| \le |z|$ this implies

$$\begin{aligned} |f(z,t) - f(z,s)| &= \left|\int_{\phi(z)}^{z} f'(\zeta,t)d\zeta\right| \\ &\le \frac{2e^t}{(1-|z|)^3}|\phi(z) - z|. \end{aligned}$$

Now to estimate $|\phi(z) - z|$. Applying Proposition 1.5 to the function p defined in Lemma 2.12, we get

$$\begin{aligned} \left[\frac{1+e^{s-t}}{1-e^{s-t}}\right]\left|\frac{z-\phi(z)}{z+\phi(z)}\right| &= |p(z,s,t)| \\ &\le \frac{1+|z|}{1-|z|}. \end{aligned}$$

Hence

$$\begin{aligned} |z-\phi(z)| &\le \left[\frac{1-e^{s-t}}{1+e^{s-t}}\right]\left[\frac{1+|z|}{1-|z|}\right]|z+\phi(z)| \\ \textbf{2.14} \qquad &\le 2|z|\left[\frac{1+|z|}{1-|z|}\right](1-e^{s-t}). \end{aligned}$$

Therefore

$$\begin{aligned} |f(z,t) - f(z,s)| &\le \frac{2e^t}{(1-|z|)^3}\, 2|z|\left[\frac{1+|z|}{1-|z|}\right](1-e^{s-t}) \\ &\le \frac{8|z|}{(1-|z|)^4}(e^t - e^s), \end{aligned}$$

proving the lemma. □

2.15 Proposition. *The set $\mathcal{L}$ of all Loewner chains is a compact subset of the metric space $C(\mathbb{D} \times [0, \infty), \mathbb{C})$.*

Proof. The inequality (2.11) implies that for every (z, t) in $\mathbb{D} \times [0, \infty)$, $\sup\{|f(z,t)| : f \in \mathcal{L}\} < \infty$. Also Lemma 2.13 shows that $\mathcal{L}$ is equicontinuous at each point of $\mathbb{D} \times [0, \infty)$. Thus the Arzela-Ascoli Theorem implies that $\mathcal{L}$ is normal in $C(\mathbb{D} \times [0, \infty))$. It remains to show that $\mathcal{L}$ is closed.

If $\{f_n\}$ is a sequence in $\mathcal{L}$ and $f_n \to f$ in $C(\mathbb{D} \times [0, \infty))$, then for each t, $f_n(z,t) \to f(z,t)$ in $H(\mathbb{D})$. Hence for each $t, z \to f(z,t)$ is analytic. Clearly $f(0,t) = 0$ and $f'(0,t) = e^t$. Since each $f_n(z,t)$ is univalent, Hurwitz's Theorem implies that $z \to f(z,t)$ is also univalent. Finally, if $0 \le s < t < \infty$, $f_n(\mathbb{D}, s) \subseteq f_n(\mathbb{D}, t)$ for all n, so $f(\mathbb{D}, s) \subseteq f(\mathbb{D}, t)$. Therefore, $f \in \mathcal{L}$ and $\mathcal{L}$ is closed. □

Note that if f is a Loewner chain, then $z \to f(z, 0)$ is a function in the class $\mathcal{S}$ defined in §14.7. A further amplification of the fact that Loewner chains exist is the next theorem, which asserts that any function in the class $\mathcal{S}$ can occur as the starting point of a Loewner chain.

2.16 Theorem. *For every function f_0 in $\mathcal{S}$ there is a Loewner chain f such that $f(z, 0) = f_0(z)$ on $\mathbb{D}$.*

Proof. First assume that f is analytic in a neighborhood of cl $\mathbb{D}$. Thus $\gamma = f(\partial \mathbb{D})$ is a closed Jordan curve. Let $g : \mathbb{C}_\infty \setminus \text{cl } \mathbb{D} \to \mathbb{C}_\infty \setminus \text{cl } [\text{ out } \gamma]$ be the conformal equivalence with $g(\infty) = \infty$ and $g'(\infty) > 0$. For $0 \le t < \infty$, put $\Omega(t) =$ the inside of the Jordan curve $g(\{z : |z| = e^t\})$. Note that $\Omega(0) = f_0(\mathbb{D})$ and $\{\Omega(t)\}$ satisfies the condition (2.3).

Letting h be as in Proposition 2.4, the uniqueness of the Riemann map implies that $h(z, 0) = f_0(z)$ and so $\beta_0 = 1$ (in the notation of (2.4)). An application of Proposition 2.4 now proves the theorem for this case.

For the general case, let f be an arbitrary function in $\mathcal{S}$, for each positive integer n put $r_n = 1 - n^{-1}$, and let $f_n(z) = r_n^{-1} f(r_n, z)$. So each $f_n \in \mathcal{S}$ and is analytic in a neighborhood of cl $\mathbb{D}$. By the first part of the proof there is a Loewner chain F_n with $F_n(z, 0) = f_n(z)$. By Proposition 2.15 some subsequence of $\{F_n\}$ converges to a Loewner chain F. It is routine to check that $F(z, 0) = f(z)$ in $\mathbb{D}$. □

Note that if f_0 maps $\mathbb{D}$ onto the complement of a Jordan arc reaching out to infinity, then the preceding theorem is just Example 2.5. It is this particular form of the theorem that will be used in the proof of de Branges's Theorem.

The study of Loewner chains continues in the next section, where we examine Loewner's differential equation.

Exercises

1. Let $f : \mathbb{D} \times [0, \infty) \to \mathbb{C}$ be a function such that for each z in $\mathbb{D}, t \to f(z,t)$ is continuous and for each t in $[0, \infty), z \to f(z,t)$ is analytic. Assume that properties (a), (b), and (c) of Definition 2.1 are satisfied. Show that Lemma 2.9 is satisfied for this function and consequently that f is a Loewner chain.

2. Let Ω be a simply connected region containing 0 such that $\mathbb{C}_\infty \setminus \Omega$ consists of two Jordan arcs that meet only at ∞. Let h_0 be the Riemann map of $\mathbb{D}$ onto Ω with $h_0(0) = 0$ and $\beta_0 = h_0(0) > 0$. Show that there are two Loewner chains f and g with $f(z,0) = g(z,0) = \beta_0^{-1} h_0(z)$ (see Proposition 2.4).

3. Let $f \in \mathcal{L}$ and let ϕ be the transition function for f. Fix $u \geq 0$ and define $g : \mathbb{D} \times [0, \infty) \to \mathbb{C}$ by $g(z,t) = e^u \phi(z,t,u)$ for $t \leq u$ and $g(z,t) = e^u z$ for $t \geq u$. Show that $g \in \mathcal{L}$.

4. If g is the Koebe function and f is the Loewner chain defined by $f(z,t) = e^t g(z) = e^t z(1-z)^{-2}$, find the transition function for f.

5. In Lemma 2.9, what can be said about the Loewner chain f if one of the inequalities is an equality?

6. If f is a Loewner chain with transition function ϕ, prove that, analogous to (2.13),

$$|\phi(z,t,u) - \phi(z,s,u)| \leq \frac{8|z|}{(1-|z|)^4}(1 - e^{s-t})$$

for $0 \leq s \leq t \leq u < \infty$ and all z in $\mathbb{D}$.

§3 Loewner's Differential Equation

In this section Loewner's differential equation and the concommitant characterization of Loewner chains is studied. There is a version of Loewner's differential equation valid for all Loewner chains, but we will only see here the version for a chain as in Example 2.5. This is all that is needed for the proof of de Branges's Theorem.

To set notation, let $\gamma : [0, \infty) \to \mathbb{C}$ be a Jordan arc with $\gamma(0) = a_0$ such that γ does not pass through 0 and $\gamma(t) \to \infty$ as $t \to \infty$. For $0 \leq t < \infty$, let γ_t be the restriction of γ to $[t, \infty)$ and put $\Omega(t) = \mathbb{C} \setminus \gamma_t$. Assume there is a Loewner chain f such that $f_t(\mathbb{D}) = \not\leq(\approx)$ for all $t \geq 0$. (The reason for the word "assume" here is that otherwise we would have to multiply the regions $\Omega(t)$ by a constant. See Proposition 2.4.b.) Let ϕ be the transition function for the chain f and let $g_t = f_t^{-1} : \Omega(t) \to \mathbb{D}$ with $g(\zeta, t) = g_t(\zeta)$. For $s \le t$ let $\phi_{st}(z) = \phi(z, s, t)$. Recall that $\phi_{st} = f_t^{-1} \circ f_s$.

Now Proposition 15.3.7 implies that f_s and f_t have continuous extensions to cl $\mathbb{D}$. Moreover Proposition 15.3.8 g_t has a continuous extension to $\Omega(t) \cup \{\gamma(t)\}$. Let $\lambda(t)$ be the unique point on the unit circle such that $f_t(\lambda(t)) = \gamma(t)$. Let C_{st} be the closed arc on $\partial\,\mathbb{D}$ defined by $C_{st} = \{z \in \partial\,\mathbb{D} : f_s(z) \in \gamma([s,t])\}$ and let $J_{st} = g_t(\gamma([s,t]))$. So J_{st} is a Jordan arc that lies in $\mathbb{D}$ except for its end point $\lambda(t)$. (The reader must draw a picture here.) Thus ϕ_{st} maps $\mathbb{D}$ conformally onto $\mathbb{D} \setminus J_{st}$. Also ϕ_{st} has a continuous extension to cl $\mathbb{D}$ that maps C_{st} onto J_{st} and the complement of C_{st} in the circle onto $\partial\,\mathbb{D} \setminus \{\lambda(t)\}$.

Observe that $\lambda(s)$ is an interior point of the arc C_{st} and C_{st} decreases to $\lambda(s)$ as $t \downarrow s$. Similarly, if t is fixed and $s \uparrow t$, then J_{st} decreases to $\lambda(t)$.

3.1 Proposition. *With the preceding notation, the function $\lambda : [0, \infty) \to \partial\,\mathbb{D}$ is continuous.*

Proof. An application of the Schwarz Reflection Principle gives an analytic continuation of ϕ_{st} to $\mathbb{C} \setminus C_{st}$. This continuation, still denoted by ϕ_{st}, is a conformal equivalence of $\mathbb{C} \setminus C_{st}$ onto $\mathbb{C} \setminus \{J_{st} \cup J_{st}^*\}$, where J_{st}^* is the reflection of J_{st} across the unit circle.

Claim. $|z^{-1}\phi_{st}(z)| \le 4e^{t-s}$ for all z in $\mathbb{C} \setminus C_{st}$.

This is shown by applying the Maximum Modulus Theorem. In fact

$$\lim_{z\to\infty} \frac{\phi(z,s,t)}{z} = \lim_{z\to 0} \frac{z}{\phi(z,s,t)} = \frac{1}{\phi'(0)} = e^{t-s}.$$

Also the Koebe 1/4-Theorem (14.7.8) implies that since J_{st} is contained in the complement of $\phi_{st}(\mathbb{D})$, $J_{st} \subseteq \mathbb{C} \setminus \{\zeta : |\zeta| < e^{s-t}/4\}$. Thus $J_{st}^* \subseteq \{\zeta : |\zeta| < 4e^{t-s}\}$. This proves the claim.

The claim shows that for any $T \ge s$, $\{z^{-1}\phi_{st} : s \le t \le T\}$ is a normal family. If $t_k \downarrow s$ and $\{\phi_{st_k}\}$ converges to an analytic function ψ, then ψ is analytic on $\mathbb{C} \setminus \{\lambda(s)\}$ and bounded there. Hence $\lambda(s)$ is a removable singularity and ψ is constant. But $\psi(0) = \lim_{t\to s} \phi'(0,s,t) = 1$. Since every convergent sequence from this normal family must converge to the constant function 1, we have that $z^{-1}\phi_{st}(z) \to 1$ (uc) on $\mathbb{C} \setminus \{\lambda(s)\}$ as $t \downarrow s$. Thus $\phi_{st}(z) \to z$ (uc) on $\mathbb{C} \setminus \{\lambda(s)\}$ as $t \downarrow s$.

Fix $s \ge 0$. We now show that λ is right continuous at s. The proof that λ is left continuous is similar and left to the reader. If $\varepsilon > 0$, choose $\delta > 0$ such that for $s < t < s + \delta$, $C_{st} \subseteq B(\lambda(s); e)$. Let C be the circle $\partial B(\lambda(s); \varepsilon)$ and put $\chi = \phi_{st}(C)$, a Jordan curve. Note that the inside of χ contains the arcs J_{st} and J_{st}^*; so in particular $\lambda(t) \in$ ins χ. Now δ can be chosen sufficiently small that for $s < t < s + \delta$, $|\phi_{st}(z) - z| < \varepsilon$ for all z in C. From here it follows that diam $\chi < 3\varepsilon$. So if we take any point z on

C, $\|\lambda(s) - \lambda(t)| \le |\lambda(s) - z| + |z\phi_{st}(z)| + |\phi_{st}(z) - \lambda(t)| < \varepsilon + \varepsilon + 3\varepsilon = 5\varepsilon$. This proves right continuity. □

For a Loewner chain f, recall that $\dot{f}(z,t) = \partial f/\partial t$ and $f'(z,t) = \partial f/\partial z$; similarly define $\dot{\phi}(z,s,t)$, $\phi'(z,s,t)$, $\dot{g}(\zeta,t)$, and $g'(\zeta,t)$.

3.2 Proposition. *Fix the notation as above. The function g has continuous partial derivatives and if $x(t) = \overline{\lambda(t)}$, then for $t \ge 0$ and ζ in $\Omega(t)$*

3.3
$$\dot{g}(\zeta,t) = -g(\zeta,t)\left[\frac{1 + x(t)g(\zeta,t)}{1 - x(t)g(\zeta,t)}\right].$$

Proof. It is left for the reader to verify that g is a continuous function. To prove that g' exists and is continuous is the easy part. In fact because f is a Loewner chain, $g'(\zeta,t)$ exists and equals $[f'(g(\zeta,t),t)]^{-1}$. Since the convergence of a sequence of analytic functions implies the convergence of its derivatives, $f' : \mathbb{D} \times [0,\infty) \to \mathbb{C}$ is continuous. Hence g' is continuous.

Note that if we can show that $\dot{g}$ exists and (3.3) holds, then the continuity of $\dot{g}$ follows from the continuity of g. To prove existence, we will show that the right partial derivative of g exists and satisfies (3.3). The proof that the left derivative exists and also satisfies (3.3) is similar and left to the reader. Return to the transition function ϕ. Remember that for $s \le t$, ϕ_{st} has a continuous extension to cl $\mathbb{D}$ and $\phi_{st}(C_{st}) = J_{st}$. Now $z^{-1}\phi_{st}$ does not vanish on $\mathbb{D}$ and so it is possible to define the analytic function

$$\Phi(z) = \log\left[\frac{\phi_{st}(z)}{z}\right],$$

where the branch of the logarithm is chosen so that $\Phi(0) = s - t$.

Thus Φ is continuous on cl $\mathbb{D}$ and analytic on $\mathbb{D}$. If $z \in \partial\mathbb{D} \setminus C_{st}$, then $\phi_{st}(z) \in \partial\mathbb{D} \setminus \{\lambda(t)\}$ and so Re $\Phi(z) = \log|z^{-1}\phi_{st}(z)| = 0$. Thus the Poisson formula gives that

$$\text{Re } \Phi(z) = \frac{1}{2\pi}\int_\alpha^\beta [\text{Re } \Phi(e^{i\theta})]P_z(e^{i\theta})\, d\theta,$$

where α and β are chosen so that $e^{i\alpha}$ and $e^{i\beta}$ are the end points of C_{st}. By the choice of the branch of the logarithm,

3.4
$$\Phi(z) = \frac{1}{2\pi}\int_\alpha^\beta [\text{Re } \Phi(e^{i\theta})]\frac{e^{i\theta} + z}{e^{i\theta} - z}\, d\theta.$$

We also have that

$$s - t = \Phi(0) = \frac{1}{2\pi}\int_\alpha^\beta \text{Re } \Phi(e^{i\theta})\, d\theta.$$

Now $f_t \circ \phi_{st} = f_s$ and so $g_t = \phi_{st} \circ g_s$. Thus letting $z = g_s(\zeta)$ in (3.4) implies that

$$\log \frac{g_t(\zeta)}{g_s(\zeta)} = \frac{1}{2\pi} \int_\alpha^\beta \text{Re } \Phi(e^{i\theta}) \left[\frac{e^{i\theta} + g_s(\zeta)}{e^{i\theta} - g_s(\zeta)} \right] d\theta.$$

Now apply the Mean Value Theorem for integrals to the real and imaginary parts of this integrand to obtain numbers u and v with $\alpha \leq u,\ v \leq \beta$ and such that

$$\begin{aligned} \log \left[\frac{g_t(\zeta)}{g_s(\zeta)} \right] &= \\ &= \left[\frac{1}{2\pi} \int_\alpha^\beta \text{Re } \Phi(e^{i\theta})\, d\theta \right] \left[\text{Re} \left\{ \frac{e^{iu} + g_s(\zeta)}{e^{iu} - g_s(\zeta)} \right\} \right. \\ &\qquad \left. + i \text{ Im} \left\{ \frac{e^{iv} + g_s(\zeta)}{e^{iv} - g_s(\zeta)} \right\} \right] \\ &= (s-t) \left[\text{Re} \left\{ \frac{e^{iu} + g_s(\zeta)}{e^{iu} - g_s(\zeta)} \right\} + i \text{ Im} \left\{ \frac{e^{iv} + g_s(\zeta)}{e^{iv} - g_s(\zeta)} \right\} \right]. \end{aligned}$$

Now divide both sides of this equation by $t - s$ and let $t \downarrow s$. When this is done, e^{iu} and e^{iv} both converge to $\lambda(s)$. Thus

$$\begin{aligned} \lim_{t \downarrow s} \frac{1}{t-s} \log \left[\frac{g_t(\zeta)}{g_s(\zeta)} \right] &= -\frac{\lambda(s) + g_s(\zeta)}{\lambda(s) - g_s(\zeta)} \\ &= -\frac{1 + x(s) g_s(\zeta)}{1 - x(s) g_s(\zeta)}. \end{aligned}$$

But the left hand side is precisely the right derivative of $\log[g_t(\zeta)/g_s(\zeta)]$ with respect to t, evaluated as $t = s$. By taking exponentials and multiplying by $g_s(\zeta)$, it follows that $t \to g_t(\zeta)$ has a right derivative at $t = s$. Elementary calculus manipulations then give that

$$\dot{g}(\zeta, s) = -g(\zeta, s) \left[\frac{1 + x(s) g(\zeta, s)}{1 - x(s) g(\zeta, s)} \right],$$

where this is actually the right derivative. As was said, the similar proof for the left derivative is in the reader's hands. □

3.5 Theorem. *If f is a Loewner chain such that f_0 is a mapping onto a slit region, then there is a continuous function $x : [0, \infty) \to \partial\mathbb{D}$ such that $\dot{f}(z, t)$ exists and satisfies*

3.6
$$\dot{f}(z, t) = \left[\frac{1 + x(s) z}{1 - x(s) z} \right] z f'(z, t).$$

Proof. The existence and continuity of f' was already shown at the beginning of the proof of Proposition 3.2. Since f_t and g_t are inverses of each other, the differentiability of f with respect to t will follow from the Inverse Function Theorem of Advanced Calculus, but this must be set up properly.

Define $F : \mathbb{D} \times [0, \infty) \to \mathbb{C} \times \mathbb{R}$ by $F(z,t) = (f(z,t), t)$. It is not hard to see that $F(\mathbb{D} \times (0,\infty))$ is the open set $\Lambda = \cup_{t>0}(\mathbb{C} \setminus \gamma_t) \times (t, \infty)$ and that F is a one-to-one mapping with inverse given by $F^{-1}(\zeta, t) = (g(\zeta,t), t)$. Thus F^{-1} is a continuously differentiable function and its Jacobian is

$$\det \begin{bmatrix} g'(\zeta,t) & 0 \\ \dot{g}(\zeta,t) & 1 \end{bmatrix} = g'(\zeta,t),$$

which never vanishes. Thus F is continuously differentiable, from which it follows that $\dot{f}(z,t)$ exists and is continuous.

Now let x be as in Proposition 3.2. Note that $\zeta = f(g_t(\zeta,t)$, so differentiating with respect to t gives that $0 = f'(g_t(\zeta),t)\dot{g}(\zeta,t) + \dot{f}(g_t(\zeta),t)$. Putting $z = g_t(\zeta)$, this shows that $0 = f'(z,t)\dot{g}(\zeta,t) + \dot{f}(z,t)$ for all z in $\mathbb{D}$. Therefore applying (3.3),

$$\begin{aligned} \dot{f}(z,t) &= -f'(z,t)\dot{g}(\zeta,t) \\ &= f'(z,t)g(\zeta,t)\left[\frac{1+x(t)g(\zeta,t)}{1-x(t)g(\zeta,t)}\right] \\ &= f'(z,t)z\left[\frac{1+x(t)z}{1-x(t)x}\right]. \end{aligned}$$

This finishes the proof. □

Equation (3.6) is *Loewner's differential equation.* There is a differential equation satisfied by all Loewner chains, not just those that begin with a mapping onto a slit region. For an exposition of this see Duren [1983] and Pommerenke [1975], two sources used in the preparation of this section and the preceding one.

This section concludes with a result valid for all Loewner chains, not just those that begin with a mapping onto a slit region. In many ways this illustrates the importance of Loewner chains in the study of the univalent functions in the class $\mathcal{S}$.

3.7 Proposition. *Let f be a Loewner chain with g_t the inverse of f_t. Then*

$$f_0(z) = \lim_{t\to\infty} e^t g(f_0(z), t)$$

uniformly on compact subsets of $\mathbb{D}$.

Proof. According to (2.11),

$$e^t \frac{|z|}{(1+|z|)^2} \le |f(z,t)| \le e^t \frac{|z|}{(1-|z|)^2}.$$

Substituting $z = g(\zeta, t)$ this becomes

$$e^t \frac{|g(\zeta,t)|}{(1+|g(\zeta,t)|)^2} \le |z| \le e^t \frac{|g(\zeta,t)|}{(1-|g(\zeta,t)|)^2}.$$

Algebraic manipulation gives that

3.8 $$[1-|g(\zeta,t)|]^2 \le e^t \left|\frac{g(\zeta,t)}{\zeta}\right| \le [1+|g(\zeta,t)|]^2.$$

Since $|g(\zeta,t)| \le 1$ for all ζ and t, $|g(\zeta,t)| \le 4e^{-t}|\zeta|$. This implies that $\{e^t g_t/\zeta t \ge 0\}$ is a normal family. But (3.8) implies that if $t_k \to \infty$ and $e^{t_k} g_{t_k}/\zeta \to h$, then h is an analytic function with $|h| \equiv 1$. Hence h is constant. But for any t, $e^t g_t'(0) = e^t/f_t'(0) = 1$. Thus $h(0) = 1$ and so $h \equiv 1$. That is, any limit point of this normal family as $t \to \infty$ must be the constant function 1. Therefore as $t \to \infty$, $e^t g_t(\zeta)/\zeta \to 1$ uniformly on compact subsets of $\mathbb{C}$. Thus $e^t g(\zeta, t) \to \zeta$ uniformly on compact subsets of $\mathbb{C}$, so that $e^t g(f_0(z), t) \to f_0(z)$ uniformly on compact subsets of $\mathbb{D}$ as $t \to \infty$. $\square$

3.9 Corollary. *If $f \in \mathcal{S}$ and g is the inverse of the Loewner chain starting at f, then $t \to e^t g(f(z), t)$ for $0 \le t \le \infty$ is a path of functions in $\mathcal{S}$ starting at z and ending at f.*

3.10 Corollary. *The family $\mathcal{S}$ of univalent functions with the relative topology of $H(\mathbb{D})$ is arcwise connected.*

Exercises

1. Let $f \in \mathcal{L}$ with g_t the inverse of f_t and put $h(z,t) = g(f_0(z), t)$. Show that h satisfies the equation

$$\dot{h}(z,t) = h'(z,t)\left[\frac{1+x(t)h(z,t)}{1-x(t)h(z,t)}\right],$$

where x is as in Proposition 3.2.

2. If $f \in \mathcal{L}$ and ϕ is its transition function, show that for all $s \ge 0$, $f(z,s) = \lim_{t\to\infty} e^t \phi(z,s,t)$ uniformly on compact subsets of $\mathbb{D}$. Compare with Proposition 3.7.

3. If g is the Koebe function and the Loewner chain f is defined by $f(z,t) = e^t g(z) = e^t z(1-z)^{-2}$, find the function x that appears in Loewner's differential equation for f.

§4 The Milin Conjecture

What will be proved in the next section is not the Bieberbach conjecture but the Milin conjecture, which is stronger than Bieberbach's conjecture. In this section the Milin conjecture will be stated and it will be shown that it implies the Bieberbach conjecture. But first the Robertson conjecture will be stated and it will be shown that it implies the Bieberbach conjecture and is implied by the Milin conjecture.

For reasons of specificity and completeness, let's restate the Bieberbach conjecture. First, as standard notation, if f is a function in $\mathcal{S}$, let

4.1 $$f(z) = z + a_2 z^2 + a_3 z^3 + \cdots .$$

4.2 Bierberbach's Conjecture. *If f belongs to the class $\mathcal{S}$ and has the power series representation (4.1), then $|a_n| \leq n$. If there is some integer n such that $|a_n| = n$, then f is a rotation of the Koebe function.*

If f is a rotation of the Koebe function, then (14.1.4) shows that $|a_n| = n$ for all the coefficients.

Recall (Proposition 14.7.4) that a function g in the class $\mathcal{S}$ is odd if and only if there is a function f in $\mathcal{S}$ such that $g(z)^2 = f(z^2)$ for all z in $\mathbb{D}$. Let $\mathcal{S}_-$ be the collection of odd functions in $\mathcal{S}$ and if $g \in \mathcal{S}_-$,let

4.3 $$g(z) = z + c_3 z^3 + c_5 z^5 + \cdots$$

be its power series. The Robertson conjecture can now be stated.

4.4 Robertson's Conjecture. *If $g \in \mathcal{S}_-$ has the power series representation (4.3), then for each $n \geq 1$*

$$1 + |c_3|^2 + \cdots + |c_{2n-1}|^2 \leq n.$$

If there is an integer n such that equality occurs, then $g(z)^2 = f(z^2)$, where f is a rotation of the Koebe function.

4.5 Theorem. *Robertson's conjecture implies Bieberbach's conjecture.*

Proof. Let $g \in \mathcal{S}_-$ satisfy (4.3) and let f be the corresponding function in $\mathcal{S}$ with $g(z)^2 = f(z^2)$ in $\mathbb{D}$. Suppose f satisfies (4.1). Thus

$$z^2 + a_2 z^4 + \cdots = (z + c_3 z^3 + \cdots)^2.$$

Expanding and identifying coefficients of the corresponding powers of z we get that for all $n \geq 1$

$$a_n = c_1 c_{2n-1} + c_3 c_{2n-3} + \cdots + c_{2n-1} c_1.$$

An application of the Cauchy-Schwarz Inequality shows that

$$|a_n| \le \sum_{k=1}^{n} |c_{2k-1}|^2,$$

whence the first part of theorem.

If equality occurs in Bieberbach's conjecture, then the preceding inequality shows that equality occurs in Robertson's conjecture. This completes the proof. □

To state the Milin conjecture is not difficult; it only requires some notation. To see that this implies the Robertson conjecture is more involved and will occupy us for most of the remainder of the section.

Let $f \in \mathcal{S}$ and let g be the corresponding function in $\mathcal{S}_-$ with $g(z)^2 = f(z^2)$ on $\mathbb{D}$. Assume (4.1) and (4.3) hold. It is easy to see that $z^{-1}f$ is an analytic function on $\mathbb{D}$ and has no zeros there. Thus there is an analytic branch of $(1/2)\log[z^{-1}f(z)]$ defined on $\mathbb{D}$; denote this function by h and let

4.6
$$h(z) = \sum_{n=1}^{\infty} \gamma_n z^n$$

be its power series representation on $\mathbb{D}$. Note that we have chosen the branch of $(1/2)\log[z^{-1}f(z)]$ that satisfies $h(0) = 0$ and with this stipulation, h is unique.

4.7 Milin's Conjecture. *If $f \in \mathcal{S}$, h is the branch of $(1/2)\log[z^{-1}f(z)]$ with $h(0) = 0$, and h satisfies (4.6), then*

$$\sum_{m=1}^{n}\sum_{k=1}^{m}(k|\gamma_k|^2 - \frac{1}{k}) \le 0.$$

If equality holds for some integer n, then f is a rotation of the Koebe function.

To show that Milin's conjecture implies the Robertson conjecture (and hence the Bieberbach conjecture), it is necessary to prove the Second Lebedev-Milin Inequality. This is the second in a collection of three inequalities that relate the power series coefficients of an analytic function with those of its exponential. All three inequalities will be stated and then it will be shown that Milin's conjecture implies Robertson's conjecture. Then the second inequality will be proved. After this the remaining inequalities will be derived for the interested reader.

Let ϕ be an analytic function in a neighborhood of 0 with $\phi(0) = 0$ and let

4.8
$$\phi(z) = \sum_{k=1}^{\infty} \alpha_k z^k$$

be its power series. Let

4.9
$$\psi(z) = e^{\phi(z)} = \sum_{k=0}^{\infty} \beta_k z^k.$$

4.10 First Lebedev-Milin Inequality. *If ϕ and ψ are as above, then*

$$\sum_{k=0}^{\infty} |\beta_k|^2 \le \exp\left\{\sum_{k=1}^{\infty} k|\alpha_k|^2\right\}.$$

If the right hand side is finite, then equality occurs if and only if there is a complex number γ with $|\gamma| < 1$ and $\alpha_k = \gamma^k/k$ for all $k \ge 1$.

4.11 Second Lebedev-Milin Inequality. *If ϕ and ψ are as above, then for all $n \ge 1$*

$$\sum_{k=0}^{n} |\beta_k|^2 \le (n+1)\exp\left\{\frac{1}{n+1}\sum_{m=1}^{n}\sum_{k=1}^{m}(k|\alpha_k|^2 - \frac{1}{k})\right\}.$$

Equality holds for a given integer n if and only if there is a complex number γ with $|\gamma| = 1$ and $\alpha_k = \gamma^k/k$ for $1 \le k \le n$.

4.12 Third Lebedev-Milin Inequality. *If ϕ and ψ are as above and $n \ne 1$,*

$$|\beta_n|^2 \le \exp\left\{\sum_{k=1}^{n}(k|\alpha_k|^2 - \frac{1}{k})\right\}.$$

Equality holds for some integer n if and only if there is a complex number γ with $|\gamma| = 1$ and $\alpha_k = \gamma^k/k$ for $1 \le k \le n$.

4.13 Theorem. *Milin's conjecture implies Robertson's conjecture.*

Proof. Let $g \in \mathcal{S}_-$ and let $f \in \mathcal{S}$ such that $g(z)^2 = f(z^2)$ on $\mathbb{D}$; assume that (4.1) and (4.3) hold. Let $h(z) = (1/2)\log[z^{-1}f(z)]$ satisfy (4.6). Note that if $z \in \mathbb{D} \setminus (-1, 0]$, $[g(\sqrt{z})/\sqrt{z}]^2 = f(z)/z$. On the other hand, $g(\sqrt{z})/\sqrt{z} = 1 + c_3 z + c_5 z^2 + \cdots$, so that $g(\sqrt{z})/\sqrt{z}$ is analytic on $\mathbb{D}$. Thus h is a branch of $\log[g(\sqrt{z})/\sqrt{z}]$ and so taking $c_1 = 1$ we get

$$\sum_{n=0}^{\infty} c_{2n+1} z^n = \exp\left\{\sum_{n=1}^{\infty} \gamma_n z^n\right\}.$$

According to the Second Lebedev-Milin Inequality, for each $n \ge 1$

$$\sum_{k=0}^{n} |c_{2k+1}|^2 \le (n+1)\exp\left\{\frac{1}{n+1}\sum_{m=1}^{n}\sum_{k=1}^{m}(k|\gamma_k|^2 - \frac{1}{k})\right\}.$$

Thus if Milin's conjecture is true, this implies that for every $n \geq 1$

$$\sum_{k=0}^{n} |c_{2k+1}|^2 \leq (n+1),$$

which we recognize as Robertson's conjecture.

Suppose $n \geq 1$ and equality holds in Robertson's conjecture. Again assuming the Milin conjecture, this implies

$$\begin{aligned} n+1 &= \sum_{k=0}^{n} |c_{2k+1}|^2 \\ &\leq (n+1)\exp\left\{\frac{1}{n+1}\sum_{m=1}^{n}\sum_{k=1}^{m}(k|\gamma_k|^2 - \frac{1}{k})\right\} \\ &\leq n+1. \end{aligned}$$

But this implies equality in Milin's conjecture and so f must be a rotation of the Koebe function. □

Now to prove the inequalities. A few preliminary observations are valid for each of the proofs. Since $\psi = e^{\phi}$, $\psi' = \phi' e^{\phi} = \phi'\psi$. Using the power series expansions of these functions we get

$$\begin{aligned} \sum_{k=1}^{\infty} k\beta_k z^{k-1} &= \left(\sum_{k=1}^{\infty} k\alpha_k z^{k-1}\right)\left(\sum_{k=0}^{\infty} \beta_k z^k\right) \\ &= \alpha_1\beta_0 + (\alpha_1\beta_1 + 2\alpha_2\beta_0)z + (\alpha_1\beta_2 + 2\alpha_2\beta_1 + 3\alpha_3\beta_0)z^2 \\ &\quad + \cdots + (\alpha_1\beta_m + \cdots + m\alpha_m\beta_0)z^m + \cdots . \end{aligned}$$

Equating corresponding coefficients gives

4.14
$$m\beta_m = \sum_{k=1}^{m} k\alpha_k\beta_{m-k}.$$

Proof of the Second Lebedev-Milin Inequality. Apply the Cauchy-Schwarz Inequality to (4.14) to get

4.15
$$\begin{aligned} m^2|\beta_m|^2 &\leq \left(\sum_{k=1}^{m} k^2|\alpha_k|^2\right)\left(\sum_{k=1}^{m} |\beta_{m-k}|^2\right) \\ &= \left(\sum_{k=1}^{m} k^2|\alpha_k|^2\right)\left(\sum_{k=0}^{m-1} |\beta_k|^2\right). \end{aligned}$$

Put

4.16
$$A_m = \sum_{k=1}^{m} k^2|\alpha_k|^2 \qquad B_m = \sum_{k=0}^{m} |\beta_k|^2.$$

So (4.15) becomes $m^2|\beta_m|^2 \leq A_m B_{m-1}$ for all $m \geq 1$. Now fix $n \geq 1$ and let's prove (4.11). Thus

$$\begin{aligned} B_n &= B_{n-1} + |\beta_n|^2 \\ &\leq B_{n-1}\left[1 + \frac{1}{n^2}A_n\right] \\ &= \frac{n+1}{n}\left[1 + \frac{A_n - n}{n(n+1)}\right] B_{n-1} \\ &\leq \frac{n+1}{n} \exp\left\{\frac{A_n - n}{n(n+1)}\right\} B_{n-1}, \end{aligned}$$

where the elementary inequality $1 + x \leq e^x$ has been used. Now apply this latest inequality to B_{n-1} and combine the two; so we have

$$B_n \leq \frac{n+1}{n-1} \exp\left\{\frac{A_n - n}{n(n+1)} + \frac{A_{n-1} - (n-1)}{(n-1)n}\right\} B_{n-2}.$$

Continuing and noting that $B_0 = |\beta_0|^2 = 1$ we get that

$$\begin{aligned} B_n &\leq (n+1)\exp\left\{\sum_{k=1}^{n} \frac{A_k - k}{k(k+1)}\right\} \\ &= (n+1)\exp\left\{\sum_{k=1}^{n} \frac{A_k}{k(k+1)} + 1 - \sum_{k=1}^{n+1}\frac{1}{k}\right\}. \end{aligned} \tag{4.17}$$

Now use the summation by parts formula (Exercise 2) with $x_k = [k(k+1)]^{-1}$ and $y_k = A_k$. Here $X_n = \sum_{k=1}^{n}[k(k+1)]^{-1} = 1 - (n+1)^{-1}$. This gives

$$\begin{aligned} \sum_{k=1}^{n} \frac{A_k}{k(k+1)} &= \sum_{k=1}^{n-1} X_k(A_k - A_{k+1}) + X_n A_n \\ &= \sum_{k=1}^{n-1}\left[1 - \frac{1}{k+1}\right](-(k+1)^2|\alpha_{k+1}^2|) \\ &\quad + \left[1 - \frac{1}{n+1}\right]\sum_{k=1}^{n} k^2|\alpha_k|^2 \\ &= \sum_{k=1}^{n} k|\alpha_k|^2 - \frac{1}{n+1}\sum_{k=1}^{n} k^2|\alpha_k|^2. \end{aligned}$$

Thus

$$\sum_{k=1}^{n} \frac{A_k}{k(k+1)} \quad + \quad 1 - \sum_{k=1}^{n+1}\frac{1}{k}$$

$$\begin{aligned}
&= \sum_{k=1}^{n} k|\alpha_k|^2 - \frac{1}{n+1}\sum_{k=1}^{n} k^2|\alpha_k|^2 + 1 - \sum_{k=1}^{n+1}\frac{1}{k} \\
&= \frac{1}{n+1}\left\{\sum_{k=1}^{n}(n+1)k|\alpha_k|^2 - \sum_{k=1}^{n} k^2|\alpha_k|^2 + n + 1\right. \\
&\qquad \left. - \sum_{k=1}^{n+1}\frac{n+1}{k}\right\} \\
&= \frac{1}{n+1}\sum_{k=1}^{n}\left\{(n+1)k|\alpha_k|^2 - k^2|\alpha_k|^2 + 1 - \frac{n+1}{k}\right\} \\
&= \frac{1}{n+1}\sum_{k=1}^{n}(n+1-k)\left(k|\alpha_K|^2 - \frac{1}{k}\right) \\
&= \frac{1}{n+1}\sum_{m=1}^{n}\sum_{k=1}^{m}\left(k|\alpha_k|^2 - \frac{1}{k}\right),
\end{aligned}$$

where we have used Exercise 3. If this is combined with (4.17), we have the Second Inequality.

Now for the case when we have equality in (4.11). (At this point the reader can go directly to the next section and begin to read the proof of the Milin conjecture and, hence, the Bieberbach conjecture. The remainder of this section is not required for that enterprise.) There were two factors that contributed to inequality in the above argument: the Cauchy-Schwarz Inequality and the inequality $1 + x \le e^x$. So if equality occurs, it must be that equality occurred whenever these two facts were used. The first such instance was when the Cauchy-Schwarz Inequality was applied to (4.14) in order to obtain (4.15). Note that for equality in (4.11) for an integer n, we need equality in (4.15) for $1 \le m \le n$. Thus there must exist constants $\lambda_1, \ldots, \lambda_n$ such that for each m, $1 \le m \le n$,

4.18 $$\beta_{m-k} = \lambda_m k \overline{\alpha_k}$$

for $1 \le k \le m$. Since $1 + x = e^x$ only when $x = 0$, an examination of the occurrence of this equality in the argument yields that $A_m = m$ for $1 \le m \le n$.

Substitution of (4.18) into (4.14) gives that $m\beta_m = \lambda_m \sum_{k=1}^{m} k^2|\alpha_k|^2 = \lambda_m A_m = m\lambda_m$. Thus $\beta_m = \lambda_m$ for $1 \le m \le n$. Since $\beta_0 = 1$, (4.18) for $k = m$ says that $\lambda_m m\overline{\alpha}_m = 1$ for $1 \le m \le n$. Thus for $m \ge 2$, $\lambda_1 = \beta_1 = \lambda_m(m-1)\overline{\alpha}_{m-1} = \lambda_m/\lambda_{m-1}$. Hence $\lambda_m = \lambda_1\lambda_{m-1}$, from which we derive that $\beta_m = \lambda_m = \lambda_1^m = \gamma^m$, where $\gamma = \lambda_1$. Equation 4.18 for $k = m$ implies that $m\alpha_m = \overline{\gamma^m}$. But for $1 \le k \le n$, $k = A_k = \sum_{m=1}^{k} m^2|\alpha_m|^2 = \sum_{m=1}^{k}|\gamma|^{2m}$. In particular it holds for $k = 1$ so that $|\gamma| = 1$. Hence (4.18) implies that for $1 \le k \le n$, $\gamma^{n-k} = \beta_{n-k} = \gamma^n k\overline{\alpha}_k$, so that $\alpha_k = \gamma^k/k$ for $1 \le k \le n$.

The proof that this condition suffices for equality is left to the reader. □

Proof of the First Inequality. Without loss of generality it can be assumed that the right hand side of (4.10) is finite. Apply the Cauchy-Schwarz Inequality to (4.14) in a different way than was done in the proof of the second inequality to get that $m^2|\beta_m|^2 \leq m\left(\sum_{k=1}^{m} m^2|\alpha_m|^2|\beta_{m-k}|^2\right)$, or

4.19
$$|\beta_m|^2 \leq \frac{1}{m}\sum_{k=1}^{m} m^2|\alpha_m|^2|\beta_{m-k}|^2.$$

Let $a_m = m|\alpha_m|^2$ and inductively define $b_0 = 1$ and

4.20
$$b_m = \frac{1}{m}\sum_{k=1}^{m} k a_k b_{m-k}.$$

An induction argument using (4.19) shows that $|\beta_m|^2 \leq b_m$ for all $m \geq 1$. If we examine how (4.14) was derived and look closely at (4.20), we see that

$$\sum_{k=0}^{\infty} b_k z^k = \exp\left\{\sum_{k=1}^{\infty} a_k z^k\right\},$$

where the hypothesis guarantees that these power series have radii of convergence at least 1. But since $a_k, b_k \geq 0$ we get

$$\begin{aligned}
\sum_{k=0}^{\infty} |\beta_k|^2 &\leq \sum_{k=0}^{\infty} b_k \\
&= \exp\left\{\sum_{k=1}^{\infty} a_k\right\} \\
&= \exp\left\{\sum_{k=1}^{\infty} k|\alpha_k|^2\right\},
\end{aligned}$$

which is the sought for inequality.

Now assume that $\sum_{k=1}^{\infty} k|\alpha_k|^2 < \infty$ and equality occurs. Clearly

$$\sum_{k=0}^{\infty} |\beta_k|^2 = \sum_{k=0}^{\infty} b_k.$$

Since $|\beta_k|^2 \leq b_k$ for all $k \geq 0$, it follows that $|\beta_k|^2 = b_k$ for all k. But this can only happen if for each $m \geq 1$ equality holds in (4.19). But this is an instance of equality holding in the Cauchy-Schwarz Inequality. Thus for every $m \geq 1$ there is a complex number λ_m such that

4.21
$$k\alpha_k\beta_{m-k} = \lambda_m \text{ for } 1 \leq k \leq m.$$

Letting $k = m$ here shows that $m\alpha_m = \lambda_m$ for all $m \geq 1$. Also if we substitute (4.21) into (4.14), we get that $\beta_m = m^{-1}(\lambda_m + \cdots + \lambda_m) = \lambda_m$. With these two identities,(4.21) becomes $\lambda_m = \lambda_{m-k}\lambda_k$. In particular, $\lambda_m = \lambda_{m-1}\lambda_1$. From here we get that $\lambda_m = \lambda_1^m$ for all $m \geq 1$. Thus putting $\gamma = \lambda_1$ we have that $\alpha_k = \gamma^k/k$ and $\beta_k = \gamma^k$ for all n. Because the right hand side of (4.10) is finite, it must be that $|\gamma| < 1$.

The proof that the condition suffices for equality is left to the reader. □

Proof of the Third Lebedev-Milin Inequality. Using the notation from the proof of the Second Lebedev-Milin Inequality, (4.15) states that

$$\begin{aligned} n^2|\beta_n|^2 &\leq A_n B_{n-1} \\ &\leq A_n n \exp\left\{\frac{1}{n}\sum_{m=1}^{n-1}\sum_{k=1}^{m}\left(k|\alpha_k|^2 - \frac{1}{k}\right)\right\}. \end{aligned}$$

Hence

$$\begin{aligned} |\beta_n|^2 &\leq \frac{A_n}{n}\exp\left\{\frac{1}{n}\sum_{m=1}^{n-1}\sum_{k=1}^{m}\left(k|\alpha_k|^2 - \frac{1}{k}\right)\right\} \\ &= \frac{A_n}{n}\exp\left\{\frac{1}{n}\sum_{k=1}^{n-1}(n-k)\left(k|\alpha_k|^2 - \frac{1}{k}\right)\right\} \\ &= \frac{A_n}{n}\exp\left\{\sum_{k=1}^{n-1}\left(k|\alpha_k|^2 - \frac{1}{k}\right) - \frac{1}{n}\sum_{k=1}^{n-1}(k^2|\alpha_k|^2 - 1)\right\} \\ &= \frac{A_n}{n}\exp\left\{\sum_{k=1}^{n-1}\left(k|\alpha_k|^2 - \frac{1}{k}\right) - \frac{A_n}{n} + \frac{1}{n}n^2|\alpha_n|^2 + \frac{n-1}{n}\right\} \\ &= e\frac{A_n}{n}\exp\left\{-\frac{A_n}{n} + \sum_{k=1}^{n}\left(k|\alpha_k|^2 - \frac{1}{k}\right)\right\}. \end{aligned}$$

Now apply the inequality $xe^{-x} \leq 1/e$ with $x = A_n/n$ and (4.12) appears.

The proof of the necessary and sufficient condition for equality in (4.12) is left to the reader. □

Exercises

1. Show that if f is a rotation of the Koebe function and $g(z)^2 = f(z^2)$, then we have equality in the Robertson and Milin conjectures for all n.

2. (The summation by parts formula) Show that if $\{x_k\}$ and $\{y_k\}$ are

two sequences of complex numbers and $X_n = \sum_{k=1}^n x_k$, then

$$\begin{aligned} \sum_{k=1}^n x_k y_k &= \sum_{k=1}^{n-1} X_k(y_k - y_{k+1}) + X_n y_n \\ &= \sum_{k=1}^{n} X_k(y_k - y_{k+1}) + X_n y_{n+1}. \end{aligned}$$

3. If $\{x_k\}$ is a sequence of complex numbers, show that $\sum_{k=1}^n (n+1-k)x_k = \sum_{m=1}^n \sum_{k=1}^m x_k$.
4. What are the functions ϕ and ψ for which equality holds in the First Lebedev-Milin Inequality?
5. Prove the necessary and sufficient condition for equality in (4.12).

§5 Some Special Functions

In this section certain special functions are introduced that were invented by de Branges for the proof of the Bieberbach conjecture. The properties of these functions are essential for the proof. This section will not be self-contained. Indeed, in order to deduce some of their crucial properties, we will also need to examine another collection of special functions, the Jacobi polynomials (defined below). Many of the properties of these functions can be found in Szegö [1959] and it will be left to the reader to ferret them out. One crucial property is a positivity result of Askey and Gasper [1976] that will not be proved here. The general attitude here will be that results about Jacobi polynomials will be quoted while the needed properties of the special functions introduced by de Branges will be proved.

The only self-contained exposition of these special functions of de Branges that I am aware of are some unpublished notes of Dov Aharonov [1984] that I used to prepare this section and for which I would like to publicly thank him.

5.1 Definition. For any choice of parameters α and $\beta > -1$, the *Jacobi polynomials* $\{p_n^{(\alpha,\beta)}\}_{n=0}^\infty$ are the unique polynomials having the following properties.

(a) $p_n^{(\alpha,\beta)}$ is a polynomial of degree n.

(b) For $w(x) = (1-x)^\alpha(a+x)^\beta$ and $n \neq m$,

$$\int_{-1}^1 p_n^{(\alpha,\beta)}(x)p_m^{(\alpha,\beta)}(x)w(x)dx = 0.$$

(c) $p_n^{(\alpha,\beta)}(0) = \binom{n+\alpha}{n} = \frac{(\alpha+1)_n}{n!}$, where for any number z and a non-negative integer n,

$$(z)_n = z(z+1)\cdots(z+n-1).$$

The fact that the Jacobi polynomials exist and are unique can be found in any standard reference. See, for example, Szegö [1959]. The proof of the next result can be found on pages 29 and 59 of that reference.

5.2 Proposition. *For all admissible α and β and $-1 \le x \le 1$,*

$$p_n^{(\alpha,\beta)}(x) = (-1)^n p_n^{(\beta,\alpha)}(-x)$$

for all $n \ge 0$.

5.3 Corollary. $p_n^{(\alpha,\beta)}(-1) = (-1)^n p_n^{(\beta,\alpha)}(1) = (-1)^n \binom{n+\beta}{n}$.

The next identity appears in the proof of Theorem 3 in Askey and Gasper [1976] (see page 717); its proof involves hypergeometric functions and won't be given here. The result following that is part of the statement of Theorem 3 in that reference.

5.4 Proposition. *If $\alpha > -1$ and $-1 \le x \le 1$, then for every $m \ge 0$*

$$\sum_{v=0}^{m} p_v^{(\alpha,0)}(x) = \sum_{j=0}^{m} \frac{\left[\frac{\alpha+1}{2}\right]_j \left[\frac{\alpha+2}{2}\right]_j (\alpha+2j+2)_{m-j}}{j!(\alpha+1)_j(m-j)!}[2(x-1)]^j.$$

5.5 Theorem. (Askey and Gasper [1976]) *If $\alpha \ge -1$ and $m \ge 0$, then*

$$\sum_{v=0}^{m} p_v^{(\alpha,0)}(x) > 0$$

for $-1 \le x \le 1$.

Now for the functions of de Branges. If $n \ge 1$ and $1 \le k \le n$, define for all $t \ge 0$

5.6 $$\tau_k(t) = k\sum_{v=0}^{n-k}(-1)^v \frac{(2k+v+1)_v(2k+2v+2)_{n-k-v}}{(k+v)v!(n-k-v)!}e^{-(v+k)t}$$

and $\tau_{n+1} \equiv 0$.

The relation between the Jacobi polynomials and the functions of de Branges is as follows.

5.7 Proposition. *For $1 \le k \le n$,*

$$\dot{\tau}_k(t) = -ke^{-kt}\sum_{v=0}^{n-k} p_v^{(2k,0)}(1-2e^{-t}).$$

Proof. From the definition of τ_k we compute

$$\frac{\dot{\tau}_k(t)}{k} = -\sum_{v=0}^{n-k}(-1)^v \frac{(2k+v+1)_v(2k+2v+2)_{n-k-v}}{v!(n-k-v)}e^{-(k+v)t}.$$

Hence

$$-\frac{\dot{\tau}_k(t)}{k}e^{kt} = \sum_{v=0}^{n-k}(-1)^v \frac{(2k+v+1)_v(2k+2v+2)_{n-k-v}}{v!(n-k-v)}e^{-vt}.$$

Now use (5.4) with $\alpha = 2k$ and $m = n-k$ to get

$$\sum_{v=0}^{n-k} p_v^{(2k,0)}(1-2e^{-t}) =$$

$$\sum_{v=0}^{n-k} \frac{\left[\frac{2k+1}{2}\right]_v \left[\frac{2k+2}{2}\right]_v (2k+2v+2)_{n-k-v}}{v!(2k+1)_v(n-k-v)!}[2(-2e^{-t})]^v.$$

But $2^{2v}\left[\frac{2k+2}{2}\right]_v (k+1)_v = (2k+1)_{2v}$ and $(2k+1)_{2v}/(2k+1)_v = (2k+v+1)_v$. Therefore

$$\begin{aligned}\sum_{v=0}^{n-k} p_v^{(2k,0)}(1-2e^{-t}) &= \sum_{v=0}^{n-k}(-1)^v \frac{(2k+v+1)_v(2k+2v+2)_{n-k-v}}{v!(n-k-v)!}e^{-vt} \\ &= -\frac{\dot{\tau}_k(t)}{k}e^{kt}.\end{aligned}$$

□

The next result contains all the information about these functions that will be used in the proof of de Branges's Theorem.

5.8 Theorem. *For the functions $\tau_1, \ldots, \tau_n$ defined in (5.6) and $\tau_{n+1} \equiv 0$, the following hold;*

5.9
$$\tau_k - \tau_{k+1} = -\left[\frac{\dot{\tau}_k}{k} + \frac{\dot{\tau}_{k+1}}{k+1}\right];$$

5.10
$$\tau_k(0) = n+1-k;$$

5.11
$$\tau_k(t) \to 0 \text{ as } t \to \infty;$$

5.12
$$\dot{\tau} < 0.$$

Proof. To prove (5.9) readers might increase their comfort level by first verifying the equality in the case that $k = n$ (and $\tau_{n+1} \equiv 0$). For $1 \le k < n$ it must be shown that $\tau_k + k^{-1}\dot{\tau}_k = \tau_{k+1} - (k+1)^{-1}\dot{\tau}_{k+1}$. To facilitate the proof, define $g_k = k^{-1}\tau_k e^{kt}$ and $h_k = k^{-1}\tau_k e^{-kt}$ for $1 \le k \le n+1$. These functions enter the picture by observing that

$$\dot{g}_k = \left[\frac{\dot{\tau}_k}{k} + \tau_k\right] e^{kt} \qquad\qquad \dot{h}_k = \left[\frac{\dot{\tau}_k}{k} - \tau_k\right] e^{-kt},$$

so to show (5.9) it suffices to show that

5.13
$$\dot{g}_k e^{kt} = -\dot{h}_{k+1} e^{(k+1)t}$$

for $1 \le k \le n$.

From the definitions of τ_k, g_k, and h_k we get that

$$\begin{aligned} g_k &= \sum_{v=0}^{n-k} (-1)^v \frac{(2k+v+1)_v (2k+2v+2)_{n-k-v}}{(k+v)v!(n-k-v)!} e^{-vt} \\ h_k &= \sum_{v=0}^{n-k} (-1)^v \frac{(2k+v+1)_v (2k+2v+2)_{n-k-v}}{(k+v)v!(n-k-v)!} e^{-vt-2kt}. \end{aligned}$$

Thus

$$\begin{aligned} e^{-kt}\dot{g}_k &= \sum_{v=1}^{n-k} (-1)^{v+1} \frac{(2k+v+1)_v (2k+2v+2)_{n-k-v}}{(k+v)(v-1)!(n-k-v)!} e^{-(k+v)t}, \\ e^{kt}\dot{h}_k &= \sum_{v=0}^{n-k} (-1)^{v+1} \frac{(v+2k)(2k+v+1)_v (2k+2v+2)_{n-k-v}}{(k+v)v!(n-k-v)!} e^{-(k+v)t} \end{aligned}$$

Now $(2k+v)(2k+v+1)_v = (2k+v)_{v+1}$. Incorporating this in the last equation and changing the index k to $k+1$, we get

$$\begin{aligned} & e^{(k+1)t}\dot{h}_{k+1} \\ &= \sum_{v=0}^{n-k-1} (-1)^{v+1} \frac{(2k+2+v)_{v+1}(2k+2v+4)_{n-k-1-v}}{(k+1+v)v!(n-k-1-v)!} e^{-(k+1+v)t} \\ &= \sum_{v=1}^{n-k} (-1)^v \frac{(2k+1+v)_v (2k+2v+2)_{n-k-v}}{(k+v)(v-1)!(n-k-v)!} e^{-(k+v)t} \end{aligned}$$

5.14
$$= e^{-kt}\dot{g}_k.$$

thus demonstrating (5.13), and hence (5.9).

To prove (5.10), first apply Corollary 5.3 and (c) of Definition 5.1 to obtain that $p_v^{(\alpha,0)}(-1) = (-1)^v$. Combine this with Proposition 5.7 to get

$$\dot{\tau}_k(0) = -k \sum_{v=0}^{n-k} p_v^{(2k,0)}(-1).$$

Thus

$$-\frac{\dot{\tau}_k(0)}{k} = \begin{cases} 1 & \text{if } n-k \text{ is even} \\ 0 & \text{if } n-k \text{ is odd.} \end{cases}$$

Now substitute this information into (5.9) to get that $\tau_k(0) - \tau_{k+1}(0) = 1$. Summing up yields (5.10).

The property (5.11) is clear from the definition of the functions and (5.12) is immediate from Theorem 5.5 and Proposition 5.7. □

§6 The Proof of de Branges's Theorem

The aim of this section is to prove the following, which is the culmination of this chapter. This approach is based on the paper of Fitzgerald and Pommerenke [1985].

6.1 Theorem. *The Milin conjecture is true. That is, if $f \in \mathcal{S}$, h is the branch of* $(1/2)\log[z^{-1}f(z)]$ *with* $h(0) = 0$, *and*

6.2
$$h(z) = \sum_{n=1}^{\infty} \gamma_n z^n,$$

then for all $n \geq 2$

6.3
$$\sum_{m=1}^{n}\sum_{k=1}^{m}\left(k|\gamma_k|^2 - \frac{1}{k}\right) \leq 0.$$

To accomplish this we first show that it suffices to prove the theorem for functions in $\mathcal{S}$ that map onto a slit region.

6.4 Proposition. *If $f \in \mathcal{S}$, then there is a sequence $\{f_n\}$ in $\mathcal{S}$ such that each f_n maps onto a slit region and $f_n \to f$ in $H(\mathbb{D})$.*

Proof. First we assume that $\Omega = f(\mathbb{D})$ is a Jordan region with its boundary parametrized by $\gamma : [0,1] \to \partial\Omega$, $\gamma(0) = \gamma(1) = \omega_0$. Replacing f by $e^{-i\theta}f(e^{i\theta})$ for a suitable θ, if necessary, we may assume that $|\omega_0| \geq |\omega|$ for all ω in cl Ω. Thus the ray $\eta = \{r\omega_0 : 1 \leq r \leq \infty\}$ in $\mathbb{C}_\infty$ meets cl Ω only at ω_0. Let $\Omega_n = \mathbb{C}_\infty \setminus [\eta \cup \{\gamma(t) : n^{-1} \leq t \leq 1\}]$ and let g_n be the Riemann map of $\mathbb{D}$ onto Ω_n with $g_n(0) = 0$ and $g_n'(0) > 0$.

Note that $\Omega_n \to \Omega$ in the sense of Definition 15.4.1. Thus Theorem 15.4.10 implies that $g_n \to f$. Thus $g_n(0) \to f(0) = 1$. So if $f_n = [g_n(0)]^{-1}g_n$, then $f_n \in \mathcal{S}$, $f_n(\mathbb{D})$ is a slit region, and $f_n \to f$.

Now assume that f is arbitrary. Put $r_n = 1 - n^{-1}$ and let $f_n(z) = r_n^{-1}f(r_n z)$. So $f_n \in \mathcal{S}$, $f_n \to f$, and $f_n(\mathbb{D})$ is a Jordan region. The proof of the special case implies that each f_n can be approximated by slit mappings in $\mathcal{S}$ and, thus, so can f. □

6.5 Corollary. *If Milin's conjecture is true for slit mappings in $\mathcal{S}$, it is true.*

Proof. If $f \in \mathcal{S}$, let $\{f_n\}$ be a sequence of slit mappings in $\mathcal{S}$ such that $f_n \to f$. For each n let $h_n(z) = (1/2)\log[f_n(z)/z]$ and let h be as in the statement of de Branges's Theorem. It is left to the reader to show that $h_n \to h$. Therefore the sequence of the k-th coefficients of the power series expansion of h_n converges to the k-th coefficient of the power series expansion of h, γ_k. Milin's conjecture now follows. □

Now to begin the path to the proof of de Branges's Theorem. To do this let us set the notation. For the remainder of the section, f is a slit mapping in $\mathcal{S}$ and F is the Loewner chain with $F_0 = f$. Thus Loewner's differential equation (3.6) holds for F. Observe that $e^{-t}F_t \in \mathcal{S}$ for all $t \geq 0$. Thus we can define

$$\begin{aligned} h(z,t) &= \frac{1}{2}\log\left[\frac{F(z,t)}{e^t z}\right] \\ &= \sum_{k=1}^{\infty} \gamma_k(t) z^k, \end{aligned} \tag{6.6}$$

where the branch of the logarithm is chosen with $h(0,t) = 0$.

The strategy of the proof is to introduce the function

$$\phi(t) = \sum_{k=1}^{n} \left[k|\gamma_k(t)|^2 - \frac{1}{k}\right] \tau_k(t) \tag{6.7}$$

for $t \geq 0$, where $\tau_1, \ldots, \tau_n$ are the special functions introduced in the preceding section. Given the function ϕ, we will prove the following.

6.8 Lemma. *If ϕ is the function defined in (6.7), then $\dot{\phi}(t) \geq 0$ for all $t > 0$.*

The proof of this lemma is the heart of the proof of the theorem. Indeed, the proof of de Branges's Theorem, except for the equality statement, easily follows once Lemma 6.8 is assumed.

Proof of (6.3). According to (5.10), $\tau_k(0) = n + 1 - k$ and so

$$\begin{aligned} \phi(0) &= \sum_{k=1}^{n} (n+1-k)(k|\gamma_k|^2 - \frac{1}{k}) \\ &= \sum_{m=1}^{n} \sum_{k=1}^{m} (k|\gamma_k|^2 - \frac{1}{k}) \end{aligned}$$

by Exercise 4.3. Also from (5.11) we know that $\tau_k(t) \to 0$ as $t \to \infty$ and so

$\phi(t) \to 0$ as $t \to \infty$. Therefore

6.9
$$\sum_{m=1}^{n}\sum_{k=1}^{m}(k|\gamma_k|^2 - \frac{1}{k}) = -\int_0^\infty \dot{\phi}(t)dt \le 0$$

by Lemma 6.8. □

The proof of the equality statement needs additional information about the function ϕ.

So now return to our assumptions for this section and the definition of the function h in (6.6).

6.10 Lemma. (a) *If* $0 < r < 1$, *then* $\sup\{|h(z,t)| : |z| \le r$ *and* $0 \le t < \infty\} < \infty$.
(b) *For each* $k \ge 1, \sup\{|\gamma_k(t)| : 0 \le t < \infty\} < \infty$.

Proof. (a) It suffices to get the bound for $|z| = r$. Using (2.11) we have for some integer N independent of r

$$\begin{aligned} |h(z,t)| &\le \frac{1}{2}\left[\log\left|\frac{F(z,t)}{ze^t}\right| + 2\pi N\right] \\ &= \frac{1}{2}\left[2\pi N - \log r + \log\left|\frac{F(z,t)}{e^t}\right|\right] \\ &\le \frac{1}{2}\left[2\pi N - \log r + \log\frac{r}{(1-r)^2}\right] \\ &= M_r. \end{aligned}$$

The Maximum Principle now gives the result.

(b) If $0 < r < 1$, then

$$\begin{aligned} \gamma_k(t) &= \frac{1}{2\pi i}\int_{|z|=r}\frac{h(z,t)}{z^{k+1}}dz \\ &= \frac{1}{2\pi r^k}\int_0^{2\pi} h(re^{i\theta},t)e^{ik\theta}d\theta. \end{aligned}$$

Thus $|\gamma_k(t)| \le r^{-k}M_r$ by part (a). □

6.11 Lemma. *For each* $k \ge 1$ *the function* $\gamma_k : [0,\infty) \to \mathbb{C}$ *is continuously differentiable and*

6.12
$$\dot{\gamma}_k(t) = \frac{1}{2\pi r^k}\int_0^{2\pi}\dot{h}(re^{i\theta},t)e^{-ik\theta}d\theta.$$

Proof. In fact this is an immediate consequence of the formula for $\gamma_k(t)$ obtained in the preceding proof and Leibniz's rule for differentiating under the integral sign. □

6.13 Lemma. *If $T < \infty$ and $0 < r < 1$, then the series $\sum_{k=1}^{\infty} \dot{\gamma}_k(t) z^k$ converges absolutely and uniformly for $|z| \le r$ and $0 \le t \le T$.*

Proof. Let $r < \rho < 1$. Equation (6.12) implies that if $|\dot{h}(w,t)| \le M$ for $0 \le t \le T$ and $|w| = \rho$, then for all $|z| \le r$, $|\dot{\gamma}_k(t) z^k| \le M(r/\rho)^k$. The result follows from the Weierstrass M-test. □

Proof of Lemma 6.8. The preceding lemma allows us to differentiate the series (6.6) for $h(z,t)$ term-by-term with respect to t. Thus, using Loewner's differential equation (3.6),

$$\begin{aligned}\sum_{k=1}^{\infty} \dot{\gamma}_k(t) z^k &= \frac{1}{2}\frac{\partial}{\partial t} \log \left| \frac{F(z,t)}{ze^t} \right| \\ &= \frac{1}{2}\left[\frac{\dot{F}(z,t)}{F(z,t)} - 1 \right] \\ &= \frac{1}{2}\left[z \frac{1 + x(t)z}{1 - x(t)z} \frac{F'(z,t)}{F(z,t)} - 1 \right]. \qquad (6.14)\end{aligned}$$

But $|x(t)z| = |z| < 1$ and so

$$\frac{1 + x(t)z}{1 - x(t)z} = 1 + 2\sum_{k=1}^{\infty} x(t)^k z^k.$$

Now we also have that

$$\begin{aligned}\sum_{k=1}^{\infty} k\gamma_k(t) z^{k-1} &= h'(z,t) \\ &= \frac{1}{2}\left[\frac{F'(z,t)}{F(z,t)} - \frac{1}{z} \right].\end{aligned}$$

Thus

$$\frac{F'(z,t)}{F(z,t)} = \frac{1}{z} + 2\sum_{k=1}^{\infty} k\gamma_k(t) z^{k-1}.$$

Substituting into (6.14) we get

$$\sum_{k=1}^{\infty} \dot{\gamma}_k(t) z^k = \frac{1}{2}\left\{ z \left[1 + 2\sum_{k=1}^{\infty} x(t)^k z^k \right] \left[\frac{1}{z} + 2\sum_{k=1}^{\infty} k\gamma_k(t) z^{k-1} \right] - 1 \right\}.$$

Therefore

$$1 + 2\sum_{k=1}^{\infty} \dot{\gamma}_k(t) z^k = \left[1 + 2\sum_{k=1}^{\infty} x(t)^k z^k \right] \left[1 + 2\sum_{k=1}^{\infty} k\gamma_k(t) z^k \right].$$

Equating coefficients gives

$$\dot{\gamma}_k(t) = k\gamma_k(t) + x(t)^k + 2\sum_{j=1}^{k-1} jx(t)^{k-j}\gamma_j(t).$$

Suppressing the dependence on t, this implies that

$$\begin{aligned}\dot{\gamma}_k &= x^k - k\gamma_k + 2\sum_{j=1}^{k} jx^{k-j}\gamma_j \\ &= x^k - k\gamma_k + 2x^k b_k,\end{aligned}$$

where $b_k(t) = \sum_{j=1}^{k} jx^{-j}(t)\gamma_j(t)$ for $k \geq 1$ (and $b_0 \equiv 0$). Now the fact that $k\gamma_k x^k = b_k - b_{k-1}$ implies

6.15 $$\dot{\gamma}_k = x^k[1 + b_k + b_{k-1}].$$

It is not hard to check that

$$\begin{aligned}\frac{d}{dt}k|\gamma_k(t)|^2 &= \frac{d}{dt}k\gamma_k(t)\overline{\gamma_k(t)} \\ &= 2\,\mathrm{Re}\ k\dot{\gamma}_k\overline{\gamma}_k \\ &= 2\,\mathrm{Re}\ kx^k[1 + b_k + b_{k-1}]\overline{\gamma}_k.\end{aligned}$$

Using the fact that $b_k - b_{k-1} = kx^{-k}\gamma_k$ we get that $kx^k\overline{\gamma}_k = (\overline{b}_k - \overline{b}_{k-1})$. Hence we can express the derivative entirely in terms of the functions b_k by

6.16 $$\frac{d}{dt}k|\gamma_k(t)|^2 = 2\,\mathrm{Re}\ [(\overline{b}_k - \overline{b}_{k-1})(1 + b_k + b_{k-1})].$$

Now consider the function ϕ defined in (6.7). Suppressing the dependence on t,

6.17 $$\dot{\phi} = \sum_{k=1}^{n} \tau_k \frac{d}{dt}[k|\gamma_k|^2] + \sum_{k=1}^{n} \dot{\tau}_k[k|\gamma_k(t)|^2 - \frac{1}{k}].$$

From (6.16) we get that

$$\begin{aligned}\psi &= \sum_{k=1}^{n} \tau_k \frac{d}{dt}[k|\gamma_k|^2] \\ &= \sum_{k=1}^{n} 2\,\mathrm{Re}\ [(\overline{b}_k - \overline{b}_{k-1})(1 + b_k + b_{k-1})]\tau_k.\end{aligned}$$

Now apply the summation by parts formula (Exercise 4.2 with $y_k = \tau_k$ and $x_k = 2\,\mathrm{Re}\ [(\overline{b}_k - \overline{b}_{k-1})(1 + b_k + b_{k-1})]$) to obtain that

6.18 $$\psi = \sum_{k=1}^{n} X_k(\tau_k - \tau_{k+1}),$$

where

$$\begin{aligned} X_m &\equiv \sum_{k=1}^{m} 2\,\mathrm{Re}\,[(\bar{b}_k - \bar{b}_{k-1})(1 + b_k + b_{k-1})] \\ &= 2\,\mathrm{Re}\sum_{k=1}^{m}(\bar{b}_k - \bar{b}_{k-1}) + 2\,\mathrm{Re}\sum_{k=1}^{m}(\bar{b}_k - \bar{b}_{k-1})(b_k + b_{k-1}). \end{aligned}$$

The first of the summands telescopes and for any complex numbers z and w, $(\bar{z} - \bar{w})(z + w) = |z|^2 - |w|^2 - 2i\,\mathrm{Im}\,(z\bar{w})$. Hence

$$\begin{aligned} X_m &= 2\,\mathrm{Re}\ b_m + 2\sum_{k=1}^{m}(|b_k|^2 - |b_{k-1}|^2) \\ &= 2[\mathrm{Re}\ b_m + |b_m|^2]. \end{aligned}$$

From (6.18) we get

$$\psi = 2\sum_{k-1}^{n}[\mathrm{Re}\ b_k + |b_k|^2](\tau_k - \tau_{k+1}).$$

Using (6.17) we now have

6.19 $$\dot{\phi} = 2\sum_{k=1}^{n}[\mathrm{Re}\ b_k + |b_k|^2](\tau_k - \tau_{k+1}) + \sum_{k=1}^{n}\dot{\tau}_k[k|\gamma_k(t)|^2 - \frac{1}{k}].$$

Focusing on the second summand, note that

$$\begin{aligned} \sum_{k=1}^{n}\dot{\tau}_k[k|\gamma_k(t)|^2 - \frac{1}{k}] &= \sum_{k=1}^{n}\frac{\dot{\tau}_k}{k}[k^2|\gamma_k|^2 - 1] \\ &= \sum_{k=1}^{n}\frac{\dot{\tau}_k}{k}[|b_k - b_{k-1}|^2 - 1]. \end{aligned}$$

For the first summand of (6.19) we use the property (5.9) of the functions τ_k to get

$$\begin{aligned} [\mathrm{Re}\ b_k + |b_k|^2](\tau_k - \tau_{k+1}) &= -[\mathrm{Re}\ b_k + |b_k|^2][\frac{\dot{\tau}_k}{k} + \frac{\dot{\tau}_{k+1}}{k+1}] \\ &= -[(\mathrm{Re}\ b_k)\frac{\dot{\tau}_k}{k} + |b_k|^2\frac{\dot{\tau}_k}{k} \\ &\quad +(\mathrm{Re}\ b_k)\frac{\dot{\tau}_{k+1}}{k+1} + |b_k|^2\frac{\dot{\tau}_{k+1}}{k+1}]. \end{aligned}$$

Now sum these terms for $1 \le k \le n$ and remember that $0 = b_0 = \tau_{n+1}$. Thus

$$\begin{aligned} \sum_{k=1}^{n}[\mathrm{Re}\ b_k + |b_k|^2](\tau_k - \tau_{k+1}) &= -\sum_{k=1}^{n}\frac{\dot{\tau}_k}{k}[\mathrm{Re}\ b_k + |b_k|^2 + \mathrm{Re}\ b_{k-1} \\ &\quad +|b_{k-1}|^2]. \end{aligned}$$

Thus

$$\begin{aligned}
\dot{\phi} &= -\sum_{k=1}^{n} \frac{\dot{\tau}_k}{k}[2\,\mathrm{Re}\; b_k + 2|b_k|^2 + 2\,\mathrm{Re}\; b_{k-1} + 2|b_{k-1}|^2 \\
&- |b_k - b_{k-1}|^2 + 1] \\
&= -\sum_{k=1}^{n} \frac{\dot{\tau}_k}{k}[|b_k|^2 + 2\,\mathrm{Re}\; b_k + 2\,\mathrm{Re}\; b_{k-1} + 2|b_{k-1}|^2 \\
&+ 2\,\mathrm{Re}\; b_k\bar{b}_{k-1} + 1]
\end{aligned}$$

6.20 $$= -\sum_{k=1}^{n} \frac{\dot{\tau}_k}{k}|b_k + b_{k-1} + 1|^2.$$

□

Proof of de Branges's Theorem. We already saw how to deduce (6.3) from Lemma 6.8. It only remains to treat the case of equality.

We show that if $f \in \mathcal{S}$ and f is not the Koebe function, then strict inequality must hold in (6.3) for all $n \geq 2$. If this is the case and $f(z) = z + a_2 z^2 + \cdots$, then $|a_2| < \alpha < 2$ (14.7.7). Now also assume that f is a slit mapping and adopt the notation used to prove Lemma 6.8. In particular, define the function h as in (6.6) and the functions b_k as in (6.15). Let $F_t(z)$ have power series expansion $e^t(z + a_2(t)z^2 + \cdots)$. So $|a_2(t)| \leq 2$ for all $t \geq 0$. A calculation shows that $\gamma_1(t) = a_2(t)/2$ and $b_k(t) = x(t)^{-1}\gamma_1(t)$. Thus (6.15) implies that

$$|\dot{\gamma}_1| = |1 + \frac{1}{2}x^{-1}a_2| \leq 2$$

and so

$$|\gamma_1(t)| = |\gamma_1 + \int_0^t \dot{\gamma}_1(s)ds| \leq \frac{\alpha}{2} + 2t.$$

Equation 6.20 and (5.12) imply that

$$\begin{aligned}
\dot{\phi}(t) &\geq (-\dot{\tau}_1)|b_1 + 1|^2 \\
&= (-\dot{\tau}_1)|x^{-1}\gamma_1 + 1|^2 \\
&\geq (-\dot{\tau}_1)(1 - \frac{\alpha}{2} - 2t)^2
\end{aligned}$$

for $0 \leq t \leq 4^{-1}(2-\alpha)$. From (6.9) we have

$$\begin{aligned}
\sum_{m=1}^{n}\sum_{k=1}^{m}(k|\gamma_k|^2 - \frac{1}{k}) &= -\int_0^\infty \dot{\phi}(t)dt \\
&\leq -\int_0^{(2-\alpha)/8} \dot{\phi}(t)dt
\end{aligned}$$

$$\leq \; -\int_0^{(2-\alpha)/8} (-\dot{\tau}_1)(1 - \frac{\alpha}{2} - 2t)^2 dt$$

6.21
$$\leq \; \left[\frac{2-\alpha}{4}\right] \int_0^{(2-\alpha)/8} \dot{\tau}_1 dt.$$

Since $\dot{\tau}_1 < 0$ everywhere, we get strict inequality in (6.3).

Now let f be an arbitrary function in the class $\mathcal{S}$ and let $\{f_j\}$ be a sequence of slit mappings in $\mathcal{S}$ that converge to f. Because $|a_2| < \alpha < 2$, it can be assumed that $|a_{j,2}| < \alpha$ for all $j \geq 1$. Thus the inequality in (6.21) holds for each function f_j (with γ_k replaced by the corresponding coefficient $\gamma_{j,k}$). This uniform bound on the sum (6.3) for the functions f_j implies the strict inequality for the limit function, f. □

Chapter 18

Some Fundamental Concepts from Analysis

Starting with this chapter it will be assumed that the reader is familiar with measure theory and something more than the basics of functional analysis. This particular chapter is an eclectic potpourri of results in analysis. Some topics fall into the category of background material and some can be labeled as material every budding analyst should know. Some of these subjects may be familiar to the reader, but we will usually proceed as though the material is new to all.

When needed, reference will be made to Conway [1990].

§1 Bergman Spaces of Analytic and Harmonic Functions

For an open subset G in $\mathbb{C}$ and $1 \le p \le \infty$, define $L^p(G)$ to be the L^p space of Lebesgue measure on G. That is, $L^p(G) = L^p(\mathcal{A}|G)$. In this section G will always denote an open subset of $\mathbb{C}$.

1.1 Definition. For $1 \le p \le \infty$ and an open subset G of $\mathbb{C}$, $L^p_a(G)$ is the collection of functions in $L^p(G)$ that are equal a.e. [Area] to an analytic function on G. Denote by $L^p_h(G)$ those elements of $L^p(G)$ that are equal a.e. [Area] to a harmonic function. These spaces are called the *Bergman spaces* of G because of the work of Bergman [1947], [1950].

Note that $L^p_h(G)$ contains $L^p_a(G)$ so anything proved about functions in $L^p_h(G)$ applies to the analytic Bergman space.

1.2 Lemma. *If f is a harmonic function in a neighborhood of the closed disk $\overline{B}(a;r)$, then*

$$f(a) = \frac{1}{\pi r^2} \int_{B(a;r)} f \, d\mathcal{A}.$$

This is, of course, a variation on the Mean Value Property of harmonic functions and can be proved by converting the integral to polar coordinates and applying that property.

1.3 Proposition. *If* $1 \le p < \infty$, $f \in L^p_h(G)$, $a \in G$, and $0 < r <$

dist$(a, \partial G)$, *then*

$$|f(a)| \leq \frac{1}{\pi^{1/p} r^{2/p}} ||f||_p.$$

Proof. Let q be the index that is conjugate to $p : 1/p + 1/q = 1$. By the preceding lemma and Hölder's Inequality, $|f(a)| = (\pi r^2)^{-1} \left| \int_{B(a;r)} f d\mathcal{A} \right| \leq (\pi r^2)^{-1} \left(\int_{B(a;r)} |f|^p d\mathcal{A} \right)^{1/p} \left(\int_{B(a;r)} 1 \, d\mathcal{A} \right)^{1/q} \leq \pi r^2)^{-1} ||f||_p (\pi r^2)^{1/q} = (\pi r^2)^{-1/p} ||f||_p$. □

1.4 Proposition. *For* $1 \leq p \leq \infty$, $L^p_h(G)$ *and* $L^p_a(G)$ *are Banach spaces and* $L^2_h(G)$ *and* $L^2_a(G)$ *are Hilbert spaces. If* $a \in G$, *the linear functional* $f \to f(a)$ *is bounded on* $L^p_h(G)$ *and* $L^p_a(G)$.

Proof. The last statement in the proposition is an immediate consequence of Proposition 1.3 for the case $p < \infty$, and it is a consequence of the definition for the case $p = \infty$. For the first statement, it must be shown that $L^p_h(G)$ and $L^p_a(G)$ are complete; equivalently, $L^p_h(G)$ and $L^p_a(G)$ are closed in $L^p(G)$. In the case that $p = \infty$, this is clear; so assume that $1 \leq p < \infty$. Only the space $L^p_h(G)$ will be treated; the analytic case will be left to the reader as it is analogous. Let $\{f_n\} \subseteq L^p_h(G)$ and suppose $f_n \to f$ in $L^p(G)$; without loss of generality we can assume that $f_n(z) \to f(z)$ a.e. Let K be a compact subset of G and let $0 < r <$ dist$(K, \partial G)$. By Proposition 1.3 there is a constant C such that $|h(z)| \leq C||h||_p$ for every h in $L^p_h(G)$ and every z in K. In particular, $|f_n(z) - f_m(z)| \leq C||f_n - f_m||_p$ for all m, n. Thus $\{f_n\}$ is a uniformly Cauchy sequence of harmonic functions on K. Since K is arbitrary, there is a harmonic function g on G such that $f_n(z) \to g(z)$ uniformly on compact subsets of G. It must be that $f(z) = g(z)$ a.e. and so $f \in L^p_h(G)$. □

The space $L^\infty(G)$ is the dual of the Banach space $L^1(G)$ and as such it has a weak-star (abbreviated weak* or wk*) topology. It can also be shown that $L^\infty_n(G)$ and $L^\infty_a(G)$ are weak* closed in $L^\infty(G)$. See Exercise 1.

This section concludes by proving some theorems on approximation by polynomials and rational functions in Bergman spaces of analytic functions.

1.5 Definition. For a bounded open set G and $1 \leq p < \infty$, let $P^p(G)$ be the closure of the polynomials in $L^p(G)$. $R^p(G)$ is the closure of the set of rational functions with poles off G that belong to $L^p(G)$.

It follows that $P^p(G) \subseteq R^p(G) \subseteq L^p_a(G)$. Note that if r is a rational function with poles off cl G, then $r \in L^p(G)$. However in the definition of $R^p(G)$, the rational functions are allowed to have poles on ∂G as long as the functions belong to $L^p(G)$. If $G =$ the punctured unit disk, then z^{-1} has its poles off G but does not belong to $L^2(G)$ even though it does belong to $L^1(G)$. If, on the other hand, $G = \{z = x + iy : 0 < x < 1 \text{ and } |y| < \exp(-x^{-2})\}$, then $z^{-1} \in L^2(G)$.

It is not difficult to construct an example of a set G for which $P^p(G) \neq L^p_a(G)$. This is the case for an annulus since $z^{-1} \in L^p_a(G)$ but cannot be approximated by polynomials. Finding a G with $R^p(G) \neq L^p_a(G)$ is a little more difficult. Indeed if $1 \leq p < 2$, then $R^p(G) = L^p_a(G)$, while there are regions G such that for $2 \leq p < \infty$ equality does not hold (Hedberg [1972 a]). See the remarks at the end of this section for more information.

If K is a compact subset of $\mathbb{C}$, then the open set $\mathbb{C} \setminus K$ has at most a countable number of components, exactly one of which is unbounded. Call the boundary of this unique unbounded component of $\mathbb{C} \setminus K$ the *outer boundary* of K. Note that the outer boundary of K is a subset of ∂K. In fact, the outer boundary of K is precisely $\partial \widehat{K}$, the boundary of the polynomially convex hull of K. For a small amount of literary economy, let's agree that for a bounded open set G the outer boundary of G is that of its closure and the polynomially convex hull of G is $\widehat{G} \equiv \widehat{\text{cl}\, G}$.

1.6 Definition. A *Carathéodory region* is a bounded open connected subset of $\mathbb{C}$ whose boundary equals its outer boundary.

1.7 Proposition. *If G is a Carathéodory region, then G is a component of* $\text{int}\{\widehat{G}\}$ *and hence is simply connected.*

Proof. Let $K = \widehat{G}$ and let H be the component of int K that contains G; it must be shown that $H = G$. Suppose there is a point z_1 in $H \setminus G$ and fix a point z_0 in G. Let $\gamma : [0,1] \to H$ be a path such that $\gamma(0) = z_0$ and $\gamma(1) = z_1$. Put $\alpha = \inf\{t : \gamma(t) \in H \setminus G\}$. Thus $0 < \alpha \leq 1$ and $\gamma(t) \notin H \setminus G$ for $0 \leq t < \alpha$. Since $H \setminus G$ is relatively closed in H, $w = \gamma(\alpha) \notin G$. Thus $w \in \partial G$. But since G is a Carathéodory region, $\partial G = \partial K$. Hence $w \in \partial K$. But $w \in H \subseteq \text{int } K$, a contradiction.

It is left as an exercise for the reader to show that the components of the interior of any polynomially convex subset of C are simply connected. (See Proposition 13.1.1.) □

There are simply connected regions that are not Carathéodory regions; for example, the slit disk. Carathéodory regions tend to be well behaved simply connected regions, however there can be some rather bizarre ones.

1.8 Example. A *cornucopia* is an open ribbon G that winds about the unit circle so that each point of $\partial\, \mathbb{D}$ belongs to ∂G. (See Figure 18.1.)

If G is the cornucopia, then cl G consists of the closed ribbon together with $\partial\, \mathbb{D}$. Hence $\mathbb{C} \setminus \text{cl } G$ has two components: the unbounded component and $\mathbb{D}$. Nevertheless G is a Carathéodory region.

1.9 Proposition. *If G is a Carathéodory region, then* $G = \text{int}\{\text{cl } G\}$. *If G is a simply connected region such that* $G = \text{int}\{\text{cl } G\}$ *and* $\mathbb{C} \setminus \text{cl } G$ *is*

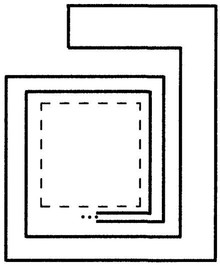

Figure 18.1.

connected, then G is a Carathéodory region.

Proof. Exercise. □

1.10 Lemma. *If $1 < p < \infty$ and $\{f_n\}$ is a sequence in $L^p_a(G)$, then $\{f_n\}$ converges weakly to f if and only if $\sup_n \|f_n\|_p < \infty$ and $f_n(z) \to f(z)$ for all z in G.*

Proof. If $f_n \to f$ weakly, then $\sup_n \|f_n\|_p < \infty$ by the Principle of Uniform Boundedness. In light of Proposition 1.4, for each z in G there is a function k_z in $L^q(G)$ such that $g(z) = \langle g, k_z\rangle$ for all g in $L^p_a(G)$. (Here q is the index that is conjugate to $p : 1/p + 1/q = 1$.) Thus $f_n(z) = \langle f_n, k_z\rangle \to \langle f, k_z\rangle = f(z)$. If τ is the topology of pointwise convergence on $L^p_a(G)$, then we have just seen that the identity map $i : (L^p_a(G), \text{weak}) \to (L^p_a(G), \tau)$ is continuous. Since (ball $L^p_a(G)$, weak) is compact and τ is a Hausdorff topology, i must be a homeomorphism. □

1.11 Theorem. (Farrell [1934] and Markusevic [1934]) *If G is a bounded Carathéodory region and $1 < p < \infty$, then $P^p(G) = L^p_a(G)$.*

Proof. Let $K = \widehat{G}$ and let $\tau : \mathbb{D} \to \mathbb{C}_\infty \setminus K$ be a Riemann map with $\tau(0) = \infty$. Put $G_n = \mathbb{C} \setminus \tau(\{z : |z| \le 1 - 1/n\})$. It is left to the reader to show that the sequence $\{G_n\}$ converges to G in the sense of Definition 15.4.1. So fix a in G and let ϕ_n be the Riemann map of G onto G_n with $\phi_n(a) = a$ and $\phi_n'(a) > 0$. By Theorem 15.4.10, $\phi_n(z) \to z$ uniformly on compact subsets of G. Let $\psi_n = \phi_n^{-1} : G_n \to G$. Fix f in $L^p_a(G)$ and put $f_n = (f \circ \psi_n)\psi_n$. Thus f_n is analytic in a neighborhood of K and so, by Runge's Theorem, f_n can be approximated uniformly on K by polynomials. Thus $f_n|G \in P^p(G)$.

Also $\|f_n\|^p \le \int_{G_n} |f_n|^p d\mathcal{A} = \int_{G_n} |f_m\psi_m|^p|\psi_n'|^p d\mathcal{A} = \int_G |f|^p d\mathcal{A} = \|f\|^p$ by the change of variables formula for area integrals. If $z \in G, \psi_n(z) \to z$ and $\psi_n'(z) \to 1$. Therefore $f_n(z) \to f(z)$ as $n \to \infty$. By Lemma 1.10, $f_n \to f$ weakly and so $f \in P^p(G)$. □

Rubel and Shields [1964] prove that if G is a bounded open set whose boundary coincides with the boundary of its polynomially convex hull and if $f \in L^\infty_a(G)$, then there exists a sequence of polynomials $\{p_n\}$ such that $||p_n||_G \leq ||f||_G$ and $p_n(z) \to f(z)$ uniformly on compact subsets of G. (Note that this condition on G is the same as the condition for a region to be a Carathéodory region, but G is not assumed to be connected here.) In particular, one can approximate with polynomials the bounded analytic function that is 1 on the open unit disk and 0 on the cornucopia. This says that Theorem 1.11 is true for $p = \infty$ if the weak* topology is used instead of the norm topology. The theorem also holds when $p = 1$ but a different proof is needed. See Bers [1965] and Lindberg [1982].

Some hypothesis is needed in Theorem 1.11 besides the simple connectedness of G. For example, if $G = \mathbb{D}\backslash(-1, 0]$, then $z^{1/2} \in L^2_a$ but $z^{1/2} \notin P^2(G)$. In fact it is not difficult to see that the functions in $P^2(G)$ are precisely those functions in $L^2_a(G)$ that have an analytic continuation to $\mathbb{D}$.

An exact description of the functions in $P^p(G)$ is difficult, though many properties of these functions can be given. Exercise 4 shows that if G is an annulus, then every f in $P^p(G)$ has an analytic extension to the open disk. In general, if U is a bounded component of $\mathbb{C} \setminus [\text{cl } G]$ such that ∂U is disjoint from the outer boundary of G, then every function in $P^p(G)$ has an analytic extension to $G \cup [\text{cl } U]$ that belongs to $P^p(G \cup [\text{cl } U])$, though the norm of the extension is larger.

What happens if U is a bounded component of $\mathbb{C} \setminus [\text{cl } G]$ and ∂U meets the outer boundary of G? The answer to this question is quite complex and the continuing subject of research. See Mergeljan [1953], Brennan [1977], and Cima and Matheson [1985].

The next theorem can be proved by reasoning similar to that used to prove Theorem 1.11. See Mergeljan [1953] for details.

1.12 Theorem. *Let G be a bounded region in $\mathbb{C}$ such that $\mathbb{C} \setminus [\text{cl } G]$ has bounded components $U_1, \ldots, U_m$. Let $K_j = \partial U_j$ and let K_0 be the outer boundary; assume $K_i \cap K_j = \emptyset$ for $i \neq j$ and fix a point z_j in U_j, $1 \leq j \leq m$. If $f \in L^p_a(G)$, then there is a sequence $\{f_n\}$ of rational functions with poles in $\{\infty, z_1, \ldots, z_m\}$ such that $f_n \to f$ in $L^p_a(G)$. In particular, $R^p(G) = L^p_a(G)$.*

We return to the subject of Bergman spaces in §21.9.

Remarks. There is a substantial literature on the subjects covered in this section. Indeed, we have only skimmed the surface of the theory. Bers [1965] shows that $R^1(G) = L^1_a(G)$. Mergeljan [1953] has the results of this section and more. Brennan [1977] discusses polynomial approximation when the underlying region is not a Carathéodory region. The interested reader can also consult Bagby [1972], Cima and Matheson [1985], Hedberg [1972a], [1972b], [1993], and Lindberg [1982].

Exercises

1. This exercise will show that $L^\infty_h(G)$ is weak* closed in $L^\infty(G)$. A similar proof works for $L^\infty_a(G)$. (a) If $a \in G$, let $0 < 2r < \text{dist}(a, \partial G)$. For $|w-a| < r$ put $g_w = (\pi r^2)^{-1}\chi_{B(w;r)}$. Show that the map $w \to g_w$ is a continuous map from $B(a;r)$ into $L^1(G)$. (b) Let $\mathcal{X}$ be the weak* closure of $L^\infty_h(G)$; so $\mathcal{X}$ is the Banach space dual of $L^1(G)/L^\infty_h(G)^\perp$. Show that if $\{f_n\}$ is a sequence in $L^\infty_h(G)$ and $f_n \to f$ weak* in $\mathcal{X}$, then $\{f_n\}$ is a uniformly Cauchy sequence on compact subsets of G and hence $f \in L^\infty_h(G)$. Now use the Krein-Smulian Theorem (Conway [1990], V.12.7) to conclude that $L^\infty_h(G)$ is weak* closed.

2. If f is analytic in the punctured disk $G = \{z : 0 < |z| < 1\}$, for which values of p does the condition $\int_G |f|^p d\mathcal{A} < \infty$ imply that f has a removable singularity at 0?

3. Give an example of a simply connected region G that is not a Carathéodory region but satisfies $G = \text{int}\{\text{cl } G\}$.

4. If G is a bounded open set in $\mathbb{C}$ and K is a compact subset of G, then every function f in $P^p(G \setminus K)$ has an analytic continuation to G that belongs to $P^p(G)$. Show that if G is connected, then the restriction map $f \to f|(G \setminus K)$ is a bijection of $P^p(G)$ onto $P^p(G \setminus K)$.

5. Let $\{a_n\}$ be an increasing sequence of positive numbers such that $1 = \lim_n a_n$. Choose $r_1, r_2, \ldots,$ such that the closed balls $B_n = \overline{B}(a_n; r_n)$ are pairwise disjoint and contained in $\mathbb{D}$; put $G = \mathbb{D} \setminus \cup_n B_n$. Show that each f in $P^p(G)$ has an analytic continuation to $\mathbb{D}$. Must this continuation belong to $L^p_a(\mathbb{D})$?

6. Let $\{a_n\}$ be a decreasing sequence of positive numbers such that $0 = \lim_n a_n$. Choose $r_1, r_2, \ldots,$ such that the closed balls $B_n = \overline{B}(a_n; r_n)$ are pairwise disjoint and contained in $\mathbb{D}$; put $G = \mathbb{D} \setminus \cup_n B_n$. Show that $R^2(G) = L^2_a(G)$.

§2 Partitions of Unity

In this section (X, d) is a metric space that will shortly be restricted. We are most interested in the case where X is an open subset of $\mathbb{C}$, but we will also be interested when X is a subset of $\mathbb{R}$ or $\partial\,\mathbb{D}$. In the next section we will examine the abstract results of this section for the case of an open subset of $\mathbb{C}$ and add some differentiable properties to the functions obtained for metric spaces.

The idea here is to use the fact that metric spaces are paracompact, terminology that will not be used here but is mentioned for the circumspective.

2.1 Definition. If $\mathcal{U}$ is an open cover of X, a *refinement* of $\mathcal{U}$ is an open cover $\mathcal{V}$ such that for each set V in $\mathcal{V}$ there is an open set U in $\mathcal{U}$ with $V \subseteq U$.

Notice that this extends the notion of a subcover. Also note that the relation of being a refinement is transitive, and that every open cover of a metric space has a refinement consisting of open balls. The typical application of this idea is to manufacture an open cover of the metric space that has certain desirable features. Because of a lack of compactness it is impossible to obtain a finite subcover, but we can always pass to a locally finite refinement (now defined).

2.2 Definition. An open cover $\mathcal{U}$ of X is said to be *locally finite* if for each $B(a;r)$ contained in X, $U \cap B(a;r) = \emptyset$ for all but a finite number of sets U in $\mathcal{U}$.

It is a standard fact from topology that every open cover of a metric space has a locally finite refinement (that is, every metric space is *paracompact*). This will be proved for metric spaces that satisfy an additional hypothesis that will facilitate the proof and be satisfied by all the examples that will occupy us in this book. See Exercise 1.

2.3 Theorem. *If X is the union of a sequence of compact sets $\{K_n\}$ such that $K_n \subseteq \operatorname{int} K_{n+1}$ for all $n \geq 1$, then every open cover of X has a locally finite refinement consisting of a countable number of open balls.*

Proof. Let $\mathcal{U}$ be the given open cover of X. For each n let $R_n = \operatorname{dist}(K_n, X \setminus \operatorname{int} K_{n+1})$. For each integer n we will manufacture a finite collection of balls $\mathcal{B}_n$ that will cover K_n and have some additional properties. These extra properties don't come into play until we reach $n = 3$.

For each point a in K_1 choose a radius r with $r < R_1$ such that $B(a;r)$ is contained in some open set from $\mathcal{U}$. By compactness we can find a finite collection $\mathcal{B}_1$ of these balls that cover K_1. Similarly let $\mathcal{B}_2$ be a finite collection of balls that cover $K_2 \setminus \operatorname{int} K_1$, with centers in $K_2 \setminus \operatorname{int} K_1$, and with radii less than R_2 and sufficiently small that the ball is contained in some set from $\mathcal{U}$. For $n \geq 3$ let $\mathcal{B}_n$ be a finite collection of balls that cover $K_n \setminus \operatorname{int} K_{n+1}$ and such that each ball in $\mathcal{B}_n$ has the form $B(a;r)$ with a in $K_n \setminus \operatorname{int} K_{n+1}$ and r chosen so that $B(a;r)$ is contained in some set from $\mathcal{U}$ and $r < \min\{R_{n-2}, R_n\}$. Note that $\mathcal{V} = \cup_n \mathcal{B}_n$ is a refinement of $\mathcal{U}$. It is left to the reader to verify that if $B \in \mathcal{B}_n$ and $B \cap K_m \neq \emptyset$, then $m = n-1, n,$ or $n+1$. Since X is the union of $\{\operatorname{int} K_n\}$, this shows that $\mathcal{V}$ is locally finite. □

2.4 Proposition. *If K is a closed subset of the metric space X, $\{U_1, \}.$ $\ldots, U_m\}$ is an open cover of K, and W is an open set containing K, then there are continuous functions $f_1, \ldots, f_m$ such that:*

(a) *for* $1 \leq j \leq m$, $0 \leq f_j \leq 1$ *and support* $f_j \subseteq U_j \cap W$*;*

(b) $\sum_{j=1}^m f_j(x) = 1$ *for all* x *in* K.

Proof. We may assume that $\{U_1, \ldots, U_m\}$ is a minimal cover of K; that is, no proper collection is a cover. The proof proceeds by induction. The case $m = 1$ is just Urysohn's Lemma. For $m = 2$, Urysohn's Lemma implies there are continuous functions f and g on X such that each takes its values in $[0, 1]$, $f(x) = 1$ for x in $K \setminus U_2$, $f(x) = 0$ for x in $K \setminus U_1$, $g(x) = 1$ for x in K, and $g(x) = 0$ for x in $X \setminus W$. Put $f_1 = fg$ and $f_2 = (1 - f)g$. It is left to the reader to verify that these functions satisfy (a) and (b) for $m = 2$.

Now suppose the proposition holds for some $m \geq 2$ and all metric spaces, and assume $\{U_1, \ldots, U_{m+1}\}$ is a minimal open cover of K. Put $F = K \setminus U_{m+1}$ and pick an open set G in X such that $F \subseteq G \subseteq \text{cl } G \subseteq U \equiv \cup_{j=1}^m U_j$. By the induction hypothesis there are continuous functions $h_1, \ldots, h_m$ such that for $1 \leq j \leq m$, $0 \leq h_j \leq 1$, support $h_j \subseteq U_j \cap W$, and $\sum_{j=1}^m h_j(x) = 1$ for all x in cl G. Also since we know the proposition holds for $m = 2$, we can find continuous functions g_1 and g_2 with $0 \leq g_1, g_2 \leq 1$, support $g_1 \subseteq G \cap W$, support $g_2 \subseteq U_{m+1} \cap W$, and $g_1(x) + g_2(x) = 1$ for all x in K. Put $f_j = g_1 h_j$ for $1 \leq j \leq m$ and $f_{m+1} = g_2$. The reader can check that these functions satisfy conditions (a) and (b). □

2.5 Definition. A collection of continuous functions $\{\phi_j\}$ on X is a *partition of unity* if:

(a) for each j, $0 \leq \phi_j \leq 1$;

(b) the collection of sets $\{\{x : \phi_j(x) > 0\}\}_j$ is a locally finite cover of X;

(c) $\sum_j \phi_j(x) = 1$ for all x in X.

If $\mathcal{U}$ is a given open cover of X, then the partition of unity $\{\phi_j\}$ is said to be *subordinate to* $\mathcal{U}$ provided the cover $\{\{x : \phi_j(x) > 0\}\}_j$ is a refinement of $\mathcal{U}$.

Two observations should be made. The first is that the collection of functions in the definition is not assumed to be countable, let alone finite, though in the applications that we will see in this book it will be at most countably infinite. The second observation is that condition (b) of the definition implies that the sum that appears in (c) has only a finite number of non-zero terms, and so no questions about convergence are necessary.

Like Theorem 2.3, the next result is valid for all metric spaces (except for the restriction that the partition of unity be countable), but we prove it here only for the metric spaces we will encounter in this book.

2.6 Theorem. *If* X *is the union of a sequence of compact sets* $\{K_n\}$ *such that* $K_n \subseteq \text{int } K_{n+1}$ *for all* $n \geq 1$ *and* $\mathcal{U}$ *is an open cover of* X*, then there is a countable partition of unity* $\{\phi_j\}$ *subordinate to* $\mathcal{U}$.

Proof. According to Theorem 2.3 there is a countable cover $\mathcal{B}$ of X by open balls that is subordinate to $\mathcal{U}$. Set $K_0 = \emptyset$ and for $n \geq 1$ let $\mathcal{B}_n$ be a finite subcollection of $\mathcal{B}$ that covers $K_n \setminus \text{int } K_{n-1}$. Arrange matters so that $\mathcal{B}_n \cap \mathcal{B}_m = \emptyset$ for $n \neq m$. Let $\mathcal{B}_n = \{U_{nk} : 1 \leq k \leq p_n\}$.

According to Proposition 2.4 for each $n \geq 1$ there are continuous functions $\{f_{nk} : 1 \leq k \leq p_n\}$ such that:

(i) for $1 \leq k \leq p_n$, $0 \leq f_{nk} \leq 1$ and support $f_{nk} \subseteq U_{nk} \cap \text{int } K_{n+1}$;

(ii) $\sum_{k=1}^{p_n} f_{nk}(x) = 1$ for all x in $K_n \setminus \text{int } K_{n-1}$.

If the set where f_{nk} is not zero is denoted by N_{nk}, then it is apparent that $\{N_{nk} : 1 \leq k \leq p_n \text{ and } n \geq 1\}$ is a locally finite cover of X that is subordinate to $\mathcal{B}$ (and hence to $\mathcal{U}$). Thus $f(x) \equiv \sum_{n=1}^{\infty} \sum_{k=1}^{p_n} f_{nk}(x)$ is a well defined continuous function on X and $f(x) \geq 1$ for all x in X.

Define $\phi_{nk}(x) = f_{nk}(x)/f(x)$ for x in X. Clearly ϕ_{nk} is continuous, $0 \leq \phi_{nk} \leq 1$, support $\phi_{nk} =$ support f_{nk}, and $\sum_{n=1}^{\infty} \sum_{k=1}^{p_n} \phi_{nk}(x) = 1$ for all x in X. That is, $\{\phi_{nk}\}$ is a partition of unity. Since $\{N_{nk}\}$ forms a locally finite cover of X, $\{\phi_{nk}\}$ is locally finite. □

Exercises

1. Show that a metric space that satisfies the hypothesis of Theorem 2.3 is locally compact. Conversely, a locally compact, σ-compact metric space satisfies the hypothesis of Theorem 2.3.

2. If X is a locally compact metric space and $\mathcal{U}$ is an open cover of X, then there is a countable partition of unity $\{\phi_j\}$ subordinate to $\mathcal{U}$ such that each function ϕ_j has compact support.

3. Suppose Z is an arbitrary Hausdorff space that is locally metrizable; that is, for each z in Z there is an open neighborhood U of z such that the relative topology on U is metrizable. Show that if every open cover of Z has a locally finite refinement, then Z is metrizable.

§3 Convolution in Euclidean Space

In this section a few basic facts about convolution in Euclidean space are presented. In the course of this book convolution on the circle also will be encountered. At the end of this section the definitions and results for the circle are presented without proof. Of course these both come under the general subject of convolution on a locally compact group, but this level of generality is inappropriate here.

Recall that an extended real-valued regular Borel measure μ is defined on all the Borel sets, is finite on compact sets, and its variation satisfies the usual regularity conditions: for any Borel set E, $|\mu|(E) = \inf\{|\mu|(U) : U \text{ is open and } E \subseteq U\} = \sup\{|\mu|(K) \text{ is compact and } K \subseteq E\}$. If μ is such a measure, then it admits a Jordan decomposition $\mu = \mu_+ - \mu_-$, where μ_+ and μ_- are positive regular Borel measures that are carried by disjoint sets. If μ is finite valued, then it is bounded with total variation $||\mu|| = |\mu|(\mathbb{R}) < \infty$. Our principal interest will be when d is 1 or 2, but specialization to these dimensions does not make the discussion simpler. If μ is extended real-valued, then either μ_+ or μ_- is bounded. An extended complex-valued regular Borel measure is one such that both its real and imaginary parts are extended real-valued regular Borel measures.

For any open subset G of $\mathbb{R}$, $C_c(G)$ denotes the linear space of continuous functions on G with compact support. Note that this is norm dense in the Banach space $C_0(G)$ of continuous functions that vanish at infinity. The space $C_c(\mathbb{R})$ will be abbreviated C_c. The extended complex-valued measures correspond to the linear functionals $L : C_c \to \mathbb{C}$ that satisfy the condition that for every compact subset K of $\mathbb{R}$ there is a constant $M = M_K$ such that $|L(\phi)| \leq M||\phi||_\infty$ for all continuous functions ϕ with support contained in K.

In the future the term "measure" will always refer to an extended complex-valued regular Borel measure. A *bounded* or *finite measure* is a measure with finite total variation and a *positive measure* is one for which $0 \leq \mu(E) \leq \infty$ for all Borel sets. Bounded measures correspond to bounded linear functionals on $C_0 = C_0(\mathbb{R})$ and positive measures correspond to positive linear functionals on C_c.

3.1 Proposition. *If μ is a measure on $\mathbb{R}$, ϕ is a continuous function with compact support, and $F : \mathbb{R} \to \mathbb{C}$ is defined by*

$$F(x) = \int \phi(x - y) d\mu(y),$$

then F is a continuous function. If μ is bounded, then F vanishes at infinity. If μ has compact support, then F has compact support.

Proof. First note that since ϕ has compact support, F is defined. If $x_n \to x$ and $\phi_n(y) = \phi(x_n - y)$, then there is a compact set K that contains the supports of all the functions ϕ_n. The Lebesgue Dominated Convergence Theorem implies that $F(x_n) \to F(x)$ and so F is continuous. If μ is a bounded measure, then the constant functions are integrable. So if $x_n \to \infty$, the fact that ϕ has compact support implies that $\phi(x_n - y) \to 0$ for all y in $\mathbb{R}$. Once again the Lebesgue Dominated Convergence Theorem implies $F(x_n) \to 0$. The statement involving compact support is left to the reader. □

3.2 Proposition. *Let μ, λ, σ, and η be measures and assume that λ and σ are bounded and η has compact support.*

(a) *There is a measure denoted by $\mu * \eta$ such that for every continuous function ϕ with compact support*

$$\begin{aligned}\int \phi\, d(\mu * \eta) &= \int \left[\int \phi(x-y)\, d\mu(y)\right] d\eta(x) \\ &= \int \left[\int \phi(x-y)\, d\eta(y)\right] d\mu(x).\end{aligned}$$

(b) *There is a bounded measure denoted by $\lambda * \sigma$ such that $||\lambda * \sigma|| \leq ||\lambda||\ ||\sigma||$ and for every continuous function ϕ with compact support,*

$$\begin{aligned}\int \phi\, d(\lambda * \sigma) &= \int \left[\int \phi(x-y)\, d\lambda(y)\right] d\sigma(x) \\ &= \int \left[\int \phi(x-y)\, d\sigma(y)\right] d\lambda(x).\end{aligned}$$

Proof. (a) If F is defined as in (3.1), then the fact that η has compact support implies that the first double integral in (a) makes sense; denote this first integral by $L(\phi)$. Clearly L is a linear functional on C_c. If K is any compact set, E is the support of η, and ϕ has its support in K, then for x in E, $\phi(x-y) \neq 0$ only if $y \in E - K = \{x - z; x \in E \text{ and } z \in K\}$. Since $E - K$ is a compact set, $M = |\mu|(E - K) < \infty$. It follows that $|L(\phi)| \leq M||\phi||_\infty\ ||\eta||$, assuring the existence of the measure $\mu * \eta$. The fact that the first integral equals the second is an exercise in the use of Fubini's Theorem.

(b) Now if $L(\phi)$ denotes the first double integral in (b), then $|L(\phi)| \leq ||\lambda||\ ||\sigma||\ ||\phi||_\infty$, thus implying the existence of the measure $\lambda * \sigma$ and the fact that $||\lambda * \sigma|| \leq ||\lambda||\ ||\sigma||$. Again the first integral equals the second by Fubini's Theorem. □

The measures $\mu * \eta$ and $\lambda * \sigma$ are called the *convolution* of the measures. Whenever we discuss the convolution of two measures it will be assumed that both measures are bounded or that one has compact support. The proof of the next proposition is left to the reader.

3.3 Proposition. *With the notation of the preceding theorem, if α is another measure with compact support and β is another bounded measure, then the following hold.*

(a) $\mu * \eta = \eta * \mu$ and $\lambda * \sigma = \sigma * \lambda$.

(b) *If all the measures are positive, then so are $\mu * \eta$ and $\lambda * \sigma$.*

(c) *The measure $\eta * \alpha$ has compact support and $(\mu * \eta) * \alpha = \mu * (\eta * \alpha)$ and $(\lambda * \sigma) * \beta$.*

(d) $\mu * (\eta + \alpha) = \mu * \eta + \mu * \alpha$ *and* $\lambda * (\eta + \beta) = \lambda * \eta + \lambda * \beta$.

(e) *If δ_0 is the unit point mass at the origin, $\mu * \delta_0 = \mu$.*

Now to specialize convolution to functions.

3.4 Definition. A Borel function f defined on some Borel subset E of $\mathbb{R}$ is said to be *locally integrable* if, for every compact subset K of E, $\int_K |f|\, d\mathcal{A} < \infty$. The set of locally integrable functions on E is denoted by $L^1_{\mathrm{loc}}(E)$; we set $L^1_{\mathrm{loc}} = L^1_{\mathrm{loc}}(\mathbb{R})$. Similarly, define $L^p_{\mathrm{loc}}(E)$ to be the linear space of all Borel functions f such that $|f|^p$ is locally integrable on E; $L^p_{\mathrm{loc}} = L^p_{\mathrm{loc}}(\mathbb{R})$.

Note that if $f \in L^1_{\mathrm{loc}}$, then $\mu(\Delta) = \int_\Delta f d\mathcal{A}$ is a measure. ($\mathcal{A}$ is used to denote Lebesgue measure on $\mathbb{R}$.) This measure is bounded if and only if f is integrable on $\mathbb{R}$. Similarly, μ is positive or has compact support if and only if $f \geq 0$ or f has compact support. This relation will be denoted by $\mu = f\mathcal{A}$.

Suppose f and g are locally integrable functions, $\mu = f\mathcal{A}$, and $\eta = g\mathcal{A}$. Assume that either both f and g are integrable or one of them has compact support. If $\phi \in C_c$, then $\int \phi(x-y)\, d\mu(y) = \int \phi(x-y) f(y)\, d\mathcal{A}(y) = \int \phi(z) f(x-z)\, d\mathcal{A}(z)$ by a change of variables. Thus

$$\begin{aligned}
\int \phi\, d(\mu * \eta) &= \int \left[\int \phi(x-y)\, d\mu(y) \right] d\eta(x) \\
&= \int \left[\int \phi(x-y) f(y)\, d\mathcal{A}(y) \right] g(x)\, d\mathcal{A}(x) \\
&= \int \left[\int \phi(z) f(x-z)\, d\mathcal{A}(z) \right] g(x)\, d\mathcal{A}(x) \\
&= \int \phi(z) \left[\int f(x-z) g(x)\, d\mathcal{A}(x) \right] d\mathcal{A}(z).
\end{aligned}$$

This leads to the following proposition.

3.5 Proposition. *If f and g are locally integrable functions and one of them has compact support (respectively, both are integrable), then the function $f * g$ defined by*

$$(f * g)(x) = \int f(x-y) g(y)\, d\mathcal{A}(y)$$

is locally integrable (respectively, integrable). If $\mu = f\mathcal{A}$ and $\eta = g\mathcal{A}$, then $\mu * \eta = (f * g)\mathcal{A}$.

The function $f * g$ is called the *convolution* of f and g. The proposition for the convolution of functions corresponding to Proposition 3.3 will not be stated but used in the sequel.

Note that for any constant c, if $\phi : \mathbb{R} \to \mathbb{R}$ is defined by letting $\phi(x) = c \exp[-(1 - |x|^2)^{-1}]$ for $|x| < 1$ and $\phi(x) = 0$ for $|x| \geq 1$, then ϕ is infinitely differentiable and non-negative. (Here $|x|$ is the usual Euclidean norm, $|x| = [x_1^2+\cdots+x_n^2]^{1/2}$.) Choose the constant c such that $\int \phi d\mathcal{A} = 1$. For $\varepsilon > 0$, let $\phi_\varepsilon(x) = \varepsilon^{-d}\phi(x/\varepsilon)$. Note that ϕ_ε is still infinitely differentiable, $\phi_\varepsilon(x) = 0$ for $|x| \geq \varepsilon$, and $\int \phi_\varepsilon d\mathcal{A} = 1$. This net $\{\phi_\varepsilon\}$ is called a *mollifier* or *regularizer* and for f in L^1_{loc}

$$(\phi_\varepsilon * f)(y) = \int \phi_\varepsilon(y-x)f(x)\, d\mathcal{A}(x)$$

is called the *mollification* or *regularization* of f. The reason for these terms will surface in part (a) of the next result.

It is important to realize, however, that the mollifier has the property that it is rotationally invariant; that is, $\phi_\varepsilon(x) = \phi_\varepsilon(|x|)$. This will be used at times in the future.

3.6 Proposition. *Let $f \in L^1_{\text{loc}}$ and let K be a compact subset of $\mathbb{R}$.*

(a) *For every $\varepsilon > 0$, $\phi_\varepsilon * f \in C^\infty$.*

(b) *If $f = 0$ off K and U is an open set containing K, then $\phi_\varepsilon * f \in C_c^\infty(U)$ for $0 < \varepsilon < \text{dist}(K, \partial U)$.*

(c) *If f is continuous on an open set that contains K, then $\phi_\varepsilon * f \to f$ uniformly on K.*

(d) *If $f \in L^p_{\text{loc}}$, $1 \leq p < \infty$, then*

$$\lim_{\varepsilon\to 0} \int_K |\phi_\varepsilon * f - f|^p\, d\mathcal{A} = 0.$$

Proof. (a) Since ϕ_ε is infinitely differentiable, the fact that $\phi_\varepsilon * f$ is infinitely differentiable follows by applying Leibniz's rule for differentiating under the integral sign.

(b) Let $0 < \varepsilon < \text{dist}(K, \mathbb{R}\setminus\mathbb{U})$. If $\text{dist}(y, K) \geq \varepsilon$, then $\phi_\varepsilon(x-y) = 0$ for all x in K. Hence if $\text{dist}(y, K) \geq \varepsilon$, then $(\phi_\varepsilon * f)(y) = \int_K \phi_\varepsilon(y-x)f(x)d\mathcal{A}(x) = 0$.

(c) Only consider $\varepsilon < \text{dist}(K, \mathbb{C}\setminus U)$. Because $\int \phi_\varepsilon\, d\mathcal{A} = 1$ and $\phi_\varepsilon = 0$ off $B(0;\varepsilon)$, for y in K we have that $|(\phi_\varepsilon * f)(y) - f(y)| = \left|\int \phi_\varepsilon(y-x)[f(x) - f(y)]d\mathcal{A}(x)\right| \leq \sup\{|f(x) - f(y)| : |y - x| < \varepsilon\}$. But since f must be uniformly continuous in a neighborhood of K, the right hand side of this inequality can be made arbitrarily small uniformly for y in K.

(d) Let U be a bounded open set containing K and let $\alpha > 0$. Let g be a continuous function with support contained in U such that $\int_U |f-g|^p d\mathcal{A} < \alpha^p$. If $0 < \varepsilon < \text{dist}(K, \mathbb{C}\setminus U)$, then

$$\begin{aligned}
&\int_K |\phi_\varepsilon * f - \phi_\varepsilon * g|^p\, d\mathcal{A} \\
&= \int_K \left| \int_U \phi_\varepsilon(x-y)^{1/q}\phi_\varepsilon(x-y)^{1/p}[f(x) - g(x)]\, d\mathcal{A}(x) \right|^p d\mathcal{A}(y)
\end{aligned}$$

$$
\begin{aligned}
&\leq \int_K \left(\int_U \phi_\varepsilon(x-y)\, d\mathcal{A}(x) \right)^{p/q} \cdot \\
&\qquad\qquad \left(\int_U \phi_\varepsilon(x-y)|f(x)-g(x)|^p\, d\mathcal{A}(x) \right) d\mathcal{A}(y) \\
&= \int_U |f(x)-g(x)|^p \left[\int_K \phi_\varepsilon(x-y)\, d\mathcal{A}(y) \right] d\mathcal{A}(x) \\
&< \alpha^p.
\end{aligned}
$$

Therefore

$$
\left[\int_K |\phi_\varepsilon * f - f|^p\, d\mathcal{A} \right]^{1/p} < 2\alpha + \left[\int_K |\phi_\varepsilon * g - g|^p\, d\mathcal{A} \right]^{1/p}.
$$

By part (c), the right hand side of this inequality can be smaller than 3α if ε is chosen sufficiently small. □

The reader might profit by now looking at Exercise 1 in the next section. Here is an application of the preceding proposition that will prove useful later.

3.7 Proposition. *If μ is a measure on $\mathbb{C}$ and U is an open subset of the plane such that $\int \phi\, d\mu = 0$ for all ϕ in $C_c^\infty(U)$, then $|\mu|(U) = 0$.*

Proof. Let f be an arbitrary function in $C_c(U)$; it suffices to show that $\int f d\mu = 0$. Let K be the support of f and put $d = 2^{-1}\text{dist}(K, \partial U)$. If $\{\phi_\varepsilon\}$ is a mollifier, then $\phi_\varepsilon * f \in C_c^\infty(U)$ for $\varepsilon < 2d$. Hence $\int \phi_\varepsilon * f\, d\mu = 0$ for $\varepsilon < 2d$. But if $L = \{z : \text{dist}(x, K) \leq d\}$, then $\phi_\varepsilon * f \to f$ uniformly on L (3.6.c) and so $\int f\, d\mu = \int_L f\, d\mu = \lim_{\varepsilon\to 0} \int_L \phi_\varepsilon * f\, d\mu = \lim_{\varepsilon\to 0} \int \phi_\varepsilon * f\, d\mu = 0$. □

3.8 Theorem. *If G is an open subset of $\mathbb{R}$ and $\mathcal{U}$ is an open cover of G, then there is a partition of unity on G that is subordinate to $\mathcal{U}$ and consists of infinitely differentiable functions.*

Proof. It is easy to see that G can be written as the union of a sequence of compact sets $\{K_n\}$ with $K_n \subseteq \text{int } K_{n+1}$ (see 7.1.2). Thus according to Theorem 2.3 there is a countable locally finite refinement $\mathcal{B}$ of $\mathcal{U}$ such that $\mathcal{B}$ consists of balls. Let $\mathcal{B} = \cup_n \mathcal{B}_n$, where $\mathcal{B}_n$ is a finite cover of $K_n \setminus \text{int } K_{n-1} (K_0 = \emptyset)$; put $\mathcal{B}_n = \{B_{nk} : 1 \leq k \leq p_n\}$. Arrange matters so that $\mathcal{B}_n \cap \mathcal{B}_m = \emptyset$ for $n \neq m$. For each $n \geq 1$ let L_n be a compact set with $K_n \setminus \text{int } K_{n-1} \subseteq \text{int } L_n \subseteq L_n \subseteq \cup_k B_{nk}$. According to Proposition 2.4 there are continuous functions $\{f_{nk} : 1 \leq k \leq p_n\}$ such that $0 \leq f_{nk} \leq 1$, support $f_{nk} \subseteq B_{nk}$, and $\sum_{k=1}^{p_n} f_{nk}(x) = 1$ for all x in L_n.

Now choose $\varepsilon_n > 0$ so that it is simultaneously less than $\text{dist}(K_n \setminus \text{int } K_{n-1}, \mathbb{R} \setminus \mathbb{L}_{\mathbb{K}})$ and $\text{dist}(\text{support } f_{nk}, \partial B_{nk})$; put $\psi_{nk} = \phi_{\varepsilon_n} * f_{nk}$, where ϕ_{ε_n} is a regularizer. The function ψ_{nk} is infinitely differentiable by the preceding proposition and it is clearly positive. Also the definition of ψ_{nk} shows that for any point x, $\psi_{nk}(x) \leq \int \phi_{\varepsilon_n} = 1$. Since

$\varepsilon_n < \operatorname{dist}(\text{support } f_{nk}, \partial B_{nk})$, the preceding proposition implies that support $\psi_{nk} \subseteq B_{nk}$. Finally, if $x \in K_n \setminus \operatorname{int} K_{n-1}$,

$$\sum_{k=1}^{p_n} \psi_{nk}(x) = \int \phi_{\varepsilon_n}(x-y) \sum_{k=1}^{p_n} f_{nk}(y)\, d\mathcal{A}(y).$$

But $\phi_{\varepsilon_n}(x-y) = 0$ unless $|x-y| < \varepsilon_n$. By the choice of x this implies that $y \in L_n$ and so $\sum_{k=1}^{p_n} f_{nk}(y) = 1$. Thus $\sum_{k=1}^{p_n} \psi_{nk}(x) = 1$ for all x in $K_n \setminus \operatorname{int} K_{n-1}$. Hence $\psi(x) = \sum_{n=1}^{\infty} \sum_{k=1}^{p_n} \psi_{nk}(x) \geq 1$ for all x in G. If $\phi_{nk} \equiv \psi_{nk}/\psi$, it is easy to check that $\{\phi_{nk} : 1 \leq k \leq p_n \text{ and } n \geq 1\}$ is the sought after partition of unity. □

We now state a somewhat abstract form of Leibniz's rule for differentiating under the integral sign.

3.9 Theorem. *Let (Y, Σ, μ) be a measure space, let G be an open subset of $\mathbb{R}$, and let e_j be the j-th basis vector in $\mathbb{R}$. Suppose $f : G \times Y \to \mathbb{C}$ is a measurable function that satisfies the following conditions:*

(a) *for each x in G the function $y \to f(x,y)$ belongs to $L^1(\mu)$;*

(b) *$\frac{\partial}{\partial x_j} f(x,y)$ exists for a.e. $[\mu]$ y in Y and all x in G;*

(c) *for each x in G there is a function g in $L^1(\mu)$ and a function θ defined for small real numbers such that $\theta(t) \to 0$ as $t \to 0$ and*

$$\frac{f(x+te_j, y) - f(x,y)}{t} - \frac{\partial f}{\partial x_j}(x,y) = g(y)\theta(t)$$

a.e. $[\mu]$. Then $F(x) \equiv \int f(x,y)\, d\mu(y)$ exists and is differentiable with respect to x_j with

$$\frac{\partial F}{\partial x_j}(x) = \int \frac{\partial f}{\partial x_j}(x,y)\, d\mu(y).$$

3.10 Corollary. *If ϕ is a continuously differentiable function with compact support and $f \in L^1(\mathcal{A})$, then $\phi * f$ is continuously differentiable and $\partial(\phi * f)/\partial x_j = (\partial\phi/\partial x_j) * f$ for $1 \leq j \leq d$.*

Now to reset the definitions and results for convolution on the circle. Since $\partial\mathbb{D}$ is compact, all regular Borel measures on the circle are finite and so the discussion of convolution is simplified. It is no longer necessary to consider locally integrable functions. Let $M(\partial\mathbb{D})$ be the space of complex-valued regular Borel measures on $\partial\mathbb{D}$. If μ and $\nu \in M(\partial\mathbb{D})$, define $L : C(\partial\mathbb{D}) \to \mathbb{C}$ by

3.11
$$L(f) = \int_{\partial\mathbb{D}} \int_{\partial\mathbb{D}} f(zw)\, d\mu(z)\, d\nu(w)$$

for all f in $C(\partial\mathbb{D})$. If $L : C(\partial\mathbb{D}) \to \mathbb{C}$ is defined as in (3.11), then L is a bounded linear functional and $||L|| \leq ||\mu||\,||\nu||$. Since this linear functional L is bounded, the Riesz Representation Theorem implies there is a unique measure on $\partial\mathbb{D}$ corresponding to it.

3.12 Definition. If μ and $\nu \in M(\partial\mathbb{D})$, then $\mu * \nu$ is the unique measure in $M(\partial\mathbb{D})$ such that

$$\int f\, d(\mu * \nu) = \int_{\partial\mathbb{D}}\int_{\partial\mathbb{D}} f(zw)\, d\mu(z)\, d\nu(w)$$

for all f in C. The measure $\mu * \nu$ is called the *convolution* of μ and ν.

3.13 Proposition. *If μ, ν, and $\eta \in M(\partial\mathbb{D})$, the following hold.*

(a) $\mu * \nu = \nu * \mu$ *and* $||\mu * \nu|| \leq ||\mu||\,||\nu||$.

(b) *If μ and ν are positive, then* $\mu * \nu \geq 0$.

(c) $(\mu * \nu) * \eta = \mu * (\nu * \eta)$.

(d) $\mu * (\nu + \eta) = \mu * \nu + \mu * \eta$.

(e) *If δ_1 is the unit point mass at 1, then* $\delta_1 * \mu = \mu = \mu * \delta_1$.

There is an equivalent way to define $\mu * \nu$ as a function defined on the Borel subsets of $\partial\mathbb{D}$. See Exercise 3.

3.14 Proposition. *If f and $g \in L^1(\partial\mathbb{D})$, $\mu = fm$, $\nu = gm$, and $h : \partial\mathbb{D} \to \mathbb{C}$ is defined by*

$$h(z) = \int f(z\overline{w})\, g(w)\, dm(w)$$

for z in $\partial\mathbb{D}$, then $h \in L^1(\partial\mathbb{D})$ and $\mu * \nu = hm$.

3.15 Definition. If f and $g \in L^1$, then the *convolution* of f and g is the function

$$(f * g)(z) = \int f(z\overline{w})\, g(w)\, dm(w).$$

Note that the preceding proposition shows that the definitions of convolution for measures and functions are consistent. Also the basic algebraic properties for the convolution of two functions can be read off from Proposition 3.13. In particular, it follows from part (a) of Proposition 3.13 that

$$||f * g||_1 \leq ||f||_1\, ||g||_1.$$

Exercises

1. Is Proposition 3.1 valid if it is not assumed that ϕ is continuous?
2. If $f \in L^1(\mu * \nu)$, show that $\int f \, d\mu * \nu = \int\int f(x-y) \, d\mu(x) \, d\nu(y)$. State and prove the analogous fact for convolution on the circle.
3. If E is a Borel subset of $\mathbb{R}^d$, show that $(\mu * \nu)(E) = (\mu \times \nu)(\{(x,y) \in \mathbb{R}^d \times \mathbb{R}^d : x+y \in E\})$. State and prove the analogous fact for convolution on the circle.
4. If μ and ν are finite measures on $\mathbb{R}^d$ and $\nu << \mathcal{A}$, then $\mu * \nu << \mathcal{A}$. State and prove the analogous fact for convolution on the circle.

§4 Distributions

In this section we will concentrate on distributions on an open subset of the complex plane. The definitions and results carry over to distributions on open sets in $\mathbb{R}$, and we will need some of these facts for distributions on $\mathbb{R}$. However we will need to see some of the relationships involving functions and distributions of a complex variable and obtain information about analytic and harmonic functions. Thus the specialization. It will be left to the reader to carry out the extension to $\mathbb{R}$.

We make the convention, in line with the mathematical community, that for any region G in $\mathbb{C}$, $\mathcal{D}(G) = C_c^\infty(G)$. The reader should regain acquaintance with the notation $\partial\phi$ and $\overline{\partial}\phi$ from §13.2.

4.1 Definition. If G is an open subset of $\mathbb{C}$, a *distribution* on G is a linear functional $L : \mathcal{D}(G) \to \mathbb{C}$ with the property that if K is any compact subset of G and $\{\phi_k\}$ is a sequence in $\mathcal{D}(G)$ with support $\phi_k \subseteq K$ for every $k \geq 1$, and if for all $m, n \geq 1$, $\partial^n \overline{\partial}^m \phi_k(z) \to 0$ uniformly for z in K as $k \to \infty$, then $L(\phi_k) \to 0$. The functions in $\mathcal{D}(G)$ are referred to as *test functions.*

It is possible to define a topology on $\mathcal{D}(G)$ such that $\mathcal{D}(G)$ becomes a locally convex topological vector space and the distributions are precisely the continuous linear functionals on this space. See §4.5 in Conway [1990]. This observation has more than psychological merit, as it means that results from functional analysis (like the Hahn-Banach Theorem) apply.

4.2 Example.

(a) If $u \in L^1_{\text{loc}}(G)$, then u defines a distribution L_u via the formula $L_u(\phi) = \int u\phi \, d\mathcal{A}$.

(b) If μ is a measure on G, then μ defines a distribution L_μ via the formula $L_\mu(\phi) = \int \phi \, d\mu$.

The verification of these statements are left to the reader. Also note that if L is a distribution on G, then $\phi \to L(\partial\phi)$ and $\phi \to L(\overline{\partial}\phi)$ also define distributions on G. This justifies the following definition.

4.3 Definition. If L is a distribution on G, let ∂L and $\overline{\partial} L$ be the distributions defined by $\partial L(\phi) = -L(\partial\phi)$ and $\overline{\partial} L(\phi) = -L(\overline{\partial}\phi)$.

The minus signs are placed here in the definition so that if u is a continuously differentiable function on G, then $\partial L_u = L_{\partial u}$ and $\overline{\partial} L_u = L_{\overline{\partial} u}$, as can be verified by an application of integration by parts.

For the most part we will be concerned with distributions that are defined by locally integrable functions, measures, and the derivatives of such distributions. Be aware, however, that the derivative of a distribution defined by a function is not necessarily a distribution defined by a function. If $u \in L^1_{\text{loc}}(G)$, we will often consider ∂u as the distributional derivative of u. That is, ∂u is the distribution ∂L_u and the caution just expressed is the reminder that ∂u is not necessarily a function. Similar statements hold for $\overline{\partial} u$ and all higher derivatives.

In Lemma 13.2.6 and Lemma 13.2.10 it was shown that for any w in $\mathbb{C}$ the functions $z \to (z-w)^{-1}$ and $z \to \log|z-w|$ are locally integrable. The derivatives in the sense of distributions of these functions are calculated below. These will see special service later in this book. For the sake of completeness, however, some additional distributions are introduced.

4.4 Lemma. *If $n \geq 1$ and $\phi \in C_c^n$, then*

$$\lim_{\varepsilon \to 0} \int_{|z| \geq \varepsilon} \frac{\phi(z)}{z^n}\, d\mathcal{A}(z)$$

exists and is finite.

Proof. Let R be sufficiently large that support $\phi \subseteq B(0;R)$. By Taylor's Formula (13.2.4) $\phi = p(z,\overline{z}) + g$, where p is a polynomial in z and $\overline{z}$ of degree $\leq n-1$, each derivative of g of order $\leq n-1$ vanishes at 0, and $|g(z)| \leq C|z|^n$ for some constant C. Thus

$$\int_{|z|\geq\varepsilon} \frac{\phi(z)}{z^n}\, d\mathcal{A}(z) = \int_{R\geq|z|\geq\varepsilon} \frac{p(z,\overline{z})}{z^n}\, d\mathcal{A}(z) + \int_{R\geq|z|\geq\varepsilon} \frac{g(z)}{z^n}\, d\mathcal{A}(z).$$

Now the first of these two integrals is a linear combination of integrals of quotients of the form $z^k\overline{z}^m/z^n$ for $k, m \geq 0$ and $k+m < n$. But

$$\begin{aligned}\int_{R\geq|z|\geq\varepsilon} \frac{z^k\overline{z^m}}{z^n}\, d\mathcal{A}(z) &= \int_0^R \left[\int_0^{2\pi} e^{i(k-m-n)\theta} d\theta\right] r^{k+m-n-1} dr \\ &= 0\end{aligned}$$

since $k-m-n \neq 0$. On the other hand g/z^n is bounded so that

$$\lim_{\varepsilon\to 0}\int_{|z|\geq\varepsilon}\frac{\phi(z)}{z^n}\,d\mathcal{A}(z)=\int_{|z|\leq R}\frac{g(z)}{z^n}d\mathcal{A}(z),$$

a finite number. □

4.5 Proposition. *For $n \geq 1$ define $PV_n : \mathcal{D} \to \mathbb{C}$ by*

$$PV_n(\phi)=\lim_{\varepsilon\to 0}\int_{|z|\geq\varepsilon}\frac{\phi(z)}{z^n}\,d\mathcal{A}(z).$$

Then PV_n defines a distribution and

$$\partial PV_n=-nPV_{n+1},\qquad \overline{\partial}PV_n=(-1)^n\frac{\pi}{(n-1)!}\partial^{n-1}\delta_0,$$

where δ_0 is the unit point mass at the origin.

Proof. The preceding lemma shows that $PV_n(\phi)$ is defined and finite. It must be shown that it is a distribution. So let R be a positive number, let $\{\phi_j\}$ be a sequence of test functions with supports all contained in $B=B(0;R)$, and assume that for all $k,m\geq 0$ $\partial^k\overline{\partial}^m\phi_j\to 0$ uniformly on B. Again use Taylor's Formula to write $\phi_j=p_j+g_j$, where p_j is a polynomial in z and $\overline{z}$ of degree $\leq n-1$ and each derivative of g_j of order $\leq n-1$ vanishes at 0. From the proof of the preceding lemma it is known that

$$\int_{|z|\geq\varepsilon}\frac{\phi_j(z)}{z^n}\,d\mathcal{A}(z)=\int_{R\geq|z|\geq\varepsilon}\frac{g_j(z)}{z^n}\,d\mathcal{A}(z)$$

and

$$PV_n(\phi_j)=\int_{|z|\leq R}\frac{g_j(z)}{z^n}\,d\mathcal{A}(z).$$

But Taylor's Formula also implies that for each $j\geq 1$,

$$\frac{|g_j(z)|}{|z|^n}\leq\frac{1}{n!}\sum_{k+j=n}\max\left\{\left|\partial^k\overline{\partial}^j\phi_j(w)\right|:|w|\leq R\right\}.$$

Thus $|g_j(z)|/|z|^n\to 0$ uniformly on B. Therefore $PV_n(\phi_j)\to 0$ and PV_n is a distribution.

To find ∂PV_n, fix a test function ϕ with support contained in $B=B(0;R)$ and write $\phi=p+g$ as in Taylor's Formula. Using the definition of the derivative of a distribution

$$\begin{aligned}\partial PV_n(\phi) &= -\lim_{\varepsilon\to 0}\int_{|z|\geq\varepsilon}\frac{\partial\phi(z)}{z^n}\,d\mathcal{A}(z)\\ &= -\lim_{\varepsilon\to 0}\int_{|z|\geq\varepsilon}\left[n\frac{\phi(z)}{z^{n+1}}+\partial(\phi z^{-n})\right]d\mathcal{A}(z)\\ &= -nPV_{n+1}(\phi)-\lim_{\varepsilon\to 0}\int_{|z|\geq\varepsilon}\partial(\phi z^{-n})\,d\mathcal{A}(z).\end{aligned}$$

So we must show this last limit is 0. To do this it suffices to show that the complex conjugate goes to 0 as $\varepsilon \to 0$. This permits the application of Green's Theorem and so

$$\begin{aligned}
\int_{|z|\geq\varepsilon} [\partial(\phi z^{-n})]^* \, d\mathcal{A}(z) &= \int_{R\geq|z|\geq\varepsilon} \bar{\partial}[\overline{\phi}/\overline{z^n}] \, d\mathcal{A}(z) \\
&= -\frac{1}{2i}\int_{|z|=\varepsilon} \frac{\overline{\phi}}{\overline{z^n}} dz \\
&= \frac{i}{2}\int_0^{2\pi} \frac{\overline{\phi(\varepsilon e^{i\theta})}}{\varepsilon^n e^{-in\theta}} i\varepsilon e^{i\theta} d\theta \\
&= -\frac{1}{2\varepsilon^{n-1}}\int_0^{2\pi} \frac{\overline{\phi(\varepsilon e^{i\theta})}}{e^{-i(n+1)\theta}} d\theta \\
&= -\frac{1}{2\varepsilon^{n-1}}\int_0^{2\pi} \overline{g(\varepsilon e^{i\theta})} e^{i(n+1)\theta} d\theta.
\end{aligned}$$

But there is a constant C such that $|g(z)| \leq C|z|^n$ for $|z| \leq R$. Therefore

$$\begin{aligned}
\left| \int_{|z|\geq\varepsilon} [\partial(\phi z^{-n})]^* \, d\mathcal{A}(z) \right| &\leq \frac{\pi}{\varepsilon^{n-1}} \max_\theta |g(\varepsilon e^{i\theta})| \\
&\leq C\pi\varepsilon,
\end{aligned}$$

which converges to 0 as $\varepsilon \to 0$.

Now to find $\bar{\partial} PV_n$. Let ϕ, p, g, and $B(0; R)$ be as above. Once again an application of the definition of the derivative of a distribution and Green's Theorem show that

$$\begin{aligned}
\bar{\partial} PV_n(\phi) &= -\lim_{\varepsilon\to 0} \frac{1}{2i}\int_{|z|=\varepsilon} \frac{\phi}{z^n} \, dz \\
&= -\lim_{\varepsilon\to 0} \frac{i}{2}\int_0^{2\pi} \frac{\phi(\varepsilon e^{i\theta})}{\varepsilon^n e^{in\theta}} i\varepsilon e^{i\theta} d\theta \\
&= \lim_{\varepsilon\to 0} \frac{1}{2\varepsilon^{n-1}}\int_0^{2\pi} \phi(\varepsilon e^{i\theta}) e^{-i(n-1)\theta} d\theta.
\end{aligned}$$

Now substitute $p+g$ for ϕ in this integral. It is left to the reader to show that the limit of the integral for g is 0. For the integral involving the polynomial p, note that the integral of all the terms involved are 0 save possibly for the term with z^{n-1}. Here we have that

$$\frac{1}{2\varepsilon^{n-1}}\int_0^{2\pi} (\varepsilon e^{i\theta})^{n-1} e^{-i(n-1)\theta} d\theta = \pi.$$

Now the coefficient of z^{n-1} in the expression for p is $\partial^{n-1}\phi(0)/(n-1)!$. Assembling these pieces produces

$$\bar{\partial} PV_n(\phi) = \frac{\pi}{(n-1)!} \partial^{n-1}\phi(0)$$

$$= \quad (-1)^{n-1}\frac{\pi}{(n-1)!}(\partial^{n-1}\delta_0)(\phi).$$

□

Two special cases of this are worth underlining. Note that in the preceding proposition when $n = 1$ the functions z^{-1} and $\overline{z^{-1}}$ are locally integrable and thus define a distribution. Combining this with Exercise 3 we get the following.

4.6 Corollary. *For any point w in $\mathbb{C}$,*

$$\begin{array}{ll} \overline{\partial}[(z-w)^{-1}] = \pi\delta_w & \partial[(z-w)^{-1}] = -PV_2 \\ \overline{\partial}[(\overline{z}-\overline{w})^{-1}] = -(PV_2)^* & \partial[(\overline{z}-\overline{w})^{-1}] = \pi\delta_{\overline{w}}, \end{array}$$

where δ_w is the unit point mass at w.

4.7 Proposition. *If $w \in G$, then, in the sense of distributions on G:*

(a) $\partial \log|z-w| = [2(z-w)]^{-1}$ *and* $\overline{\partial}\log|z-w| = [2(\overline{z}-\overline{w})]^{-1}$*;*

(b) $\Delta \log|z-w| = 2\pi\delta_w$*, where δ_w is the unit point mass at w.*

Proof. (a) If $\phi \in \mathcal{D}(G)$, let Γ be a smooth positive Jordan system in G such that support $\phi \subseteq U \equiv \text{ins}\,\Gamma \subseteq G$ and $w \in U$. Let ε be a positive number such that $B_\varepsilon = \overline{B}(w;\varepsilon) \subseteq U$ and let $U_\varepsilon = U \setminus B_\varepsilon$. So using Green's Theorem

$$\begin{aligned} [\overline{\partial}\log|z-w|](\phi) &= -\int(\overline{\partial}\phi)\log|z-w|\,d\mathcal{A}(z) \\ &= -\lim_{\varepsilon\to 0}\int_{U_\varepsilon}(\overline{\partial}\phi)\log|z-w|\,d\mathcal{A}(z) \\ &= -\lim_{\varepsilon\to 0}\int_{U_\varepsilon}\{\overline{\partial}[\phi(z)\log|z-w|] \\ &\qquad - \phi(z)\overline{\partial}\log|z-w|\}\,d\mathcal{A}(z) \\ &= \lim_{\varepsilon\to 0}\left\{\frac{1}{2i}\int_{\partial U_\varepsilon}\phi(z)\log|z-w|dz \right. \\ &\qquad \left. + \frac{1}{2}\int_{U_\varepsilon}\frac{\phi(z)}{\overline{z}-\overline{w}}\,d\mathcal{A}(z)\right\}. \end{aligned}$$

But

$$\begin{aligned} \int_{\partial U_\varepsilon}\phi(z)\log|z-w|dz &= -\int_{\partial B_\varepsilon}\phi(z)\log|z-w|\,dz \\ &= -\varepsilon\log\varepsilon\int_0^{2\pi}\phi(w+\varepsilon e^{i\theta})ie^{i\theta}d\theta \end{aligned}$$

and this converges to 0 as $\varepsilon \to 0$. This proves the second half of (a). To get the first half, just apply Exercise 3.

(b) This is a consequence of (a), the preceding corollary, and the fact that $\Delta = 4\partial\bar{\partial}$. □

Note that it is not possible to define the product of two distributions, but if L is a distribution on G and ϕ is a test function, then $\phi L(\psi) \equiv L(\phi\psi)$ defines another distribution on G. This produces a product rule.

4.8 Proposition. *If L is a distribution on G and ϕ is a test function, then $\partial(\phi L) = (\partial\phi)L + \phi(\partial L)$ and $\bar{\partial}(\phi L) = (\bar{\partial}\phi)L + \phi(\bar{\partial}L)$.*

The proof is left as an exercise.

Say that a distribution L on G is *positive* if $L(\phi) \geq 0$ whenever ϕ is a non-negative test function. An example of a positive distribution is one defined by a positive measure. The next proposition provides a converse.

4.9 Proposition. *The distribution L on G is positive if and only if there is a positive measure μ on G such that $L(\phi) = \int \phi d\mu$.*

Proof. To prove this proposition it must be shown that for any compact subset K of G there is a constant $C = C_K$ such that $|L(\phi)| \leq C||\phi||_\infty$ for every real-valued test function ϕ with support contained in K. Indeed, if this is done, then the fact that the infinitely differentiable functions with support contained in K are dense in $C_0(\text{int } K)$ allows us to extend L to a bounded linear functional on $C_0(\text{int } K)$. From here we produce the measure on G (details?).

Let ψ be a test function with compact support in G such that $0 \leq \psi \leq 1$ and $\psi = 1$ on K. (See Exercise 1.) If ϕ is a real-valued test function with support included in K and $||\phi||_\infty \leq 1$, then $-\psi \leq \phi \leq \psi$. Because L is positive, this implies that $-L(\psi) \leq L(\phi) \leq L(\psi)$. That is, $|L(\phi)| \leq C$, where $C \equiv L(\psi)$. Thus $|L(\phi)| \leq C||\phi||_\infty$ for any real-valued test function with support included in K. □

4.10 Weyl's Lemma. *If $u \in L^1_{\text{loc}}(G)$ and $\partial\bar{\partial}u = 0$ as a distribution, then there is a harmonic function f on G such that $u = f$ a.e. [Area].*

Proof. Let $\{\phi_\varepsilon\}$ be a mollifier. Fix δ with $0 < \varepsilon < \delta$ and put $G_\delta = \{z \in G : \text{dist}(z, \partial G) > \delta\}$. Note that $\frac{\partial}{\partial \bar{z}}(\phi_\varepsilon(w - z)) = -\frac{\partial}{\partial \bar{w}}(\phi_\varepsilon(w - z))$. Hence if $\psi \in \mathcal{D}(G_\delta)$, then

$$\begin{aligned}\int \psi\partial\bar{\partial}(\phi_\varepsilon * u)\, d\mathcal{A} &= \int \partial\bar{\partial}\psi(w)(\phi_\varepsilon * u)(w)\, d\mathcal{A}(w) \\ &= \int u(z)\left[\int \partial\bar{\partial}\psi(w)\phi_\varepsilon(w - z)\, d\mathcal{A}(w)\right] d\mathcal{A}(z)\end{aligned}$$

$$
\begin{aligned}
&= \int u(z)\left[\int \psi(w)\frac{\partial}{\partial w}\frac{\partial}{\partial \overline{w}}\left[\phi_\varepsilon(w-z)\right]\,d\mathcal{A}(w)\right]d\mathcal{A}(z)\\
&= \int u(z)\left[\int \psi(w)\frac{\partial}{\partial z}\frac{\partial}{\partial \overline{z}}\left[\phi_\varepsilon(w-z)\right]\,d\mathcal{A}(w)\right]d\mathcal{A}(z)\\
&= \int u(z)\frac{\partial}{\partial z}\frac{\partial}{\partial \overline{z}}\left[\int \psi(w)\phi_\varepsilon(w-z)\,d\mathcal{A}(w)\right]d\mathcal{A}(z)\\
&= \int u(z)\partial\overline{\partial}(\phi_\varepsilon * \psi)(z)\,d\mathcal{A}(z)\\
&= \partial\overline{\partial}L_u(\phi_\varepsilon * \psi)\\
&= 0.
\end{aligned}
$$

That is, $\partial\overline{\partial}(\phi_\varepsilon * u) = 0$ on G_δ. Since $\phi_\varepsilon * u \in C^\infty$, $\phi_\varepsilon * u$ is harmonic on G_δ when $0 < \varepsilon < \delta$. By part (d) of Proposition 3.6, $\int_K |\phi_\varepsilon * u - u|\,d\mathcal{A} \to 0$ as $\varepsilon \to 0$ for any compact subset K of G_δ. Since Bergman spaces are complete, $u \in L^1_h(U)$ for any open set U with cl $U \subseteq G_\delta$. Since δ was arbitrary, the result follows. □

4.11 Corollary. *If u is a locally integrable function on G and $\overline{\partial}u = 0$ in the sense of distributions, then there is an analytic function f on G such that $u = f$ a.e. [Area].*

Proof. Since $\overline{\partial}u = 0$, $\partial\overline{\partial}u = 0$. But Weyl's Lemma implies that u is harmonic on G. In particular, u is infinitely differentiable. It now follows that u is analytic (13.2.1). □

This corollary is also referred to as Weyl's Lemma. Indeed there is the mother of all Weyl Lemmas, which states that if L is a distribution on G and D is an elliptic differential operator such that $D(L) = 0$, then L is given by an infinitely differential function u that satisfies $Du = 0$. Both $\overline{\partial}$ and $\partial\overline{\partial}$ are examples of elliptic differential operators.

Exercises

1. If K is a compact subset of the open set U in $\mathbb{C}$, this exercise shows how to construct an infinitely differentiable function that is 1 on K and has compact support inside U. (a) Define g_1 on $\mathbb{R}$ by $g_1(x) = \exp(-x^{-2})$ for $x \geq 0$ and $g_1(x) = 0$ for $x \leq 0$. Show that g_1 is a C^∞ function. (b) Put $g(x) = g_1(x)g_1(1-x)$ for all x in $\mathbb{R}$ and show that g is a C^∞ function, $g \geq 0$, and $g(x) = 0$ for $x \notin (0,1)$. (c) If $M = \int g(x)\,dx$ and $h(x) = M^{-1}\int_0^x g(t)\,dt$, then h is a C^∞ function, $0 \leq h \leq 1$, $h(x) = 0$ for $x \leq 0$, and $h(x) = 1$ for $x \geq 1$. (d) Define

$k(x) = 1 - h(2x-1)$ for $x \geq 0$ and extend k to the negative real axis by letting $k(x) = k(-x)$ for $x \leq 0$. Show k is a C^∞ function, $0 \leq k \leq 1$, $k(x) = 1$ for $|x| \leq 1/2$, and $k(x) = 0$ for $|x| \geq 1$. (e) Now define $f : \mathbb{C} \to \mathbb{R}$ by $f(z) = k(|z|)$ and show that f is a C^∞ function, $0 \leq f \leq 1$, $f(z) = 1$ for $|z| \leq 1/2$, and $f(z) = 0$ for $|z| \geq 1$. If $\varepsilon > 0$, define $f_\varepsilon(z) = f(z/\varepsilon)$ and put $C = \int f(z)\, d\mathcal{A}(z)$. Check by using the change of variables formula that $\int f_\varepsilon(z)\, d\mathcal{A}(z) = C\varepsilon^2$. If $\phi_\varepsilon(z) = [C\varepsilon^2]^{-1} f_\varepsilon(z)$, then ϕ_ε is a C^∞ function, $\phi_\varepsilon \geq 0$, $\int \phi_\varepsilon(z)\, d\mathcal{A}(z) = 1$, and $\phi_\varepsilon(z) = 0$ for $|z| \geq \varepsilon$. So $\{\phi_\varepsilon\}$ is a mollifier as in (3.6). (f) Let K be a compact set, let U be open, and suppose that $K \subseteq U$. Let ψ be any continuous function with $0 \leq \psi \leq 1$, $\psi(z) = 1$ for z in K, and $\psi(z) = 0$ for $z \notin U$. Show that for an appropriate choice of ε, $\phi \equiv \phi_\varepsilon * \psi$ is a C^∞ function with compact support contained in U, $0 \leq \phi \leq 1$, and $\phi(z) = 1$ for all z in K.

2. Show that if u is locally integrable such that ∂u exists a.e. [Area] on G and ∂u is locally integrable, then $\partial L_u = L_{\partial u}$.

3. Show that for any distribution L, $L^*(\phi) \equiv \overline{L(\overline{\phi})}$ defines a distribution and $(\partial L)^* = \bar{\partial} L^*$, $(\bar{\partial} L)^* = \partial L^*$.

§5 The Cauchy Transform

In this section we introduce and give the elementary properties of the Cauchy transform of a compactly supported measure on the plane. This is a basic tool in the study of rational approximation, a fact we will illustrate by using it to give an independent proof of Runge's Theorem.

If μ is any compactly supported measure on $\mathbb{C}$, let

$$\tilde{\mu}(w) = \int \frac{1}{|w-z|}\, d|\mu|(z)$$

when the integral converges, and let $\tilde{\mu}(w) = \infty$ when the integral diverges. It follows from Proposition 3.2 that $\tilde{\mu}$ is locally integrable with respect to area measure. Thus $\tilde{\mu}$ is finite a.e. [Area]. (Also see Lemma 13.2.6.) Since $\tilde{\mu} \in L^1_{\text{loc}}$, the following definition makes sense.

5.1 Definition. If μ is a compactly supported measure on the plane, the *Cauchy transform* of μ is the function $\hat{\mu}$ defined a.e. [Area] by the equation

$$\hat{\mu}(w) = \int \frac{1}{z-w}\, d\mu(z).$$

In fact the Cauchy transform is the convolution of the locally integrable function z^{-1} and the compactly supported measure μ.

5.2 Proposition. *If μ is a compactly supported measure, the following statements hold.*

(a) *$\hat{\mu}$ is locally integrable.*

(b) *$\hat{\mu}$ is analytic on $\mathbb{C}_\infty \setminus \text{support}(\mu)$.*

(c) *For w in $\mathbb{C} \setminus \text{support}(\mu)$ and $n \geq 0$,*

$$\partial^n \hat{\mu}(w) = (-1)^n n! \int (z-w)^{-n-1} d\mu(z).$$

(d) *$\hat{\mu}(\infty) = 0$ and the power series of $\hat{\mu}$ near ∞ is given by*

$$\hat{\mu}(w) = \sum_{n=0}^{\infty} \left(\int z^n \, d\mu(z) \right) \frac{1}{w^{n+1}}.$$

Proof. The proof of (a) follows the lines of the discussion preceding the definition. Part (c), and hence the proof that $\hat{\mu}$ is analytic on $\mathbb{C} \setminus \text{support}(\mu)$, follows by differentiating under the integral sign. Note that as $w \to \infty$, $(z - w)^{-1} = w^{-1}(z/w - 1)^{-1} \to 0$ uniformly for z in any compact set. Hence $\hat{\mu}$ has a removable singularity at ∞ and $\hat{\mu}(\infty) = 0$.

It remains to establish (d). This is done by choosing R so that $\text{support}(\mu) \subseteq B(0; R)$, expanding $(z-w)^{-1} = -w^{-1}(1-z/w)^{-1}$ in a geometric series for $|w| > R$, and integrating term-by-term. □

A particular Cauchy transform deserves special consideration.

5.3 Proposition. *If K is a compact set having positive area and*

$$f(z) \equiv \int_K (\zeta - z)^{-1} dA(\zeta)$$

for all z in $\mathbb{C}$ and $f(\infty) = 0$, then $f : \mathbb{C}_\infty \to \mathbb{C}$ is a continuous function that is analytic on $\mathbb{C}_\infty \setminus K$ with $f'(\infty) = -\text{Area}\, K$. In addition,

$$|f(z)| \leq [\pi \ \text{Area}(K)]^{1/2}.$$

Proof. The fact that f is analytic on $\mathbb{C} \setminus K$ and $f(\infty) = 0$ follows from the preceding proposition. That f is a continuous function on $\mathbb{C}_\infty$ is left as an exercise for the reader (see Exercise 1). Since f is continuous at ∞, ∞ is a removable singularity; because $f(\infty) = 0$, $f'(\infty)$ is the limit of $zf(z)$ as $z \to \infty$. But $zf(z) = \int_K (\zeta/z - 1)^{-1} dA(\zeta) \to -\text{Area}\, K$ as $z \to \infty$ since $\zeta/z \to 0$ uniformly for ζ in K.

It remains to prove the inequality for $|f(z)|$. This inequality is due to Ahlfors and Beurling [1950], though the proof here is from Gamelin and Khavinson [1989]. From the properties we have already established and the

Maximum Modulus Theorem, f attains its maximum value at some point of K. By translating the set K, we may assume that $0 \in K$ and f attains its maximun at 0. In addition, if K is replaced by a suitable unimodular multiple of itself, we may assume that $f(0) > 0$. Thus

$$|f(z)| \leq f(0) = \int_K \operatorname{Re} \frac{1}{\zeta}\, d\mathcal{A}(\zeta).$$

Let $c = \frac{1}{2}\left[\frac{\pi}{\operatorname{Area} K}\right]^{1/2}$ and let $a = 1/2c$. It is elementary to see that the closed disk $D = \overline{B}(a;a)$ is $\{z : \operatorname{Re}(1/z) \geq c\}$ and that D and K have the same area. Thus $\mathcal{A}(D \cap K) + \mathcal{A}(D \setminus K) = \mathcal{A}(D) = \mathcal{A}(K) = \mathcal{A}(D \cap K) + \mathcal{A}(K \setminus D)$; hence $\mathcal{A}(D \setminus K) = \mathcal{A}(K \setminus D)$. On the other hand, $\operatorname{Re}(1/\zeta) \leq c$ for ζ in $K \setminus D$ and $\operatorname{Re}(1/\zeta) \geq c$ for ζ in $D \setminus K$. Therefore

$$\begin{aligned}
f(0) &\leq \int_{K\cap D} \operatorname{Re} \frac{1}{\zeta}\, d\mathcal{A} + c\mathcal{A}(K \setminus D) \\
&= \int_{K\cap D} \operatorname{Re} \frac{1}{\zeta}\, d\mathcal{A} + c\mathcal{A}(D \setminus K) \\
&\leq \int_{K\cap D} \operatorname{Re} \frac{1}{\zeta}\, d\mathcal{A} + \int_{D\setminus K} \operatorname{Re} \frac{1}{\zeta} \\
&= \int_D \operatorname{Re} \frac{1}{\zeta}\, d\mathcal{A}.
\end{aligned}$$

We leave it to the reader to show that for $0 < r < a$,

$$\int_0^{2\pi} \frac{1}{a + re^{i\theta}}\, d\theta = \frac{2\pi}{a},$$

by converting this to an integral around the circle $z = a + re^{i\theta}$. Hence converting to polar coordinates, we get that

$$\begin{aligned}
\int_D \operatorname{Re} \frac{1}{\zeta}\, d\mathcal{A} &= \int_0^a \int_0^{2\pi} \operatorname{Re} \frac{1}{a + re^{i\theta}}\, d\theta\, r\, dr \\
&= \int_0^a \operatorname{Re} \int_0^{2\pi} \frac{1}{a + re^{i\theta}}\, d\theta\, r\, dr \\
&= \int_0^a \frac{2\pi}{a}\, r\, dr \\
&= \pi a \\
&= \pi \left[\frac{\operatorname{Area} K}{\pi}\right]^{1/2} \\
&= [\pi \operatorname{Area} K]^{1/2}.
\end{aligned}$$

□

Note that if $\mu = \delta_a$, then $\hat{\mu}(z) = (a-z)^{-1}$. So in general the Cauchy transform of a compactly supported measure is not continuous.

Since $\hat{\mu}$ is locally integrable, it defines a distribution on $\mathbb{C}$ and so it can be differentiated.

5.4 Theorem. *If μ is a compactly supported measure on $\mathbb{C}$, then*

$$\bar{\partial}\hat{\mu} = -\pi\mu.$$

Moreover, $\hat{\mu}$ is the unique solution to this differential equation in the sense that if $h \in L^1_{\text{loc}}$ such that $\bar{\partial}h = -\pi\mu$, h is analytic in a neighborhood of ∞, and $h(\infty) = 0$, then $h = \hat{\mu}$ a.e. [Area].

Proof. If $\phi \in C_c^\infty$, then

$$\begin{aligned} \bar{\partial}\hat{\mu}(\phi) = -\int \hat{\mu}\bar{\partial}\phi\, d\mathcal{A} &= -\int \bar{\partial}\phi(z)\left[\int (w-z)^{-1} d\mu(w)\right] d\mathcal{A}(z) \\ &= \int \left[\int \bar{\partial}\phi(z)(z-w)^{-1}\, d\mathcal{A}(z)\right] d\mu(w). \end{aligned}$$

By Corollary 13.2.9 this becomes $\bar{\partial}\hat{\mu}(\phi) = -\pi\int \phi\, d\mu$, whence the first part of the theorem.

For the uniqueness statement, suppose h is such a function. It follows that $\bar{\partial}(\hat{\mu} - h) = 0$. By Weyl's Lemma (4.11), $\hat{\mu} - h$ is almost everywhere equal to an entire function f. But $f = \hat{\mu} - h$ has a removable singularity at ∞ and is 0 there. Hence $f \equiv 0$. □

5.5 Corollary. *If G is an open set, μ is a compactly supported measure on the plane, and $\hat{\mu} = 0$ a.e.* [Area] *on G, then $|\mu|(G) = 0$.*

Proof. If $\phi \in C_c^\infty(G)$, then $\int \phi\, d\mu = -\pi^{-1}\bar{\partial}\hat{\mu}(\phi) = 0$. It follows by Proposition 3.7 that $|\mu|(G) = 0$. □

5.6 Corollary. *If G is an open set, μ is a compactly supported measure on the plane, and $\hat{\mu}$ is analytic on G, then $|\mu|(G) = 0$.*

The Cauchy transform is the premier tool in uniform rational approximation and in the next section this statement will be borne out.

Exercises

1. If f is a function that is integrable with respect to Lebesque measure and has compact support, show that the Cauchy transform of $\mu = f\mathcal{A}$ is a continuous function on $\mathbb{C}_\infty$.

2. When does equality occur in the inequality in Proposition 5.3?

3. Using the method used in proving Proposition 5.3, show that for any compact set K, $\int_K |z-w|^{-1} d\mathcal{A}(w) \leq 2[\pi \text{Area}(K)]^{1/2}$. When does equality occur?

4. Let μ be a measure with compact support and suppose g is a continuously differentiable function with compact support. If $\nu = g\mu - \pi^{-1}(\partial g)\hat{\mu} \cdot \mathcal{A}$, then $\hat{\nu} = g\hat{\mu}$.

5. (a) If μ and ν are measures with compact support such that $\hat{\mu}$ and $\hat{\nu}$ are continuous functions, show that $\hat{\mu}\hat{\nu}$ is the Cauchy transform of $\hat{\nu}\mu + \hat{\mu}v$.

   (b) Show that if K is a compact set, then $\{\hat{h} : h$ is a bounded$\}$. Borel function with compact support and $h = 0$ a.e. on $K\}$ is a dense subalgebra of $R(K)$.

   (c) If K is compact and E is a Borel subset of K, define $R(K, E)$ to be the closure in $C(K)$ of $\{\hat{h} : h$ is a bounded Borel$\}$. function with compact support and $h = 0$ a.e. on $E\}$. Show that $R(K, E)$ is a subalgebra of $C(K)$ with the following properties: (i) $R(K) \subseteq R(K, E)$; (ii) a measure μ supported on K is orthogonal to $R(K, E)$ if and only if $\hat{\mu} = 0$ a.e. on $\mathbb{C} \setminus E$; (iii) if Area$(E) = 0$, $R(K, E) = C(K)$; and if Area$(K \setminus E) = 0$, $R(K, E) = R(K)$.

§6 An Application: Rational Approximation

In this section the Cauchy transform will be applied to prove two theorems in rational approximation: the Hartogs-Rosenthal Theorem and Runge's Theorem. But first a detour into some general material is required. The next result shows that by means of the Hahn-Banach Theorem questions of rational approximation in the supremun norm can be reduced to questions of weak approximation.

6.1 Definition. If K is a compact set in the plane, $R(K)$ is the uniform closure in $C(K)$ of Rat(K).

Note that $R(K)$ is a Banach algebra.

6.2 Theorem. *If K is a compact subset of $\mathbb{C}$ and μ is a measure on K, then $\mu \perp R(K)$ if and only if $\hat{\mu}(w) = 0$ a.e. [Area] on $\mathbb{C} \setminus K$.*

Proof. Assume $\mu \perp R(K)$ and note that for $w \notin K$, $(z-w)^{-1} \in R(K)$. Hence $\hat{\mu}(w) = 0$ off K. Conversely, assume that $\hat{\mu} = 0$ a.e. [Area] off K; since $\hat{\mu}$ is analytic off K, $\hat{\mu}$ is identically 0 off K. This implies that all the derivatives of $\hat{\mu}$ vanish on $\mathbb{C}_\infty \setminus K$. From (5.2.c) and (5.2.d) we get that μ annihilates all polynomials and all rational functions with poles off K. Hence $\mu \perp R(K)$. $\square$

Here is a classical theorem on rational approximation obtained before the introduction of the Cauchy transform. Note that it extends the Weierstrass Approximation Theorem.

6.3 Hartogs-Rosenthal Theorem. *If* $\text{Area}(K) = 0$, *then* $R(K) = C(K)$.

Proof. Let $\mu \in M(K)$ such that $\mu \perp R(K)$; so $\hat{\mu} = 0$ off K. Since $\text{Area}(K) = 0$, this implies that $\hat{\mu} = 0$ a.e. [Area] on $\mathbb{C}$. By Corollary 5.5, $\mu = 0$. By the Hahn-Banach Theorem $R(K) = C(K)$. □

One of the main ways in which Cauchy transforms are used is the following device. Assume μ is supported on the compact set K and let f be analytic on an open set G with $K \subseteq G$. Let Γ be a positively oriented smooth Jordan system in G such that $K \subseteq \text{ins}\,\Gamma$. Thus for every z in K, $f(z) = (2\pi i)^{-1} \int_\Gamma f(w)(w-z)^{-1} dw$. An application of Fubini's Theorem now implies that

6.4
$$\int f(z)\, d\mu(z) = -\frac{1}{2\pi i} \int_\Gamma f(w)\, \hat{\mu}(w)\, dw.$$

6.5 Runge's Theorem. *Let* K *be a compact subset of* $\mathbb{C}$ *and let* E *be a subset of* $\mathbb{C}_\infty \setminus K$ *that meets each component of* $\mathbb{C}_\infty \setminus K$. *If* f *is analytic in a neighborhood of* K, *then there are rational functions* $\{f_n\}$ *whose only poles are in the set* E *such that* $f_n \to f$ *uniformly on* K.

Proof. Let μ be a measure on K such that $\int g\, d\mu = 0$ for every rational function g with poles contained in E. It suffices to show that $\int f\, d\mu = 0$ for every function f that is analytic in a neighborhood of K.

Fix a component U of $\mathbb{C}_\infty \setminus K$ and let $w \in U \cap E$; assume for the moment that $w \neq \infty$. Using (5.2.c) we get that every derivative of $\hat{\mu}$ at w is 0. Thus $\hat{\mu}(z) \equiv 0$ on U. If $w = \infty$, then the assumption on μ implies that $\int p\, d\mu = 0$ for all polynomials. Thus using (5.2.d) we get that $\hat{\mu}(z) \equiv 0$ on U. Hence $\hat{\mu}$ vanishes on the complement of K (and so $\mu \perp R(K)$). If G and Γ are chosen as in the discussion prior to the statement of the theorem, (6.4) shows that $\int f\, d\mu = 0$. □

Uniform rational approximation is a subject unto itself; Conway [1991], Gamelin [1969], Stout [1971] are a few references.

Exercises

1. Let K be a compact subset of $\mathbb{C}$ and let $\mu \in M(K)$. If ϕ is any smooth function with compact support, put $\mu_\phi = \phi\mu - \pi^{-1}\hat{\mu}\,\bar{\partial}\phi\,\mathcal{A}$. Prove the following. (a) $\widehat{\mu_\phi} = \phi\hat{\mu}$. (b) $\mu \perp R(K)$ if and only if $\mu_\phi \perp R(K)$ for all smooth function ϕ with compact support. (c) $R(K) = C(K)$ if and only if for every closed disk D, $R(K \cap D) = C(K \cap D)$.

2. If K is a closed disk or an annulus, show that $R(K) = A(K)$, the algebra of continuous functions on K that are analytic on its interior.

3. Show that there are open disks $\{\Delta_j\}$ of radius r_j having the following properties: (i) cl $\Delta_j \subseteq \mathbb{D}$ and cl $\Delta_j \cap$ cl $\Delta_i = \emptyset$ for $i \neq j$; (ii) $\sum_j r_j < \infty$; (iii) $K = \text{cl}\,\mathbb{D} \backslash \bigcup_j \Delta_j$ has no interior. The set K is called a *Swiss cheese*. For $j \geq 1$, let γ_j be the boundary of Δ_j with positive orientation and let ν_j be the measure on K such that $\int f\,d\nu_j = -\int_{\gamma_j} f$ for every f in $C(K)$. Note that $\|\nu_j\| = r_j$. Let γ_0 be the positively oriented boundary of $\mathbb{D}$ and let ν_0 be the measure on K such that $\int f\,d\nu_0 = \int_{\gamma_0} f$ for every f in $C(K)$. Let $\mu \equiv \nu_0 + \sum_j \nu_j$, a measure in $M(K)$. Show that $\mu \perp R(K)$ and, since $\mu \neq 0$, $R(K) \neq C(K)$ even though K has no interior.

§7 Fourier Series and Cesàro Sums

Throughout this section, normalized Lebesgue measure on $\partial\mathbb{D}$ will be denoted by m, and the Lebesgue spaces of this measure will be denoted by $L^p(\partial\mathbb{D})$ or simply L^p. Note that since m is a finite measure, $L^p(\partial\mathbb{D}) \subseteq L^1(\partial\mathbb{D})$ for $1 \leq p \leq \infty$. So results obtained for functions that belong to L^1 are valid for functions in L^p. We also will be concerned with the space of continuous functions on $\partial\mathbb{D}$, $C = C(\partial\mathbb{D})$, and its dual, the space of complex-valued regular Borel measures on $\partial\mathbb{D}$, $M = M(\partial\mathbb{D})$.

7.1 Definition. If $\mu \in M(\partial\mathbb{D})$, then the *Fourier transform* of μ is the function $\hat{\mu} : \mathbb{Z} \to \mathbb{C}$ defined by

$$\hat{\mu}(n) = \int_{\partial\mathbb{D}} \overline{z^n}\,d\mu.$$

(First our apologies for using the same notation for the Fourier transform of μ as for the Cauchy transform, but here is an instance where tradition is best followed.) Now if $f \in L^1$, then $d\mu = f\,dm$ is a measure and so its Fourier transform can also be defined. Here the notation used is $\hat{f} = \hat{\mu}$.

For any measure μ we call $\{\hat{\mu}(n) : n \in \mathbb{Z}\}$ the *Fourier coefficients* of μ. The series $\sum_{n=-\infty}^{\infty} \hat{\mu}(n) z^n$ is called the *Fourier series* for the measure μ.

The idea here is that we would like to know if the measure or function can be recaptured from its Fourier series. That we should have any right to have such a hope stems from the density of a certain set of functions. A *trigonometric polynomial* is a function in $C(\partial\mathbb{D})$ of the form $\sum_{k=-m}^{n} a_k z^k$.

7.2 Proposition. *The trigonometric polynomials are uniformly dense in $C(\partial\mathbb{D})$ and hence dense in $L^p(\partial\mathbb{D})$ for $1 \leq p < \infty$; they are weak* dense in $L^\infty(\partial\mathbb{D})$. Thus $\{z^n : n \in \mathbb{Z}\}$ is an orthonormal basis for $L^2(\partial\mathbb{D})$.*

Proof. The first part is an easy consequence of the Stone-Weierstrass Theorem. The last statement only needs the calculation necessary to show that the functions z^n are orthonormal. $\square$

7.3 Corollary. *If $\mu \in M$ and $\hat{\mu}(n) = 0$ for all n in $\mathbb{Z}$, then μ is the zero measure.*

The preceding corollary says that a measure is completely determined by its Fourier coefficients. Thus we have the hope that the measure can be recaptured from its Fourier series. At least in the case of functions in the space L^2 this hope is borne out.

7.4 Theorem. *If $f \in L^2(\partial\mathbb{D})$, then $\hat{f} \in \ell^2(\mathbb{Z})$. If $V : L^2(\partial\mathbb{Z}) \to \ell^2(\mathbb{Z})$ is defined by $Vf = \hat{f}$, then V is an isomorphism of the Hilbert spaces.*

Proof. The first part, that $\hat{f} \in \ell^2(\mathbb{Z})$, as well as the statement that V is an isometry, is a direct consequence of Parseval's Identity and the fact that $\{z^n\}$ is a basis for $L^2(\partial\mathbb{D})$. If $f = z^m$, then it is straightforward to check that $\hat{f}(n) = 0$ if $n \neq m$ and $\hat{f}(m) = 1$. Thus the range of V is dense and so V must be an isomorphism. $\square$

Theorem 7.4 says that, at least in the case of an L^2 function, the Fourier series converges to the function in the L^2 norm. This is not the case for other functions and measures, but the intricacies of this theory are more appropriately handled by themselves. Instead we will concentrate on what is true and will have value for later in this book.

The reader interested in pursuing convergence of Fourier series can see Chapter II of Katznelson [1976]. In particular it is proved there that the Fourier series of a function in L^p, $1 < p < \infty$, converges to the function in norm and that this is false for L^1. He also gives an example of a continuous function whose Fourier series diverges at a point and thus cannot converge uniformly to the function. The story for pointwise convergence is much more complicated. It was proved in Carleson [1966] that the Fourier series of a function in L^2 converges a.e. This was extended in Hunt [1967] to L^p, $p > 1$. An exposition of the Carleson and Hunt work can be found in Mozzochi [1971]. In Katznelson [1976] is the proof of a result of Kolmogorov that there is a function in L^1 whose Fourier series diverges everywhere.

For any formal Fourier series $\sum_{n=-\infty}^{\infty} c_n z^n$, z in $\partial\mathbb{D}$, let $s_n(z)$ be the n-th partial sum of the series, $s_n(z) = \sum_{k=-n}^{n} c_k z^k$. The n-th *Cesàro means* of the series is defined for z in $\partial\mathbb{D}$ by

$$\sigma_n(z) = \frac{1}{n} \sum_{k=0}^{n-1} s_k(z).$$

It is worth noting that a Cesàro mean is a trigonometric polynomial. By the n-th partial sum and the n-th Cesàro mean for the measure μ, we mean the corresponding quantity for the associated series. To indicate the dependence on μ, these sums are denoted by $s_n(\mu, z)$ and $\sigma_n(\mu, z)$; if $f \in L^p$, $s_n(f, z)$ and $\sigma_n(f, z)$ are the corresponding sums.

Recall that $M = C^*$, and hence M has a natural weak* topology.

Here is the main result of this section.

7.5 Theorem.

(a) *If $f \in L^p$, $1 \le p < \infty$, then $\sigma_n(f, z) \to f$ in the L^p norm.*

(b) *If $f \in C$, $\sigma_n(f, z) \to f$ uniformly on $\partial\mathbb{D}$.*

(c) *If $f \in L^\infty$, $\sigma_n(f, z) \to f$ in the weak* topology of L^∞.*

(d) *If $\mu \in M$, $\sigma_n(\mu, z) \to \mu$ in the weak* topology of M. (Here we think of the Cesàro mean $\sigma_n(\mu, z)$ as the measure $\sigma_n(\mu, z) \cdot m$.)*

The proof will be obtained by a recourse to operator theory on a Banach space. If $\mathcal{X}$ is one of the Banach spaces under consideration (that is, $\mathcal{X} = L^p$, C, or M), define $\sigma_n : \mathcal{X} \to \mathcal{X}$ by letting $\sigma_n(x) =$ the n-th Cesàro sum of x. (If $\mathcal{X} = M$ and $\mu \in M$, then $\sigma_n(\mu)$ is the measure that is absolutely continuous with respect to m whose Radon-Nikodym derivative is the trigonometric polynomial $\sigma_n(\mu)$.) To prove the theorem, it must be shown that if $\mathcal{X} = L^p$, $1 \le p < \infty$, or C, then $\|\sigma_n(z) - x\| \to 0$, and if $\mathcal{X} = L^\infty$ or M, then $\sigma_n(x) \to x$ weak* for every x in $\mathcal{X}$. Actually we will see that the last part follows from the first part and a duality argument. But first we will see that σ_n is actually an integral operator.

If $f \in L^1$ with Fourier series $\sum_{n=-\infty}^{\infty} c_n z^n$, then

$$\begin{aligned} s_n(f, z) &= \sum_{k=-n}^{n} \left[\int f(w)\overline{w^k} dm(w) \right] z^k \\ &= \int f(w) \left[\sum_{k=-n}^{n} (\overline{w}z)^k \right] dm(w). \end{aligned}$$

Therefore

7.6 $$\sigma_n(f, z) = \int f(w) K_n(\overline{w}z)\, dm(w),$$

where K_n is the n-th Cesàro mean of the formal series $\sum_{n=-\infty}^{\infty} \zeta^n$. This kernel K_n is called *Fejer's kernel.* The same type of formula holds for a measure μ:

7.7 $$\sigma_n(\mu, z) = \int K_n(\overline{w}z)\, d\mu(w).$$

To properly study σ_n, we need to get a better hold on the kernel K_n. To do this, let's first look at the n-th partial sum of the series $\sum_{n=-\infty}^{\infty} \zeta^n$. So if $\zeta = e^{i\theta}$,

$$\begin{aligned} s_n(\zeta) &= \sum_{k=-n}^{n} \zeta^k \\ &= \sum_{k=0}^{n} \zeta^k + \sum_{k=1}^{n} \overline{\zeta^k} \\ &= \frac{1-\zeta^{n+1}}{1-\zeta} + \frac{1-\overline{\zeta^{n+1}}}{1-\overline{\zeta}} - 1 \\ &= \frac{\operatorname{Re}\zeta^n - \operatorname{Re}\zeta^{n+1}}{1-\operatorname{Re}\zeta} \\ &= \frac{\cos n\theta - \cos(n+1)\theta}{1-\cos\theta}. \end{aligned}$$

From here it follows that

7.8 $$K_n(\zeta) = \frac{1}{n}\left[\frac{1-\operatorname{Re}\zeta^n}{1-\operatorname{Re}\zeta}\right] = \frac{1}{n}\left[\frac{1-\cos n\theta}{1-\cos\theta}\right].$$

7.9 Lemma. *For each $n \geq 1$, $K_n \geq 0$, $K_n(\zeta) = K_n(\overline{\zeta})$, and $\int K_n dm = 1$.*

Proof. Applying the half angle formulas from trigonometry to (7.8), it follows that $K_n(\zeta) = n^{-1}[(\sin(n\theta/2))/\sin(\theta/2)]^2 \geq 0$. It is also clear from (7.8) that the second part is valid. For the last part, use (7.6) with $f \equiv 1$. □

7.10 Lemma.

(a) *If $1 \leq p < \infty$ and q is the index dual to p, then the adjoint of the operator $\sigma_n : L^p \to L^p$ is the operator $\sigma_n : L^q \to L^q$.*

(b) *The adjoint of the operator $\sigma_n : C \to C$ is the operator $\sigma_n : M \to M$.*

Proof. Only part (a) will be proved. Let $f \in L^p$ and $g \in L^q$. By interchanging the order of integration and using the preceding lemma, it follows that

$$\begin{aligned} \langle \sigma_n(f), g\rangle &= \int g(z)\left[\int f(w)K_n(\overline{w}z)\, dm(w)\right] dm(z) \\ &= \int f(w)\left[\int g(z)K_n(\overline{w}z)\, dm(z)\right] dm(w) \\ &= \int f(w)\left[\int g(z)K_n(\overline{z}w)\, dm(z)\right] dm(w) \\ &= \langle f, \sigma_n(g)\rangle. \end{aligned}$$

□

We can now state a general Banach space result that, when combined with the preceding lemma, will show how parts (c) and (d) of Theorem 7.5 follow from parts (a) and (b).

7.11 Proposition. *Let $\mathcal{X}$ be a Banach space and let $\{T_n\}$ be a sequence of bounded operators from $\mathcal{X}$ into $\mathcal{X}$.*

(a) *If $\sup_n \|T_n\| < \infty$, D is a dense subset of $\mathcal{X}$, and $\|T_n x - x\| \to 0$ for all x in D, then $\|T_n x - x\| \to 0$ for all x in $\mathcal{X}$.*

(b) *If $\|T_n(x) - x\| \to 0$ for all x in $\mathcal{X}$, then for every x^* in $\mathcal{X}^*$, $T_n^* x^* \to x^*$ weak* in $\mathcal{X}^*$.*

Proof. (a) If $x \in \mathcal{X}$ and ε is a positive number, let $y \in D$ such that $\|x-y\| < \min(\varepsilon/2c, \varepsilon/2)$, where $c = \sup_n \|T_n\|$. Then $\|T_n x - x\| \le \|T_n(x-y)\| + \|T_n y - y\| + \|y - x\| \le 2\varepsilon/3 + \|T_n y - y\|$. This can be made less than ε for all sufficiently large n.
(b) This is easy: $|\langle x, T_n^* x^* - x^*\rangle| = |\langle T_n x - x, x^*\rangle| \le \|T_n x - x\| \, \|x^*\| \to 0$.
□

We can now prove the main theorem.

Proof of Theorem 7.5. (a) Let $f \in L^p$ and let $g \in L^q$, where q is dual to p. First note that, by a change of variables, $\sigma_n(f,z) = \int f(w) K_n(\overline{w}z)dm(w) = \int f(\overline{\zeta}z)\, K_n(\zeta)\, dm(\zeta)$. Hence

$$\begin{aligned}
|\langle \sigma_n(f), g\rangle| &= \left|\int g(z)\left[\int f(w)K_n(\overline{w}z)\, dm(w)\right] dm(z)\right| \\
&= \left|\int g(z)\left[\int g(\overline{\zeta}z)K_n(\zeta)\, dm(\zeta)\right] dm(z)\right| \\
&= \left|\int K_n(\zeta)\left[\int g(z)f(\overline{\zeta}z)\, dm(z)\right] dm(\zeta)\right| \\
&\le \int K_n(\zeta) \int \left|g(z)f(\overline{\zeta}z)\right|\, dm(z)\; dm(\zeta).
\end{aligned}$$

Applying Hölder's inequality and using the fact that $\int f(\overline{\zeta}z)\, dm(z) = \int f\, dm$, we get

$$\begin{aligned}
|\langle \sigma_n(f), g\rangle| &\le \int K_n(\zeta)\|g\|_q\|f\|_p dm(\zeta) \\
&\le \|g\|_q\|f\|_p
\end{aligned}$$

since $\int K_n dm = 1$.

Hence $\|\sigma_n\| \le 1$ for all n. It is easy to check that for any integer k, $\sigma_n(z^k) \to z^k$ uniformly as $n \to \infty$. Thus for a trigonometric polynomial f, $\sigma_n(f) \to f$ uniformly as $n \to \infty$. Since the trigonometric polynomials are dense in L^p for all finite p, part (a) follows from part (a) of the preceding proposition.

(b) This is easier than part (a). If $f \in C$, then $|\sigma_n(f,z)| \leq \int |f(w)|\, K_n(\overline{w}z\, dm(w) \leq ||f||_\infty$ and so $||\sigma_n|| \leq 1$. Since convergence holds for the trigonometric polynomials, part (b) follows.

As mentioned before, parts (c) and (d) follow from parts (a) and (b) via the second part of the preceding proposition. □

Exercises

1. Compute the Cesàro means of δ_1, the unit point mass at 1.

2. If μ and $\nu \in M$, show that $\widehat{\mu * \nu}(n) = \hat{\mu}(n)\hat{\nu}(n)$ for all n in $\mathbb{Z}$.

3. In this and succeeding exercises a concept is presented that can be used to give some of the results of this section a unifying treatment. For details see Katznelson [1976], p. 14. For any function f on $\partial\mathbb{D}$ and z in $\partial\mathbb{D}$, define $f_z(w) = f(w\overline{z})$. A linear manifold $\mathcal{X}$ in L^1 is a *homogeneous* space if: (i) $\mathcal{X}$ has a norm $||\cdot||$ such that $||f|| \geq ||f||_1$ for all f in $\mathcal{X}$ and with this norm $\mathcal{X}$ is a Banach space; (ii) $f_z \in \mathcal{X}$ whenever $f \in \mathcal{X}$; (iii) $||f|| = ||f_z||$ for all f in $\mathcal{X}$ and all z in $\partial\mathbb{D}$; (iv) for each f in $\mathcal{X}$, the map $z \to f_z$ is a continuous function from $\partial\mathbb{D}$ into $\mathcal{X}$. (a) Show that $C(\partial\mathbb{D})$ and L^p, $1 \leq p < \infty$, are homogeneous spaces. (b) Show that L^∞ satisfies properties (i) through (iii) in the definition, but is not a homogeneous space. (c) If $\mathcal{X}$ is a linear manifold in L^1 that satisfies conditions (i) through (iii) in the definition of a homogeneous space and $\mathcal{X}_c$ is defined as the set of all f in $\mathcal{X}$ such that the function $z \to f_z$ is a continuous function from $\partial\mathbb{D}$ into $\mathcal{X}$, show that $\mathcal{X}_c$ is a closed subspace of $\mathcal{X}$ (and hence a homogeneous space). If $\mathcal{X} = L^\infty$, show that $\mathcal{X}_c = C$.

4. This exercise continues the preceding one; $\mathcal{X}$ denotes a homogeneous space. (a) If $f \in \mathcal{X}$ and $g \in L^1$, show that $f * g \in \mathcal{X}$. (b) If $\sigma_n(f)$ is as in (7.6), show that $||\sigma_n(f) - f|| \to 0$ for every f in $\mathcal{X}$. (c) Show that the trigonometric polynomials are dense in $\mathcal{X}$.

5. If $\mathcal{X}$ and $\mathcal{X}_c$ are as in part (c) of Exercise 3, show that $\sigma_n(f) \in \mathcal{X}_c$ for all f in $\mathcal{X}$. Prove that $\mathcal{X}_c$ is the closure in $\mathcal{X}$ of the trigonometric polynomials.

6. This exercise continues Exercise 4 and maintains its notation. If $x^* \in \mathcal{X}^*$ and $z \in \partial\mathbb{D}$, define $x_z^*(f) = x^*(f_z)$ for all f in $\mathcal{X}$. (a) Show that $||x_z^*|| = ||x^*||$ and $z \to x_z^*$ is a continuous function from $\partial\mathbb{D}$ into $(\mathcal{X}^*, wk^*)$. (b) If σ_n is as in (7.6) and $\sigma_n^* : \mathcal{X}^* \to \mathcal{X}^*$ is its dual map, show that $\sigma_n^*(x^*) \to x^*(wk^*)$ in $\mathcal{X}^*$ for all x^* in $\mathcal{X}^*$.

Chapter 19

Harmonic Functions Redux

In this chapter a treatment of the Dirichlet problem for sets in the plane is presented. This topic will be continued in Chapter 21 when harmonic measure and logarithmic capacity are introduced and applied. Some material from Chapter 10 must be restated in the more general setting needed for the more extensive study of harmonic functions. In Chapter 10 all functions considered were continuous; here measure theory will be used to broaden the class of functions. The attitude taken will be that usually results from Chapter 10 will be restated in the more inclusive context, but proofs will be furnished only if there is a significant difference between the proof needed at present and the one given for continuous functions.

The chapter begins by returning to a closer examination of functions defined on the unit disk, $\mathbb{D}$.

§1 Harmonic Functions on the Disk

The notation from §18.7 remains in force. Recall the definition of the Poisson kernel: for w in $\partial\,\mathbb{D}$ and $|z| < 1$,

$$\begin{aligned} P_z(w) &= \operatorname{Re}\left(\frac{w+z}{w-z}\right) \\ &= \operatorname{Re}\left(\frac{1+\overline{w}}{1-z\overline{w}}\right) \\ &= \sum_{n=0}^{\infty}(z\overline{w})^n + \sum_{n=1}^{\infty}(\overline{z}w)^n. \end{aligned}$$

The reader should review the properties of the Poisson kernel from Proposition 10.2.3.

If $\mu \in M$ and $|z| < 1$, define $\tilde{\mu}(z) = \int P_z(w)\,d\mu(w)$. Similarly, if $f \in L^1$, define $\tilde{f}(z) = \int f\,P_z\,dm$. These definitions are consistent since $\tilde{fm} = \tilde{f}$. It is not difficult to prove the following.

1.1 Proposition. *If $\mu \in M(\partial\,\mathbb{D})$, then $\tilde{\mu}$ is a harmonic function in $\mathbb{D}$.*

Note that we are dealing with complex valued harmonic functions here. If $f \in C(\partial\,\mathbb{D})$, then we know that $\tilde{f}$ is the solution of the Dirichlet problem with boundary values f (see Theorem 10.2.4). Indeed, this is given another proof in Theorem 1.4.a below.

If $u : \mathbb{D} \to \mathbb{C}$ and $0 < r < 1$, define $u_r : \partial\,\mathbb{D} \to \mathbb{C}$ by $u_r(w) = u(rw)$ for w in $\partial\,\mathbb{D}$. We will sometimes want to consider the function u_r as defined on $\mathbb{D}$ or cl $\mathbb{D}$ by the same formula, but no separate notation will be employed.

If $\mu \in M$ and $0 < r < 1$, then $\tilde{\mu}_r$ is an element of $C(\partial\,\mathbb{D})$ by the preceding proposition. Thus if $1 \le p \le \infty$ and $f \in L^p$, $\tilde{f}_r \in L^p$ and so we can define the operator $T_r : L^p \to L^p$ by $T_r f = \tilde{f}_r$; similarly, we can define $T_r : C \to C$ by $T_r f = \tilde{f}_r$ and $T_r\mu : M \to M$ by $T_r\mu = \tilde{\mu}_r \cdot m$.

1.2 Proposition.

(a) *For $\mathcal{X} = L^p$, C, or M, $T_r : \mathcal{X} \to \mathcal{X}$ is a bounded linear operator with $\|T_r\| \le 1$ for all r.*

(b) *If $1 \le p < \infty$ and q is conjugate to p, the dual of the map $T_r : L^p \to L^p$ is the map $T_r : L^q \to L^q$.*

(c) *The dual of the map $T_r : C \to C$ is the map $T_r : M \to M$.*

Proof. (a) This will only be proved for $\mathcal{X} = L^p$. Let $f \in L^p$ and let $h \in L^q$, where q is the index that is conjugate to p. Thus

$$
\begin{aligned}
\langle T_r f, h\rangle &= \int \tilde{f}_r(\zeta)\, h(\zeta)\, dm(\zeta) \\
&= \int \left[\int f(w)\, P_{r\zeta}(w)\, dm(w)\right] h(\zeta)\, dm(\zeta). \qquad \textbf{1.3}
\end{aligned}
$$

Substitute $w\zeta$ for w in this equation and use the following two facts: this change of variables does not change the value of the integral and $P_{r\zeta}(w\zeta) = P_r(w)$. This gives

$$
\begin{aligned}
|\langle T_r f, h\rangle| &= \left|\int \left[\int f(w\zeta)\, P_r(w)\, dm(w)\right] h(\zeta)\, dm(\zeta)\right| \\
&\le \int P_r(w) \left[\int |f(w\zeta)|\, |h(\zeta)|\, dm(\zeta)\right] dm(w) \\
&\le \|f\|_p\, \|h\|_q,
\end{aligned}
$$

since $\int P_r(w)\, dm(w) = 1$. This shows that $\|T_r\| \le 1$ for all r. It is easy to see that T_r is linear.

(b) If $f \in L^p$ and $h \in L^q$, then use the fact that $P_{r\zeta}(w) = P_{r\overline{w}}(\overline{\zeta}) = P_{rw}(\zeta)$ for $|w| = |\zeta| = 1$ to obtain

$$
\begin{aligned}
\langle T_r f, h\rangle &= \int \left[\int f(w)\, P_{rw}(\zeta)\, dm(w)\right] h(\zeta)\, dm(\zeta) \\
&= \int \left[\int h(\zeta)\, P_{rw}(\zeta)\, dm(\zeta)\right] f(w)\, dm(w) \\
&= \langle f, T_r h\rangle.
\end{aligned}
$$

The proof of (c) is similar. □

1.4 Theorem.

(a) *If $f \in C(\partial\,\mathbb{D})$, then $\|f - \tilde{f}_r\|_{\partial\mathbb{D}} \to 0$ as $r \to 1-$.*

(b) *If $1 \le p < \infty$ and $f \in L^p(\partial\,\mathbb{D})$, then $\|f - \tilde{f}_r\|_p \to 0$.*

(c) *If $\mu \in M(\partial\,\mathbb{D})$ and if μ_r is the element of $M(\partial\,\mathbb{D})$ defined by $\mu_r = (\tilde{\mu})_r \cdot m$, then $\mu_r \to \mu$ weak* in $M(\partial\,\mathbb{D})$.*

(d) *If $f \in L^\infty(\partial\,\mathbb{D})$, then $\tilde{f}_r \to f$ weak* in $L^\infty(\partial\,\mathbb{D})$.*

Proof. It is easy to check that $T_r z^n = r^n z^n$ for all n in $\mathbb{Z}$. Hence for a trigonometric polynomial f, $T_r f \to f$ uniformly on $\partial\,\mathbb{D}$ as $r \to 1$. Since $\|T_r\| \le 1$ for all r, Proposition 18.7.2 implies that (a) and (b) hold. Parts (c) and (d) follow by applying Proposition 18.7.11.b and Proposition 1.2.c. □

As was said before, part (a) of the preceding theorem shows that, for f in $C(\partial\,\mathbb{D})$, $\tilde{f}$ has a continuous extension to cl $\mathbb{D}$, thus solving the Dirichlet problem with boundary values f. For f in $L^p(\partial\,\mathbb{D})$ we can legitimately consider $\tilde{f}$ as the solution of the Dirichlet problem with the non-continuous boundary values f. Indeed, such a perspective is justified by the last theorem. Further justification is furnished in the next section when we show that f can be recaptured as the radial limit of $\tilde{f}$.

Now suppose that $u : \mathbb{D} \to \mathbb{C}$ is a harmonic function. What are necessary and sufficient conditions that $u = \tilde{\mu}$ for some measure μ? Before providing an answer to this question, it is helpful to observe the consequences of a few elementary manipulations with the basic properties of a harmonic function.

If u is harmonic and real valued, there is an analytic function $f : \mathbb{D} \to \mathbb{C}$ with $u = \text{Re}\ f$. Let $f(z) = \sum_n a_n z^n$ be the power series expansion of f in $\mathbb{D}$. So for w in $\partial\,\mathbb{D}$ and $0 < r < 1$ the series $\sum_n a_n r^n w^n$ converges absolutely and thus

$$\begin{aligned} u_r(w) &= \frac{1}{2}[f(rw) + f(rw)^*] \\ &= \frac{1}{2}\left[\sum_{n=0}^{\infty} a_n r^n w^n + \sum_{n=0}^{\infty} \overline{a_n} r^n \overline{w^n}\right] \\ &= \sum_{n=-\infty}^{\infty} c_n r^{|n|} w^n, \end{aligned}$$

where $c_0 = \text{Re}\ a_0$, $c_n = \frac{1}{2}a_n$ for n positive, and $c_n = \frac{1}{2}\overline{a_n}$ for $n < 0$. For a complex valued harmonic function $u : \mathbb{D} \to \mathbb{C}$, a consideration of its real and imaginary parts shows that for $|w| = 1$ and $r < 1$,

1.5
$$u_r(w) = \sum_{n=-\infty}^{\infty} c_n r^{|n|} w^n$$

for some choice of constants c_n. Moreover the series (1.5) converges uniformly and absolutely for w in $\partial\,\mathbb{D}$. It follows that (1.5) is the Fourier series of the function u_r. The next lemma formally states this and gives the analogous fact for measures.

1.6 Lemma.

(a) *If $u : \mathbb{D} \to \mathbb{C}$ is a harmonic function and $0 < r < 1$, the Fourier series for u_r is given by the formula (1.5).*

(b) *If $\mu \in M(\partial\,\mathbb{D})$, then the Fourier series of the function $\tilde{\mu}_r$ is given by*

$$\sum_{n=-\infty}^{\infty} r^{|n|}\hat{\mu}(n)w^n$$

and the convergence is uniform and absolute for w on $\partial\,\mathbb{D}$.

The next theorem is the principal result of this section and characterizes the harmonic functions that can arise as the Poisson transform of a measure or a function from one of the various classes.

1.7 Theorem. *Suppose $u : \mathbb{D} \to \mathbb{C}$ is a harmonic function.*

(a) *There is a μ in $M(\partial\,\mathbb{D})$ with $u = \tilde{\mu}$ if and only if $\sup_r ||u_r||_1 < \infty$.*

(b) *If $1 < p \leq \infty$, there is a function f in $L^p(\partial\,\mathbb{D})$ with $u = \tilde{f}$ if and only if $\sup_r ||u_r||_p < \infty$.*

(c) *There is a function f in $L^1(\mathbb{D})$ with $u = \tilde{f}$ if and only if $\{u_r\}$ is L^1 convergent.*

(d) *There is a function f in $C(\partial\,\mathbb{D})$ with $u = \tilde{f}$ if and only if $\{u_r\}$ is uniformly convergent.*

Proof. (a) If $u = \tilde{\mu}$, then $||u_r||_1 = \int |u_r|\, dm \leq \int \left|\int P_{rz}(w)\, d\mu(w)\right| dm(z) \leq \int\int P_{rz}(w)\, d|\mu|(w)\, dm(z) = \int[\int P_{rz}(w)\, dm(z)]d|\mu|(w) = ||\mu||$.

Now assume that u is a harmonic function on $\mathbb{D}$ and L is a constant such that $||u_r||_1 \leq L$ for all $r < 1$. Put $\nu_r =$ the measure $u_r \cdot m$ in $M(\partial\,\mathbb{D})$. So $\{\nu_r\}$ is a uniformly bounded net of measures on $\partial\,\mathbb{D}$. By Alaoglu's Theorem there exists a measure μ in $M(\partial\,\mathbb{D})$ that is a weak* cluster point of this net. Hence

$$\hat{\nu}_r(n) = \int \overline{w^n} d\nu_r = \int u_r \overline{w^n} dm \to_{\text{cl}} \hat{\mu}(n).$$

But Lemma 1.6 implies that $\hat{\nu}_r(n) = \hat{\mu}_r(n) = r^{|n|}c_n \to c_n$ as $r \to 1$. Hence $\hat{\mu}(n) = c_n$. This implies that the weak* cluster point of $\{\nu_r\}$ is unique. Hence$\nu_r \to \mu$ weak* in $M(\partial\,\mathbb{D})$. An examination of the series in Lemma 1.6 shows that $u = \tilde{\mu}$.

(b) This proof is like that of part (a). For $1 < p < \infty$, the weak compactness of bounded sets in $L^p(\partial\,\mathbb{D})$ is used instead of weak* compactness. The weak* topology on $L^\infty(\partial\,\mathbb{D})$ is used when $p = \infty$.

The proofs of (c) and (d) are left as exercises. □

Part of the proof of this theorem needs to be made explicit.

1.8 Corollary. *Suppose $u : \mathbb{D} \to \mathbb{C}$ is a harmonic function.*

(a) *If $\sup_r ||u_r||_1 < \infty$, then the measures $u_r \cdot m \to \mu$ weak* in $M(\partial\,\mathbb{D})$, where μ is the measure such that $u = \tilde{\mu}$.*

(b) *If $1 < p \leq \infty$ and $\sup_r ||u_r||_p < \infty$, then $u_r \to f$ weakly in L^p (weak* in L^∞ if $p = \infty$), where f is the function in $L^p(\partial\,\mathbb{D})$ with $u = \tilde{f}$.*

If the proof of Theorem 10.2.4 is examined closely, a "point" theorem results. This next result not only improves (10.2.4) but provides a means of obtaining various estimates for harmonic functions as the subsequent corollary illustrates.

1.9 Theorem. *If $f \in L^1$ and f is continuous at the point a, then the function that is defined to be $\tilde{f}$ on $\mathbb{D}$ and f on $\partial\,\mathbb{D}$ is continuous at a.*

1.10 Corollary. *If $f \in L^1$ and $a \in \partial\,\mathbb{D}$, then*

$$\limsup_{z \to a} \tilde{f}(z) \leq \limsup_{\substack{\zeta \to a \\ \zeta \in \partial\,\mathbb{D}}} f(\zeta).$$

Proof. Let α be the right hand side of this inequality. If $\alpha = \infty$, there is nothing to prove; thus it may be assumed that $\alpha < \infty$. By definition, for every $\varepsilon > 0$ there is a $\delta > 0$ such that for ζ in $\partial\,\mathbb{D}$ and $|\zeta - a| < \delta$, $f(\zeta) < \alpha + \varepsilon$. Define f_1 on $\partial\,\mathbb{D}$ by letting $f_1(\zeta) = f(\zeta)$ for $|z - a| \geq \delta$ and $f_1(\zeta) = \alpha + \varepsilon$ for $|\zeta - a| < \delta$. So $f_1 \in L^1$ and $f \leq f_1$; thus $\tilde{f} \leq \tilde{f}_1$. Using the preceding theorem,

$$\limsup_{z \to a} \tilde{f}(z) \leq \limsup_{z \to a} \tilde{f}_1(z) = \alpha + \varepsilon.$$

Since ε was arbitrary, the proof is complete. □

The final result can be taken as a corollary of Theorem 1.7.

1.11 Herglotz's Theorem. *If u is a harmonic function on $\mathbb{D}$, then $u = \tilde{\mu}$ for a positive measure μ on $\partial\,\mathbb{D}$ if and only if $u \geq 0$ on $\mathbb{D}$.*

Proof. It is easy to see that since the Poisson kernel is positive, for any positive measure μ, $u = \tilde{\mu} \geq 0$ on $\mathbb{D}$. Conversely, assume that $u(z) \geq 0$ for $|z| < 1$. Then $||u_r||_1 = \int u_r(w)\, dm(w) = u(0)$ by the Mean Value Theorem.

By Corollary 1.8, there is a μ in $M(\partial\,\mathbb{D})$ such that $u = \tilde{\mu}$ and $u_r \cdot m \to \mu$ weak*. Since $u_r \geq 0$, it must be that $\mu \geq 0$. □

Exercises

1. Let $p(z, \bar{z})$ be a polynomial in z and $\bar{z}$ and find a formula for the function u that is harmonic on $\mathbb{D}$, continuous on cl $\mathbb{D}$, and equal to $p(z, \bar{z})$ on $\partial\,\mathbb{D}$.

2. If $\mu \in M(\partial\,\mathbb{D})$ has Fourier coefficients $\{\hat{\mu}(n)\}$, show that for $|z| < 1$, $\tilde{\mu}(z) = \sum_{n=0}^{\infty} \hat{\mu}(n)\, z^n + \sum_{n=1}^{\infty} \hat{\mu}(-n)\overline{z^n}$. Examine Exercise 1 in light of this.

3. Let u be a real-valued harmonic function on $\mathbb{D}$ and show that there is a real-valued measure μ on $\partial\,\mathbb{D}$ such that $u = \tilde{\mu}$ if and only if u is the difference of two positive harmonic functions.

4. Prove the following equivalent formulation of Herglotz's Theorem. If f is an analytic function on $\mathbb{D}$, then f takes its values in the right half plane and satisfies $f(0) > 0$ if and only if there is a positive measure μ on $\partial\,\mathbb{D}$ such that
$$f(z) = \int_{\partial\,\mathbb{D}} \frac{w + z}{w - z}\, d\mu(w).$$

5. Let C be the set of analytic functions on $\mathbb{D}$ such that Re $f \geq 0$ and $f(0) = 1$. Show that C is a compact convex subset of $H(\mathbb{D})$ and characterize its extreme points. (Hint: Use Exercise 4.)

§2 Fatou's Theorem

We have seen in the preceding section that for a measure μ in $M = M(\partial\,\mathbb{D})$, the measure μ can be recaptured from $\tilde{\mu}$, the solution of the Dirichlet problem with boundary values μ, by examining the weak* limit of the measures $(\tilde{\mu})_r \cdot m$ (1.4). In this section we will look at the radial limit of the function $\tilde{\mu}(z)$. For an arbitrary measure we recapture μ if and only if μ is absolutely continuous with respect to m. This will essentially prove the results stated in §13.5.

There is a standard temptation for all who first see Fatou's Theorem. If $f \in L^1$, then we know that $\tilde{f}_r \to f$ in the L^1 norm. Thus there is a sequence $\{r_n\}$ that converges to 1 such that $\tilde{f}_{r_n}(\zeta) \to f(\zeta)$ a.e. on $\partial\,\mathbb{D}$. That is, $\tilde{f}(r_n\zeta) \to f(\zeta)$ a.e. This is sufficiently close to the existence of radial limits a.e. on $\partial\,\mathbb{D}$ that it seems that a proof of their existence is just ahead of us. Unfortunately no one has ever been able to parlay this into a proof. It remains only an intuitive argument that makes the result believable.

Recall some measure theory, or rather a part of measure theory that is not universally exposed in courses on measure theory. If $\mu \in M(\partial\,\mathbb{D})$, there is a corresponding measure on $[0, 2\pi]$, which will also be denoted by μ, such that $\int f d\mu = \int f(e^{it})d\mu(t)$ for every f in $C(\partial\,\mathbb{D})$. The corresponding measure on $[0, 2\pi]$ is not unique. For example, if $\mu = \delta_1$ in $M(\partial\,\mathbb{D})$, then either δ_0 or $\delta_{2\pi}$ can be chosen as the corresponding measure on $[0, 2\pi]$. This is, however, essentially the only way in which uniqueness fails. (What does this mean?)

For a measure μ on $[0, 2\pi]$ there is a function of bounded variation u on $[0, 2\pi]$ such that $\int f\,d\mu = \int f(t)\,du(t)$ for every continuous function f, where this second integral is a Lebesgue-Stieltjes integral. It might be worthwhile to recall how this correspondence is established, though no proofs will be given here. The proofs can be found in many of the treatments of integration theory.

If μ is a positive measure on $[0, 2\pi]$, define a function $u : [0, 2\pi] \to \mathbb{R}$ by letting $u(0) = 0$ and $u(t) = \mu([0, t))$ for $t > 0$. The function u is left continuous, increasing, and $\int f\,d\mu = \int f(t)\,du(t)$ for all continuous functions f on $[0, 2\pi]$. If μ is an arbitrary complex-valued Borel measure on $[0, 2\pi]$, let $\mu = \mu_1 - \mu_2 + i(\mu_3 - \mu_4)$ be the Jordan decomposition and let $u = u_1 - u_2 + i(u_3 - u_4)$, where u_i is the increasing function corresponding to the positive measure μ_i. This establishes a bijective correspondence between complex-valued measures μ on $[0, 2\pi]$ and left continuous functions u of bounded variation that are normalized by requiring that $u(0) = 0$.

The next proposition gives the basic properties of this correspondence between measures and functions of bounded variation.

2.1 Proposition. *Let $\mu \in M[0, 2\pi]$ and let u be the corresponding normalized function of bounded variation.*

(a) *The function u is continuous at t_0 if and only if $\mu(\{t_0\}) = 0$.*

(b) *The measure μ is absolutely continuous with respect to Lebesgue measure if and only if u is an absolutely continuous function, in which case $\int f\,d\mu = \int f(t)\,u'(t)\,dt$ for every continuous function f.*

(c) *If $E = \{t : u'(t)$ exists and is not $0\}$, then E is measurable, $\mu|E$ is absolutely continuous with respect to Lebesgue measure, and $\mu|([0, 2\pi] \setminus E)$ is singular with respect to Lebesgue measure.*

In §13.5 the concept of non-tangential limit was introduced; namely for w_0 in $\partial\,\mathbb{D}$, $z \to w_0$ (n.t.) if z approaches w_0 through a Stolz angle with vertex w_0 and opening α, $0 < \alpha < \pi/2$.

2.2 Lemma. *Given a Stolz angle with vertex $w_0 = e^{i\theta_0}$ and opening α,*

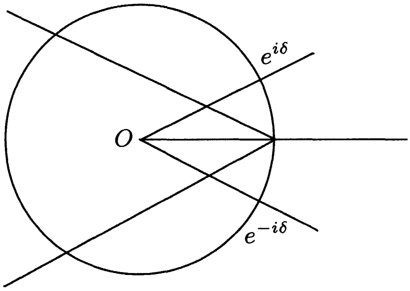

Figure 19.1.

there is a constant C and a $\delta > 0$ such that if $z = re^{i\theta}$ belongs to the Stolz angle and $|z - w_0| < \delta$, then $|\theta - \theta_0| \leq C(1 - r)$.

Proof. It suffices to assume that $\theta_0 = 0$ so that $w_0 = 1$. If L is the straight line that forms an edge of the Stolz angle, then a reference to Figure 19.1 will show that for $z = re^{i\theta}$ on L, $\sin(\alpha + \theta) = \frac{\sin\alpha}{r}$. Hence as $\theta \to 0+$

$$\frac{1-r}{\theta} = \frac{1}{\sin(\alpha+\theta)}\left[\frac{\sin(\alpha+\theta) - \sin\alpha}{\theta}\right] \to \cot\alpha.$$

Thus $\frac{\theta}{1-r} \to \tan\alpha$ as $\theta \to 0+$. Since the tangent function is increasing, the lemma now follows. $\square$

2.3 Fatou's Theorem. *Let $\mu \in M[0, 2\pi]$ and let u be the corresponding function of bounded variation; extend u to be defined on $\mathbb{R}$ by making u periodic with period 2π. If u is differentiable at θ_0, then $\tilde{\mu}(z) \to 2\pi\, u'(\theta_0)$ as $z \to e^{i\theta_0}$ (n.t.).*

(Note: We are identifying $M(\partial\, \mathbb{D})$ and $M[0, 2\pi]$. Also the only reason for extending u to be defined on $\mathbb{R}$ is to facilitate the discussion at 0 and 2π.)

Proof. It suffices to only consider the case where $\theta_0 = 0$, so we are assuming that $u'(0)$ exists. We may also assume that $u'(0) = 0$. In fact, if $u'(0) \neq 0$, let $\nu = \mu - 2\pi\, u'(0)m$. The function of bounded variation corresponding to ν is $v(\theta) = u(\theta) - u'(0)\theta$, since m is normalized Lebesgue measure. So $v'(0)$ exists and $v'(0) = 0$. If we know that $\tilde{\nu}(z) \to 0$ as $z \to 1$ (n.t.), then $\tilde{\mu}(z) = \tilde{\nu}(z) + u'(0) \to u'(0)$ as $z \to 1$ (n.t.).

So assume that $u'(0) = 0$. We want to show that

2.4 $$\int_{-\pi}^{\pi} \frac{1-r^2}{1 - 2r\cos(t-\theta) + r^2}\, d\mu(t) \to 0$$

as $z = re^{i\theta} \to 1$ (n.t.). Using the preceding lemma, it suffices to show that (2.4) holds if, for some fixed positive constant C, $\theta \to 0$ and $r \to 1$ while

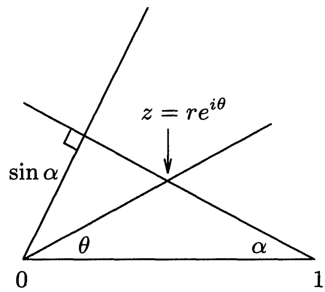

Figure 19.2.

satisfying

2.5 $$|\theta| \leq C(1-r).$$

Let Γ be the set of $z = re^{i\theta}$ satisfying (2.5).

Recall that if the Poisson kernel is considered as a function of θ with r fixed, then $P_r(\theta) = (1-r^2)/(1-2r\cos\theta + r^2)$ and so differentiation with respect to θ gives that

$$P_r'(\theta) = \frac{2r(1-r^2)\sin\theta}{(1-2r\cos\theta+r^2)^2}.$$

Fix $\varepsilon > 0$. Since $u'(0)$ exists and equals 0, there is a $\delta > 0$ such that $|u(t)| \leq \varepsilon|t|$ for $|t| < \delta$. Thus if $z = re^{i\theta} \in \Gamma$,

$$\tilde{\mu}(z) = \int_{-\pi}^{\pi} P_r(t-\theta)\,d\mu(t) = \left[\int_{-\delta}^{\delta} + \int_{\pi \geq |t| \geq \delta}\right] P_r(t-\theta)\,d\mu(t).$$

Examining Figure 19.2 we see that if $0 < \delta_1 < \delta$, there is a neighborhood U_1 of 1 such that if $z \in \Gamma \cap U_1$ and $|t| \geq \delta$, then $|t-\theta| \geq \delta_1$. Thus Proposition 10.2.3.d implies δ_1 and U_1 can be chosen so that $P_r(t-\theta) < \varepsilon$ for $|t| \geq \delta$ and $z \in \Gamma \cap U_1$. Therefore

2.6 $$|\tilde{\mu}(z)| \leq \varepsilon\|\mu\| + \left|\int_{-\delta}^{\delta} P_r(t-\theta)\,d\mu(t)\right|.$$

Using integration by parts, for $z = re^{i\theta}$ in $\Gamma \cap U_1$,

$$\begin{aligned}
\left|\int_{-\delta}^{\delta} P_r(t-\theta)\,d\mu(t)\right| &\leq |u(t)\,P_r(t-\theta)|_{-\delta}^{\delta} + \left|\int_{-\delta}^{\delta} u(t)\,P_r'(t-\theta)\,dt\right| \\
&= |u(\delta)P_r(\delta-\theta) - u(-\delta)P_r(-\delta-\theta)| \\
&\quad + \left|\int_{-\delta}^{\delta} u(t)\,P_r'(t-\theta)\,dt\right|
\end{aligned}$$

$$\leq \quad 2\delta\varepsilon^2 + \left| \int_{-\delta}^{\delta} u(t)\, P_r'(t-\theta)\, dt \right| .$$

From (2.6) we infer that

2.7 $$|\tilde{\mu}(z)| \leq \varepsilon ||\mu|| + 2\delta\varepsilon^2 + \left| \int_{-\delta}^{\delta} u(t)\, P_r'(t-\theta)\, dt \right| .$$

Now fix z and assume that $\theta \geq 0$. The case where $\theta \leq 0$ is treated similarly and will be left to the reader. Also assume that U_1 is sufficiently small that $\theta < \delta/2$ for $z = re^{i\theta}$ in U_1. Hence

$$\begin{aligned} \int_{-\delta}^{\delta} u(t)\, P_r'(t-\theta)\, dt &= \left(\int_{-\delta}^{0} + \int_{0}^{2\theta} + \int_{2\theta}^{\delta} \right) P_r'(t-\theta) u(t)\, dt \\ &= X + Y + Z. \end{aligned}$$

2.8

Now since $|u(t)| \leq \varepsilon |t|$ for $|t| \leq \delta$,

$$\begin{aligned} |Y| &= \left| \int_0^{2\theta} \frac{2r(1-r^2)\sin(t-\theta)u(t)}{(1-2r\cos(t-\theta)+r^2)^2}\, dt \right| \\ &\leq 2r(1-r^2) \int_0^{2\theta} \frac{|\sin(t-\theta)|\varepsilon t}{(1-2r\cos(t-\theta)+r^2)^2}\, dt. \end{aligned}$$

But $(1-2r\cos(t-\theta)+r^2) \geq 1-2r+r^2 = (1-r)^2$ and $|\sin(t-\theta)| \leq |t-\theta| \leq \theta$ for $0 \leq t \leq 2\theta$. Hence

$$\begin{aligned} |Y| &\leq \frac{2\varepsilon r(1-r^2)}{(1-r)^4}\, \theta \int_0^{2\theta} t\, dt \\ &= \frac{\varepsilon r(1+r)\,\theta(4\theta^2)}{(1-r)^3} \\ &\leq \frac{8\varepsilon\theta^3}{(1-r)^3}. \end{aligned}$$

By (2.5) we get

$$|Y| \leq 8\varepsilon C^3.$$

Now for the term Z in (2.8). If $2\theta \leq t \leq \delta$, then $0 \leq t - 2\theta = 2(t-\theta) - t$ and so $t \leq 2(t-\theta)$. Hence $|u(t)| \leq 2\varepsilon(t-\theta)$. Thus

$$\begin{aligned} |Z| &= \left| \int_{2\theta}^{\delta} P_r'(t-\theta)\, u(t)\, dt \right| \\ &\leq 2\varepsilon \int_{2\theta}^{\delta} (-P_r'(t-\theta))\,(t-\theta)\, dt \end{aligned}$$

$$
\begin{aligned}
&= 2\varepsilon \int_{\theta}^{\delta-\theta} (-P_r'(t))\, t\, dt \\
&\leq 2\varepsilon \int_0^{\pi} (-P_r'(t))\, t\, dt \\
&= 2\varepsilon(-tP_r(t))|_0^{\pi} + 2\varepsilon \int_0^{\pi} P_r(t)\, dt \\
&\leq 2\varepsilon\pi \left(\frac{1-r^2}{(1+r)^2} \right) + 2\varepsilon\pi \\
&< 4\pi\varepsilon,
\end{aligned}
$$

provided U_1 is chosen sufficiently small. (That is, we force r to be very close to 1.)

For the term X in (2.8), observe that

$$\int_{-\delta}^{0} P_r'(t-\theta)\, u(t)\, dt = \int_{\theta}^{\theta+\delta} P_r'(t)\, u(\theta - t)\, dt.$$

Now for $\theta \leq t \leq \theta + \delta$, $0 \leq t - \theta \leq \delta$ and so

$$|X| \leq \varepsilon \int_{\theta}^{\theta+\delta} (-P_r'(t))\, (t-\theta)\, dt.$$

Using the preceding methods we obtain the fact that for some constant M, $|X| \leq M\varepsilon$.

Referring to (2.7) and (2.8), we get that there is a constant C' that is independent of ε such that for all z in Γ and in a sufficiently small neighborhood U_1 of 1, $|\tilde{\mu}(z)| \leq C'\varepsilon$. □

2.9 Corollary. *If $\mu \in M(\partial\,\mathbb{D})$, then $\tilde{\mu}$ has non-tangential limits a.e. $[m]$ on $\partial\,\mathbb{D}$.*

Proof. Functions of bounded variation have finite derivatives a.e. □

The reader might wonder if it could be concluded that $\tilde{\mu}$ can have a limit at points of the circle. In other words, is it really necessary to impose the restriction in the preceding corollary and its ancestors that the limits be non-tangential? The answer is emphatically no; only the non-tangential limits are guaranteed. This is sketched in Exercise 1.

2.10 Corollary. *If u is a non-negative harmonic function on $\mathbb{D}$, then $\lim_{r\to 1-} u(re^{i\theta})$ exists and is finite a.e. on $[0, 2\pi]$.*

Proof. According to Herglotz's Theorem, $u = \tilde{\mu}$ for some positive measure. □

2.11 Corollary. *If μ is a measure on $\partial\,\mathbb{D}$ that is singular with respect to Lebesgue measure, then the non-tangential limits of $\tilde{\mu}$ are 0 a.e. on $\partial\,\mathbb{D}$.*

Why doesn't this contradict the Maximum Principle for harmonic functions?

The next result is also a corollary of Fatou's Theorem but it is sufficiently important to merit a more proclamatory label.

2.12 Theorem. *If $1 \le p \le \infty$ and $u : \mathbb{D} \to \mathbb{C}$ is a harmonic function such that* $\sup_{r<1} ||u_r||_p < \infty$, *then*

$$f(w) \equiv \lim_{r \to 1-} u(rw)$$

exists and is finite a.e. $[m]$ on $\partial\,\mathbb{D}$. If $1 < p \le \infty$, then $f \in L^p(m)$ and $u = \tilde{f}$. If $p = 1$, then $u = \tilde{\mu}$ for some measure μ in $M[0, 2\pi]$ and f is the Radon-Nikodym derivative of the absolutely continuous part of μ.

Proof. This proof is actually a collage of several preceding results. First assume that $1 < p \le \infty$. By Theorem 1.7 there is a function g in L^p such that $u = \tilde{g}$. By Fatou's Theorem, $g = f$ a.e. $[m]$. Now suppose $p = 1$. Again Theorem 3.8 implies that $u = \tilde{\mu}$ for some μ in $M[0, 2\pi]$. Let $\mu = \mu_a + \mu_s$ be the Lebesgue decomposition of μ with respect to m. Let g be the Radon-Nikodym derivative of μ_a with respect to m. Thus if w is the function of bounded variation on $[0, 2\pi]$ corresponding to μ, then $w' = g$ a.e. It follows by Fatou's Theorem that $g = f$ a.e. $\square$

Note that the preceding theorem contains Theorem 13.5.2 as a special case.

2.13 Example. If $\mu = \delta_1$, the unit point mass at 1 on $\partial\,\mathbb{D}$, $\tilde{\mu}(z) = \int P_z d\mu = P_z(1) = \text{Re}\left(\frac{1+z}{1-z}\right)$. Here the conclusion of Fatou's Theorem can be directly verified.

Exercises

1. Let $f \in L^1$ and put $g(z) = \tilde{f}(z)$ for $|z| < 1$ and $g(e^{i\theta}) = \lim_{r \to 1-} \tilde{f}(re^{i\theta}$ when this limit exists; so g and f agree a.e. on $\partial\,\mathbb{D}$. Let E be the set of points on $\partial\,\mathbb{D}$ where g is defined. (a) Show that if $\tilde{f}(z) \to g(e^{i\theta})$ as $z \to e^{i\theta}$ with z in $\mathbb{D}$ (tangential approach allowed), then g, as a function defined on $\mathbb{D} \cup E$, is continuous at $e^{i\theta}$. Let E_c be the set of points in $\partial\,\mathbb{D}$ where g is continuous. (b) Show that E_c has measure zero if and only if f is not equivalent to any function whose points of continuity have positive measure. (Two functions are equivalent if they agree on a set of full measure and thus define the same element of L^1.) The rest of this exercise produces a function f in L^1 that is not equivalent to any function whose points of continuity have positive measure. (Note that the characteristic function of the irrational numbers is equivalent to the constantly 1 function.) Once this is done, the

harmonic function $\tilde{f}$ will fail to have a limit a.e. on $\partial\,\mathbb{D}$ even though it has a non-tangential limit a.e. (c) Let K be a Cantor subset of $\partial\,\mathbb{D}$ with positive measure and show that the set of points of continuity of χ_K is $\partial\,\mathbb{D} \setminus K$. (d) Construct a sequence $\{K_n\}$ of Cantor sets in $\partial\,\mathbb{D}$ that are pairwise disjoint and such that $\partial\,\mathbb{D} \setminus \cup_n K_n$ has zero measure. Show that $\cup_n K_n$ contains no interval. (e) Define f on $\partial\,\mathbb{D}$ by letting $f(z) = 1/2^n$ for z in K_n and $f(z) = 0$ for z in $\partial\,\mathbb{D} \setminus \cup_n K_n$. Show that if g is any function equivalent to f, then the set of points of continuity of g has measure zero.

2. Give an example of an analytic function on $\mathbb{D}$ that fails to have a non-tangential limit at almost every point of $\partial\,\mathbb{D}$.

3. Suppose $f \in L^1$ and f is real-valued. Show that if $a \in \partial\,\mathbb{D}$ and $\lim_{z\to a} f(z) = +\infty$, then $\lim_{r\to 1} \tilde{f}(ra) = +\infty$.

§3 Semicontinuous Functions

In this section we will prove some basic facts about semicontinuous functions (lower and upper). Most readers will have learned at least some of this material, but we will see here a rather complete development as it seems to be a topic that most modern topology books judge too specialized for inclusion and most analysis books take for granted as known by the reader. We will, of course, assume that the reader has mathematical maturity and omit many details from the proofs.

3.1 Definition. If X is a metric space and $u : X \to [-\infty, +\infty)$, then u is *upper semicontinuous* (usc) if, for every c in $[-\infty, +\infty)$, the set $\{x \in X : u(x) < c\}$ is an open subset of X. Similarly, $u : X \to (-\infty, +\infty]$ is *lower semicontinuous* (lsc) if, for every c in $(-\infty, +\infty]$, the set $\{x \in X : u(x) > c\}$ is open.

Note that the constantly $-\infty$ and $+\infty$ functions are upper and lower semicontinuous, respectively. This is not standard in the literature. Also a function u is upper semicontinuous if and only if $-u$ is lower semicontinuous. In the sequel, results will be stated and proved for upper semicontinuous functions. The correct statements and proofs for lower semicontinuous functions are left to the reader. Throughout the section (X, d) will be a metric space.

The reason for using the words "upper" and "lower" here comes from considerations on the real line. If $X = \mathbb{R}$ and u is a continuous function except for jump discontinuities, u will be upper semicontinuous if and only if at each discontinuity x_0, $u(x_0)$ is the upper value.

3.2 Proposition. *If X is a metric space and $u : X \to [-\infty, \infty)$, then the* *following statements are equivalent.*

(a) u *is usc.*

(b) *For every* c *in* $[-\infty, \infty)$ *the set* $\{x \in X : u(x) \geq c\}$ *is closed.*

(c) *If* $x_0 \in X$ *and* $u(x_0) > -\infty$, *then for every* $\varepsilon > 0$ *there is a* $\delta > 0$ *such that* $u(x) < u(x_0) + \varepsilon$ *whenever* $d(x, x_0) < \delta$; *if* $u(x_0) = -\infty$ *and* $M < 0$, *then there is a* $\delta > 0$ *such that* $u(x) < M$ *for* $d(x, x_0) < \delta$.

(d) *If* $x_0 \in X$, *then* $u(x_0) \geq \limsup_{x \to x_0} u(x)$.

3.3 Proposition. *If* K *is a compact subset of* X, u *is an usc function on* X, *and* $u(x) < \infty$ *for all* x *in* K, *then there is an* x_0 *in* K *such that* $u(x_0) \geq u(x)$ *for all* x *in* K.

Proof. First let $U_n = \{x \in X : u(x) < n\}$. Then $K \subseteq \cup_n U_n$ and each U_n is open. By the compactness of K, there is an n such that $K \subseteq U_n$. Thus $\alpha \equiv \sup\{u(x) : x \in K\} < \infty$. Now put $K_n = \{x \in K : u(x) \geq \alpha - n^{-1}\}$. Each K_n is a compact and non-empty subset of K and $K_{n+1} \subseteq K_n$. So there is an x_0 that belongs to each K_n and it must be that $u(x_0) \geq u(x)$ for all x in K. □

3.4 Proposition.

(a) *If* u_1 *and* u_2 *are usc functions, then* $u_1 + u_2$ *and* $u_1 \vee u_2 \equiv \max\{u_1, u_2\}$ *are usc functions.*

(b) *If* $\{u_i\}$ *is a collection of usc functions, then* $\wedge u_i \equiv \inf_i u_i$ *is usc.*

Proof. (a) Let $u = u_1 + u_2$, fix c in $[-\infty, \infty)$, and let $U = \{x : u(x) < c\}$. If $x_0 \in U$, then $u_1(x_0) < c$ and $u_2(x_0) < c - u_1(x_0)$. Hence $U_1 = \{x : u_1(x) < c\}$ and $U_2 = \{x : u_2(x) < c - u_1(x_0)\}$ are both open neighborhoods of x_0 and $U_1 \cap U_2 \subseteq U$. Since x_0 was arbitrary, U must be open.

Now if $u = u_1 \vee u_2$ and $c \in [-\infty, \infty)$, then $\{x : u(x) < c\} = \{x : u_1(x) < c\} \cap \{x : u_2(x) < c\}$, and so u is usc.

(b) If $u = \wedge u_i$ and $c \in [-\infty, \infty)$, then $\{x : u(x) < c\} = \cup\{x : u_i(x) < c\}$. □

3.5 Corollary. *If* $\{u_n\}$ *is a sequence of usc functions on* X *such that for every* x, $\{u_n(x)\}$ *is decreasing, then* $u(x) \equiv \lim u_n(x)$ *is usc.*

A sequence of functions satisfying the hypothesis of this corollary is called a *decreasing sequence of functions.*

3.6 Theorem. *If* $u : X \to [-\infty, \infty)$ *is usc and* $u \leq M < +\infty$ *on* X, *then there is a decreasing sequence of uniformly continuous functions* $\{f_n\}$ *on* X *such that* $f_n \leq M$ *and for every* x *in* X, $f_n(x) \downarrow u(x)$.

Proof. If u is constantly equal to $-\infty$, then the result is trivial. So assume that this is not the case. If d is the metric on X, define $f_n : X \to \mathbb{R}$ by

$$f_n(x) = \sup\{u(y) - nd(x,y) : y \in X\}.$$

For any x and y in X, $f_n(x) \geq u(y) - nd(x,y) \geq u(y) - (n+1)d(x,y)$; so $f_n(x) \geq f_{n+1}(x) \geq u(x)$ for all x in X. Also $f_n \leq M$ for all n.

Let ε be a positive number, let $x \in X$, and fix $n \geq 1$. By definition, there is a y in X such that $f_n(x) < u(y) - nd(x,y) + \varepsilon$. If $d(z,x) < \varepsilon$, then $d(z,y) < d(x,y) + \varepsilon$. Thus $f_n(z) \geq u(y) - nd(z,y) > u(y) - nd(x,y) - n\varepsilon > f_n(x) - (n+1)\varepsilon$ whenever $d(z,x) < \varepsilon$. That is, $f_n(z) > f_n(x) - (n+1)\varepsilon$ whenever $d(z,x) < \varepsilon$. Now interchange the roles of x and z in the preceding argument to get that $f_n(x) > f_n(z) - (n+1)\varepsilon$ when $d(x,z) < \varepsilon$. Therefore $|f_n(x) - f_n(z)| < (n+1)\varepsilon$ when $d(x,z) < \varepsilon$. That is, f_n is uniformly continuous.

It remains to show that $f_n(x) \to u(x)$ for all x in X. So fix x_0 in X and let ε be a positive number. Assume that $u(x_0) > -\infty$. (The case in which $u(x_0) = -\infty$ is left as an exercise.) Since u is usc, there is a $\delta > 0$ such that $u(y) < u(x_0) + \varepsilon$ for $d(y,x_0) < \delta$. Thus $u(y) - nd(y,x_0) < u(x_0) + \varepsilon$ whenever $d(y,x_0) < \delta$. Now suppose that $d(y,x_0) \geq \delta$; here $u(y) - nd(y,x_0) \leq u(y) - n\delta \leq M - n\delta$. Choose n_0 such that $M - n\delta < u(x_0) + \varepsilon$ for $n \geq n_0$. Thus for $n \geq n_0$ and for all y in X, $u(y) - nd(y,x_0) \leq u(x_0) + \varepsilon$. Therefore $u(x_0) \leq f_n(x_0) \leq u(x_0) + \varepsilon$ for $n \geq n_0$. $\square$

Exercises

1. Give an example of a family $\mathcal{U}$ of usc functions such that $\sup \mathcal{U}$ is not usc.

2. If u is a monotone function on an interval (a,b) in $\mathbb{R}$, show that u is usc if and only if for each discontinuity x of u, $u(x) = \lim_{t \to x+} u(t)$.

3. Show that the uniform limit of a sequence of usc functions is usc.

4. If X is a metric space, E is a subset of X, and u is the characteristic function of the set E, show that u is upper semicontinuous if and only if E is closed.

5. Let X be a metric space. Show that u is an upper semicontinuous function on X if and only if $A = \{(x,t) \in X \times \mathbb{R} : t \leq u(x)\}$ is a closed subset of $X \times \mathbb{R}$.

6. Suppose A is any closed subset of $X \times \mathbb{R}$ such that for each x in X the set $A_x \equiv \{t \in \mathbb{R} : (x,t) \in A\}$ is either empty or bounded above. Define $u : X \to [-\infty, \infty)$ by $u(x) = -\infty$ if $A_x = \emptyset$ and $u(x) = \sup A_x$ otherwise, and show that u is upper semicontinuous. Show that $A = \{(x,t) \in X \times \mathbb{R} : t \leq u(x)\}$.

7. Let $f : X \to [-\infty, \infty)$ be any function and let $\Gamma = \{(x, f(x)) : x \in X \text{ and } f(x) > -\infty\} \subseteq X \times \mathbb{R}$. If $A = \text{cl}\, \Gamma$, show that if $u(x)$ is defined as in the preceding exercise, then $u(x) = \limsup_{y \to x} f(y)$ for all x in X.

8. Suppose G is an open subset of a metric space X and $f : G \to [-\infty, \infty)$ is any function such that for each ζ in ∂G, $u(\zeta) \equiv \limsup\{f(x$ $x \in G$ and $x \to \zeta\} < \infty$. Show that u is upper semicontinuous.

§4 Subharmonic Functions

Subharmonic and superharmonic functions were already defined in 10.3.1, but it was assumed there that these functions were continuous. This was done to avoid assuming that the reader's background included anything other than basic analysis. In particular, it was assumed that the reader did not know the Lebesgue integral and thus could not discuss the integral of a semicontinuous function. It is desirable to go beyond this and extend the definition to semicontinuous functions. Propositions for semicontinuous subharmonic functions that were stated for the continuous version in Chapter 10 will sometimes be restated here. If the proof given in Chapter 10 extends naturally to the present situation, it will not be repeated and the reader will be referred to the appropriate result from the first volume of this work.

4.1 Definition. If G is an open subset of $\mathbb{C}$, a function $u : G \to [-\infty, \infty)$ is *subharmonic* if u is upper semicontinuous and, for every closed disk $\overline{B}(a; r)$ contained in G, we have the inequality

4.2
$$u(a) \le \frac{1}{2\pi} \int_0^{2\pi} u(a + re^{i\theta})\, d\theta.$$

A function $u : G \to \mathbb{R} \cup \{+\infty\}$ is *superharmonic* if $-u$ is subharmonic.

Some remarks are in order here. Since u is upper semicontinuous, the fact that $u(a + re^{i\theta}) < \infty$ for all θ implies u is uniformly bounded above on this circle. Thus it is not being assumed that the integral in the definition is finite, but the integral is defined with the possibility that it is $-\infty$. In fact it may be that the function is constantly equal to $-\infty$ on some or all of the components of G. We will see below (Proposition 4.11) that this is the only way that a subharmonic function can fail to be integrable on such circles. There is a slight difference between this definition of a subharmonic function and that given by many authors in that the function that is identically equal to $-\infty$ is allowed to be subharmonic. In fact, since G is not assumed to be connected, u may be constantly equal to $-\infty$ on some components and finite valued on others. This is also usually excluded as a possibility in the definition of a subharmonic function.

Results will usually be stated and proved only for subharmonic functions. The statements and proofs for superharmonic functions will be left to the reader.

4.3 Example. If $f : G \to \mathbb{C}$ is an analytic function, then $\log|f|$ is a subharmonic function on G. In fact this is an immediate consequence of Jensen's Formula (11.1.2).

For a compact subset K of $\mathbb{C}$, let $C_h(K)$ denote the continuous functions on K that are harmonic on int K.

4.4 Definition. Say that a function $u : G \to [-\infty, \infty)$ *satisfies the Maximum Principle* if, for every compact set K contained in G and every h in $C_h(K)$, $u \le h$ on K whenever $u \le h$ on ∂K.

4.5 Theorem. *If $u : G \to [-\infty, \infty)$ is an upper semicontinuous function, then the following statements are equivalent.*

(a) *u is subharmonic.*

(b) *u satisfies the Maximum Principle.*

(c) *If D is a closed disk contained in G and $h \in C_h(D)$ with $u \le h$ on ∂D, then $u \le h$ on D.*

(d) *If $D = \overline{B}(a; r) \subseteq G$, then*

$$u(a) \le \frac{1}{\pi r^2} \int_D u \, dA.$$

Proof. (a) implies (b). (This part of the proof is like the proof of Theorem 10.3.3.) Let K be a compact subset of G and assume that $h \in C_h(K)$ with $u \le h$ on ∂K. By replacing u with $u - h$, it is seen that it must be shown that $u \le 0$ on K whenever u is subharmonic and satisfies $u \le 0$ on ∂K. Suppose there is a point z_0 in int K such that $u(z_0) > 0$ and let $0 < \varepsilon < u(z_0)$. Put $A = \{z \in K : u(z) \ge \varepsilon\}$. Because u is usc, A is compact. Also, if $b \in \partial K$, $0 \ge u(b)$. By standard compactness arguments, this implies there is a neighborhood V of ∂K with $V \subseteq G$ and $u(z) < \varepsilon$ for all z in V. Hence, $A \subseteq \text{int}K$. Again the fact that u is usc implies that there is a point a in A with $u(a) \ge u(z)$ for all z in A. It follows that $u(a) \ge u(z)$ for all z in K. Let H be the component of $\text{int}K$ that contains a and put $B = \{z \in H : u(z) = u(a)\}$.

Clearly B is a relatively closed non-empty subset of H. If $w \in B$, let $\overline{B}(w; r) \subseteq G$. So $\overline{B}(w; r) \subseteq H$ and, for $0 < \rho < r$, $2\pi u(w) \le \int_0^{2\pi} u(w + \rho e^{i\theta}) d\theta \le 2\pi u(w)$ since $u(w + \rho e^{i\theta}) \le u(a) = u(w)$ for all θ. That is, the integral of the non-negative function $u(a) - u(w + \rho e^{i\theta})$ is 0 and so

$\overline{B}(w,r) \subseteq B$. Thus B is open and so $B = H$. But this implies that if $b \in \partial H \subseteq \partial K$, then $0 \geq u(b) \geq \limsup\{u(z) : z \in H, z \to b\} = u(a) > 0$, a contradiction.

(b) *implies* (c). Clear.

(c) *implies* (d). Suppose $\overline{B}(a;r) \subseteq G$ and let $\{f_n\}$ be a sequence of continuous functions on $\partial B(a;r)$ such that $f_n(z) \downarrow u(z)$ (Theorem 3.6). Let $D = \overline{B}(a;r)$ and let $h_n \in C_h(D)$ such that $h_n = f_n$ on ∂D. By (c), $u \leq h_n$ on D. By Exercise 10.1.6,

$$\begin{aligned} u(a) &\leq h_n(a) \\ &= \frac{1}{\pi r^2}\int_D f_n\, d\mathcal{A}. \end{aligned}$$

But $\int_D f_n d\mathcal{A} \to \int u\, d\mathcal{A}$ by monotone convergence.

(d) *implies* (b). This is like the proof that (a) implies (b) and is left to the reader.

(b) *implies* (a). This is like the proof that (c) implies (d) and is left to the reader. □

An examination of the appropriate results in Chapter 10 shows that the newly defined versions of subharmonic and superharmonic functions also satisfy the various versions of the Maximum Principle given there. These will not be stated explicitly.

4.6 Proposition.

(a) *If u_1 and u_2 are subharmonic functions on G, then $u_1 + u_2$ and $u_1 \vee u_2$ are subharmonic.*

(b) *Let $\mathcal{U}$ be a family of subharmonic functions on G that is locally bounded above and let $v \equiv \sup\mathcal{U}$. If $u(z) \equiv \limsup_{w\to z} v(w)$, then u is subharmonic. If v is upper semicontinuous, then $v = u$.*

(c) *If $\mathcal{U}$ is a family of subharmonic functions such that for all u_1 and u_2 in $\mathcal{U}$ there is a u_3 in $\mathcal{U}$ with $u_3 \leq u_1 \wedge u_2$, then $\inf\mathcal{U}$ is subharmonic. In particular, if $\{u_n\}$ is a sequence of subharmonic functions such that $u_n \geq u_{n+1}$ for all n, then $\lim u_n$ is subharmonic.*

(d) *If $\{u_n\}$ is a sequence of positive subharmonic functions such that $u = \sum_n u_n$ is upper semicontinuous, then on each component of G either u is subharmonic or $u \equiv \infty$.*

(e) *If $\{u_n\}$ is a sequence of negative subharmonic functions, then $u = \sum_n u_n$ is subharmonic.*

Proof. (a) It is clear that $u_1 + u_2$ is a subharmonic function. Since $u_1 \vee u_2$ is upper semicontinuous, the proof that this function is subharmonic is given in the proof of part (b).

(b) First note that $u(z) < \infty$ for all z in G because $\mathcal{U}$ is locally bounded above. It follows from Exercise 3.8 that u is upper semicontinuous. To show that u is subharmonic, we will show that u satisfies the Maximum Principle. So let K be a compact subset of G and let $h \in C_h(K)$ such that $u \le h$ on ∂K. Fix u' in $\mathcal{U}$; now $u \ge v \ge u'$ on G. Hence $h \ge u'$ on ∂K and so $h \ge u'$ on K since u' is subharmonic. Since u' was arbitrary, $h \ge v$ on K, whence we have that $h \ge u$. By Theorem 4.5, u is subharmonic.

If v is upper semicontinuous, then the same proof shows that v is subharmonic. The fact that $v = u$ is a consequence of Proposition 4.8 below, which is worth separating out.

(c) By Proposition 3.4, $u = \inf \mathcal{U}$ is usc and clearly $u(z) < \infty$ for all z. Suppose $D = \overline{B}(a;r) \subseteq G$. By Theorem 3.6 there is a sequence of continuous functions $\{f_n\}$ on ∂D such that $f_n(z) \downarrow u(z)$ for z in ∂D. It may be assumed that $u(z) < f_n(z)$ for all n and all z in ∂D. Fix n for the moment. So for every z_1 in ∂D there is a u_1 in $\mathcal{U}$ such that $u_1(z_1) < f_n(z_1)$. But $\{z : u_1(z) - f_n(z) < 0\}$ is an open neighborhood of z_1. A compactness argument shows that there are functions $u_1, \ldots, u_m$ in $\mathcal{U}$ such that for every z in ∂D there is a u_k, $1 \le k \le m$, with $u_k(z) < f_n(z)$. By hypothesis, there is one function v in $\mathcal{U}$ with $v(z) < f_n(z)$ for all z on ∂D. Thus $2\pi\, u(a) \le 2\pi\, v(a) \le \int_0^{2\pi} v(a + re^{i\theta})d\theta \le \int_0^{2\pi} f_n(a + re^{i\theta})\, d\theta$. By monotone convergence of the integrals, it follows that u is subharmonic.

(d) This proof is like the proof of part (b).

(e) This is immediate from part (c). $\square$

4.7 Example. If $f : G \to \mathbb{C}$ is an analytic function, then $\log^+ |f|$ is a subharmonic function on G. In fact, $\log^+ |f| = 0 \vee \log |f|$.

Following are two results that will prove useful as we progress.

4.8 Proposition. *If G is any open set and u is a subharmonic function on G, then for any a in G, $u(a) = \limsup_{z \to a} u(z)$.*

Proof. Since u is usc, $u(a) \ge \limsup_{z \to a} u(z)$. On the other hand, if $r > 0$ such that $B_r = B(a;r) \subseteq G$,

$$\begin{aligned} u(a) &\le \frac{1}{\pi r^2} \int_{B_r} u(z)\, d\mathcal{A}(z) \\ &\le \sup\{u(z) : z \in B_r, z \ne a\}. \end{aligned}$$

By definition this says that $u(a) \le \limsup_{z \to a} u(z)$. $\square$

4.9 Proposition. *If u is a subharmonic function on G and $\overline{B}(a;r) \subseteq G$, then $(2\pi)^{-1} \int_0^{2\pi} u(a + \rho e^{i\theta})\, d\theta \downarrow u(a)$ as $\rho \downarrow 0$. Similarly, $(\pi \rho^2)^{-1} \int_{B(a;\rho)} u\, d\mathcal{A} \downarrow u(a)$ as $\rho \downarrow 0$.*

Proof. Let $I_\rho = (2\pi)^{-1} \int_0^{2\pi} u(a + \rho e^{i\theta})\, d\theta$. If $\sigma < \rho < r$, let $\{f_n\}$ be a sequence of continuous functions on $\partial B(a;\rho)$ such that $f_n \downarrow u$. Let $h_n \in$

$C_h(\overline{B}(a;\rho))$ such that $h_n = f_n$ on $\partial B(a;\rho)$. Thus $u \le h_n$ on $\overline{B}(a;\rho)$ and so, by monotone convergence,

$$\begin{aligned} I_\sigma &\le \frac{1}{2\pi}\int_0^{2\pi} h_n(a+\sigma e^{i\theta})\,d\theta \\ &= h_n(a) \\ &= \frac{1}{2\pi}\int_0^{2\pi} h_n(a+\rho e^{i\theta})\,d\theta \\ &\downarrow I_\rho \end{aligned}$$

as $n \to \infty$. Therefore I_ρ decreases as ρ decreases. But $u(a) \le I_\rho$ for all ρ and so, by Fatou's Lemma, $u(a) \le \lim_{\rho\to 0} I_\rho \le \frac{1}{2\pi}\int_0^{2\pi} \limsup_{\rho\to 0} u(a + \rho e^{i\theta})\,d\theta \le u(a)$. $\square$

After the next lemma, it will be shown that a subharmonic function on a region G belongs to $L^1_{\text{loc}}(G)$ and $\theta \to u(a+re^{i\theta})$ is integrable if $\overline{B}(a;r) \subseteq G$.

4.10 Lemma. *If G is connected, u is a subharmonic function on G, and $\{z \in G : u(z) = -\infty\}$ has non-empty interior, then $u(z) \equiv -\infty$.*

Proof. Let $B = \{z : u(z) = -\infty\}$ and let $A = \operatorname{int} B$. So, by hypothesis, A is a non-empty open subset of G. It will be shown that A is also relatively closed in G and hence $A = G$.

Let $a \in G \cap \operatorname{cl} A$ and let r be a positive number such that $\overline{B}(a;r) \subseteq G$. If b is any element of $B(a;r/4)$, then $\overline{B}(b;r/2) \subseteq G$ and, because $a \in B(b;r/2)$, $B(b;r) \cap A \ne \emptyset$. Thus there is a ρ, $0 < \rho < r/2$, such that $\partial B(b;\rho) \cap A \ne \emptyset$. Since u is subharmonic, $u(b) \le \frac{1}{2\pi}\int_0^{2\pi} u(b+\rho e^{i\theta})\,d\theta$. But the fact that $\partial B(b;\rho)$ meets the open set A implies that $b + \rho e^{i\theta} \in A$ for some interval of θ's. Therefore $\int_0^{2\pi} u(b+\rho e^{i\theta})\,d\theta = -\infty$ and so $u(b) = -\infty$ whenever $|a-b| < r/4$. That is, $B(a;r/4) \subseteq B$ and so $a \in \operatorname{int} B = A$. $\square$

4.11 Proposition. *If G is connected and u is a subharmonic function on G that is not identically $-\infty$, then:*

(a) $u \in L^1_{\text{loc}}(G)$;

(b) $\int_0^{2\pi} u(a+re^{i\theta})\,d\theta > -\infty$ *whenever* $\overline{B}(a;r) \subseteq G$. *That is,* $\theta \to u(a+re^{i\theta})$ *is integrable with respect to Lebesgue measure on* $[0,2\pi]$ *whenever* $\overline{B}(a;r) \subseteq G$.

Proof. Let $A = \{z \in G : u(z) = -\infty\}$. By Lemma 4.10, $\operatorname{int} A = \emptyset$.

(a) Let K be a compact subset of G. Since u is usc, and consequently bounded above on K, to show that $u \in L^1(K, \mathcal{A})$ it suffices to show that $\int u\,d\mathcal{A} > -\infty$. But since $\operatorname{int} A = \emptyset$, there is a finite number of disks $B(a_k;r)$ with $\overline{B}(a_k;r) \subseteq G$, $u(a_k) > -\infty$, and $K \subseteq \cup_k B(a_k;r)$. Thus $-\infty < \pi r^2 u(a_k) \le \int_{B(a_k;r)} u\,d\mathcal{A}$. Hence $\int_K u\,d\mathcal{A} > -\infty$.

(b) Suppose that $D = \overline{B}(a;r) \subseteq G$ and put $I_\rho = \frac{1}{2\pi}\int_0^{2\pi} u(a+\rho e^{i\theta})\,d\theta$. By Proposition 4.9, I_ρ is decreasing as ρ decreases. So if there is an $\sigma < r$ such that $I_\sigma = -\infty$, then $I_\rho = -\infty$ for all $\rho \leq \sigma$. Hence $\int_D u d\mathcal{A} = \int_0^r I_\rho \rho\, d\rho = -\infty$, contradicting part (a). □

Recall the definition of a convex function (6.3.1). We extend the definition a little.

4.12 Definition. If $-\infty \leq a < b \leq +\infty$, a function $\phi : [a,b] \to [-\infty, \infty)$ is *convex* if:

(a) ϕ is continuous on $[a,b]$;

(b) $\phi(x) \in \mathbb{R}$ if $x \in (a,b)$;

(c) $\phi(\alpha x + (1-\alpha)y) \leq \alpha\phi(x) + (1-\alpha)\phi(y)$ for x, y in (a,b) and $0 \leq \alpha \leq 1$.

It is not hard to show that a twice differentiable ϕ function is convex if and only if $\phi'' \geq 0$. In particular the exponential function is convex while the logarithm is not.

4.13 Proposition. (Jensen's Inequality) *If (X, Ω, μ) is a probability measure space, $f \in L^1(\mu)$, $-\infty \leq a < f(x) < b \leq \infty$ a.e. $[\mu]$, and $\phi : [a,b] \to [-\infty, \infty)$ is a convex function, then $\phi(\int f\, d\mu) \leq \int \phi \circ f d\mu$.*

Proof. It may be assumed that f is not constant. Thus $a < I = \int f\, d\mu < b$. Put $S = \sup\{[\phi(I) - \phi(t)]/[I - t] : a < t < I\}$. By Exercise 4, $S < \infty$. So for $a < t < I$,

$$\phi(I) + S(t - I) \leq \phi(t). \tag{4.14}$$

If $I < t < b$, then Exercise 4 implies that $S \leq [\phi(t) - \phi(I)]/[t - I]$; equivalently, (4.14) holds for $I < t < b$ and hence for all t in (a,b). In particular, letting $t = f(x)$ implies that $0 \leq \phi(f(x)) - \phi(I) - S(f(x) - I)$ a.e. $[\mu]$. Since μ is a probability measure, $0 \leq \int \phi \circ f\, d\mu - \phi(I) - S\left[\int f\, d\mu - I\right] = \int \phi \circ f d\mu - \phi(I)$. □

The reader should be warned that in the literature there is more than one inequality that is called "Jensen's Inequality."

4.15 Theorem. *If u is a subharmonic function on G with $-\infty \leq a \leq u(z) \leq b \leq \infty$ for all z in G and $\phi : [a,b] \to [-\infty, \infty)$ is an increasing convex function, then $\phi \circ u$ is subharmonic.*

Proof. It is immediate that $\phi \circ u$ is upper semicontinuous. On the other hand, if $\overline{B}(a;r) \subseteq G$, then Proposition 4.13 and the fact that ϕ is increasing imply that

$$\phi(u(a)) \leq \frac{1}{2\pi}\int_0^{2\pi} (\phi \circ u)(a + re^{i\theta})\,d\theta.$$

Hence $\phi \circ u$ is subharmonic. □

4.16 Example. If $f : G \to \mathbb{C}$ is an analytic function, then $|f|^p$ is a subharmonic function on G for $0 < p < \infty$. From Example 4.3 we know that $\log|f|$ is subharmonic. If $\phi(t) = e^{pt}$, then ϕ is increasing and convex. By the preceding proposition, $|f(z)|^p = \phi(\log|f(z)|)$ is subharmonic.

4.17 Proposition. *Let H be an open subset of the open set G and let u be a subharmonic function on G. If v is an usc function on G that is subharmonic on H and satisfies $v \geq u$ on H and $v = u$ on $G \setminus H$, then v is subharmonic on G.*

Proof. If $b \in \partial H \cap G$ and $\overline{B}(b;\rho) \subseteq G$, then $\int_0^{2\pi} v(b+\rho e^{i\theta})\,d\theta \geq \int_0^{2\pi} u(b+\rho e^{i\theta})\,d\theta \geq 2\pi\, u(b) = 2\pi\, v(b)$. The details of the rest of the proof are left to the reader. □

Recall that for a function u defined on the boundary of a disk B, $\tilde{u}$ denotes the solution of the Dirichlet problem on B with boundary values u (§1). The next corollary is an extension of Corollary 10.3.7 to the present definition of a subharmonic function.

4.18 Corollary. *Let u be a subharmonic function on G and $\overline{B}(a;r) \subseteq G$. If $v : G \to \mathbb{R}$ is defined by letting $v = u$ on $G \setminus B(a;r)$ and $v = \tilde{u}$ on $B(a;r)$, then v is subharmonic.*

Proof. Let $B = B(a;r)$. First note that Proposition 4.11 implies that u is integrable on ∂B and so $\tilde{u}$ is well defined on B. To apply the preceding proposition it must be shown that v is upper semicontinuous and $v \geq u$. To show that $v \geq u$, first use Theorem 3.6 to get a decreasing sequence of continuous functions $\{\phi_n\}$ on ∂B that converges to $u|\partial B$ pointwise. Thus $\tilde{\phi}_n \geq u$ on B by the Maximum Principle. But $\tilde{\phi}_n(z) \to \tilde{u}(z)$ for each z in B by the Monotone Convergence Theorem for integrals. Hence $\tilde{u} \geq u$ on B. Also for z in cl B, $v(z) = \lim_n \tilde{\phi}_n(z)$. Since $\tilde{\phi}_n$ is continuous on cl B, v is usc on cl B (3.5). It follows that v is usc on G and thus is subharmonic. □

With the notation of the preceding corollary, the subharmonic function v is called the *harmonic modification of* u on $B(a;r)$.

In the next result we will use the fact that the mollifier $\{\phi_\varepsilon\}$ has the property that $\phi_\varepsilon(z) = \phi_\varepsilon(|z|)$.

4.19 Proposition. *Let u be a subharmonic function on the open set G and for a mollifier $\{\phi_\varepsilon\}$ let $u_\varepsilon = u * \phi_\varepsilon$. Then:*

(a) *u_ε is a C^∞ function on $\mathbb{C}$;*

(b) *if K is a compact subset of G, then $\int_K |u - u_\varepsilon|\,d\mathcal{A} \to 0$;*

(c) *u_ε is subharmonic on $\{z \in G : d(z,\partial G) > \varepsilon\}$;*

(d) *for each z in G, $u_\varepsilon(z) \downarrow u(z)$ as $\varepsilon \downarrow 0$.*

Proof. Parts (a) and (b) follow from Proposition 3.6.

(c) If $z \in G_\varepsilon = \{w \in G : d(w, \partial G) > \varepsilon\}$, then a change of variables in the definition of u_ε gives that

4.20 $$u_\varepsilon(z) = \int_{\mathbb{D}} u(z - \varepsilon w)\, \phi(w)\, d\mathcal{A}(w).$$

So if $\overline{B}(a; r) \subseteq G_\varepsilon$,

$$\begin{aligned}
\int_0^{2\pi} u_\varepsilon(a + re^{i\theta})\, d\theta &= \int_0^{2\pi} \int_{\mathbb{D}} u(a + re^{i\theta} - \varepsilon w)\, \phi(w)\, d\mathcal{A}(w)\, d\theta \\
&= \int_{\mathbb{D}} \phi(w) \int_0^{2\pi} u(a + re^{i\theta} - \varepsilon w)\, d\theta\, d\mathcal{A}(w) \\
&\geq 2\pi \int_{\mathbb{D}} \phi(w)\, u(a - \varepsilon w)\, d\mathcal{A}(w) \\
&= 2\pi\, u_\varepsilon(a).
\end{aligned}$$

Therefore u_ε is subharmonic on G_ε.

(d) Let $\varepsilon < \delta$; we will show that $u_\varepsilon \leq u_\delta$ on G_δ and $u_\varepsilon(z) \to u(z)$ for all z in G. So fix δ and fix a point z in G_δ. Using (4.20) and Proposition 4.9, we have that for $r < \varepsilon$

$$\begin{aligned}
u_\varepsilon(z) &= \int_{\mathbb{D}} u(z - \varepsilon w)\, \phi(w)\, d\mathcal{A}(w) \\
&= \int_0^1 r\, \phi(r) \left\{ \int_0^{2\pi} u(z - \varepsilon re^{i\theta})\, d\theta \right\} dr \\
&\leq \int_0^1 r\, \phi(r) \left\{ \int_0^{2\pi} u(z - \delta re^{i\theta})\, d\theta \right\} dr \\
&= u_\delta(z).
\end{aligned}$$

To show convergence, again according to (4.9), $\int_0^{2\pi} u(z - \varepsilon re^{i\theta})\, d\theta \to 2\pi\, u(z)$ as $\varepsilon \to 0$. Thus using the just concluded display of equations and the Monotone Convergence Theorem, we get that as $\varepsilon \to 0$, $u_\varepsilon(z) = \int u(z - \varepsilon w)\, \phi(w)\, d\mathcal{A}(w) \to \int u(z)\, \phi(w)\, d\mathcal{A}(w) = u(z)$. □

4.21 Corollary. *If u is subharmonic on $\mathbb{C}$, there is a decreasing sequence $\{u_n\}$ of continuous subharmonic functions that converges pointwise to u.*

The next corollary can be interpreted as saying that a locally integrable function that is locally subharmonic almost everywhere is globally subharmonic.

4.22 Corollary. *If $u \in L^1_{\text{loc}}(G)$ such that whenever $\overline{B}(a; r) \subseteq G$ there is a subharmonic function v on $B(a; r)$ that is equal to u a.e.* [Area], *then there is a subharmonic function U on G such that $u = U$ a.e.* [Area].

Proof. Define u_ε as in the preceding proposition. Fix $\delta > 0$ and let K be a compact subset of $\{z \in G : d(z, \partial G) > 2\delta\}$; so $K_\varepsilon = \{z : d(z, K) \le \delta\}$ is a compact subset of G. It follows from the hypothesis that there is a subharmonic function v on int K_ε such that $u = v$ a.e. [Area]. Hence $u_\varepsilon(z) = v_\varepsilon(z)$ for all z in K and $\varepsilon < \delta$. This implies several things. First, $u_\varepsilon(z)$ decreases with ε and, by the preceding proposition, $u_\varepsilon(z) \to v(z)$ as $\varepsilon \to 0$. Also, since K was arbitrary, it follows that $U(z) \equiv \lim_{\varepsilon \to 0} u_\varepsilon(z)$ exists for all z in G. Moreover, by what has just been proved, U is subharmonic.

On the other hand, $\int_K |u_\epsilon - u| \, d\mathcal{A} \to 0$ for every compact subset K of G. So for any compact set K there is a sequence $\{\varepsilon_j\}$ such that $\varepsilon_j \to 0$ and $u_{\varepsilon_j}(z) \to u(z)$ a.e. [Area] on K. Thus $u = U$ a.e. [Area] on G. □

Subharmonic functions have been found to be quite useful in a variety of roles in analysis. This will be seen in this book. The discussion of subharmonic functions will continue in the next section, where the logarithmic potential is introduced and used to give another characterization of these functions, amongst other things. The work Hayman and Kennedy [1976] gives a full account of these functions.

Exercises

1. Show that if u is a subharmonic function on G, $\overline{B}(a;r) \subseteq G$, and μ is a probability measure on $[0,r]$ that is not the unit point mass at 0, then $u(a) \le \frac{1}{2\pi} \int_{[0,r]} \int_0^{2\pi} u(a + te^{i\theta}) \, d\theta \, d\mu(t)$.

2. Show that for any complex number z_0, the function $u(z) = |z - z_0|$ is subharmonic.

3. Show that Definition 4.12 extends Definition 6.3.1.

4. A function $\phi : [a,b] \to [-\infty, \infty)$ is convex if and only if conditions (a) and (b) from Definition 4.12 hold as well as

   4.23 $$\frac{\phi(u) - \phi(x)}{u - x} \le \frac{\phi(y) - \phi(u)}{y - u}$$

   for $a < x < u < y < b$.

5. A function $\phi : [a,b] \to [-\infty, \infty)$ is convex if and only if conditions (a) and (b) from Definition 4.12 hold as well as the condition that $\{x + iy : a < x < b \text{ and } \phi(x) \ge y\}$ is a convex subset of $\mathbb{C}$.

6. Show that if u is subharmonic on $\mathbb{C}$ and μ is a positive measure with compact support, then $u * \mu$ is subharmonic.

7. Let G be an open set and let $a \in G$. If u is a subharmonic function on $G \setminus \{a\}$ that is bounded above on a punctured neighborhood of

a, show that u can be defined at the point a in such a way that the resulting function is subharmonic on G.

8. If u is a subharmonic function on Ω and $\tau : G \to \Omega$ is an analytic function, then $u \circ \tau$ is subharmonic on G. (Hint: Use the Chain Rule to show that $\partial\bar{\partial}u \geq 0$ as a distribution.)

9. Let G_1 and G_2 be open sets with $\partial G_1 \subseteq G_2$ and put $G = G_1 \cup G_2$. If u_j is a subharmonic function on $G_j, j = 1, 2$, and $u_1 \leq u_2$ on $G_1 \cap G_2$, show that the function u defined on G by $u = u_1$ on G_1 and $u = u_2$ on $G_2 \setminus G_1$ is subharmonic.

§5 The Logarithmic Potential

Recall from Lemma 13.2.10 that the logarithm is locally integrable on the plane. Thus it is possible to form the convolution of the logarithm and any compactly supported measure.

5.1 Definition. For a compactly supported measure μ on $\mathbb{C}$, the *logarithmic potential* of μ is the function

$$L_\mu(z) = \int \log |z - w|^{-1} \, d\mu(w).$$

So for any compactly supported measure μ, the function L_μ is defined at every point of the plane. The elementary properties of the logarithmic potential are in the next proposition, where it is shown, in particular, that L_μ is a locally integrable function. Recall that $\hat{\mu}$ is the Cauchy transform of a compactly supported measure (18.5.1)

5.2 Proposition. *For a compactly supported measure μ, the following hold.*

(a) *L_μ is a locally integrable function.*

(b) *L_μ is harmonic on the complement of the support of μ.*

(c) *If μ is positive, L_μ is a superharmonic function on $\mathbb{C}$.*

(d) *$\partial L_\mu = -2^{-1}\hat{\mu}$ and $\bar{\partial} L_\mu = -2^{-1}\hat{\bar{\mu}}$, the Cauchy transform of the conjugate of μ.*

Proof. Part (a) is a consequence of Proposition 18.3.2 and the fact that a measure is finite on compact sets. Part (b) follows by differentiating under the integral sign. Part (c) is a consequence of Exercise 4.6. Part (d) follows by an application of Proposition 18.4.7; the details are left to the reader. □

Let's fix the measure μ and let K = support μ. Note that

$$L_\mu(z) + \mu(K)\log|z| = \int_K \log|1 - \frac{w}{z}|^{-1} d\mu(w)$$

and hence $L_\mu(z) + \mu(K)\log|z| \to 0$ as $z \to \infty$. This proves part of the next theorem.

5.3 Theorem. *If μ is a compactly supported measure, then*

$$\Delta L_\mu = -2\pi\,\mu$$

and $L_\mu(z)+\mu(K)\log|z| \to 0$ as $z \to \infty$. Moreover L_μ is the unique solution of this equation in the sense that if h is a locally integrable function such that $\Delta h = -2\pi\,\mu$ and $h(z) + \mu(K)\log|z| \to 0$ as $z \to \infty$, then $h = L_\mu$ a.e. [Area].

Proof. The fact that $\Delta L_\mu = -2\pi\,\mu$ is just a combination of (5.2.d) and Theorem 18.5.4. It only remains to demonstrate uniqueness. So let h be a locally integrable function as in the statement of the theorem. Thus $\Delta(L_\mu - h) = 0$ as a distribution. By Weyl's Lemma, $L_\mu - h$ is equal a.e. to a function v that is harmonic on $\mathbb{C}$. But we also have that $v(z) \to 0$ as $z \to \infty$. Thus v must be a bounded harmonic function on the plane and hence constant. Thus $v \equiv 0$, or $h = L_\mu$. □

5.4 Corollary. *If μ is a compactly supported measure, G is an open set, and $L_\mu(z) = 0$ a.e.* [Area] *on G, then $|\mu|(G) = 0$.*

Proof. If $\phi \in \mathcal{D}(G)$, then $\int \phi\, d\mu = -(2\pi)^{-1}\int \Delta\phi d\mathcal{A} = 0$. □

5.5 Corollary. *If μ is a compactly supported measure and G is an open set such that L_μ is harmonic on G, then $|\mu|(G) = 0$.*

Why define the logarithmic potential? From the preceding theorem we have that for a fixed w, $\Delta(log|z - w|) = -2\pi\,\delta_w$, where δ_w is the unit point mass at w. Thus the logarithmic potential is the convolution of the measure μ with this fundamental solution of the Laplacian and L_μ solves the differential equation $\Delta T = -2\pi\,\mu$.

The reader may have noticed several analogies between the Cauchy transform and the logarithmic potential. Indeed Proposition 5.2 shows that there is a specific relation between the two. Recall that a function u is harmonic if and only if $\partial\, u$ is analytic. This will perhaps bring (5.2) a little more into focus. Also as we progress it will become apparent that the logarithmic potential plays a role in the study of approximation by harmonic functions like the role played by the Cauchy transform in rational approximation. For example, see Theorem 18.6.2. But for now we will concentrate on an application to the characterization of subharmonic functions.

5.6 Theorem. *If u is a locally integrable function on the open set G, then the following statements are equivalent.*

(a) *The function u is equal to a subharmonic function a.e.* [Area] *in G.*

(b) *The Laplacian of u in the sense of distributions is positive.*

(c) *There is a positive extended real-valued measure ν on G such that if G_1 is a bounded open set with* cl $G_1 \subseteq G$ *and $\mu \equiv \nu | G_1$, then there is a harmonic function h on G_1 with*

$$u|G_1 = h - 2\pi\, L_\mu \text{ a.e.[Area]}.$$

Proof. (a) implies (b). Assume that u is a subharmonic function on G; it must be shown that, for every ϕ in $\mathcal{D}(G)$ such that $\phi \geq 0$, $\int u\, \Delta\phi\, d\mathcal{A} \geq 0$. So fix such a ϕ and let $0 < r_0 < 2^{-1}\text{dist}(\text{supp}\,\phi, \partial G)$. Let $K = \{z : \text{dist}(z, \text{supp}\,\phi) \leq r_0\}$. Let r be arbitrary with $0 < r < r_0$. So if $z \in K$, then $\overline{B}(z;r) \subseteq G$ and $2\pi\, u(z) \leq \int_0^{2\pi} u(z + re^{i\theta})\, d\theta$. Therefore

$$\begin{aligned} 2\pi \int_G u(z)\, \phi(z)\, d\mathcal{A}(z) &\leq \int_K \phi(z) \int_0^{2\pi} u(z + re^{i\theta})\, d\theta\, d\mathcal{A}(z) \\ &= \int_K u(z) \int_0^{2\pi} \phi(z - re^{i\theta})\, d\theta\, d\mathcal{A}(z). \end{aligned}$$

Hence

5.7 $$0 \leq \int_K u(z) \left[\int_0^{2\pi} \phi(z - re^{i\theta})\, d\theta - 2\pi\, \phi(z) \right] d\mathcal{A}(z).$$

Now look at the Taylor expansion of ϕ about any point z in supp ϕ to get

$$\begin{aligned} \int_0^{2\pi} \phi(z - re^{i\theta})\, d\theta &= \int_0^{2\pi} [\phi(z) + [(\bar\partial\, \phi)(z)][-re^{-i\theta}] + [(\partial\, \phi)(z)][re^{i\theta}] \\ &\quad + \frac{1}{2}[(\partial\bar\partial\phi)(z)]|re^{i\theta}|^2 + \frac{1}{2}[(\partial^2\phi)(z)][re^{i\theta}]^2 \\ &\quad + \frac{1}{2}[(\bar\partial^2\phi)(z)][re^{-i\theta}]^2 + \Psi(z - re^{i\theta})]\, d\theta \\ &= 2\pi\, \phi(z) + 2\pi\, r^2[(\partial\bar\partial\phi)(z)] + \Psi_1(z, r). \end{aligned}$$

Now the nature of the remainder term $\Psi_1(z,r)$ is such that there is a constant M with $|\Psi_1(z,r)| \leq M\, r^3$ for all z in K. Substituting into (5.7) and dividing by r^2, we get

$$\begin{aligned} 0 &\leq \int_K u(z) \left[\frac{\pi}{2}\Delta\phi(z) + r^{-2}\Psi_1(z,r) \right] d\mathcal{A}(z) \\ &= \frac{\pi}{2} \int_K u(z)\, \Delta\phi(z)\, d\mathcal{A}(z) + \int_K u(z)\, r^{-2}\Psi_1(z,r)\, d\mathcal{A}(z). \end{aligned}$$

Since $|\Psi_1(z,r)| \leq M\, r^3$ for all z in K, this second integral can be made as small as desired for a suitably small choice of r. So letting $r \to 0$, we get that

$$0 \leq \int u\, \Delta\phi\, d\mathcal{A}.$$

(b) *implies* (c). Since Δu is a positive distribution on G, there is a positive regular Borel measure ν on G such that $\Delta u = \nu$ (18.4.9). (It may be that ν is unbounded.) That is, for every ϕ in $\mathcal{D}(G)$, $\int u\, \Delta\phi\, d\mathcal{A} = \int \phi\, d\nu$. Let G_1 be any bounded open set with cl $G_1 \subseteq G$ and put $u_1 = u|G_1$ and $\mu = \nu|G_1$. So μ has compact support and L_μ is well defined. Also, as distributions on G_1, $\Delta(u_1 + 2\pi\, L_\mu) = 0$. By Weyl's Lemma, there is a harmonic function h on G_1 such that $u_1 + 2\pi\, L_\mu = h$. That is, $u_1 = h - 2\pi\, L_\mu$.

(c) *implies* (a). Since L_μ is superharmonic, (c) implies that, for every closed disk B contained in G, u equals a subharmonic function a.e. [Area] on the interior of B. Part (a) now follows by Corollary 4.22. □

Part (c) of the preceding theorem is called the Riesz Decomposition Theorem for subharmonic functions. Another version of this will be seen in Theorem 21.4.10 below.

5.8 Corollary. *If $u \in C^2(G)$, then u is subharmonic if and only if $\Delta u \geq 0$.*

The reader can now proceed, if desired, to the next section. A few essential properties of the logarithmic potential are appropriately obtained here, however. We will see these used in Chapter 21. The key to the proof of each of these results is the following lemma.

5.9 Lemma. *If μ is a positive measure with support contained in the compact set K and $a \in K$ such that $L_\mu(a) < \infty$, then for any $\varepsilon > 0$ there is a $\delta > 0$ with the property that for $|z - a| < \delta$ and κ any point in K with* $\text{dist}(z, K) = |z - \kappa|$ *we have*

$$L_\mu(z) < \varepsilon + L_\mu(\kappa).$$

Proof. Notice that the fact that $L_\mu(a) < \infty$ implies that μ does not have an atom at a. So if $\varepsilon > 0$, there is a $\rho > 0$ such that for $D = \overline{B}(a;\rho)$, $\mu(D) < \varepsilon$. Look at the two logarithmic potentials

$$u_1(z) = \int_D \log|w - z|^{-1}\, d\mu(w), \quad u_2(z) = \int_{K\setminus D} \log|w - z|^{-1}\, d\mu(w).$$

So u_2 is harmonic on int D (5.3). For each z in $\mathbb{C}$ let $\kappa(z)$ be any point in K such that $|z - \kappa(z)| = \text{dist}(z, K)$. (So $\kappa(z)$ is not a function.) Now if $z \to a$, $\kappa(z) \to a$. Therefore $\delta > 0$ can be chosen so that $\delta < \rho$ and for $|z - a| < \delta$ we have $|\kappa(z) - a| < \rho$ and $|u_2(z) - u_2(\kappa(z))| < \varepsilon$.

For all w in K, $|w - \kappa(z)| \leq |w - z| + |z - \kappa(z)| \leq 2|w - z|$. Hence

$$\begin{aligned} u_1(z) &\leq \int_D \log \left| \frac{w - \kappa(z)}{2} \right|^{-1} d\mu(w) \\ &= \log 2\mu(D) + u_1(\kappa(z)) \\ &< \varepsilon \log 2 + u_1(\kappa(z)). \end{aligned}$$

Thus for $|z - a| < \delta$,

$$\begin{aligned} L_\mu(z) &= u_1(z) + u_2(z) \\ &\leq \varepsilon \log 2 + u_1(\kappa(z)) + u_2(\kappa(z)) + \varepsilon \\ &\leq \varepsilon(1 + \log 2) + L_\mu(\kappa(z)). \end{aligned}$$

□

The next result is often referred to as the Maximum Principle for the logarithmic potential.

5.10 Proposition. *If μ is a positive measure with support contained in the compact set K and if $L_\mu(z) \leq M$ for all z in K, then $L_\mu(z) \leq M$ for all z in $\mathbb{C}$.*

Proof. Without loss of generality we may assume that $M < \infty$. By the Maximum Principle, $L_\mu \leq M$ in all the bounded components of the complement of K. Let G be the unbounded component of $\mathbb{C} \setminus K$.

Fix $\varepsilon > 0$. The preceding lemma implies that for any point a in K there is a $\delta > 0$ with the property that whenever $|z - a| < \delta$ there is a point κ in $B(a; \delta) \cap K$ for which $L_\mu(z) < \varepsilon + L_\mu(\kappa) \leq \varepsilon + M$. That is, $L_\mu \leq \varepsilon + M$ on $B(a; \delta)$. The collection of all such disks $B(a; \delta)$ covers K. Extracting a finite subcover we get an open neighborhood U of K such that $L_\mu \leq \varepsilon + M$ on U. But $L_\mu(z) \to -\infty$ as $z \to \infty$. Therefore $L_\mu(z) \leq \varepsilon + M$ in G by the Maximum Principle. Since ε was arbitrary, this completes the proof. □

Here are two important points about the last proposition. First, the upper bound M in this result may be infinite. For example, if K is the singleton $\{a\}$, then the only positive measures μ supported on K are of the form $\mu = \beta\delta_a$ for some $\beta \geq 0$. In this case $L_\mu = \infty$ on K. There are non-trivial examples of compact sets K for which each measure μ supported on K has $\sup_K L_\mu(z) = \infty$. See (21.8.17) below. The other point is that this result says nothing about the lower bound of the logarithmic potential. In fact, for a compactly supported measure μ, L_μ is always bounded below on its support (Exercise 1) while, according to Theorem 5.3, $L_\mu(z) \to -\infty$ as $z \to \infty$.

The next result is called the Continuity Principle for the logarithmic potential.

5.11 Proposition. *If μ is a positive measure with support contained in the compact set K, $a \in K$ such that $L_\mu(a) < \infty$, and if $L_\mu|K$ is continuous at a as a function from K into the extended real numbers, then L_μ is continuous at a as a function defined on $\mathbb{C}$.*

Proof. Fix $\varepsilon > 0$ and choose $\delta_0 > 0$ such that if $\kappa \in K$ and $|a - \kappa| < \delta_0$, then $|L_\mu(\kappa) - L_\mu(a)| < \varepsilon$. Now apply Lemma 5.9 to get a δ with $0 < \delta < \delta_0$ such that whenever $|z - a| < \delta$ there is a point κ in $B(a;\delta) \cap K$ for which $L_\mu(z) < \varepsilon + L_\mu(\kappa)$. But this implies $L_\mu(z) < \varepsilon + L_\mu(a) + \varepsilon$. What this says is that $\limsup_{z\to a} L_\mu(z) \le L_\mu(a)$. But L_μ is lsc and so $L_\mu(a) \le \liminf_{z\to a} L_\mu(z)$. $\square$

Potential theory is a vast subject. If other kernel functions are used besides $\log|z-w|^{-1}$, other potentials are defined and the theory carries over to n-dimensional Euclidean space. Indeed, the logarithmic potential is peculiar and annoying in that it is not positive; a cause of pain and extra effort later in this book. Various properties and uses of these general potentials are found in the literature. In particular, most have properties analogous to those for the logarithmic potential found in the last two propositions. The reader interested in these matters can consult Brelot [1959], Carleson [1967], Choquet [1955], Frostman [1935], Hedberg [1972b], Helms [1975], Landkof [1972], and Wermer [1974].

Exercises

1. For a positive compactly supported measure μ, show that $L_\mu(z) > -\infty$ for all z in $\mathbb{C}$, and for any compact set K there is a constant m such that $m \le L_\mu(z)$ for all z in K.

2. If $\mu, \mu_1, \mu_2, \dots$ are measures whose supports are contained in the compact set K and $\mu_n \to \mu$ weak* in $M(K)$, show that $L_\mu(z) \le \liminf_n L_{\mu_n}(z)$ for all z in $\mathbb{C}$.

3. Let G be an open set and K be a compact subset of G. Show that if u is a real-valued harmonic function on $G \setminus K$, then $u = u_0 + u_1$, where u_1 is harmonic on G, u_0 is harmonic on $\mathbb{C}\setminus K$, and there is a constant a such that $u_0(z) + a\log|z| \to 0$ as $z \to \infty$. (Hint: Let $\phi \in C_C^\infty$ such that $0 \le \phi \le 1$, supp$\phi \subseteq G$, and $\phi \equiv 1$ on $K_\delta \equiv \{z : \text{dist}(z, K) \le \delta\} \subseteq G$. Define the function $\Psi : G \to \mathbb{R}$ by $\Psi = (1-\phi)u$ on $G\setminus K$ and $\Psi = 0$ on intK_δ. Let μ be the measure $(\Delta\Psi)\chi_{G\setminus K}$ and put $V = -(2\pi)^{-1}L_\mu$. Show that V is a subharmonic function that is locally integrable, $\Delta V = \mu$, and, if $a = -(2\pi)^{-1}\mu(L)$, then $V(z) + a\log|z| \to 0$ as $z \to \infty$. Define the function $u_0 : \mathbb{C} \setminus K \to \mathbb{R}$ by letting $u_0 = u\phi + V$ on $G\setminus K$ and $u_0 = V$ on the complement of supp ϕ. Define $u_1 : G \to \mathbb{R}$ by letting $u_1 = \Psi - V$ on $G \setminus K$ and $u_1 = -V$ on intK_δ.)

4. Show that the functions u_0 and u_1 in Exercise 3 are unique.

5. In Exercise 3, show that if both G and u are bounded, then so is u_1.

§6 An Application: Approximation by Harmonic Functions

In this section we will see that the logarithmic potential plays a role in approximation by harmonic functions similar to the role played by the Cauchy transform in approximation by rational functions (see §18.6).

6.1 Definition. If K is a compact subset of $\mathbb{C}$, let $H(K)$ be the uniform closure in $C(K)$ of the functions that are harmonic in a neighborhood of K. $H_{\mathbb{R}}(K)$ denotes the real-valued functions in $H(K)$.

Note that $H(K)$ is a Banach space but is not a Banach algebra. On the other hand, if a function belongs to $H(K)$, then so do its real and imaginary parts. This will be helpful below, since a measure annihilates $H(K)$ if and only if its real and imaginary parts also annihilate $H(K)$. This was not the case when we studied the algebra $R(K)$.

6.2 Theorem. *If K is a compact subset of $\mathbb{C}$, μ is a real-valued measure on K, and $\{a_1, a_2, \ldots\}$ is chosen so that each bounded component of the complement of K contains one element of this sequence, then the following are equivalent.*

(a) $\mu \perp H(K)$.

(b) $L_\mu(z) = 0$ *for all* z *in* $\mathbb{C} \setminus K$.

(c) $\mu \perp R(K)$ *and* $L_\mu(a_j) = 0$ *for* $j = 1, 2, \ldots$

Proof. (a) *implies* (b). If $z \notin K$, then $w \to \log|z - w|$ is harmonic in a neighborhood of K. Thus (b).

(b) *implies* (c). Since L_μ vanishes identically off K, $\hat{\mu} = -2\partial L_\mu \equiv 0$ on $\mathbb{C} \setminus K$. By Theorem 18.6.2, $\mu \perp R(K)$.

(c) *implies* (a). Since μ is real-valued, Theorem 18.6.2 together with Proposition 5.2.d imply that ∂L_μ and $\bar{\partial} L_\mu$ vanish on the complement of K. Thus L_μ is constant on components of $\mathbb{C} \setminus K$; the other condition in part (c) implies that L_μ vanishes on all the bounded components of $\mathbb{C} \setminus K$. On the other hand, for $|a|$ sufficiently large there is a branch of $\log(z - a)$ defined in a neighborhood of K so that this function belongs to $R(K)$ by Runge's Theorem. Thus L_μ also is identically 0 in the unbounded component of $\mathbb{C} \setminus K$. □

Note that the only place in the proof of the preceding theorem where the fact that μ is real-valued was used was in the proof that (c) implies (a). Conditions (a) and (b) are equivalent for complex measures. In fact, if $L_\mu \equiv 0$ on $\mathbb{C} \setminus K$, then $L_\nu \equiv 0$ there, where ν is either the real or imaginary part of μ.

The next theorem can be taken as the analogue of Runge's Theorem. To set the notation, fix K and let E be a subset of $\mathbb{C} \setminus K$ that meets each bounded component of $\mathbb{C} \setminus K$. Let $H_{\mathbb{R}}(K, E)$ be the uniform closure of all the functions of the form $\operatorname{Re} f + \sum_{k=1}^n c_k \log|z - a_k|$, where f is analytic in a neighborhood of K, $n \geq 1$, the constants c_k are real, and $a_k \in E$ for $1 \leq k \leq n$.

6.3 Theorem. *If K is a compact subset of $\mathbb{C}$ and E is a set that meets each bounded component of $\mathbb{C} \setminus K$, then $H_{\mathbb{R}}(K, E) = H(K)$, where $H_{\mathbb{R}}(K, E)$ is defined as above.*

Proof. Clearly $H_{\mathbb{R}}(K, E) \subseteq H(K)$. It suffices to show that if μ is a real-valued measure supported on K and $\mu \perp H_{\mathbb{R}}(K, E)$, then $\mu \perp H(K)$. But since $\int \operatorname{Re} f \, d\mu = 0$ for every function f analytic in a neighborhood of K and μ is real-valued, $\mu \perp R(K)$. Also $L_\mu(a) = 0$ for all a in E. By the preceding theorem, $\mu \perp H(K)$. □

The proof of the next corollary is similar to the proof of Corollary 8.1.14.

6.4 Corollary. *If G is an open subset of $\mathbb{C}$ and E is a subset of $\mathbb{C} \setminus G$ that meets every component of $\mathbb{C} \setminus G$, then every real-valued harmonic function u on G can be approximated uniformly on compact subsets of G by functions of the form*

$$\operatorname{Re} f(z) + \sum_{k=1}^{n} b_k \log|z - a_k|,$$

where f is analytic on G, $n \geq 1$, and $a_1, \ldots, a_n$ are points from E.

We will return to the logarithmic potential when we take up the study of potential theory in the plane in Chapter 21.

Exercises

1. Let K be a compact subset of $\mathbb{C}$. For μ in $M(K)$ and ϕ a smooth function with compact support, let $\mu_\phi \equiv \phi\mu + \pi^{-1}(\hat{\mu}\partial\phi + \hat{\bar{\mu}}\bar{\partial}\phi)\mathcal{A} - (2\pi)^{-1} L_\mu \Delta\phi \mathcal{A}$. Prove the following. (a) $L_{\mu_\phi} = \phi L_\mu$. (b) $\mu \perp H(K)$ if and only if $\mu_\phi \perp H(K)$ for all smooth functions ϕ with compact support. (c) $H(K) = C(K)$ if and only if for every closed disk D, $H(D \cap K) = C(D \cap K)$. (Compare with Exercise 18.6.1.)

2. If K is either a closed disk or a closed annulus, show that $H(K)$ is the space of continuous functions on K that are harmonic on int K.

§7 The Dirichlet Problem

This topic was discussed in §10.4 and we will return to examine various aspects of the subject in the remainder of this book. We begin by recalling the following definition.

7.1 Definition. A *Dirichlet set* is an open subset of $\mathbb{C}_\infty$ with the property that for each continuous function $u : \partial_\infty G \to C$ there is a continuous function $h : \text{cl}\ _\infty G \to \mathbb{C}$ such that h is harmonic on G and $h|\partial_\infty G = u$. The function h is called the *classical solution of the Dirichlet problem* with boundary values u. A connected Dirichlet set is called a *Dirichlet region.*

The sets G in this definition are allowed to contain the point at infinity, so let's be clear about what is meant by being harmonic (or subharmonic) at infinity. Suppose K is a compact subset of $\mathbb{C}$ and $u : \mathbb{C} \setminus K \to \mathbb{R}$ is harmonic. Thus $u(z^{-1})$ is harmonic in a deleted neighborhood of 0. We say that u is *harmonic at infinity* if $u(z^{-1})$ has a harmonic extension to a neighborhood of 0. According to Theorem 15.1.3, there is a real constant c and an analytic function h defined in this punctured neighborhood such that $u(z^{-1}) = c \log|z| + \operatorname{Re} h(z)$. Thus $u(z) = -c \log|z| + \operatorname{Re} h(z^{-1})$. So u is harmonic at ∞ precisely when this constant c is 0 and h has a removable singularity at 0. In most of the proofs we will assume that G is contained in the finite plane. Usually the most general case can be reduced to this one by an examination of the image of G under an appropriate Möbius transformation chosen so that this image is contained in $\mathbb{C}$.

Some of the results of §10.4 will be used here. It is shown there that if each component of $\partial_\infty G$ consists of more that one point, then G is a Dirichlet set. It is also shown there that the punctured disk is not a Dirichlet set.

In this section we will not concentrate on the classical solution of the Dirichlet problem. This topic will be encountered in §10 below, when we discuss regular points. Even though there are open subsets of $\mathbb{C}$ that are not Dirichlet sets, each function on $\partial_\infty G$ gives rise to a "candidate" for the solution of the Dirichlet problem as we saw in Theorem 10.3.11. We recall this result in a somewhat different form to accommodate our new definition of a subharmonic function and to extend the functions that are admissible as boundary values. Let $\mathbb{R}_\infty = \mathbb{R} \cup \{\pm\infty\}$ with the obvious topology; that is, $\mathbb{R}_\infty$ is the "two point" compactification of $\mathbb{R}$. An extended real-valued function is one that takes its values in $\mathbb{R}_\infty$.

7.2 Definition. If u is any extended real-valued function defined on $\partial_\infty G$, let

$$\begin{aligned}\hat{\mathcal{P}}(u,G) \quad = \quad & \{\phi : \phi \text{ is subharmonic on } G,\ \phi \text{ is bounded above, and} \\ & \left.\limsup_{z\to a} \phi(z) \leq u(a) \text{ for every } a \text{ in } \partial_\infty G\right\},\end{aligned}$$

$$\check{\mathcal{P}}(u,G) = \Big\{\psi : \psi \text{ is superharmonic on } G,\ \psi \text{ is bounded below, and } \liminf_{z \to a} \psi(z) \geq u(a) \text{ for every } a \text{ on } \partial_\infty G\Big\}.$$

These collections of functions are sometimes called the *lower* and *upper Perron families*, respectively, associated with u and G. It is somewhat useful to observe that $-\check{\mathcal{P}}(-u,G) = \hat{\mathcal{P}}(u,G)$. Also define functions $\hat{u}$ and $\check{u}$ on G by

$$\begin{aligned} \hat{u}(z) &= \sup\{\phi(z) : \phi \in \hat{\mathcal{P}}(u,G)\}, \\ \check{u}(z) &= \inf\{\psi(z) : \psi \in \check{\mathcal{P}}(u,G)\}. \end{aligned}$$

These functions are called the *lower* and *upper Perron functions*, respectively, associated with u and G. They will also be denoted by $\hat{u}_G$ and $\check{u}_G$ if the dependence on G needs to be emphasized, which will be necessary at certain times.

Note that $\hat{\mathcal{P}}(u,G)$ contains the identically $-\infty$ function, so that it is a non-empty family of functions. Thus $\hat{u}$ is well defined, though it may be the identically $-\infty$ function. Of course if u is a bounded function, then $\hat{\mathcal{P}}(u,G)$ contains some finite constant functions and so $\hat{u}$ is a bounded function. Similar comments apply to $\check{\mathcal{P}}(u,G)$ and $\check{u}$. Usually results will be stated for the lower Perron family or function associated with u and G; the corresponding statements and proofs for the upper family and function will be left to the reader. The proof of the next proposition is left to the reader. (See Corollary 4.18.)

7.3 Proposition. *If ϕ_1 and $\phi_2 \in \hat{\mathcal{P}}(u,G)$, then so is $\phi_1 \vee \phi_2$. If $\phi \in \hat{\mathcal{P}}(u,G)$ and $\overline{B}(a;r) \subseteq G$, then the harmonic modification of ϕ on $B(a;r)$ also belongs to $\hat{\mathcal{P}}(u,G)$.*

7.4 Lemma. *Assume that $\mathcal{P}$ is a family of subharmonic functions on G with the properties:*

(i) *if ϕ_1 and $\phi_2 \in \mathcal{P}$, then so is $\phi_1 \vee \phi_2$;*

(ii) *if $\phi \in \mathcal{P}$ and $\overline{B}(a;r) \subseteq G$, then the harmonic modification of ϕ on $B(a;r)$ also belongs to $\mathcal{P}$.*

If $h(z) = \sup\{\phi(z) : \phi \in \mathcal{P}\}$, then on each component of G either $h \equiv \infty$ or h is harmonic.

Proof. The proof of this lemma is like the proof of Theorem 10.3.11, but where Corollary 4.18 is used rather than Corollary 10.3.7. The details are left to the reader. □

7.5 Proposition.

(a) *If u is any function, $\hat{u}$ and $\check{u}$ are harmonic functions on any component of G on which they are not identically $\pm\infty$.*

(b) *If c is a non-negative real number, then* $\widehat{cu} = c\hat{u}$ *and* $\widecheck{cu} = c\check{u}$.

(c) *If c is a non-positive real number, then* $\widehat{cu} = c\check{u}$ *and* $\widecheck{cu} = c\hat{u}$.

(d) *If v is a finite-valued function on $\partial_\infty G$ such that $\hat{v}$ and $\check{v}$ are finite on G, then* $\hat{u} + \hat{v} \leq \widehat{(u+v)} \leq \widecheck{(u+v)} \leq \check{u} + \check{v}$.

(e) *If u is a bounded function,* $-||u||_\infty \leq \hat{u} \leq \check{u} \leq ||u||_\infty$.

Proof. Part (a) is immediate from Proposition 7.3 and the preceding lemma. The proofs of parts (b) and (c) are routine. The proof of (d) is a combination of basic mathematics and an application of the Maximum Principle, which implies that if $\phi \in \hat{\mathcal{P}}(u, G)$ and $\psi \in \check{\mathcal{P}}(u, G)$, then $\phi \leq \psi$. The proof of (e) follows by observing that certain constant functions belong to the appropriate families of subharmonic and superharmonic functions. □

If $u \in C_\mathbb{R}(\partial_\infty G)$ and there is a classical solution h of the Dirichlet problem with boundary values u, then $\hat{u} = \check{u} = h$ (Exercise 1). So in order to solve the Dirichlet problem, it must be that $\hat{u} = \check{u}$. Suppose that this does indeed happen and $\hat{u} = \check{u}$. Define $h : \text{cl}_\infty G \to \mathbb{R}$ by letting $h = \hat{u} = \check{u}$ on G and $h = u$ on $\partial_\infty G$. There are two difficulties here. First, how do we know that h is continuous on $\text{cl}_\infty G$? Indeed, it will not always be so since we cannot always solve the Dirichlet problem. Second, how can we decide whether $\hat{u} = \check{u}$? The first question was discussed in §10.4 and will receive a more complete discussion as we progress. The second question has an affirmative answer in most situations and will be addressed shortly. But first we show that it is only necessary to consider connected open sets. As is customary, results will be stated and proved for lower Perron functions, while the analogous results for the upper Perron functions will be left to the reader to state and verify.

7.6 Proposition. *If G is any open set, H is the union of some collection of components of G, $u : \partial_\infty G \to [-\infty, \infty]$, and $v = u|\partial_\infty H$, then $\hat{u}_G = \hat{v}_H$ on H.*

Proof. It is easy to see that if $\phi \in \hat{\mathcal{P}}(u, G)$, then $\phi|H \in \hat{\mathcal{P}}(v, H)$; thus $\hat{u}_G \leq \hat{v}_H$ on H. Now suppose $\phi \in \hat{\mathcal{P}}(u, G)$ and $\phi_0 \in \hat{\mathcal{P}}(v, H)$ and define ϕ_1 on G by $\phi_1(z) = \phi(z) \vee \phi_0(z)$ for z in H and $\phi_1(z) = \phi(z)$ otherwise. The reader can check that $\phi_1 \in \hat{\mathcal{P}}(u, G)$ and, of course, $\phi_0(z) \leq \phi_1(z)$ for z in H. Thus $\hat{v}_H(z) \leq \hat{u}_G(z)$ for all z in H. □

The previous proposition is generalized in Proposition 21.1.13 below.

7.7 Corollary. *If G is any open set and $u : \partial_\infty G \to [-\infty, \infty]$, then $\hat{u} = \check{u}$ if and only if for each component H of G, $\hat{v}_H = \check{v}_H$ on H, where $v = u|\partial_\infty H$.*

7.8 Proposition. *Let G be an arbitrary open set and assume u is an* *arbitrary extended real-valued function on $\partial_\infty G$.*

(a) *There is a sequence $\{\phi_n\}$ in $\hat{\mathcal{P}}(u,G)$ such that for every z in G, $\{\phi_n(z)\}$ is increasing and $\phi_n(z) \to \hat{u}(z)$. If α is a real constant such that $u(z) \geq \alpha$ for all z in $\partial_\infty G$, the sequence $\{\phi_n\}$ can be chosen such that $\phi_n \geq \alpha$ for all n.*

(b) *There is an increasing sequence $\{u_n\}$ of upper semicontinuous functions on $\partial_\infty G$ such that each u_n is bounded above, $u_n \leq u$, and $\hat{u}_n(z) \to \hat{u}(z)$ for all z in G. If α is a real constant such that $u(z) \geq \alpha$ for all z in $\partial_\infty G$, the sequence $\{u_n\}$ can be chosen such that $u_n \geq \alpha$ for all n.*

Proof. (a) This will only be proved under the additional assumption that G is connected. The proof of the general case is left to the reader. If $\hat{\mathcal{P}}(u,G)$ only contains the constantly $-\infty$ function, $\hat{u} \equiv -\infty$ and we can put $\phi_n \equiv -\infty$ for all n. Assume that $\hat{u}$ is not identically $-\infty$. Fix a point a in G and let $\{\phi_n\}$ be a sequence in $\hat{\mathcal{P}}(u,G)$ such that $\phi_n(a) \to \hat{u}(a)$; it can be assumed that ϕ_1 is not identically $-\infty$. Replacing ϕ_n by $\phi_1 \vee \ldots \vee \phi_n$, we may assume that for each z in $G, \{\phi_n(z)\}$ is an increasing sequence. Now write G as the union of regions $\{G_n\}$ such that for every $n \geq 1$, $a \in G_n$, $\operatorname{cl} G_n \subseteq G_{n+1}$, and ∂G_n is a Jordan system. Let $\tilde{\phi}_n$ be the function on G that agrees with ϕ_n on $G \setminus G_n$ and on G_n is the solution of the Dirichlet problem with boundary values ϕ_n. It follows from the Maximum Principle that $\{\tilde{\phi}_n\}$ is also an increasing sequence, each $\tilde{\phi}_n \in \hat{\mathcal{P}}(u,G)$, and $\phi_n \leq \tilde{\phi}_n$. Thus $\tilde{\phi}_n(a) \to \hat{u}(a)$.

Let $h(z) = \lim_n \tilde{\phi}_n(z) = \sup_n \tilde{\phi}_n(z)$. So $h \leq \hat{u}$ and by Harnack's Theorem either $h \equiv \infty$ or h is harmonic. If $h \equiv \infty$, then so is $\hat{u}$ and we are done. Otherwise, $h \leq \hat{u}$ and $h(a) = \hat{u}(a)$ and so, by the Maximum Principle, $h = \hat{u}$.

If $u \geq \alpha$, then replace each ϕ_n by $\phi_n \vee \alpha$.

(b) First assume that u is bounded above. Let $\{\phi_n\}$ be as in part (a) and define $u_n(\zeta) = \limsup_{z\to\zeta} \phi_n(z)$ for all ζ in $\partial_\infty G$. According to Exercise 3.8, u_n is upper semicontinuous and clearly $u_n \leq u$; since ϕ_n is bounded above, so is u_n. Also $\phi_n \in \hat{\mathcal{P}}(f,G)$ and so $\phi_n \leq \hat{u}_n \leq \hat{u}$. Thus $\hat{u}_n(z) \to \hat{u}(z)$ for all z in G.

If u is not bounded above, consider $u \wedge m$, obtain a sequence $\{u_n^m\}$ of usc functions as in the preceding paragraph, and put $u_n = u_n^n$. $\square$

To facilitate the discussion we make the following definition.

7.9 Definition. Say that a function $u : \partial_\infty G \to \mathbb{R}_\infty$ is *solvable* if $\hat{u} = \check{u}$ and this function is finite-valued on G. If u is a solvable function, then $\hat{u}$ will be called the solution of the Dirichlet problem with boundary values u. If G is an open set such that every continuous function on $\partial_\infty G$ is solvable, say that G is a *solvable set.*

Because $\hat{u}$ and $\check{u}$ are harmonic functions and $\hat{u} \leq \check{u}$ on G, in order for u to be solvable it suffices, by the Maximum Principle, that the functions are equal and finite at one point of each component of G. In Theorem 21.1.12 below the solvable functions will be characterized for solvable open sets G. Solvable sets will be dealt with at length as we progress through this chapter and given another characterization in §10. From Chapter 10 we know that all regions whose boundary has no trivial components is a solvable set. It is also the case that $\mathbb{C}$ is solvable; a somewhat unfortunate turn of events as most results on solvable sets will have to omit this particular one.

The tiresome aspect of considering open sets G that are not connected is that a function u on $\partial_\infty G$ can be solvable on some components while not being solvable on others. However in light of Corollary 7.7 we have that the function u is solvable on G if and only if it is solvable for each component of G.

7.10 Proposition. *If $\mathcal{S}(G)$ is the collection of all finite-valued solvable functions on G, then $\mathcal{S}(G)$ is a real linear space. If u and $v \in \mathcal{S}(G)$ and $\alpha, \beta \in \mathbb{R}$, then $\widehat{\alpha u + \beta v} = \alpha\hat{u} + \beta\hat{v}$. If $\mathcal{S}^\infty(G)$ is the collection of all bounded solvable functions on $\partial_\infty G$, and $\mathcal{S}^\infty(G)$ is endowed with the supremum norm, then $\mathcal{S}^\infty(G)$ is a real Banach space. If $\{u_n\}$ is a sequence in $\mathcal{S}^\infty(G)$ and $u_n \to u$ in the supremum norm, then $\hat{u}_n \to \hat{u}$ uniformly on G.*

Proof. The fact that $\mathcal{S} = \mathcal{S}(G)$ is a real linear space is a consequence of (b), (c), and (d) of Proposition 7.5. If Proposition 7.5 is massaged in the appropriate manner, it will produce the fact that, if u and $v \in \mathcal{S}$ and $\alpha, \beta \in \mathbb{R}$, then $\widehat{\alpha u + \beta v} = \alpha\hat{u} + \beta\hat{v}$. The massaging is left to the reader.

To show that $\mathcal{S}^\infty = \mathcal{S}^\infty(G)$ is complete, let $\{u_n\}$ be a sequence of functions in $\mathcal{S}^\infty$ and assume $u : \partial G \to \mathbb{R}$ is a bounded function such that $||u_n - u||_{\partial G} \to 0$. Using parts (d) and (e) of Proposition 7.5, we have that $\check{u} = \widecheck{(u - u_n + u_n)} \leq \widecheck{(u - u_n)} + \check{u}_n \leq ||u - u_n|| + \check{u}_n$; also $\hat{u} = \widehat{u - u_n + u_n} \geq \widehat{u - u_n} + \hat{u}_n \geq -||u - u_n|| + \hat{u}_n$. Since $\hat{u}_n = \check{u}_n$, $0 \leq \hat{u} - \check{u} \leq 2||u - u_n||$. Hence $u \in \mathcal{S}^\infty$.

Using (7.5.e) we have that $|\hat{u}(z) - \hat{u}_n(z)| = |\widehat{u - u_n}(z)| \leq ||u - u_n||_\infty$. Thus $\hat{u}_n \to \hat{u}$ uniformly on G. □

We will retain this notation. For any open set G, $\mathcal{S}^\infty = \mathcal{S}^\infty(G)$ will denote the bounded solvable functions defined on $\partial_\infty G$ furnished with the supremum norm and $\mathcal{S} = \mathcal{S}(G)$ will denote the collection of all solvable functions.

7.11 Proposition. *Assume G is connected. If for each $n \geq 1$, $u_n \in \mathcal{S}(G)$ with $u_n \geq 0$ and if $u \equiv \sum_n u_n$, then either:*

(a) *$u \in \mathcal{S}(G)$ and $\hat{u} = \sum_n \hat{u}_n$, where the convergence of this series is uniform on compact subsets of G; or*

(b) $\hat{u} = \check{u} \equiv \infty$ *and* $\sum_n \hat{u}_n(z)$ *converges to* ∞ *uniformly on compact subsets of* G.

Proof. Let $v_n = \sum_{k=1}^n u_k$; so $v_n \in \mathcal{S}(G)$ and $\check{v}_n = \sum_{k=1}^n \check{u}_k$. Since $v_n \leq u$, $\check{v}_n \leq \check{u}$. Thus $\sum_n \check{u}_n \leq \check{u}$. Similarly, $\sum_n \hat{u}_n \leq \hat{u}$.

Note that since $u_n \geq 0$, Proposition 7.8 implies we may restrict our attention to the positive functions in $\check{\mathcal{P}}(u_n, G)$. Fix a point z_0 in G and an $\varepsilon > 0$. For each $n \geq 1$, let $\psi_n \in \check{\mathcal{P}}(u_n, G)$ such that ψ_n is positive and $\psi_n(z_0) < \check{u}_n(z_0) + \varepsilon/2^n$. According to Proposition 4.6.c, $\psi \equiv \sum_n \psi_n$ is superharmonic. If $a \in \partial_\infty G$, then

$$\liminf_{z \to a} \psi(z) \geq \sum_{n=1}^{\infty} \liminf_{z \to a} \psi_n(z) \geq \sum_{n=1}^{\infty} u_n(a) = u(a).$$

Thus $\psi \in \check{\mathcal{P}}(u, G)$. Hence $\check{u}(z_0) \leq \psi(z_0) \leq \sum_n \check{u}_n(z_0) + \varepsilon$. Since z_0 and ε were arbitrary, this implies that $\check{u} \leq \sum_n \check{u}_n$. Since the reverse inequality was already established, this implies that $\sum_n \hat{u}_n \leq \hat{u} \leq \check{u} = \sum_n \check{u}_n = \sum_n \hat{u}_n$. Thus $\hat{u} = \check{u}$.

If u is solvable, then because the harmonic functions in the sum $\hat{u}(z) = \sum_n \hat{u}_n(z)$ are all positive, the convergence is uniform on compact sets. Part (b) also follows by Harnack's Theorem. □

7.12 Corollary. *Assume* G *is connected. If* $\{u_n\}$ *is an increasing sequence of solvable functions on* $\partial_\infty G$ *and* $u(\zeta) \equiv \lim_n u_n(\zeta)$, *then either* u *is solvable and* $\hat{u}_n(z) \to \hat{u}(z)$ *uniformly on compact subsets of* G *or* $\hat{u} = \check{u} \equiv \infty$ *and* $\hat{u}_n(z) \to \infty$ *uniformly on compact subsets of* G.

This allows us to prove a useful fact about solvable open sets as defined in (7.9).

7.13 Corollary. *If* G *is a solvable set, then every bounded real-valued Borel function on* $\partial_\infty G$ *is solvable.*

Proof. The hypothesis implies that $\mathcal{S}^\infty$ is a collection of bounded functions on $\partial_\infty G$ that contains the continuous functions and, according to Corollary 7.12, contains the bounded pointwise limit of any increasing or decreasing sequence in $\mathcal{S}^\infty$. This implies that $\mathcal{S}^\infty$ contains the bounded Borel functions. (A result from measure theory says that if we take the increasing limits of continuous functions, then take the decreasing limits of the resulting functions, and continue to repeat this transfinitely up to the first uncountable ordinal, the resulting class of functions is precisely the Borel functions.) □

Now to get a rich supply of examples of solvable sets. The next lemma extends Exercise 1.

7.14 Lemma. *If ϕ is a real-valued function that is continuous on $\text{cl}_\infty G$ and subharmonic on G, then $\phi|\partial_\infty G$ is solvable.*

Proof. Let $u = \phi|\partial_\infty G$ and $\phi_1 = \phi|G$; it follows that $\phi_1 \in \hat{\mathcal{P}}(u, G)$ and hence that $\phi_1 \leq \hat{u}$ on G. Thus $\liminf_{z\to a} \hat{u}(z) \geq \liminf_{z\to a} \phi_1(z) = \phi(a) = u(a)$. That is, $\hat{u} \in \check{\mathcal{P}}(u, G)$. Therefore $\check{u} \leq \hat{u}$. But we always have that $\hat{u} \leq \check{u}$ by (7.5.e). Thus u is solvable. □

The next theorem is due to Wiener.

7.15 Theorem. *If G is a bounded open set, then G is solvable.*

Proof. Let u be a real-valued continuous function on ∂G and let $\{p_n(z, \bar{z})\}$ be a sequence of polynomials in z and $\bar{z}$ that converges to u uniformly on ∂G. (The boundedness of G is being used here.) Let c_n be a positive constant that is sufficiently large that $\Delta(p_n + c_n|z|^2) > 0$ on cl G (again the boundedness of G). By Corollary 5.8, $p_n + c_n|z|^2$ is subharmonic on G. The preceding lemma implies that $p_n + c_n|z|^2$ is solvable. Thus both $p_n + c_n|z|^2$ and $c_n|z|^2$ are solvable. Therefore p_n is solvable by Proposition 7.10. This same proposition implies that u, the limit of $\{p_n\}$, is solvable. □

With a little work this collection of solvable sets can be enlarged. This requires a preliminary result that has independent interest.

7.16 Proposition. *If G is a solvable set and $\tau : \text{cl}_\infty G \to \text{cl}_\infty \Omega$ is a homeomorphism such that τ is analytic on G, then Ω is a solvable set.*

Proof. It is not difficult to see that with the assumptions on τ, for any function $v : \partial_\infty \Omega \to \mathbb{R}$ and $u = v \circ \tau$, $\hat{\mathcal{P}}(u, G) = \{\phi \circ \tau : \phi \in \hat{\mathcal{P}}(v, \Omega)\}$ and $\check{\mathcal{P}}(u, G) = \{\psi \circ \tau : \psi \in \check{\mathcal{P}}(v, \Omega)\}$. The rest readily follows from this. □

7.17 Theorem. *If G is not dense in $\mathbb{C}$ or if G is a Dirichlet set, then G is a solvable set.*

Proof. Suppose G is not dense. So there is a disk $B(a; r)$ that is disjoint from the closure of G. But then the image of G under the Möbius transformation $(z - a)^{-1}$ is contained in $B(a; r)$. By Theorem 7.15 and the preceding proposition, G is a solvable set.

If G is a Dirichlet set, then for every continuous function u on $\partial_\infty G$ there is a classical solution h of the Dirichlet problem. According to Lemma 7.14 (or Exercise 1) this implies that u is solvable. □

7.18 Corollary. *If G is not connected, then G is solvable.*

Proof. According to Corollary 7.7, G is solvable if and only if each component of G is solvable. But if G is not connected, then no component of G is dense and hence each component is a solvable set. □

An examination of the preceding theorem and its corollary leads one to believe that an open subset of a solvable set is also solvable. This is the case and will follow after we characterize solvable sets in Theorem 10.7 below.

This last result leaves the candidates for non-solvable sets to those of the form $G = \mathbb{C} \setminus F$, where F is a closed set without interior in $\mathbb{C}$. If no component of F is a singleton, then G is a Dirichlet set (10.4.17) and so G is solvable. So a typical example would be to take $G = \mathbb{C}_\infty \setminus K$, where K is a Cantor set. Note that such open sets are necessarily connected.

It is not difficult to get an example of a non-solvable set. In fact the punctured plane is such an example. To see this we first prove a lemma that will be used in a later section as well.

7.19 Lemma. *If ϕ is a subharmonic function on the punctured plane, $\mathbb{C}_0$, such that $\phi \leq 0$ and $\limsup_{z\to 0} \phi(z) \leq \alpha$ for some α, then $\phi \leq \alpha$.*

Proof. Let $\varepsilon > 0$; by hypothesis, there is a $\delta > 0$ such that $\phi(z) < \alpha + \varepsilon$ for $|z| \leq \delta$. Take b to be any point in $\mathbb{C}_0$ and let r be an arbitrary number with $r > |b|$. If $h_r(z) = [\log(\delta/r)]^{-1}[\alpha + \varepsilon] \log |z/r|$, then h_r is harmonic in a neighborhood of the closure of the annulus $A = \text{ann}(0; \delta, r)$. For $|z| = \delta$, it is easy to see that $h_r(z) \geq \phi(z)$. For $|z| = r, h_r(z) = 0 \geq \phi(z)$. Thus the Maximum Principle implies that $h_r \geq \phi$ on cl A; in particular, $\phi(b) \leq h_r(b)$. Since r was arbitrary, $\phi(b) \leq \lim_{r\to\infty} h_r(b) = \alpha + \varepsilon$. Since ε was arbitrary, $\phi(b) \leq \alpha$. □

7.20 Example. The punctured plane $\mathbb{C}_0$ is not a solvable set. Define $u(0) = 0$ and $u(\infty) = 1$. If $\phi \in \hat{\mathcal{P}}(u, \mathbb{C}_0)$, then $\limsup_{z\to 0} \phi(z) \leq 0$ and $\limsup_{z\to\infty} \phi(z) \leq 1$. Applying the preceding lemma to the subharmonic function $\phi - 1$, we have that $\phi - 1 \leq -1$ so that $\phi \leq 0$; thus $\hat{u} \leq 0$. If $\psi \in \check{\mathcal{P}}(u, \mathbb{C}_0)$ and the preceding lemma is applied to $-\psi(1/z)$, we deduce that $\psi \geq 1$; thus $\check{u} \geq 1$. Therefore $\mathbb{C}_0$ is not solvable.

Exercises

1. Suppose h is a continuous function on $\text{cl}_\infty G$ that is harmonic on G and put $u = h|\partial_\infty G$. Show that $\hat{u} = \check{u} = h$.

2. In Proposition 7.8 show that the sequence $\{\phi_n\}$ converges to $\hat{u}$ uniformly on compact subsets of G.

3. If α is a real constant, $\phi \in \mathcal{P}(u, G)$ such that $\phi(z) \geq \alpha$ for all z in G, and $z_0 \in \partial_\infty G$ with $u(z_0) = \alpha$, then $\phi(z) \to \alpha$ as $z \to z_0$, z in G.

4. Show that if u and v are solvable functions, then so are $\max(u, v)$ and $\min(u, v)$.

5. Show that if $F = \{a_n\}$, where $a_n \to \infty$, then $G = \mathbb{C} \setminus F$ is not a solvable set.

6. Show that $\mathbb{C}$ is a solvable set.

§8 Harmonic Majorants

This short section presents a result that will be useful to us in this chapter as well as when we begin the study of Hardy spaces on arbitrary regions in the plane.

8.1 Theorem. *Let G be an arbitrary open set in the plane. If u is a subharmonic function on G that is not identically $-\infty$ on any of its components, and if there is a harmonic function g on G such that $g \geq u$ on G, then there is a unique harmonic function h on G such that:*

(a) *$h \geq u$ on G;*

(b) *if f is any harmonic function on G such that $f \geq u$, then $f \geq h$.*

Proof. Clearly it suffices to assume that G is connected. Let $\{\Gamma_n\}$ be a sequence of smooth positively oriented Jordan systems in G such that if $G_n = \text{ins}\,\Gamma_n$, then $\text{cl}\,G_n \subseteq G_{n+1}$ and $G = \cup_n G_n$. Let h_n be the solution of the Dirichlet Problem on G_n with boundary values u. Note that if $\psi \in \check{\mathcal{P}}(u, G_n)$, then $u \leq \psi$ on G_n; hence $u \leq h_n$ on G_n. For any n, h_{n+1} is continuous on $\text{cl}\,G_n$ and dominates u there. Thus $h_{n+1} \geq h_n$ on G_n. If n is fixed, this implies that $h_n \leq h_{n+1} \leq h_{n+2} \leq \ldots$ on G_n. This same type argument also shows that on $G_n, h_{n+k} \leq g$ for all $k \geq 0$, where g is the harmonic function hypothesized in the statement of the theorem. Thus Harnack's Theorem, and a little thought, shows that $h(z) \equiv \lim_n h_n(z)$ defines a harmonic function on G such that $h \leq g$. Also since $h_{n+k} \geq u$ on G_n, $h \geq u$ on G.

Now suppose f is as in condition (b). The Maximum Principle shows that for each $n \geq 1$, $f \geq h_n$ on G_n. Thus $f \geq h$.

The uniqueness of the function h is a direct consequence of part (b). □

8.2 Definition. If G is an open set and u is a subharmonic function on G that is not identically $-\infty$ on any of its components and if there is a harmonic function h on G that satisfies conditions (a) and (b) of the preceding theorem, then h is called the *least harmonic majorant* of u. In a similar way, the *greatest harmonic minorant* of a superharmonic function is defined.

So Theorem 8.1 can be rephrased as saying that a subharmonic function that has a harmonic majorant has a least harmonic majorant. Particular information about the form of the least harmonic majorant of a subharmonic function can be gleaned from the proof of Theorem 8.1 and is usefully recorded.

8.3 Corollary. *With the notation of the preceding theorem, if G is connected and $\{G_n\}$ is any sequence of Dirichlet regions such that* $\operatorname{cl} G_n \subseteq G_{n+1}$ *and* $G = \cup G_n$*, and if h_n is the solution of the Dirichlet Problem on G_n with boundary values $u|G_n$, then* $h(z) = \lim_n h_n(z)$ *is the least harmonic majorant of u.*

Exercises

1. If G is a bounded Dirichlet region and $u : \operatorname{cl} G \to [-\infty, \infty)$ is an upper semicontinuous function that is subharmonic on G and not identically $-\infty$, then u has a least harmonic majorant.

2. Fix an open set G in $\mathbb{C}$ and let $H_b(G)$ denote the real-valued bounded harmonic functions on G. For u and v in $H_b(G)$, let $u \vee v$ be the least harmonic majorant of $\max\{u, v\}$ and let $u \wedge v$ be the greatest harmonic minorant of $\min\{u, v\}$. Show that, with these definitions of join and meet, $H_b(G)$ is a Banach lattice.

§9 The Green Function

We extend the definition of a Green function that was given in §10.5.

9.1 Definition. For an open subset G of $\mathbb{C}_\infty$ a *Green function* is a function $g : G \times G \to (-\infty, \infty]$ having the following properties:

(a) for each a in G the function $g_a(z) = g(z, a)$ is positive and harmonic on $G \setminus \{a\}$;

(b) for each $a \neq \infty$ in G, $z \to g(z, a) + \log|z - a|$ is harmonic in a neighborhood of a; if $\infty \in G$, $z \to g(z, \infty) - \log|z|$ is harmonic in a neighborhood of ∞;

(c) g is the smallest function from $G \times G$ into $(-\infty, \infty]$ that satisfies properties (a) and (b).

Several observations and notes are needed before we proceed. If in part (b) $\infty \notin G$, then the function $g(z, a) + \log|z - a|$ is harmonic throughout G; the restriction to a neighborhood of a is only needed when G contains the point at infinity. Also the reason for the statement that if $\infty \in G$, $z \to g(z, \infty) - \log|z|$ is harmonic in a neighborhood of ∞ rather than throughout G is due to the possibility that $0 \in G$. After all if $0 \in G$, then $g(z, \infty)$ is harmonic near 0 and $\log|z|$ is not. See Exercise 1.

Why the minus sign in $g(z, \infty) - \log|z|$? Don't forget that to say that $g(z, \infty) - \log|z|$ is harmonic near ∞ is to say that $g(z^{-1}, \infty) - \log|z^{-1}| = g(z^{-1}, \infty) + \log|z|$ is harmonic near 0.

The next point to be made is that this definition extends Definition 10.5.1. So let G be a region in $\mathbb{C}$ (we need not consider the case that $\infty \notin G$) and let $g(z,a) = g_a(z)$ be as in Definition 10.5.1. Note that $g_a(z) \to \infty$ as $z \to a$ while $g_a(z) \to 0$ as z approaches any point on the extended boundary of G. Thus g is a positive harmonic function on $G \setminus \{a\}$. Clearly condition (b) above is also satisfied. It remains to show that g is the smallest such function. So let $f : G \times G \to (-\infty, \infty]$ be a function having properties (a) and (b) in Definition 9.1. If $f_a(z) = f(z,a)$, then $f_a(z) - g_a(z) = [f_a(z) + \log|z-a|] - [g_a(z) + \log|z-a|]$ is harmonic on G. Since f_a is positive and for any w in $\partial_\infty G$, $g_a(z) \to 0$ as $z \to w$, $\liminf_{z \to w}[f_a(z) - g_a(z)] \geq 0$. Therefore the Maximum Principle implies that $f_a(z) \geq g_a(z)$ for all z in G.

The essence of the generality that the above definition has over Definition 10.5.1 lies in its applicability to arbitrary open sets, not just Dirichlet regions. The fact that G is not assumed to be connected does not truly represent a gain in generality. The Green function for an open set is obtained by piecing together Green functions for each of the components of G. In most of the proofs only regions will be considered as this obviates the need for certain awkward phrasings.

9.2 Proposition. *Let G be a region in the extended plane.*

(a) *If there is a Green function for G, then it is unique. In fact if a is a fixed point in G and f is a positive harmonic function on $G \setminus \{a\}$ such that $f(z) + \log|z-a|$ is harmonic near a and f is the smallest such function, then $f(z) = g(z,a)$ for all z in G.*

(b) *If g is the Green function for G and $a \in G$, then $g_a(z) = g(z,a)$ for $z \neq a$ and $g_a(a) = \infty$ defines a superharmonic function on G.*

(c) *If Ω is another region in $\mathbb{C}_\infty$, $\tau : \Omega \to G$ is a conformal equivalence, and g is the Green function for G, then $h(\zeta, \alpha) = g(\tau(\zeta), \tau(\alpha))$ is the Green function for Ω.*

(d) *If G is a simply connected region in the extended plane and for each a in G, $\tau_a : G \to \mathbb{D}$ is the Riemann map with $\tau_a(a) = 0$ and $\tau_a'(0) > 0$, then $g(z,a) = -\log|\tau_a(z)|$ is the Green function for G.*

(e) *The greatest harmonic minorant of the superharmonic function g_a defined in* (b) *is the zero function.*

Proof. The point here is that to get uniqueness it is only necessary to verify that the properties hold for one fixed choice of the singularity a. If f is as described, then, as in the argument that Definition 9.1 extends Definition 10.5.1, we have that $f \geq g_a$ on G. Since f is the smallest such function, $f = g_a$.

Part (b) is an easy exercise. To prove part (c) it suffices to assume that neither G nor Ω contains ∞. If h is as in (c), then clearly h is positive and for each α in Ω, $\zeta \to h(\zeta,\alpha)$ is harmonic on $\Omega \setminus \{\alpha\}$. If $\tau(\alpha) = a$, $h(\zeta,\alpha) + \log|\zeta - \alpha|$ is the composition of $g_a + \log|z-a|$ with τ and therefore must be harmonic on Ω. It remains to verify condition (c) of the definition. But if f is another function on $\Omega \times \Omega$ satisfying conditions (a) and (b), $f \circ (\tau^{-1} \times \tau^{-1})$ is such a function on $G \times G$. The fact that g is the smallest such function on $G \times G$ implies that h is the smallest such function on $\Omega \times \Omega$.

We note that $-\log[|z-\alpha|/|1-\overline{\alpha}z|]$ is the Green function for $\mathbb{D}$. If we fix a point a in G, then for any point b in G, $\tau_b(z) = c[\tau_a(z) - \tau_a(b)]/[1 - \overline{\tau_a(b)}\tau_a(z)]$ for some constant c (dependent on b) with $|c| = 1$. With these observations, part (d) follows from part (c).

To prove (e) first note that 0 is a harmonic minorant of g_a. Let h_a be the greatest harmonic minorant of g_a; so $0 \le h_a \le g_a$. Thus $g_a - h_a$ is a positive superharmonic function on G that is harmonic on $G \setminus \{a\}$ and such that $(g_a - h_a) + \log|z-a|$ is harmonic on G. From the definition of the Green function we have that $g_a - h_a \ge g_a$, so that $h_a \le 0$. □

Part (d) of the preceding proposition can be used to compute the Green function for a simply connected region provided we know the conformal equivalence. See Exercise 2.

Of course the Green function may not exist for a specific region G. This question of existence is crucial and is, as we shall see, intricately connected to other questions about harmonic functions. Note that if there is a Green function g for G, then there are non-constant negative subharmonic functions on G; viz, $-g_a$ for each a in G. This leads us to the following classification of regions.

9.3 Definition. An open set G in $\mathbb{C}_\infty$ is *hyperbolic* if there is a subharmonic function on G that is bounded above and not constant on any component of G. Otherwise G is called *parabolic.*

The comments preceding the definition show that if G has a Green function, then G is hyperbolic. By subtracting a constant function, we see that to say that G is hyperbolic is equivalent to the statement that there is a negative subharmonic function on G that is not constant on any component of G. Every bounded region is hyperbolic since $\operatorname{Re} z$ is a bounded subharmonic function. Since any region that is conformally equivalent to a hyperbolic region is clearly hyperbolic, this says that any region that is not dense in the extended plane is hyperbolic. Thus the only candidates for a parabolic region are those of the form $\mathbb{C}_\infty \setminus F$, where F is some closed subset of $\mathbb{C}_\infty$.

9.4 Proposition. *If G is an open set that is not connected, then G is hyperbolic. An open subset of a hyperbolic set is hyperbolic.*

Proof. If G is not connected, then no component of G is dense. Hence on each component we can find a non-constant negative subharmonic function. The proof of the second statement is trivial. □

9.5 Proposition.

(a) *If G is a solvable region and ∂G is not a singleton, then G is hyperbolic.*

(b) *If G is parabolic, then $K = \mathbb{C}_\infty \setminus G$ is totally disconnected and $K = \partial K = \hat{K}$.*

(c) *$\mathbb{C}$ is a parabolic region.*

Proof. (a) Assume that $\infty \notin G$. Since G is not the whole plane, $\partial_\infty G$ has at least 2 points (7.20). Let $u : \partial_\infty G \to [0,1]$ be a continuous function that assumes both the values 0 and 1. If $\phi \in \hat{\mathcal{P}}(u, G)$, then $\phi - 1$ is a negative subharmonic function. If G were parabolic, then it would follow that each function in $\hat{\mathcal{P}}(u, G)$ is a constant and that $\hat{u} \equiv 0$. Similarly, if G were parabolic, it would follow that $\check{u} \equiv 1$ and so G is not solvable.

(b) Assume that $\infty \notin G$. If $K \neq \partial K$, then G is not dense and therefore solvable; by (a) G cannot be parabolic. Similarly, if $K \neq \hat{K}$, G is not connected and is thus hyperbolic (9.4). If K is not totally disconnected, it contains a component X that has infinitely many points. Put $H = \mathbb{C}_\infty \setminus X$. Let $\tau : H \to \mathbb{D}$ be the Riemann map with $\tau(\infty) = 0$ and put $\phi = \log|\tau|$. So ϕ is a subharmonic function on H with $\phi \leq 0$ and H must be hyperbolic. Therefore G is hyperbolic, a contradiction.

(c) Let ϕ be a negative finite-valued subharmonic function on $\mathbb{C}$ and fix two points a and b. Since $\limsup_{z\to a} \phi(z) \leq \phi(a)$, Lemma 7.19 implies that $\phi(a) \geq \phi(z)$ for all z in $\mathbb{C}$; in particular, $\phi(a) \geq \phi(b)$. Reversing the roles of a and b, we get that $\phi(a) = \phi(b)$. Therefore ϕ is constant. □

In Theorem 10.7 in the next section it will be shown that a hyperbolic region is solvable, thus establishing the converse to part (a) of the preceding proposition. But now we concentrate on the connection of hyperbolic regions to the existence of the Green function.

9.6 Proposition. *Let G be a region in $\mathbb{C}_\infty$ and let γ be a Jordan curve in G such that $H = \operatorname{ins}\gamma \subseteq G$. Let u denote the characteristic function of γ defined on $\partial_\infty(G \setminus \operatorname{cl} H)$ with $\hat{u}$ the corresponding Perron function on $G \setminus \operatorname{cl} H$.*

(a) *$\hat{u}(z) < 1$ for all z in $G \setminus \operatorname{cl} H$ if and only if G is hyperbolic.*

(b) *$\hat{u}(z) \equiv 1$ on $G \setminus \operatorname{cl} H$ if and only if G is parabolic.*

Proof. By the Maximum Principle $\hat{u}(z) \leq 1$ on $\Omega \equiv G \setminus \operatorname{cl} H$. Also Ω is connected (why?) and so if $\hat{u}(z) = 1$ for any point in Ω, it is identically 1. Thus the statements (a) and (b) are equivalent. So assume that G is hyperbolic and let us show that $\hat{u}(z) < 1$ on Ω. Let ϕ be a negative subharmonic function on G that is not constant. Without loss of generality we may assume that $\sup\{\phi(z) : z \in G\} = 0$. Put $M = \max\{\phi(z) : z \in \gamma\}$; so $M \leq 0$. By the Maximum Principle $M < 0$. By normalization, it may be assumed that $M = -1$.

By Theorems 10.4.3 and 10.4.9, $\hat{u}(z) \to 1$ as z approaches any point on γ. Since u is positive and $\phi(z) \leq -1$ on γ, $-\phi \in \check{\mathcal{P}}(u, \Omega)$; thus $\hat{u} \leq -\phi$ on Ω. Since $\sup_G \phi(z) = 0$, there is a point a in G with $\phi(a) > -1$. But $\phi(z) \leq -1$ for all z in cl H (why?) and so $a \in \Omega$. Thus $\hat{u}(a) < 1$ and so $\hat{u}(z) < 1$ for all z in Ω.

Now assume that $\hat{u}(z) < 1$ for all z in Ω and define $\phi : G \to \mathbb{R}$ by letting $\phi(z) = -\hat{u}(z)$ for z in Ω and $\phi(z) = -1$ for z in cl H. It is left to the reader to verify that ϕ is subharmonic on G. Clearly it is negative and, by assumption, it is not constant. $\square$

Let us underscore the fact used in the preceding proof that the function $\hat{u}$ has a continuous extension to γ where it is constantly 1.

9.7 Theorem. *The region G in $\mathbb{C}_\infty$ has a Green function if and only if G is hyperbolic.*

Proof. We have already observed that a necessary condition for G to have a Green function is that G be hyperbolic; so assume that G is hyperbolic. Fix a in G, $a \neq \infty$; we will produce the Green function with singularity at a. (The case where $a = \infty$ is left to the reader.)

Define u on $\partial_\infty[G \setminus \{a\}] = \partial_\infty G \cup \{a\}$ by $u(a) = \infty$ and $u(\zeta) = 0$ for ζ in $\partial_\infty G$. Let g be the upper Perron function $\check{u}$ on $G \setminus \{a\}$ corresponding to the boundary function u. Clearly g is a non-negative harmonic function on $G \setminus \{a\}$. To see that g satisfies condition (b) of Definition 9.1 takes a bit of cleverness.

Choose $r > 0$ such that $B = \overline{B}(a;r) \subseteq G$ and let $R > r$ such that $\overline{B}(a;R) \subseteq G$. Observe that $G \setminus B$ is a solvable set, let χ be the characteristic function of the circle ∂B, and put $h = \hat{\chi}$, the Perron function of χ on $G \setminus B$. So $0 \leq h \leq 1$; according to the preceding proposition, $h(z) < 1$ for all z in $G \setminus B$. Thus $m = \max\{h(z) : |z - a| = R\} < 1$. Choose a positive constant M with $mM < M + \log(r/R)$ and define a function ψ_0 on G by

$$\psi_0(z) = \begin{cases} M\,h(z) & z \notin G \setminus B \\ M + \log \frac{r}{|z-a|} & z \in B. \end{cases}$$

Claim. ψ_0 is a superharmonic function on G.

It is clear that ψ_0 is lower semicontinuous; in fact it is a continuous function from G into $\mathbb{R}_\infty$. The only place we have to check the integral condition is when $|z - a| = r$. To check this we first verify that $M\,h(w) <$

$M+\log[r/|w-a|]$ for $r < |w-a| < R$. Indeed, if $|w-a| = r$, $h(w) = 1$ and so $Mh(w) = M = M + \log[r/|w-a|]$; if $|w-a| = R$, $Mh(w) \leq Mm < M + \log[r/|w-a|]$. The desired inequality now holds by the Maximum Principle. Now if $|z-a| = r$ and $\delta < R - r$, this implies that

$$\frac{1}{2\pi}\int_0^{2\pi} \psi_0(z+\delta e^{i\theta})\,d\theta \leq M + \log\frac{r}{|z-a|} = \psi_0(z).$$

This proves the claim.

Note that the function ψ_0 belongs to the Perron family $\check{\mathcal{P}}(u, G\setminus\{a\})$ and so the harmonic function g satisfies $g(z) \leq \psi_0(z)$ for all z in $G\setminus\{a\}$. If $z \in B$ and $z \neq a$, $g(z) + \log|z-a| \leq \psi_0(z) + \log|z-a| = M + \log r$, a constant. Therefore $g(z) + \log|z-a|$ admits a harmonic extension to $B(a;r)$. This says that g satisfies conditions (a) and (b) of Definition 9.1.

Now to verify condition (c) of the definition. Assume that k is another function having properties (a) and (b). So k is a superharmonic function on G. Now $\lim_{z\to a} k(z) = \infty = u(a)$. Since k is positive, $\liminf_{z\to\zeta} k(z) \geq 0 = u(\zeta)$ for ζ in $\partial_\infty G$. Therefore $k \in \check{\mathcal{P}}(u, G\setminus\{a\})$ and so $k \geq g$. Hence g is the Green function for G with singularity at a. □

The next result can be viewed as a computational device, but in reality it is only a theoretical aid.

9.8 Proposition. *Let $\{G_n\}$ be a sequence of open sets such that $G_n \subseteq G_{n+1}$ and $G = \cup_n G_n$ is hyperbolic. If g_n is the Green function for G_n and g is the Green function for G, then for each a in G, $g_n(z,a) \uparrow g(z,a)$ uniformly on compact subsets of $G\setminus\{a\}$.*

Proof. Fix the point a in G; we will only consider those n for which $a \in G_n$. Note that the restriction of g_{n+1} to G_n satisfies conditions (a) and (b) of Definition 9.1. Hence $g_{n+1}(z,a) \geq g_n(z,a)$ for all z in G_n. Thus $k(z,a) = \lim_n g_n(z,a)$ exists uniformly on compact subsets of $G\setminus\{a\}$ and either k_a is harmonic or $k_a \equiv \infty$. But the same type of argument shows that $g(z,a) \geq g_n(z,a)$ for all n and all z in G_n. Thus $g(z,a) \geq k(z,a)$ and so k_a is a harmonic function on $G\setminus\{a\}$. Clearly $k \geq 0$ and $k(z,a) + \log|z-a|$ is harmonic near a. That is, k satisfies properties (a) and (b) in the definition of a Green function. Therefore, $k(z,a) \geq g(z,a)$. Since we have already shown that $g \geq k$, we get that $g = k$, proving the proposition. □

Recall the information obtained about the Green function for an analytic Jordan region in §15.2. In particular, Theorem 15.2.5 gives a formula for the solution of the Dirichlet problem for such regions involving the conjugate differential of the Green function. We can use this material together with the preceding proposition to obtain information about the Green function for any hyperbolic region.

9.9 Theorem. *If G is a hyperbolic region, a and b are points in G, and g*

is the Green function for G, then $g(a,b) = g(b,a)$.

Proof. First assume that G is an analytic Jordan region with positively oriented boundary $\Gamma = \{\gamma_0, \dots, \gamma_n\}$, with γ_0 as the outer boundary. Let $a_1 = a$ and $a_2 = b$. For $j = 1, 2$, put $g_j(z) = g(z, a_j)$ and $\eta_j = \partial B(a_j; r)$ with positive orientation. (Here r is chosen so that $\overline{B}(a_j; r) \subseteq G$ and $\eta_1 \cap \eta_2 = \emptyset$.) So $\Lambda = \{\gamma_0, \dots, \gamma_n, -\eta_1, -\eta_2\}$ is a positively oriented Jordan system and both g_1 and g_2 are harmonic on the inside of Λ. By Corollary 15.2.3, $0 = \int_\Lambda (g_1^* dg_2 - g_2^* dg_1)$. Since g_1 and g_2 vanish on Γ, Lemma 15.2.4 implies that

$$\begin{aligned} 0 &= -\int_{\eta_1+\eta_2} (g_1^* dg_2 - g_2^* dg_1) \\ &= -2\pi\, g_2(a_1) + 2\pi\, g_1(a_2) + \beta(r), \end{aligned}$$

where $\beta(r) \to 0$ as $r \to 0$.

This proves the theorem for the case of an analytic Jordan region. The proof of the general case follows by taking a sequence $\{G_n\}$ of analytic Jordan regions such that $G_n \subseteq G_{n+1}$ and $G = \cup_n G_n$ and applying Proposition 9.8. □

We finish with a result that will prove useful in further developments.

9.10 Proposition. *If G is a hyperbolic region in $\mathbb{C}$, $a \in G$, and g is the Green function for G, then*

$$\begin{aligned} g(z,a) &= \inf\{\psi(z) : \psi \text{ is a positive superharmonic function on } G \text{ such} \\ &\qquad \text{that } \psi(z) + \log|z-a| \text{ is superharmonic near } a\}. \end{aligned}$$

Proof. Let ψ be a positive superharmonic function on G such that $v(z) \equiv \psi(z) + \log|z-a|$ is superharmonic on G. Since $z \to g(z,a)$ is a superharmonic function having these properties, it suffices to show that $\psi(z) \geq g_a(z) = g(z,a)$ for all z in G. To this end, first note that $v(a) \leq \liminf_{z\to a} [\psi(z) + \log|z-a|]$ and so $\psi(z) \to \infty$ as $z \to a$. Thus $\liminf_{z\to a} \psi(z) = \infty = \lim_{z\to a} g_a(z)$.

Now let $\{G_n\}$ be a sequence of bounded Dirichlet regions, each of which contains the point a and such that $G_n \subseteq G_{n+1}$ and $G = \cup_n G_n$. Let g_n be the Green function for G_n with pole at a. According to Proposition 9.8, $g_n(z) \to g_a(z)$ uniformly on compact subsets of $G \setminus \{a\}$. If $\psi_n = \psi|G_n$, then for any ζ in ∂G_n, $\liminf_{z\to\zeta} \psi_n(z) \geq 0 = \limsup_{z\to\zeta} g_n(z)$. By the Maximum Principle, $\psi_n \geq g_n$ on G_n and the proposition follows. □

Additional results on the Green function can be found in Theorem 21.1.19 and through the remainder of Chapter 21.

Exercises

1. If $\infty \in G$ and $a \notin G$, then $g(z, \infty) - \log|z - a|$ is harmonic in G.

2. Use Proposition 9.2 to find the Green function for $\mathbb{D}$ as well as the upper half plane.

3. If G is a hyperbolic region in $\mathbb{C}_\infty$ and $\infty \in G$, show that $g(z, \infty) = \inf\{\psi(z) : \psi$ is a positive superharmonic function on G such that $\psi(z) - \log|z|$ is superharmonic near $\infty\}$.

4. Let $p(z) = z^n + a_1 z^{n-1} + \ldots + a_n$ and put $G = \mathbb{C}_\infty \setminus \{z \in \mathbb{C} : |p(z)| \leq R\}$. Show that G is connected and $g(z, \infty) = n^{-1}[\log|p(z)| - \log R]$.

§10 Regular Points for the Dirichlet Problem

Here we return to some of the ideas connected with the solution of the Dirichlet problem in the classical sense. Specifically, we will study the boundary behavior of the solution of the Dirichlet problem and how this behavior collates with the actual values of the boundary function. In §10.4 the concept of a barrier was introduced and it was shown that each point of $\partial_\infty G$ has a barrier if and only if G is a Dirichlet region. We begin by examining more closely the idea of a barrier; in the process some of the results from §10.4 will be reproved.

Recall the notation that for a in $\partial_\infty G$ and $r > 0$, $G(a; r) = B(a; r) \cap G$.

10.1 Theorem. *If G is an open subset of $\mathbb{C}_\infty$ and $a \in \partial_\infty G$, then the following are equivalent.*

(a) *There is a barrier for G at a.*

(b) *There is a negative subharmonic function ϕ on G such that $\phi(z) \to 0$ as $z \to a$ and for every open neighborhood U of a, $\sup\{\phi(z) : z \in G \setminus U\} < 0$.*

(c) *For every $r > 0$ there is a negative non-constant subharmonic function ϕ on $G(a; r)$ such that $\phi(z) \to 0$ as $z \to a$ in $G(a; r)$.*

(d) *For every $r > 0$, if u is the solution of the Dirichlet problem on $G(a; r)$ with boundary values $\zeta \to d(a, \zeta)$ (d is the metric), then $u(z) \to 0$ as $z \to a$ in $G(a; r)$.*

(e) *There is an $r > 0$ and a negative non-constant subharmonic function ϕ on $G(a; r)$ such that $\phi(z) \to 0$ as $z \to a$ in $G(a; r)$.*

(f) *There is an $r > 0$ such that if u is the solution of the Dirichlet problem on $G(a; r)$ with boundary values $\zeta \to d(a, \zeta)$ (d is the metric), then $u(z) \to 0$ as $z \to a$ in $G(a; r)$.*

Note: if $a \neq \infty$, then the metric in (d) and (f) can be taken to be the usual absolute value; otherwise it is the metric of the extended plane.

Proof. In this proof it will be assumed that a is a finite boundary point of G; the proof of the case where $a = \infty$ is left to the reader.

(a) *implies* (b). Let $\{\psi_r\}$ be a barrier for G at a and let $\{r_n\}$ be any sequence of radii converging to 0. Extend ψ_r to G by letting it be identically 1 on $G \setminus G(a;r)$. Put $\psi_n = \psi_{r_n}$ and $\phi = \sum_n (-2^{-n})\psi_n$. By (4.6.e) ϕ is subharmonic and clearly $0 \geq \phi \geq -1$. Since $\psi_n(z) \to 0$ as $z \to a$ for each n, it is easy to check that $\phi(z) \to 0$ as $z \to a$. Finally, if $r > 0$ and n is chosen so that $r_n < r$, then for z in $G \setminus B(a;r)$, $\phi(z) \leq (-2^{-n})\psi_n(z) = -2^{-n}$.

(b) *implies* (c). This is trivial.

(c) *implies* (d). Let $0 < \varepsilon < r^2$; put $B_\varepsilon = B(a;\varepsilon)$. If $G \cap \partial B_\varepsilon \neq \emptyset$, write $G \cap \partial B_\varepsilon$ as the disjoint union of a non-empty compact set K_ε and a relatively open set U whose arc length as a subset of the circle is less than $2\pi\varepsilon/r$. Let g be the solution of the Dirichlet problem on the disk B_ε with boundary values $r\chi_U$. Observe that $0 \leq g \leq r$ and g is continuous at the points of U with $g(\zeta) = r$ there. Writing g as the Poisson integral on the disk B_ε it is easy to see that $g(a) < \varepsilon$. If $G \cap \partial B_\varepsilon = \emptyset$, let $g \equiv 0$.

Let ϕ be as in the statement of (c) and define $\theta : G(a;\varepsilon) \to \mathbb{R}$ by $\theta = r(\phi/\alpha) + g + \varepsilon$, where $\alpha = \max\{\phi(z) : z \in K_\varepsilon\}$. Since ϕ is negative and usc, $\alpha < 0$. Note that θ is superharmonic on $G(a;\varepsilon)$ and $\theta \geq \varepsilon$ there. Now let η be any function in $\hat{\mathcal{P}}(|\zeta - a|,\ G(a;r))$.

Claim. For any ζ in $\partial G(a;\varepsilon)$, $\limsup_{z\to\zeta} \eta(z) \leq \liminf_{z\to\zeta} \theta(z)$.

In fact, $\partial G(a;\varepsilon) = (B_\varepsilon \cap \partial G) \cup (G \cap \partial B_\varepsilon)$. If $\zeta \in B_\varepsilon \cap \partial G$, then $\limsup_{z\to\zeta} \eta(z) \leq |\zeta - a| < \varepsilon \leq \liminf_{z\to\zeta} \theta(z)$. If $\zeta \in G \cap \partial B_\varepsilon$, then $\limsup_{z\to\zeta} \eta(z) \leq r$ and $\liminf_{z\to\zeta} \theta(z) \geq r \liminf_{z\to\zeta}(\phi(z)/\alpha) + \liminf_{z\to\zeta} g(z) + \varepsilon$. If $\zeta \in K_\varepsilon$, then this gives that $\liminf_{z\to\zeta} \theta(z) \geq r + \liminf_{z\to\zeta} g(z) + \varepsilon \geq r$. If $\zeta \in U$, then we get that $\liminf_{z\to\zeta} \theta(z) \geq 0 + \liminf_{z\to\zeta} g(z) + \varepsilon = r + \varepsilon \geq r$. This proves the claim.

The Maximum Principle implies $\eta \leq \theta$ on $G(a;\varepsilon)$; so if u is as in (d), then $u \leq \theta$ there. Since $\phi(z) \to 0$ as $z \to a$, $\limsup_{z\to a} \theta(z) = \lim_{z\to a} \theta(z) = g(a) + \varepsilon < 2\varepsilon$. Thus $0 \leq \limsup_{z\to a} u(z) < 2\varepsilon$. Since ε was arbitrary, $u(z) \to 0$ as $z \to a$.

(d) *implies* (f). This is trivial.

(f) *implies* (a). Let r and u be as in the statement of (f). Note that $u(z) = r$ when $z \in G \cap \partial B_r$, so that if $\psi_r = r^{-1}u$ on $G(a;r)$ and $\psi_r \equiv 1$ on $G \setminus B(a;r)$, then ψ_r is superharmonic on G, $0 \leq \psi_r \leq 1$, $\psi_r(z) \to 0$ as $z \to a$, and $\psi_r(z) \to 1$ as $z \to \zeta$ on $G \cap \partial B_r$.

Now let $0 < s < r$ and let v be the solution of the Dirichlet problem on $G(a;s)$ with boundary values $|\zeta - a|$. Since $|z - a|$ is a subharmonic function, $|z - a| \in \hat{\mathcal{P}}(|\zeta - a|,\ G(a;r))$; therefore $|z - a| \leq u(z)$ on $G(a;r)$. Thus $u \in \check{\mathcal{P}}(|\zeta - a|,\ G(a;s))$ and so $v(z) \leq u(z)$ for all z in $G(a;s)$. But this implies that $v(z) \to 0$ as $z \to a$. But s was arbitrary, so if the argument of the preceding paragraph is applied to the function v, we get a barrier $\{\psi_s : 0 < s \leq r\}$.

It is trivial that (c) implies (e) and the proof that (c) implies (d) given above shows that (e) implies (f). □

Note that the existence of a barrier for G at any of its boundary points implies the existence of a non-constant negative subharmonic function on G. Thus a region is hyperbolic if there is a barrier at one of its boundary points.

In the next lemma it is important to emphasize that the set G is not assumed to be solvable.

10.2 Lemma. *Assume there is a barrier for G at a and f is a real-valued function defined on $\partial_\infty G$.*

(a) *If f is bounded below, then*

$$\liminf_{z\to a} \hat{f}(z) \geq \min\left\{ \liminf_{\substack{\zeta\to a\\ \zeta\in\partial_\infty G}} f(\zeta), f(a) \right\}.$$

(b) *If f is bounded above, then*

$$\limsup_{z\to a} \check{f}(z) \leq \max\left\{ \limsup_{\substack{\zeta\to a\\ \zeta\in\partial_\infty G}} f(\zeta), f(a) \right\}.$$

(The reason for taking the minimum and maximum in parts (a) and (b), respectively, is that the definition of the limits inferior and superior of f as z approaches a does not use the value of f at a.)

Proof. It suffices to prove only part (a). Let ρ be the right hand side of the inequality in (a) and let m be a constant such that $\inf\{f(\zeta) : \zeta \in \partial_\infty G\} > m$. We will only consider the case that $a \neq \infty$. If $\rho > \beta > m$, let $r > 0$ such that $f(\zeta) > \beta$ for ζ in $\partial G \cap \overline{B}_r$, $B_r = B(a; r)$. Put $u =$ the solution of the Dirichlet problem on $G(a; r)$ with boundary values $|\zeta - a|$; so $u(z) \to 0$ as $z \to a$ by the preceding theorem. If u is extended to all of G by setting $u = r$ on $G \setminus B_r$, then u is superharmonic (verify!). It is easy to check that $\beta - r^{-1}(\beta - m)u$ belongs to the lower Perron family for (f, G). Thus $\hat{f} \geq \beta - r^{-1}(\beta - m)u$ and so

$$\liminf_{z\to a} \hat{f}(z) \geq \liminf_{z\to a} \left[\beta - \frac{\beta - m}{r} u\right] = \beta.$$

Since β was an arbitrary number smaller than ρ, we have proved the lemma. □

10.3 Definition. Say that a point a on the extended boundary of an open set G is a *regular boundary point* for G if for every bounded function $f : \partial_\infty G \to \mathbb{R}$ that is continuous at a,

$$\lim_{z\to a} \hat{f}(z) = f(a) = \lim_{z\to a} \check{f}(z).$$

So the open set G is a Dirichlet set if and only if every extended boundary point of G is a regular boundary point. The next proposition was essentially proved in Theorem 10.4.3. Half the proof was given there; the other half comes from the preceding lemma.

10.4 Proposition. *There is a barrier for G at the point a in $\partial_\infty G$ if and only if a is a regular boundary point of G.*

The next small result is an emphasis of the fact that the condition that a point a is a regular point for G is a local property. That is, this condition is only affected by the disposition of G near a. This is clear from Theorem 10.1.

10.5 Proposition. *If a is a regular point for the open set G and H is an open subset of G such that $a \in \partial_\infty H$, then a is a regular point for H.*

10.6 Lemma. *If G is a region and there is a negative subharmonic function ϕ on G whose least harmonic majorant is 0, then there is a negative subharmonic function ϕ_0 on G that is not identically $-\infty$ such that if $\zeta \in \partial_\infty G$ and $\limsup_{z\to\zeta} \phi(z) < 0$, then $\phi_0(z) \to -\infty$ as $z \to \zeta$.*

Proof. Let $\{G_n\}$ be a sequence of smooth Jordan regions with $\operatorname{cl} G_n \subseteq G_{n+1}$ and $G = \cup_n G_n$. For each n define the function ϕ_n on G by $\phi_n = \phi$ on $G \setminus G_n$ and $\phi_n =$ the solution of the Dirichlet problem on G_n with boundary values $\phi|\partial G_n$. According to Corollary 8.3 and the hypothesis, $\phi_n(z) \to 0$ as $n \to \infty$. Fix a point a in G; passing to a subsequence if necessary, it can be assumed that $\sum_n |\phi_n(a)| < \infty$. By Proposition 4.6.e, $\phi_0(z) \equiv \sum_n \phi_n(z)$ is a negative subharmonic function; by construction it is not identically $-\infty$.

Now fix a point ζ in $\partial_\infty G$ and assume that $\limsup_{z\to\zeta} \phi(z) < 0$. Let $r > 0$ such that $\phi(z) \le -\delta < 0$ for z in $G(\zeta; r)$. Let n be an arbitrary but fixed positive integer. There is an $r_n < r$ such that $B(\zeta; r_n) \cap \operatorname{cl} G_n = \emptyset$. If $z \in G(\zeta; r_n)$, then for $1 \le m \le n$, $\phi_m(z) = \phi(z) \le -\delta$. Thus $\phi_0(z) \le \sum_{m=1}^n \phi_m(z) \le -n\delta$ for z in $G(\zeta; r_n)$. Since n was arbitrary, $\phi_0(z) \to -\infty$ as $z \to \zeta$. □

Now for the promised converse to Proposition 9.5.

10.7 Theorem. *If G is a region in $\mathbb{C}_\infty$ such that $\partial_\infty G$ has at least two points, then G is solvable if and only if it is hyperbolic.*

Proof. It suffices to assume that G is hyperbolic and prove that it is solvable. Since G is hyperbolic, it has a Green function g; let $\phi(z) = -g(z,a)$ for some a in G. So ϕ is a negative subharmonic function with 0 as its least harmonic majorant (9.2). Let ϕ_0 be the negative subharmonic function as in the preceding lemma.

To show that G is solvable, let $f : \partial_\infty G \to \mathbb{R}$ be a continuous function and fix an arbitrary positive ε.

Claim. If $\zeta \in \partial_\infty G$, $\lim_{z\to\zeta}\sup[\check{f}(z) + \varepsilon\phi_0(z)] \le f(\zeta)$.

If $\phi(z) \to 0$ as $z \to \zeta$, then ζ is a regular boundary point (10.1) and so $\check{f}(z) \to f(\zeta)$ as $z \to \zeta$. Since $\phi_0 \le 0$, this proves the claim in this case. In the other case, $\limsup_{z\to\zeta}\phi(z) < 0$ and so the lemma implies that $\phi_0(z) \to -\infty$ as $z \to \zeta$. This proves the claim in this case also.

But the claim implies that $\check{f} + \varepsilon\phi_0 \in \hat{\mathcal{P}}(f,G)$ and so we have that $\check{f} + \varepsilon\phi_0 \le \hat{f} \le \check{f}$. Since ε was arbitrary, $\hat{f} = \check{f}$ and so G is solvable. $\square$

From this point on the term "solvable set" will be dropped in favor of "hyperbolic set" since this is standard in the literature.

Now let's turn our attention to a consideration of the irregular points of an open set. Recall that at the end of §10.3 it was shown that the origin is not a regular point of the punctured disk. That same proof can be used to prove the following.

10.8 Proposition. *If G is an open set and a is an isolated point of $\partial_\infty G$, then a is an irregular point of G.*

10.9 Proposition. *If G is an open set and $a \in \partial_\infty G$, then a is a regular point if and only if for every component H of G either a is a regular point of H or $a \notin \mathrm{cl}_\infty H$.*

Proof. Assume a is a regular point of G; let $r > 0$ and let ϕ be a negative non-constant subharmonic function on $G(a;r)$ such that $\phi(z) \to 0$ as $z \to a$. If H is a component such that $a \in \mathrm{cl}_\infty H$, then $a \in \partial_\infty H$ and $\phi_1 = \phi | H(a;r)$ is a non-constant negative subharmonic function such that $\phi_1(z) \to 0$ as $z \to a$.

For the converse, it only makes sense to assume that G is not connected. So each component of G is a hyperbolic region. Let $\{H_n\}$ be the components of G and for each n let h_n be defined on $\partial_\infty H_n$ by $h_n(\zeta) = \min\{d(\zeta,a), 1/n\}$. Let u_n be the solution of the Dirichlet problem on H_n with boundary values h_n. Note that if $a \in \partial_\infty H_n$, then a is a regular point of H_n by assumption and so $u_n(z) \to 0$ as $z \to a$. Define u on G by letting $u = -u_n$ on H_n. So u is a non-constant negative harmonic function.

Claim. $\lim_{z\to a} u(z) = 0$.

To see this, let $\varepsilon > 0$ and choose $n_0 > \varepsilon^{-1}$. For $1 \le n \le n_0$, let $r_n > 0$ such that either $B(a; r_n) \cap \text{cl}_\infty H_n = \emptyset$ or $u_n(z) < \varepsilon$ for z in $H_n(a; r_n)$. Thus $|u(z)| < \varepsilon$ for z in $G(a; r)$, where $r = \min\{r_n : 1 \le n \le n_0\}$. By Theorem 10.1, a is a regular point of G. □

10.10 Corollary. *If $\{H_n\}$ are the components of G, then every point in $\partial_\infty G \setminus \cup_n \partial_\infty H_n$ is a regular point.*

So in a search for irregular points it suffices to examine the boundary points of the components of G. The next proposition says that in such a search we can ignore those points that belong to the boundary of more than one component.

10.11 Proposition. *If $a \in \partial_\infty G$ and a is an irregular point, then there is a unique component H of G such that $a \in \partial_\infty H$.*

Proof. Assume a is an irregular point and let $\{H_k\}$ be the collection of those components of G such that a belongs to their boundary. The preceding corollary says that this is a non-empty collection. Suppose there is more than one such component. If k is arbitrary and $j \ne k$, then $a \in \text{cl}_\infty H_j \subseteq \mathbb{C}_\infty \setminus H_k$ and $\text{cl}_\infty H_j$ is connected. By Theorem 10.4.9 this implies a is a regular point of H_k for each $k \ge 1$. By Proposition 10.9, a is a regular point of G, a contradiction. □

10.12 Corollary. *If $a \in \partial_\infty G$ and there is a component H of G such that a is a regular point of H, then a is a regular point of G.*

Proof. If a were an irregular point of G, then the preceding proposition would imply that H is the unique component of G that has a as a boundary point. Thus the assumption that a is regular for H contradicts the assumption that it is irregular for G. □

Later (§21.6) we will return to an examination of irregular points.

Exercises

1. Let $G = \mathbb{D} \setminus [\{0\} \cup \bigcup_{n=1}^\infty D_n]$, where $D_n = \overline{B}(a_n; r_n)$ with $0 < a_n < 1$ and the radii r_n chosen so that $r_n < a_n$ and the disks $\{D_n\}$ are pairwise disjoint. Show that G is a Dirichlet region.

2. Let $G = \mathbb{D} \setminus \bigcup_{n=1}^\infty D_n$, where $\{D_n\}$ is a sequence of pairwise disjoint closed disks such that the accumulation points of their centers is precisely the unit circle, $\partial\mathbb{D}$. Show that G is a Dirichlet region.

§11 The Dirichlet Principle and Sobolev Spaces

In this section a classical approach to the Dirichlet problem is explored. The idea is that if G is a region and if f is a function on ∂G, then the solution of the Dirichlet problem with boundary values f is the function u on G that "equals" f on the boundary and minimizes the integral $\int |\nabla u|^2 \, d\mathcal{A}$. This will be made precise and proven as the section develops.

We begin by defining a Sobolev space that is suitable for our needs. We will only scratch the surface of the theory of Sobolev spaces; indeed, we will use little of this subject other than some of the elementary language. A fuller introduction can be found in Adams [1975] and Evans and Gariepy [1992].

Recall that C_c^∞ is the collection of infinitely differentiable functions on $\mathbb{C}$ that have compact support.

11.1 Definition. If $\phi,\ \psi \in C_c^\infty$, define

$$\langle \phi, \psi \rangle = \int \phi \overline{\psi} \, d\mathcal{A} + \int \partial\phi \overline{\partial\psi} \, d\mathcal{A} + \int \overline{\partial}\phi \overline{\overline{\partial}\psi} \, d\mathcal{A}$$

and let $W_1^2 = W_1^2(\mathbb{C})$ be the completion of C_c^∞ with respect to the norm defined by this inner product.

Of course it must be shown that $\langle \cdot, \cdot \rangle$ defines an inner product on C_c^∞, but this is a routine exercise for the reader to execute. (Recall that $\overline{\partial \psi} = \overline{\partial}\, \overline{\psi}$ and $\partial \overline{\psi} = \overline{\overline{\partial} \psi}$). To record the notation, note that

$$||\phi||^2 = ||\phi||_{W_1^2}^2 = \int [|\phi|^2 + |\partial\phi|^2 + |\overline{\partial}\phi|^2] \, d\mathcal{A}.$$

The space W_1^2 is an example of a Sobolev space. The superscript 2 in W_1^2 is there because we have used an L^2 type norm. The subscript 1 refers to the fact that only one derivative of the functions is used in defining the norm. The first task is to get an internal characterization of the functions that belong to W_1^2. The term *weak derivative* of a function u will mean the derivative of u in the sense of distributions.

11.2 Theorem. *A function u belongs to W_1^2 if and only if $u \in L^2$ and each of the weak derivatives ∂u and $\overline{\partial} u$ is a function in L^2.*

Proof. First assume that $u \in W_1^2$ and let $\{\phi_n\}$ be a sequence in C_c^∞ such that $||u - \phi_n|| \to 0$ as $n \to \infty$. This implies that $\phi_n \to u$ in L^2 and that $\{\partial \phi_n\}$ and $\{\overline{\partial} \phi_n\}$ are Cauchy sequences in L^2. Let u_1 and $u_2 \in L^2$ such that $\partial \phi_n \to u_1$ and $\overline{\partial} \phi_n \to u_2$ in the L^2 norm. If $\psi \in C_c^\infty$, then $\int u_1 \, \psi \, d\mathcal{A} = \lim_n \int (\partial \phi_n) \, \psi \, d\mathcal{A} = - \lim_n \int \phi_n \partial \psi \, d\mathcal{A} = - \int u \, \partial \psi \, d\mathcal{A}$. Hence $u_1 = \partial u$, the weak derivative of u. Similarly $u_2 = \overline{\partial} u$.

Conversely, suppose that u, ∂u, and $\bar{\partial} u$ belong to L^2. For each $n \geq 1$, let $f_n \in C_c^\infty$ such that $0 \leq f_n \leq 1$, $f_n(z) = 1$ for $|z| \leq n$, $f_n(z) = 0$ for $|z| \geq n+1$, and $|\partial f_n|$, $|\bar{\partial} f_n| \leq 2$ throughout $\mathbb{C}$. Let $u_n = u\, f_n$; so u_n has compact support, $u_n \in L^2$, and $u_n \to u$ in L^2. Also the weak derivative $\partial u_n = (\partial f_n)\, u + f_n(\partial u)$ and $f_n(\partial u) \to \partial u$ in L^2. On the other hand

$$\begin{aligned}\int |(\partial f_n)\, u|^2 d\mathcal{A} &= \int_{n \leq |z| \leq n+1} |(\partial f_n)\, u|^2 d\mathcal{A} \\ &\leq 4 \int_{n \leq |z| \leq n+1} |u|^2 d\mathcal{A} \\ &\to 0\end{aligned}$$

as $n \to \infty$. Hence $\partial u_n \to \partial u$ in L^2. Similarly $\bar{\partial} u_n \to \bar{\partial} u$ in L^2. This argument shows that it suffices to show that $u \in W_1^2$ under the additional assumption that u has compact support.

Now let $\{\phi_\varepsilon\}$ be a mollifier. Since u has compact support, $\phi_\varepsilon * u \to u$ in L^2 and $\phi_\varepsilon * u \in C_c^\infty$ (18.3.6). Since $\partial u \in L^2$ and has compact support, $\partial(\phi_\varepsilon * u) = \phi_\varepsilon * \partial u \to \partial u$ in L^2. Similarly $\bar{\partial}(\phi_\varepsilon * u) \to \bar{\partial} u$ in L^2. Hence $\phi_\varepsilon * u \to u$ in W_1^2. $\square$

Now to give W_1^2 another inner product that is equivalent to the original when we restrict the supports to lie in a fixed bounded set. This is done in the next proposition, though the definition of this new inner product comes after this result.

11.3 Proposition. *If G is a bounded open set in $\mathbb{C}$, then there is a constant $M > 0$ such that for all ϕ in $C_c^\infty(G)$*

$$||\phi||_{W_1^2}^2 \leq M^2 \left[\int |\partial \phi|^2\, d\mathcal{A} + \int |\bar{\partial} \phi|^2 d\mathcal{A}\right].$$

Proof. Clearly it suffices to show that there is a constant M such that

$$\int |\phi|^2\, d\mathcal{A} \leq M^2 \left[\int |\partial \phi|^2\, d\mathcal{A} + \int |\bar{\partial} \phi|^2\, d\mathcal{A}\right].$$

To this end, let Q be an open square, $Q = \{z : |\text{Re } z| < R,\ |\text{Im } z| < R\}$, that contains the closure of G. If $\phi \in C_c^\infty(G)$ and $\partial_1 \phi$ is its derivative with respect to the first variable, then for $z = x + iy$

$$\begin{aligned}|\phi(z)| &= \left|\int_{-R}^{x} \partial_1 \phi(t+iy)\, dt\right| \\ &\leq \left[\int_{-R}^{x} 1^2\, dt\right]^{1/2} \left[\int_{-r}^{x} |\partial_1 \phi(t+iy)|^2\, dt\right]^{1/2} \\ &\leq \sqrt{2R} \left[\int_{-R}^{R} |\partial_1 \phi(t+iy)\, dt|^2\right]^{1/2}.\end{aligned}$$

Thus

$$\int_{-R}^{R} |\phi(x+iy)|^2 dx \le 4R^2 \int_{-R}^{R} |\partial_1 \phi(x+iy)|^2 \, dx,$$

from which we get that

$$\begin{aligned} \int |\phi|^2 d\mathcal{A} &= \int_Q |\phi|^2 d\mathcal{A} \\ &\le 4R^2 \int_Q \left[|\partial_1 \phi|^2 + |\partial_2 \phi|^2 \right] d\mathcal{A} \\ &= 8R^2 \int \left[|\partial \phi|^2 + |\overline{\partial} \phi|^2 \right] d\mathcal{A}. \end{aligned}$$

□

The preceding inequality is called the *Poincaré Inequality.*

11.4 Definition. For a bounded open set G, let $\overset{\circ}{W}{}_1^2(G) =$ the closure of $C_c^\infty(G)$ in W_1^2. The inner product on $\overset{\circ}{W}{}_1^2(G)$ is defined by

11.5 $$\langle u, v \rangle_G = \int \partial u \overline{\partial v} d\mathcal{A} + \int \overline{\partial} u \, \overline{\overline{\partial} v} \, d\mathcal{A}.$$

The norm on $\overset{\circ}{W}{}_1^2(G)$ is denoted by $||u||_G^2 = \langle u, u \rangle_G$.

By the preceding proposition, the inner product defined in (11.5) is equivalent to the one inherited from W_1^2. When discussing $\overset{\circ}{W}{}_1^2(G)$ we will almost always use the inner product (11.5).

Now to prove a few facts about the functions in W_1^2 and $\overset{\circ}{W}{}_1^2(G)$. Most of these results are intuitively clear, but they do require proofs since their truth is not obvious.

11.6 Lemma. *If $u \in W_1^2$ with weak derivative $\partial_1 u$, then there is a function u^* in W_1^2 such that:*

(i) $u^* = u$ *a.e.* [Area];

(ii) $\partial_1 u^* = \partial_1 u$ *a.e.* [Area];

(iii) *u^* is absolutely continuous when restricted to a.e. line parallel to the real axis.*

Proof. As in the proof of Theorem 11.2, it can be assumed that u has compact support. Since $\partial_1 u \in L^2$ and has compact support, $\partial_1 u \in L^1$. This implies that for a.e. y in $\mathbb{R}$, the function $x \to \partial_1 u(x+iy)$ is an integrable function on $\mathbb{R}$. Define u^* on $\mathbb{C}$ by

$$u^*(x+iy) = \int_{-\infty}^{x} \partial_1 u(t+iy) \, dt$$

when this integral exists. Thus u^* is absolutely continuous on almost every line parallel to the real axis and $\partial_1 u^* = \partial_1 u$ a.e. [Area]. It remains to show that $u^* = u$ a.e. [Area].

Let f and g be functions in $C_c^\infty(\mathbb{R})$ with $f(x) = 1$ in a neighborhood of Re(supp u) and g arbitrary. Let $\phi(x+iy) = f(x)g(y)$; so $\phi \in C_c^\infty(C)$. Thus $\int \phi(\partial_1 u) = -\int(\partial_1\phi)u = 0$ since $\partial_1\phi = f'(x)g(y) = 0$ in a neighborhood of Re(supp$u$). Therefore

$$\begin{aligned} 0 &= \int \phi(\partial_1 u)\, d\mathcal{A} \\ &= \int_{-\infty}^{\infty} g(y) \left[\int_{-\infty}^{\infty} \partial_1 u(x+iy)\, dx\right] dy. \end{aligned}$$

But g was arbitrary so we get that

$$\textbf{11.7} \qquad \int_{-\infty}^{\infty} \partial_1 u(x+iy)\, dx = 0$$

for almost all y in $\mathbb{R}$.

Now if ϕ is any function in $C_c^\infty(C)$, then using integration by parts with $\partial_1 \Phi = \phi$,

$$\begin{aligned} \int \phi u^* d\mathcal{A} &= \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} \phi(x+iy) \int_{-\infty}^{x} \partial_1 u(t+iy)\, dt\, dx\right] dy \\ &= \int_{-\infty}^{\infty} \left\{ \left[\Phi(x+iy) \int_{-\infty}^{x} \partial_1 u(t+iy)\, dt\right]_{x=-\infty}^{\infty} \right. \\ &\qquad \left. - \int_{-\infty}^{\infty} \Phi(x+iy)\partial_1 u(x+iy)\, dx \right\} dy. \end{aligned}$$

Now apply (11.7) to get

$$\int \phi u^* d\mathcal{A} = -\int \Phi \partial_1 u\, d\mathcal{A}.$$

If ψ is a function in $C_c^\infty(C)$ such that $\psi = 1$ in a neighborhood of supp u, then

$$\begin{aligned} \int \phi u^* d\mathcal{A} &= -\int \Phi\psi\partial_1 u\, d\mathcal{A} \\ &= \int \partial_1(\Phi\psi) u\, d\mathcal{A} \\ &= \int \phi u\, d\mathcal{A}. \end{aligned}$$

Since ϕ was arbitrary, $u^* = u$ a.e. [Area]. $\square$

11.8 Lemma. *If $u \in W_1^2$ and u is real-valued, then for every constant $s \geq 0$ the function $u_s \equiv \min(u, s) \in W_1^2$ and $||u_s|| \leq ||u||$. In particular $u^+ = \max(u, 0) = -\min(-u, 0)$, $u^- = -\min(u, 0)$, and $|u| = u^+ + u^-$ belong to W_1^2.*

Proof. Let u^* be the function obtained in Lemma 11.6. Thus $u_s^* = \min(u^*, s)$ is absolutely continuous on the same lines as u^*. Also $\partial_1 u_s^* = \partial_1 u$ a.e. [Area] for $u_s^* < s$ and $\partial_1 u_s^* = 0$ a.e. [Area] for $u_s^* > s$. On the set where $u_s^* = s$ it is not difficult to show that $\partial_1 u_s^* = 0$ a.e. [Area]. Thus $\partial_1 u_s^* \in L^2$. But $u^* = u$ a.e. [Area] so $u^* = u_s$ a.e. [Area]. Thus $\partial_1 u_s \in L^2$; similarly $\partial_2 u_s \in L^2$. Since u_s clearly belongs to L^2, we have that $u_s \in W_1^2$ by Theorem 11.2.

The proof of the remainder of the lemma is left as an exercise. □

11.9 Proposition. *Let G be a bounded open set. If $u \in \overset{\circ}{W_1^2}(G)$, then $u = 0$ a.e. [Area] on $\mathbb{C} \setminus G$. If $u \in W_1^2$ and u is zero off some compact subset of G, then $u \in \overset{\circ}{W_1^2}(G)$.*

Proof. First assume that $u \in \overset{\circ}{W_1^2}(G)$ and let $\{\phi_n\}$ be a sequence in $C_c^\infty(G)$ such that $||\phi_n - u|| \to 0$. So in particular, $\int |\phi_n - u|^2 d\mathcal{A} \to 0$. By passing to a subsequence if necessary, it can be assumed that $\phi_n \to u$ a.e. [Area]. Since each ϕ_n vanishes off G, $u = 0$ a.e. [Area] off G.

Now assume that $u \in W_1^2$ and $u = 0$ off K, a compact subset of G. It suffices to assume that u is real-valued. By Lemma 11.8 it also can be assumed that $u \geq 0$ and u is bounded. If $\{\phi_\varepsilon\}$ is a mollifier, then $\phi_\varepsilon * u \in C_c^\infty(G)$ for all sufficiently small ε. It is left to the reader to use Proposition 18.3.6 and the fact that $\partial(\phi_\varepsilon * u) = \phi_\varepsilon * \partial u$ and $\overline{\partial}(\phi_\varepsilon * u) = \phi_\varepsilon * \overline{\partial} u$ to show that $||\phi_\varepsilon * u - u|| \to 0$. Hence $u \in \overset{\circ}{W_1^2}(G)$. □

The preceding result can be improved siginificantly to give a characterization of those functions in W_1^2 that belong to $\overset{\circ}{W_1^2}(G)$. Using the notion of capacity (§21.7) it can be shown that if $u \in W_1^2$, then $u \in \overset{\circ}{W_1^2}(G)$ if and only if u is zero off G except for a set having capacity zero. This result exceeds the purpose of this book and will be avoided except for its role in justifying a certain point of view, which we now examine. The interested reader can see Bagby [1972] or Aleman, Richter, and Ross [preprint] for another description.

Now we return to the study of the Dirichlet problem. The idea here is to use the preceding proposition to replace or weaken the idea of two functions having the same boundary values. Specifically we will say that two functions u and v in W_1^2 agree on ∂G if $u - v \in \overset{\circ}{W_1^2}(G)$. (Of course the functions will be restricted in some further way.) Since Proposition 11.9

says that $u - v$ is 0 a.e. on $\mathbb{C} \setminus G$, this is safe. In light of the referred result that $u-v$ must vanish on the complement of G except for a set of capacity 0, the stability of the ground on which this statement is based is increased.

Another essential ingredient in this mathematical stew is the next proposition. A small abuse of the language will be employed here; for a function u in W_1^2 we will say that u is harmonic (or analytic) on the open set G if u is equal a.e. [Area] on G to a function that is harmonic (or analytic) on G.

11.10 Proposition. *If $u \in W_1^2$ and G is a bounded open set, then u is harmonic on G if and only if for every v in $\overset{\circ}{W_1^2}(G)$*

$$\langle u, v \rangle_G = \int [\partial u \overline{\partial v} + \overline{\partial} u \partial \overline{v}] d\mathcal{A} = 0.$$

Proof. First notice that in the above integral the fact that the function v belongs to $\overset{\circ}{W_1^2}(G)$ implies that it is 0 off G and so the integral can be taken over G. Consider u as a distribution on G. For any ϕ in $C_c^\infty(G)$

$$\langle u, \phi \rangle_G = -2\partial\overline{\partial} u(\phi).$$

Thus the condition in the proposition is equivalent to the condition that $\partial\overline{\partial} u = 0$ as a distribution on G. According to Weyl's Lemma (18.4.10) this is equivalent to the condition that u is harmonic on G. □

11.11 Corollary. *If G and Ω are bounded open sets with* cl $G \subseteq \Omega$ *and $u \in \overset{\circ}{W_1^2}(\Omega)$, then u is harmonic on G if and only if $u \perp \overset{\circ}{W_1^2}(G)$.*

11.12 Dirichlet Principle. *Let G and Ω be bounded open sets with* cl $G \subseteq \Omega$ *and let $f \in \overset{\circ}{W_1^2}(\Omega)$.*

(a) *If $u \in \overset{\circ}{W_1^2}(\Omega)$ such that u is harmonic on G and $u - f \in \overset{\circ}{W_1^2}(G)$, then*

$$\int_G [|\partial u|^2 + |\overline{\partial} u|^2] d\mathcal{A} \leq \int_G [|\partial v|^2 + |\overline{\partial} v|^2] d\mathcal{A}$$

for all v in $\overset{\circ}{W_1^2}(\Omega)$ such that $v - f \in \overset{\circ}{W_1^2}(G)$.

(b) *If u is the orthogonal projection of f onto $\overset{\circ}{W_1^2}(G)^\perp$, then u is harmonic on G and $u - f \in \overset{\circ}{W_1^2}(G)$.*

Proof. Let P denote the orthogonal projection of $\overset{\circ}{W_1^2}(\Omega)$ onto $\overset{\circ}{W_1^2}(G)^\perp$. Note that for functions w in $\overset{\circ}{W_1^2}(G)$, $||w||_G = ||w||_\Omega$. We therefore drop the subscripts G and Ω from the notation for the norm and inner product while in this proof.

(a) Since u is harmonic on G, the preceding corollary implies $u \perp \overset{\circ}{W}{}_1^2(G)$. Thus $Pf = P(f-u) + Pu = u$. Hence if v is as described in (a), $||u|| = ||Pf|| = ||P(f-v) + Pv|| = ||Pv|| \leq ||v||$, proving (a).

(b) Now assume that $u = Pf \in \overset{\circ}{W}{}_1^2(G)^{\perp}$. By Corollary 11.11, u is harmonic on G. Also $P(u-f) = 0$, so $u - f \in \overset{\circ}{W}{}_1^2(G)$. □

Can the Dirichlet Principle be used to solve the Dirichlet problem? The answer is yes if the region G and the boundary values are suitably restricted. Indeed this was the classical way in which the Dirichlet problem was solved. Let's look more closely at this. First, as we mentioned before, the condition that $u - f \in \overset{\circ}{W}{}_1^2(G)$ says that u and f agree on ∂G. So if we are given a continuous function g on ∂G, we would like to get a region Ω containing cl G and a function f in $\overset{\circ}{W}{}_1^2(\Omega)$ that agrees with g on ∂G. This is not always possible if g is only assumed continuous.

Let $g \in C(\partial\mathbb{D})$; we want to get a region Ω that contains cl $\mathbb{D}$ and a function f in $\overset{\circ}{W}{}_1^2(\Omega)$ with $f|\partial\mathbb{D} = g$. Clearly we can take Ω to be an open disk about 0 with radius > 1. Thus $\int_0^{2\pi} |\partial f(re^{i\theta})|^2 d\theta < \infty$ for almost all r, $0 \leq r < 1+\varepsilon$. It is left as an exercise for the reader to show that this implies (is equivalent to?) the statement that $g(e^{i\theta})$ has a Fourier series $\sum c_n e^{i\theta}$ with $\sum |n||c_n|^2 < \infty$. Clearly there are functions in $C(\partial\mathbb{D})$ that do not have such a Fourier series, so without a restriction on the boundary functions the Dirichlet Principle cannot be used to solve the Dirichlet problem. Once the boundary function is restricted, however, the Dirichlet Principle does give the corresponding solution of the Dirichlet problem. Before seeing this, the stage must be set.

11.13 Lemma. *If G is a Jordan region and $\delta < \delta_0 = \min\{\text{diam}\,\gamma :$ γ is a boundary curve of $G\}$, then for every a in ∂G and all ϕ in $\overset{\circ}{W}{}_1^2(G)$,*

$$\int_{B(a;\delta)} |\phi|^2 d\mathcal{A} \leq 8\pi^2\delta^2 \int_{B(a;\delta)} [|\partial\phi|^2 + |\bar{\partial}\phi|^2] d\mathcal{A}.$$

Proof. Without loss of generality we may prove the lemma for functions ϕ that belong to $C_c^\infty(G)$. Fix a in ∂G and let γ be the component of ∂G that contains a. If $0 < \delta < \delta_0$, $\partial B(a;\delta) \cap \gamma \neq \emptyset$. For ϕ in $C_c^\infty(G)$, since support $\phi \subseteq G$ for each r, $0 < r < \delta$, there is a θ_r with $\phi(a + re^{i\theta}r) = 0$. Hence

$$\phi(a + re^{i\theta}) = \int_{\theta_r}^{\theta} \frac{\partial\phi}{\partial\theta}(a + re^{it})dt.$$

Applying the Cauchy-Schwarz Inequality and extending the interval of integration gives

$$|\phi(a+re^{i\theta})|^2 \leq 2\pi \int_0^{2\pi} \left|\frac{\partial \phi}{\partial \theta}(a+re^{it})\right|^2 dt.$$

Thus

$$\int_0^{2\pi} |\phi(a+re^{i\theta})|^2 d\theta \leq 4\pi^2 \int_0^{2\pi} \left|\frac{\partial \phi}{\partial \theta}(a+re^{it})\right|^2 dt.$$

Observe that $2[|\partial\phi|^2 + |\overline{\partial}\phi|^2] = r^{-2}[|\partial\phi/\partial\theta|^2 + |\partial\phi/\partial r|^2]$. Now for $r < \delta$,

$$\int_0^{2\pi} |\phi(a+re^{i\theta})|^2 d\theta \leq 4\pi^2\delta^2 \int_0^{2\pi} \left|\frac{1}{r^2}\frac{\partial \phi}{\partial \theta}(a+re^{it})\right|^2 dt.$$

Now integrate with respect to $r\,dr$ with $0 < r < \delta$ and we get

$$\begin{aligned}\int_{B(a;\delta)} |\phi|^2 d\mathcal{A} &\leq 4\pi^2\delta^2 \int_{B(a;\delta)} \left|\frac{1}{r^2}\frac{\partial\phi}{\partial\theta}\right|^2 d\mathcal{A} \\ &\leq 8\pi^2\delta^2 \int_{B(a;\delta)} [|\partial\phi|^2 + |\overline{\partial}\phi|^2] d\mathcal{A}. \quad \square\end{aligned}$$

The reader will note similarities between the proof of the next lemma and that of Theorem 7.15.

11.14 Lemma. *Let G be a bounded region, let $\{G_n\}$ be a sequence of Jordan regions such that* cl $G_n \subseteq G_{n+1}$ *for all n and $G = \cup_n G_n$, and let f be a continuous function on* cl G. *If u is the solution of the Dirichlet problem for G with boundary values $f|\partial G$ and, for each n, u_n is the solution of the Dirichlet problem for G_n with boundary values $f|\partial G_n$, then $u_n(z) \to u(z)$ for all z in G.*

Proof. Fix a_0 in G_1; it will be shown that $u_n(a_0) \to u(a_0)$. The proof is obtained by considering several special classes of functions f. In each such case u and u_n are as in the statement of the lemma.

First assume that $f = p$, a polynomial in z and $\overline{z}$ that is subharmonic in a neighborhood of cl G. For each $n \geq 1$ define $v_n : \mathbb{C} \to \mathbb{C}$ by letting $v_n = p$ on $\mathbb{C} \setminus G_n$ and letting $v_n = u_n$ on G_n. So v_n is subharmonic on $\mathbb{C}$; since G_n is a Dirichlet region, v_n is also continuous. By the Maximum Principle, $v_n \geq p$ on G_n.

Because cl $G_n \subseteq G_{n+1}$, another application of the Maximum Principle implies that $v_n \leq v_{n+1}$ on $\mathbb{C}$. Moreover, each v_n belongs to the Perron family $\hat{\mathcal{P}}(p, G)$ and so, on G, $v_n \leq \hat{p}$, the solution of the Dirichlet problem on G with boundary values p. Thus $v(z) \equiv \lim_n v_n(z) = \lim_n u_n(z)$ is a harmonic function on G and $v \leq \hat{p} = u$. On the other hand, we have that $v_n \geq p$ on G_n so that $v \geq p$ on G and, hence, $v \in \check{\mathcal{P}}(p, G)$. Thus

$v \geq u$ so that, in fact, $v = u$. This proves the lemma whenever f is such a polynomial.

Now let f be any polynomial in z and $\bar{z}$ and choose a constant $c > 0$ such that $f + c|z|^2$ is subharmonic in a neighborhood of cl G (see the proof of Theorem 7.15). Let v_n (respectively, v) be the solution of the Dirichlet problem on G_n (respectively, G) with boundary values $c|z|^2$. According to the preceding paragraph, $u_n(z) + v_n(z) \to u(z) + v(z)$ for all z in G. But $c|z|^2$ is also subharmonic, so again the preceding paragraph implies $v_n(z) \to v(z)$ for all z in G. Thus the lemma holds for all polynomials.

Now let f be arbitrary and let $\varepsilon > 0$; pick a polynomial p such that $|f(z) - p(z)| < \varepsilon$ for all z in cl G. Let v_n (respectively, v) be the solution of the Dirichlet problem on G_n (respectively, G) with boundary values p. Now on ∂G_n, $p - \varepsilon \leq f \leq p + \varepsilon$; hence $v_n - \varepsilon \leq u_n \leq v_n + \varepsilon$ on G_n. Similarly, $v - \varepsilon \leq u \leq v + \varepsilon$ on G. It follows that $|u - u_n| \leq |v - v_n| + 2\varepsilon$ on G_n and so the lemma holds. $\square$

The preceding lemma will be generalized in Theorem 21.10.9 below.

11.15 Theorem. *Let G be a bounded region, let Ω be an open set with* cl $G \subseteq \Omega$, *and assume that $f \in \overset{\circ}{W}{}^2_1(\Omega)$ such that f is continuous on* cl G. *If u is the orthogonal projection of f onto $\overset{\circ}{W}{}^2_1(\Omega) \cap \overset{\circ}{W}{}^2_1(G)^{\perp}$, then u is the solution of the Dirichlet problem on G with boundary values f.*

Proof. We already know from the Dirichlet Principle that u is harmonic on G. We first prove this theorem for the case that G is a Jordan region. Then Lemma 11.14 can be used to prove it for arbitrary bounded regions.

The assumption that G is a Jordan region guarantees that every point of ∂G is a regular point; so we need to show that, for each a in ∂G, $u(z) - f(z) \to 0$ as $z \to a$ with z in G. So fix a in ∂G and let δ_0 be as in Lemma 11.13; let $\varepsilon > 0$ be arbitrary. Since f is uniformly continuous on cl G, there is a $\delta_1 > 0$ such that $|f(z) - f(w)| < \varepsilon$ for z, w in cl G with $|z - w| < \delta_1$. Now let δ be less than both δ_1 and $\delta_0/2$; δ will be further restricted later in such a way that it only depends on the point a. Fix z in G with $|z-a| < \delta$. Letting $\delta_2 = \operatorname{dist}(z, \partial G) \leq |z - a|$, the Mean Value Property of harmonic functions implies

$$
\begin{aligned}
u(z) - f(z) &= \frac{1}{\pi\,\delta_2^2} \int_{B(z:\delta_2)} u \, d\mathcal{A} - f(z) \\
&= \frac{1}{\pi\,\delta_2^2} \int_{B(z:\delta_2)} (u - f) \, d\mathcal{A} - \frac{1}{\pi\delta^2} \int_{B(z:\delta_2)} [f - f(z)] \, d\mathcal{A}. \qquad \textbf{11.16}
\end{aligned}
$$

Now $|f(w) - f(z)| < \varepsilon$ for all w in $B(z; \delta_2)$ so the absolute value of the last of these two summands is less than ε. Also the Cauchy-Schwarz Inequality gives that

$$\left|\int_{B(z:\delta_2)} (u-f)d\mathcal{A}\right|^2 \leq \pi\delta_2^2 \int_{B(z:\delta_2)} |u-f|^2 d\mathcal{A}.$$

Now $B(z;\delta_2) \subseteq B(a;2\delta)$ and $2\delta < \delta_0$. Using Lemma 11.13, this transforms the preceding inequality into

$$\begin{aligned}\left|\int_{B(z:\delta_2)} (u-f)d\mathcal{A}\right|^2 &\leq \pi\delta_2^2 \int_{B(a:2\delta)} |u-f|^2 d\mathcal{A} \\ &\leq 32\pi^3\delta_2^4 \int_{B(a;2\delta)} [|\partial(u-f)|^2 + |\overline{\partial}(u-f)|^2] d\mathcal{A}.\end{aligned}$$

Using (11.16) this gives

$$|u(z) - f(z)| \leq \sqrt{32\pi}\left[\int_{B(a;2\delta)} [|\partial(u-f)|^2 + |\overline{\partial}(u-f)|^2] d\mathcal{A}\right]^{\frac{1}{2}} + \varepsilon.$$

Because $\partial(u-f)$ and $\overline{\partial}(u-f)$ are square integrable, δ can be chosen sufficiently small, depending on a, that $|u(z)-f(z)| < 2\varepsilon$ when $|z-a| < \delta$.

This proves the theorem for the case that G is a Jordan region. The details in applying Lemma 11.14 to obtain the arbitrary case are left to the reader. □

Exercises

1. Show that $\partial\phi\overline{\partial\psi} + \overline{\partial}\phi\overline{\overline{\partial}\psi} = \frac{1}{2}(\nabla\phi \cdot \nabla\overline{\psi})$ for ϕ and ψ in C_c^∞.

2. Show that if $u \in W_1^2$ and $\phi \in C_c^\infty(G)$, then $\phi u \in \overset{\circ}{W_1^2}(G)$.

3. Prove a version of Lemma 11.8 for functions in $\overset{\circ}{W_1^2}(G)$.

Chapter 20

Hardy Spaces on the Disk

In this chapter the classical theory of the Hardy spaces on the open unit disk will be explored. The structure of the functions belonging to the spaces H^p will be determined, and this will be applied to characterize the invariant subspaces of multiplication by the independent variable on a Hardy space.

§1 Definitions and Elementary Properties

Here we introduce the Hardy spaces H^p of analytic functions on the open unit disk.

1.1 Definition. If $f : \mathbb{D} \to \mathbb{C}$ is a measurable function and $1 \leq p < \infty$, define

$$M_p(r, f) \equiv \left[\frac{1}{2\pi}\int_0^{2\pi} |f(re^{i\theta})|^p d\theta\right]^{1/p};$$

also define

$$M_\infty(r, f) \equiv \sup\{|f(re^{i\theta})| : 0 \leq \theta \leq 2\pi\}.$$

For any value of p, $1 \leq p \leq \infty$, let H^p denote the space of all analytic functions on $\mathbb{D}$ for which $||f||_p \equiv \sup_{r<1} M_p(r, f) < \infty$.

If $f : \mathbb{D} \to \mathbb{C}$ and $0 < r < 1$, denote by f_r the function defined on $\partial\,\mathbb{D}$ by $f_r(z) = f(rz)$ (as in §19.1). Thus for any such r, $M_p(r, f)$ is the L^p norm of f_r. From this observation it follows that $||\cdot||_p$ is a norm on H^p. In fact, we will see that H^p is a Banach space (1.5).

Also from standard L^p space theory, $H^p \subseteq H^r \subseteq H^1$ if $1 \leq r \leq p$. In particular H^∞ is the space of bounded analytic functions on $\mathbb{D}$ and $H^\infty \subseteq H^p$ for all p.

The same definition applies for $0 < p < 1$, though the "norm" for this range of values of p must be redefined (without taking the p-th root). This will not be pursued here. The interested reader can consult Duren [1970], Hoffman [1962], or Koosis [1980] for this topic.

1.2 Proposition. *If $f : \mathbb{D} \to \mathbb{C}$ is an analytic function and $1 \leq p \leq \infty$, then $||f||_p = \lim_{r\to 1-} M_p(r, f)$.*

Proof. Assume that p is finite. From Example 19.4.16 we know that $z \to$

$|f(z)|^p$ is a continuous subharmonic function and so (19.4.9) $M_p(r, f)$ is an increasing function of r. Therefore the supremum must equal the limit.

By the Maximum Modulus Theorem $M_\infty(r, f)$ is also an increasing function of r. Thus the proposition also holds for $p = \infty$. □

Note that if $f \in H^1$, then, by Theorem 19.2.12, there is a measure μ on $\partial\,\mathbb{D}$ such that $f = \tilde{\mu}$. Also if $1 < p \le \infty$ and $f \in H^p$, then $f = \tilde{g}$ for some g in $L^p(\partial\,\mathbb{D})$. Moreover in the case that $p > 1$, $f_r \to g$ a.e. $[m]$ as $r \to 1$. In particular, each function in H^p, $1 \le p \le \infty$, has non-tangential limits at almost every point of $\partial\,\mathbb{D}$.

There are two questions that occur here. First, when $p > 1$, which functions g in L_p can arise in this way? Note that when we answer this question we will have identified H^p with a certain subspace of L^p and thus have the possibility of combining measure theory with the theory of analytic functions.

The second question concerns the case when $p = 1$. Here the theory becomes more subtle and difficult. If $f \in H^1$, then Theorem 19.2.12 says that the radial limit function g for f exists and $g =$ the Radon-Nikodym derivative of μ with respect to m, where $f = \tilde{\mu}$. It will turn out that μ is absolutely continuous with respect to Lebesgue measure so that, indeed, $f = \tilde{g}$. This is the F and M Riesz Theorem proved in §3 below. For now we content ourselves with complete information when $p > 1$ and partial information when $p = 1$. Recall (18.7.1) that for any function f (respectively, measure μ), $\hat{f}$ (respectively, $\hat{\mu}$) denotes the Fourier transform of f (respectively, μ).

1.3 Theorem. *If $1 \le p \le \infty$ and $f \in L^p$ such that $\hat{f}(n) = 0$ for $n < 0$, then $\tilde{f}$, the Poisson integral of f, belongs to H^p. Moreover:*

(a) $||f||_p = ||\tilde{f}||_p$;

(b) *if $1 \le p < \infty$, $||\tilde{f}_r - f||_p \to 0$ as $r \to 1-$;*

(c) *if $p = \infty$, $\tilde{f}_r \to f$ weak* in L^∞ as $r \to 1-$.*

Proof. If $z = re^{i\theta} \in \mathbb{D}$ and $|w| = 1$, then $P_z(w) = \sum_{n=-\infty}^{\infty} r^{|n|} e^{in\theta}\overline{w^n}$. Hence

$$\begin{aligned}
\tilde{f}(z) &= \int f(w)\, P_z(w)\, dm(w) \\
&= \int f(w) \sum_{n=-\infty}^{\infty} r^{|n|} e^{in\theta} \overline{w^n}\, dm(w) \\
&= \sum_{n=-\infty}^{\infty} r^{|n|} e^{in\theta} \int f(w)\overline{w^n}\, dm(w),
\end{aligned}$$

since the series converges uniformly in w. Therefore

$$\tilde{f}(z) = \sum_{n=-\infty}^{\infty} r^{|n|} e^{i\theta} \hat{f}(n) = \sum_{n=0}^{\infty} \hat{f}(n) z^n$$

since, by hypothesis, $\hat{f}(n) = 0$ when $n < 0$. Thus $\tilde{f}$ has a power series expansion in $\mathbb{D}$ and so $\tilde{f}$ is analytic on $\mathbb{D}$.

Also for $1 \le p < \infty$, $||\tilde{f}_r - f||_p \to 0$ as $r \to 1-$ (19.1.4). If $p = \infty$, then $||\tilde{f}_r||_\infty \le ||f||_\infty$. Hence, in either case, $\sup_r M_p(r, \tilde{f}) = \sup_r ||\tilde{f}_r||_p < \infty$ and so $\tilde{f} \in H^p$. The remaining details are easily deduced from Theorem 19.1.4. □

What about the converse of the preceding theorem? Here is where we must assume that $p > 1$ and postpone consideration of the case where $p = 1$ until later. Suppose that $f \in H^p$, $1 \le p \le \infty$, and let g be the non-tangential limit function of f. That is, $g(w) = \lim f(rw) = \lim f_r(w)$ a.e. $[m]$ on $\partial\mathbb{D}$. By Theorem 19.2.12, $g \in L^p$. Now if $1 < p < \infty$, $||f_r - g||_p \to 0$. Thus

$$\begin{aligned} \hat{g}(n) &= \int g(w)\, \overline{w^n}\, dm(w) \\ &= \lim_{r\to 1-} \int f_r(w)\, \overline{w^n}\, dm(w) \\ &= \lim_{r\to 1-} \hat{f}_r(n). \end{aligned}$$

But $f(z) = \sum_0^\infty a_n z^n$ for $|z| < 1$, with convergence uniform on proper subdisks of $\mathbb{D}$. It follows (how?) that $\hat{f}_r(n) = a_n r^n$ if $n \ge 0$ and $\hat{f}_r(n) = 0$ if $n < 0$. Therefore $\hat{g}(n) = a_n$ if $n \ge 0$ and $\hat{g}(n) = 0$ if $n < 0$. By Theorem 1.3, $\tilde{g} \in H^p$. But also (see the proof of Theorem 1.3) for $|z| < 1$,

$$\tilde{g}(z) = \sum_0^\infty \hat{g}(n) z^n = \sum_0^\infty a_n z^n.$$

Hence $\tilde{g} = f$ and the desired converse is obtained.

If $p = \infty$, then it is not necessarily true that $f_r \to g$ in L^∞ norm, but it is true that $f_r \to g$ weak* as $r \to 1-$. Since $\overline{w^n} \in L^1$ for every n, the argument of the preceding paragraph shows that $\hat{g}(n) = 0$ for $n < 0$ and $\tilde{g} = f$. This discussion can be summarized as follows.

1.4 Theorem. *If $1 < p \le \infty$ and $f \in H^p$, then $g(w) = \lim_{r\to 1-} f(rw)$ defines a function g in L^p with $\hat{g}(n) = 0$ for $n < 0$ and $\tilde{g} = f$.*

These last two theorems establish a correspondence between functions in H^p (defined on $\mathbb{D}$) and the functions in

$$\mathcal{H}^p \equiv \{f \in L^p : \hat{f}(n) = 0 \text{ for } n < 0\},$$

but only when $p > 1$. Now it is easy to see that $\mathcal{H}^p$ is a closed subspace of L^p and we are led to the following.

1.5 Theorem. *If $1 < p \leq \infty$, then the map that takes f in H^p to its boundary values establishes an isometry of H^p onto the closed subspace $\mathcal{H}^p$ of L^p. Thus H^p is a Banach space. For $p = 2$, H^p is a Hilbert space with $\{1, z, z^2, \dots\}$ as an orthonormal basis. For $p = \infty$, H^∞ is the dual of a Banach space since $\mathcal{H}^\infty$ is a weak* closed subspace of L^∞.*

Because of this result we can identify the functions in H^p, $1 < p \leq \infty$, with their radial or non-tangential limits. Henceforward this identification will be made without fanfare. In §3 we will prove this assertion for $p = 1$; but first, in the next section, we will investigate another class of analytic functions on $\mathbb{D}$. This information will be used to derive the correspondence between H^1 and $\mathcal{H}^1$.

Exercises

1. Supply the details in the proof of Corollary 1.5.

2. Let $\mathcal{A} = \{f \in C(\partial\,\mathbb{D}) : \hat{f}(n) = 0 \text{ for } n < 0\}$. If $f \in \mathcal{A}$, show that its Poisson integral, $\tilde{f}$, is an analytic function on $\mathbb{D}$. Also if $g : \text{cl } \mathbb{D} \to \mathbb{C}$ is defined by $g(z) = \tilde{f}(z)$ for $|z| < 1$ and $g(z) = f(z)$ for $|z| = 1$, then g is continuous. Conversely, if g is a continuous function on cl $\mathbb{D}$ that is analytic on $\mathbb{D}$ and $f = g|\partial\,\mathbb{D}$, then $g(z) = \tilde{f}(z)$ for $|z| < 1$. (See §4.)

3. Give a direct proof that H^1 is a Banach space.

4. Show that $(1 - z)^{-1} \notin H^1$.

5. If $1 \leq p < \infty$ and $|a| < 1$, define $L_a : H^p \to \mathbb{C}$ by $L_a(f) = f(a)$ for all f in H^p. Show that $L_a \in (H^p)^*$ and $||L_a|| = (1 - |a|^2)^{-1/p}$. For $p = 2$, find the unique function k_a in H^2 such that $L_a(f) = \langle f, k_a \rangle$ for all f in H^2.

6. Prove Littlewood's Subordination Theorem: If f and g are analytic functions on $\mathbb{D}$, f is subordinate to g (17.1.1), and $g \in H^p$, then $f \in H^p$ and $||f||_p \leq ||g||_p$.

§2 The Nevanlinna Class

In this section we will study another collection of analytic functions which is not a Banach space but includes all the Hardy spaces H^p.

2.1 Definition. A function f is in the *Nevanlinna class* (or of *bounded characteristic*) if f is an analytic function on $\mathbb{D}$ and

$$\sup_{r<1} \frac{1}{2\pi} \int_0^{2\pi} \log^+ |f(re^{i\theta})|\, d\theta < \infty.$$

The Nevanlinna class is denoted by N.

Note that since $\log^+ |f(z)|$ is a subharmonic function (Example 19.4.7) and is also continuous, $\frac{1}{2\pi} \int_0^{2\pi} \log^+ |f(re^{i\theta})|\, d\theta$ is an increasing function of r (19.4.9). Thus the definition of a function in the Nevanlinna class can be weakened by only stipulating the finiteness of the supremum over a sequence $\{r_n\}$ with $r_n \to 1$. Also

$$\sup_{r<1} \frac{1}{2\pi} \int_0^{2\pi} \log^+ |f(re^{i\theta})|\, d\theta = \lim_{r\to 1-} \frac{1}{2\pi} \int_0^{2\pi} \log^+ |f(re^{i\theta})| d\theta.$$

Since $\log x \leq x^p$ for $x \geq 1$, we have the following.

2.2 Proposition. *If $1 \leq p \leq \infty$, then $H^p \subseteq N$.*

Thus every result for the Nevanlinna class is a result about functions belonging to all the Hardy classes. Our immediate goal is to give a characterization of functions in N (Theorem 2.10 below) and to study the zeros of these functions. In the next section we will obtain a factorization theorem for this class. We begin with an elementary but important result.

2.3 Lemma. *If $\{a_n\}$ is a sequence in $\mathbb{D}$, the following statements are equivalent.*

(a) $\sum_{n=1}^{\infty} (1 - |a_n|) < \infty$.

(b) $\prod_{n=1}^{\infty} |a_n|$ *converges.*

(c) $\sum_{n=1}^{\infty} \log |a_n| < \infty$.

(d) $\prod_{n=1}^{\infty} \frac{|a_n|}{a_n} \left(\frac{a_n - z}{1 - \overline{a}_n z} \right)$ *converges uniformly and absolutely on compact subsets of $\mathbb{D}$.*

Proof. The proof that (a), (b), and (c) are equivalent can be found in §7.5. The fact that (a) implies (d) is Exercise 7.5.4. By evaluating the infinite product in (d) at $z = 0$, (b) can be deduced from (d). □

2.4 Definition. A sequence $\{a_n\}$ satisfying one of the equivalent conditions in the preceding lemma is called a *Blaschke sequence.* If $\{a_n\}$ is a Blaschke sequence and m is an integer, $m \geq 0$, then the function

2.5 $$b(z) = z^m \prod_{n=1}^{\infty} \frac{|a_n|}{a_n} \left(\frac{a_n - z}{1 - \overline{a_n} z} \right)$$

is called a *Blaschke product.*

The factor z^m in the definition of a Blaschke product is there to allow b to have a zero at the origin.

2.6 Proposition. *If b is the Blaschke product defined by (2.5), then $b \in H^\infty$ and $|b(w)| = 1$ a.e. $[m]$ on $\partial\,\mathbb{D}$. The zeros of b are precisely the points $a_1, a_2, \ldots$ and, provided $m > 0$, the origin.*

Proof. Let b_n denote the product of z^m with the first n factors in (2.5). It is easy to see that $|b_n(z)| \leq 1$ on cl $\mathbb{D}$ and $|b_n(w)| = 1$ for w in $\partial\,\mathbb{D}$. By (2.3) $|b(z)| \leq 1$ on $\mathbb{D}$ and so $b \in H^\infty$. Also for $n > k$,

$$\begin{aligned}\int |b_n - b_k|^2\, dm &= 2\left[1 - \operatorname{Re}\int b_n \overline{b_k}\, dm\right] \\ &= 2\left[1 - \operatorname{Re}\int \frac{b_n}{b_k}\, dm\right].\end{aligned}$$

Because $n > k$, b_n/b_k is analytic on $\mathbb{D}$. Thus the mean value property implies that

$$\int \frac{b_n}{b_k}\, dm = \frac{b_n}{b_k}(0) = \prod_{j=k+1}^{n} |a_j|.$$

Hence

$$\int |b_n - b_k|^2 dm = 2[1 - \prod_{j=k+1}^{n} |a_j|].$$

But Lemma 2.3 implies that the right hand side of this last equation can be made arbitrarily small for sufficiently large n and k. Therefore $\{b_n\}$ is a Cauchy sequence in H^2 and must converge to some function f. If $|z| < 1$ and P_z is the Poisson kernel, then $P_z \in L^2$ and so $b_n(z) = \int P_z b_n\, dm \to \int P_z\, f\, dm$. Hence it must be that $f = b$. That is, $b_n \to b$ in H^2.

Because we have convergence in the L^2 norm, there is a subsequence $\{b_{n_k}\}$ such that $b_{n_k}(w) \to b(w)$ a.e. $[m]$. Since $|b_{n_k}| = 1$ a.e. on $\partial\,\mathbb{D}$, it follows that $|b| = 1$ a.e. on $\partial\,\mathbb{D}$. Finally, the statement about the zeros of the Blaschke product follows from (7.5.9). □

We obtained a useful fact in the course of the preceding proof that is worth recording.

2.7 Corollary. *If $\{a_n\}$ is a Blaschke sequence, b is the corresponding Blaschke product, and b_n is the finite Blaschke product with zeros $a_1, \ldots, a_n$, then there is a sequence of integers $\{n_k\}$ such that $b_{n_k} \to b$ a.e. on $\partial\,\mathbb{D}$.*

The next result is the reason for our concern with Blaschke sequences.

2.8 Theorem. *If f is in the Nevanlinna class and f is not identically 0, then the zeros of f form a Blaschke sequence. Moreover, if b is the Blaschke product with the same zeros as f, then $f/b \in N$. If $f \in H^p$, then $f/b \in H^p$.*

Proof. It suffices to assume that $f(0) \neq 0$. By Jensen's Formula (11.1.2)

2.9 $$\frac{1}{2\pi}\int_0^{2\pi} \log|f(re^{i\theta})|\,d\theta = \log|f(0)| + \sum_{|a_k|<r} \log\left(\frac{r}{|a_k|}\right),$$

where $a_1, a_2, \ldots$ are the zeros of f, repeated as often as their multiplicity, and r is chosen so that $|a_k| \neq r$ for any k. But $\log x \leq \log^+ x$ and so, since $f \in N$, there is a finite constant $M > 0$ such that $\frac{1}{2\pi}\int_0^{2\pi} \log|f(re^{i\theta})|\,d\theta \leq M$ for all $r < 1$. This implies, with a small argument, that $\sum_{k=1}^{\infty} \log\left(\frac{1}{|a_k|}\right) < \infty$. By Lemma 2.3, $\{a_k\}$ is a Blaschke sequence.

Now to show that $f/b \in N$, when b is the Blaschke product with the same zeros as f. It is left as an exercise to show that if f has a zero at $z = 0$ of order m, then $f/z^m \in N$. Thus we may assume that $f(0) \neq 0$ and hence $b(0) \neq 0$. If $g = f/b$, then g is an analytic function on $\mathbb{D}$ and never vanishes. If $z \in \mathbb{D}$ with $|g(z)| \leq 1$, then $|f(z)| \leq |b(z)| \leq 1$ and so $\log^+|g(z)| = 0 < -\log|b(z)| = \log^+|f(z)| - \log|b(z)|$. If $|g(z)| > 1$, then $\log^+|g(z)| = \log|g(z)| = \log|f(z)| - \log|b(z)| \leq \log^+|f(z)| - \log|b(z)|$. Thus we have that, for all z in $\mathbb{D}$,

$$\log^+|g(z)| \leq \log^+|f(z)| - \log|b(z)|.$$

But $\log|b|$ is a subharmonic function and so for $0 < r < 1$,

$$\begin{aligned}\frac{1}{2\pi}\log^+|g(re^{i\theta})|d\theta &\leq \frac{1}{2\pi}\int_0^{2\pi} \log^+|f(re^{i\theta})|d\theta - \frac{1}{2\pi}\int_0^{2\pi} \log|b(re^{i\theta})|d\theta \\ &\leq \frac{1}{2\pi}\int_0^{2\pi} \log^+|f(re^{i\theta})|d\theta - \log|b(0)|.\end{aligned}$$

Since $b(0) \neq 0$, this implies that $g \in N$.

Now to show that $f/b \in H^p$ when $f \in H^p$. Let $a_1, a_2, \ldots$ be the zeros of f, repeated as often as their multiplicities, and let b_n be the Blaschke product with zeros $a_1, \ldots, a_n$. Put $g_n = f/b_n$ and let $M \geq \int |f(rw)|^p dm(w)$ for all $r < 1$. If $\varepsilon > 0$, there is an r_0 such that for $r_0 \leq |z| \leq 1$, $|b_n(z)| > 1 - \varepsilon$. Therefore

$$\begin{aligned}\int |g_n(rw)|^p dm(w) &\leq \frac{1}{(1-\varepsilon)^p}\int |f(rw)|^p\,dm(w) \\ &\leq \frac{M}{(1-\varepsilon)^p}\end{aligned}$$

for $r_0 < r < 1$. Letting $r \to 1$, we get that g_n belongs to H^p. Since ε was arbitrary, we can let $\varepsilon \to 0$ to see that $\|g_n\|_p \leq M$ for all n. By Corollary 2.7 there is a subsequence $\{g_{n_k}\}$ such that $g_{n_k} \to g \equiv f/b$ a.e. on $\partial\mathbb{D}$.

Therefore $g_{n_k} \to g$ weakly in L^p. Since each g_n belongs to H^p (why?), $g \in H^p$. $\square$

Here is the promised characterization of functions in the Nevanlinna class.

2.10 Theorem. (F and R Nevanlinna) *If f is an analytic function on $\mathbb{D}$, then f belongs to the Nevanlinna class if and only if $f = g_1/g_2$ for two bounded analytic functions g_1 and g_2.*

Proof. First assume that f is the quotient of the two bounded analytic functions g_1 and g_2. It can be further assumed that $|g_i(z)| \leq 1$ for z in $\mathbb{D}$, $i = 1, 2$. Since f must be analytic, it also can be assumed that g_2 has no zeros in $\mathbb{D}$. It follows that $\log|g_i(z)| \leq 0$ and so $\log|f| = \log|g_1| - \log|g_2| \leq -\log|g_2|$; hence $\log^+|f| \leq -\log|g_2|$ on $\mathbb{D}$. This implies that

$$\begin{aligned}\frac{1}{2\pi}\int_0^{2\pi} \log^+|f(re^{i\theta})|d\theta &\leq -\frac{1}{2\pi}\int_0^{2\pi} \log|g_2(re^{i\theta})|d\theta \\ &= -\log|g_2(0)|\end{aligned}$$

since $\log|g_2|$ is harmonic. Thus $f \in N$.

Now assume that $f \in N$. By Theorem 2.8 we may assume that f does not vanish on $\mathbb{D}$. Hence $u = \log|f|$ is a harmonic function. Thus

$$\begin{aligned}u(0) &= \int \log|f_r|\, d_m \\ &= \int \log^+|f_r|\, dm - \log^-|f_r|\, dm.\end{aligned}$$

(Here $\log^- x = -\min\{\log x, 0\}$.) Since $f \in N$ and the left hand side of this equation is independent of r, it follows that $\sup_{r<1}\int \log^-|f_r|dm < \infty$. Therefore

$$\begin{aligned}\sup_{r<1}\|u_r\|_1 &= \sup_{r<1}\int |\log|f_r||\, dm \\ &= \sup_{r<1}\left\{\log^+|f_r|\, dm + \int \log^-|f_r|dm\right\} \\ &< \infty.\end{aligned}$$

By Theorem 19.1.7, there is a measure μ on $\partial\,\mathbb{D}$ such that $u(z) = \tilde{\mu}(z) = \int P_z d\mu$. Since u is real-valued, μ is a real-valued measure. Let $\mu = \mu_+ - \mu_-$ be the Hahn decomposition of μ and put $u_\pm = \tilde{\mu_\pm}$; so u_+ and u_- are non-negative harmonic functions on $\mathbb{D}$.

Now $\mathbb{D}$ is simply connected and f is a non-vanishing analytic function, so there is an analytic function h on $\mathbb{D}$ such that $f = e^h$; stipulate that $h(0) = \log|f(0)|$ so that h is unique. Thus $u = \operatorname{Re} h$ and it follows (by uniqueness) that

$$h(z) = \int \frac{w+z}{w-z}\, d\mu(w).$$

Also let

$$h_\pm(z) = \int \frac{w+z}{w-z}\, d\mu_\pm(w).$$

Thus $h = h_+ - h_-$ and Re $h_\pm = u_\pm \geq 0$. Put $g_1 = e^{-h_-}$ and $g_2 = e^{-h_+}$; so g_1 and g_2 are analytic functions on $\mathbb{D}$ with no zeros. Also $|g_i(z)| = \exp(-\operatorname{Re} h_\pm(z)) = \exp(-u_\pm(z)) \leq 1$. That is, g_1 and g_2 belong to H^∞. Finally, $g_1/g_2 = \exp(-h_-)/\exp(-h_+) = \exp(h_+ - h_-) = f$. □

2.11 Theorem. *If $f \in N$ and f is not constantly* 0, *then f has a non-tangential limit a.e. on $\partial\,\mathbb{D}$ and* $\log|f(e^{i\theta})| \in L_1$.

Proof. We begin by proving the corollary if $f \in H^\infty$. So assume that $||f||_\infty \leq 1$; we may also assume that $f(0) \neq 0$. By Fatou's Theorem, f has non-tangential limits a.e. on $\partial\,\mathbb{D}$; thus $f_r(w) \to f(w)$ a.e. on $\partial\,\mathbb{D}$. By Fatou's Lemma (from real variables)

$$\begin{aligned} \int_{\partial\,\mathbb{D}} |\log|f(w)||\, dm(w) &\leq \liminf_{r\to 1-} \int |\log|f_r(w)||\, dm(w) \\ &= \liminf_{r\to 1-} \left[- \int \log|f_r(w)|\, dm(w) \right]. \end{aligned}$$

But $\log|f|$ is subharmonic, so $\log|f(0)| \leq \int \log|f_r|\, dm$. Hence

$$\int_{\partial\,\mathbb{D}} |\log|f(w)||\, dm(w) \leq -\log|f(0)| < \infty.$$

Thus $\log|f| \in L_1$.

Now let $f \in N$. By the preceding theorem, $f = g_1/g_2$ for two bounded analytic functions g_1 and g_2. Since $\log|g_i| \in L_1$, neither g_1 nor g_2 can vanish on a subset of $\partial\,\mathbb{D}$ with positive measure. Thus the fact that both g_1 and g_2 have non-zero non-tangential limits a.e. on $\partial\,\mathbb{D}$ implies that the same is true of f. Also, because $\log|g_i| \in L_1$, it follows that $\log|f| = \log|g_1| - \log|g_2| \in L_1$. □

The condition that $\log|f(e^{i\theta})|$ is integrable is to say that f is *log integrable.*

2.12 Corollary. *If $f \in N$ and $f(w) = 0$ on a subset of $\partial\,\mathbb{D}$ having positive measure, then f is identically* 0.

2.13 Corollary. *If $f \in H^p$ and $f(w) = 0$ on a subset of $\partial\,\mathbb{D}$ having positive measure, then f is identically* 0.

Exercise

1. Suppose f is an analytic function on $\{z : \operatorname{Re} z > 0\}$ and $f(n) = 0$ for all $n \geq 1$. Show that if f is bounded, then $f = 0$.

§3 Factorization of Functions in the Nevanlinna Class

In this section we will give a canonical factorization of functions in the class N. This will also factor functions in the Hardy spaces H^p. Actually the strategy is to factor bounded analytic functions and then use Theorem 2.10 to factor functions in N.

3.1 Definition. An *inner function* is a bounded analytic function ϕ on $\mathbb{D}$ such that $|\phi(w)| = 1$ a.e. $[m]$.

It follows that every Blaschke product is an inner function, but there are some additional ones.

3.2 Proposition. *If μ is a positive singular measure on $\partial\mathbb{D}$ and*

3.3
$$\phi(z) = \exp\left(-\int \frac{w+z}{w-z}\, d\mu(w)\right),$$

then ϕ is an inner function.

Proof. It is easy to see that ϕ is well defined and analytic on $\mathbb{D}$. Let $u(z) = -\tilde{\mu}(z) = -\int P_z(w)\, d\mu(w) = -\operatorname{Re}\int \frac{w+z}{w-z}\, d\mu(w)$; so $|\phi(z)| = e^{u(z)}$. Since μ is a positive measure, $u(z) \leq 0$ for all z in $\mathbb{D}$. Also the fact that μ is a singular measure implies, by Fatou's Theorem, that $u(rw) \to 0$ as $r \to 1-$ a.e. $[m]$. Hence if $|w| = 1$ and both ϕ and u have a non-tangential limit at w, then $|\phi(w)| = \lim_{r\to 1-} |\phi(rw)| = \lim_{r\to 1-} e^{-u(rw)} = 1$ a.e. $[m]$. Therefore ϕ is inner. □

An inner function ϕ as defined in (3.3) is called a *singular inner function.* It will turn out that singular functions are the inner functions with no zeros in $\mathbb{D}$. At this point the reader might be advised to work Exercise 12 to see that the correspondence between singular inner functions and positive singular measures μ as described by (3.3) is bijective.

We come now to another class of analytic functions that are, in a certain sense, complementary to the inner functions. The idea here is to use formula (3.3) but with an absolutely continuous measure. It is also not required that the measure be positive.

3.4 Definition. An analytic function $f : \mathbb{D} \to \mathbb{C}$ is an *outer function* if there is a real-valued function h on $\mathbb{D}$ that is integrable with respect to

Lebesgue measure and such that

3.5
$$f(z) = \exp\left(\int \frac{w+z}{w-z}\, h(w)\, dm(w)\right)$$

for all z in $\mathbb{D}$.

It is clear that (3.5) defines an analytic function on $\mathbb{D}$ that has no zeros. Also the fact that h is real-valued implies that the harmonic function $\log|f|$ is precisely the Poisson transform of h.

3.6 Proposition. *If f is the outer function defined by (3.5), then f is in the Nevanlinna class and $h = \log|f|$ a.e. $[m]$ on $\partial\mathbb{D}$. Moreover, $f \in H^p$ if and only if $e^h \in L^p$.*

Proof. As we already observed, $\log|f| = \tilde{h}$ on $\mathbb{D}$. Thus

$$\begin{aligned}\frac{1}{2\pi}\int_0^{2\pi} \log^+|f(re^{i\theta})|\, d\theta &= \frac{1}{2\pi}\int_0^{2\pi} \max\{\tilde{h}_r(e^{i\theta}), 0\}\, d\theta \\ &\le ||\tilde{h}_r||_1.\end{aligned}$$

But Theorem 19.1.7 implies that this last term is uniformly bounded in r. Therefore $f \in N$. Also Fatou's Theorem implies that f has non-tangential limits a.e. on $\partial\mathbb{D}$ and so, when the limit exists, $h(w) = \lim_{r\to 1-} \tilde{h}(rw) = \lim_{r\to 1-} \log|f(rw)|$. This proves the first part of the proposition.

Since $|f|^p = (e^{\tilde{h}})^p = e^{p\tilde{h}}$,

$$\begin{aligned}\int |f(rw)|^p\, dm(w) &= \int e^{p\tilde{h}_r(w)}\, dm(w) \\ &= \int \exp\left(\int ph(z)\, P_{rw}(z)\, dm(z)\right) dm(w).\end{aligned}$$

Since the exponential function is convex and $P_{rw}(z)\, dm(z)$ is a probability measure, Proposition 19.4.13 implies that

$$\begin{aligned}\int |f(rw)|^p\, dm(w) &\le \int\int \exp(ph(z))\, P_{rw}(z)\, dm(z)\, dm(w) \\ &= \int \exp(ph(z)) \left(\int P_{rw}(z)\, dm(w)\right) dm(z) \\ &= \int \exp(ph(z))\, dm(z) \\ &= \int (e^h)^p\, dm,\end{aligned}$$

since $\int P_{rw}(z)\, dm(w) = \int P_{rz}(w)\, dm(w) = 1$. So if $e^h \in L^p$, then $f \in H^p$.

Conversely, if $f \in H^p$, then $f(e^{i\theta}) \in L^p$ and, since $e^h = |f|$, $e^h \in L^p$. □

We are now in a position to prove one of the main results of this section.

3.7 Factorization Theorem. *If $f \in H^\infty$, then*

$$f(z) = c\,b(z)\,\phi(z)\,F(z),$$

where c is a constant with $|c| = 1$, b is a Blaschke product, ϕ is a singular inner function, and F is an outer function in H^∞. Conversely, any function having this form belongs to H^∞.

Proof. Assume that f is a bounded analytic function with $||f||_\infty \le 1$. By Theorem 2.8, $f = c\,b\,g$, where b is a Blaschke product, g is a bounded analytic function on $\mathbb{D}$ with $g(0) > 0$, and c is a constant with $|c| = 1$. It also follows that $||g||_\infty \le 1$. Let $g = e^{-k}$ for a unique analytic function $k : \mathbb{D} \to \mathbb{C}$ with $k(0) = \log|g(0)|$. Thus $u \equiv \operatorname{Re} k = -\log|g| \ge 0$. That is, u is a non-negative harmonic function on $\mathbb{D}$. By Herglotz's Theorem, $u(z) = \int P_z(w)\,d\mu(w)$ for some positive measure μ on $\partial\,\mathbb{D}$. Therefore

$$k(z) = \int \frac{w+z}{w-z}\,d\mu(w).$$

(Why?) Let $\mu = \mu_a + \mu_s$ be the Lebesgue decomposition of μ with respect to m, with $\mu_a \ll m$ and $\mu_s \perp m$, and let h be the Radon-Nikodym derivative of μ_a with respect to m; so $\mu_a = h\,m$. It follows that $h \ge 0$ a.e. $[m]$ since μ is a positive measure. Define

$$F(z) = \exp\left(\int \frac{w+z}{w-z}\,[-h(w)]\,dm(w)\right)$$

and let ϕ be the singular inner function corresponding to the measure μ_s. It is easy to see that $g = e^{-k} = \phi\,F$. By (3.6), $-h = \log|F|$ a.e. $[m]$. Also $\log|g| = -\operatorname{Re} k = -u$ and $u(rw) \to h(w)$ a.e. $[m]$ by Theorem 19.2.12. Thus $-h = \log|g|$ on $\partial\,\mathbb{D}$ and this implies that $|g| = |F|$ on $\partial\,\mathbb{D}$. Therefore $F \in H^\infty$.

The converse is clear. $\square$

In light of Theorem 2.10 we can now factor functions in N.

3.8 Corollary. *If f is a function in the Nevanlinna class, then*

$$f(z) = c\,b(z)\left[\frac{\phi_1(z)}{\phi_2(z)}\right]\,F(z),$$

where c is a constant with $|c| = 1$, b is a Blaschke product, ϕ_1 and ϕ_2 are singular inner functions, and F is an outer function in N.

Now concentrate on H^p for finite p. It will be shown that Theorem 3.7 extends to this situation with the factor F an outer function in H^p. To do this we will first prove a result that has some independent interest.

3.9 Proposition. *If $f \in H^1$, then $f = hk$ for two functions h and k in H^2.*

Proof. Let b be the Blaschke product with the same zeros as f. By Theorem 2.8, $f = bg$, where $g \in H^1$ and g has no zeros in $\mathbb{D}$. Since $\mathbb{D}$ is simply connected, there is an analytic function h on $\mathbb{D}$ such that $h^2 = g$. Since $g \in H^1$, $h \in H^2$. Also $k = bh \in H^2$. Clearly, $f = hk$. □

3.10 Theorem. *If $1 \le p \le \infty$ and $f \in H^p$, then*

$$f(z) = c\, b(z)\, \phi(z)\, F(z),$$

where c is a constant with $|c| = 1$, b is a Blaschke product, ϕ is a singular inner function, and F is an outer function in H^p. Conversely, any function having this form belongs to H^p.

Proof. In light of Theorem 3.7 it remains to consider the case where $p < \infty$. Before getting into the proper part of the proof, the reader is asked to establish the inequality

$$|\log^+ a - \log^+ b| \le |a - b|,$$

valid for all positive numbers a and b. (Just consider various cases.) Hence

$$\begin{aligned} \int \left|\log^+|f(rw)| - \log^+|f(w)|\right| dm(w) &\le \int |f(rw) - f(w)|\, dm(w) \\ &\le \left[\int |f(rw) - f(w)|^p dm(w)\right]^{\frac{1}{p}}. \end{aligned}$$

Since this last quantity converges to 0 as $r \to 1-$, we obtain that

3.11 $$\lim_{r\to 1-} \int \left|\log^+|f(rw)| - \log^+|f(w)|\right| dm(w) = 0.$$

3.12 Claim. If $f \in H^p$, then $\log|f(z)| \le \int P_z(w) \log|f(w)|\, dm(w)$ for $|z| < 1$.

To see this, fix z in $\mathbb{D}$. By Theorem 2.8, $f = bg$, where b is a Blaschke product and g is a function in H^p with no zeros. Hence, $\log|f| = \log|b| + \log|g| \le \log|g|$. Since $|f(w)| = |g(w)|$ on $\partial\,\mathbb{D}$, it suffices to prove the claim for the function g.

Because g does not vanish on $\mathbb{D}$, $\log|g|$ is a harmonic function on $\mathbb{D}$, and so, for $0 < r < 1$, $\log|g_r|$ is harmonic in a neighborhood of cl $\mathbb{D}$. Thus $\log|g_r(z)| = \int P_z(w) \log|g_r(w)|\, dm(w)$. But for $|z| < 1$, $P_z \in L^\infty$. Therefore (3.11) implies that

$$\lim_{r\to 1-} \int P_z(w) \log^+|g_r(w)|\, dm(w) = \int P_z(w) \log^+|g(w)|\, dm(w).$$

Now Fatou's Lemma implies that $\int P_z(w)\ \log^-|g(w)|\,dm(w) \leq \liminf_{r\to 1-} \int P_z(w)\ \log^-|g_r(w)|\,dm(w)$. Hence

$$\begin{aligned}\log|g(w)| &= \lim_{r\to 1-} \log|g_r(w)| \\ &= \lim_{r\to 1-} \int P_z(w)\left[\log^+|g_r(w)| - \log^-|g_r(w)|\right] dm(w) \\ &\leq \int P_z \log^+|g|\,dm - \int P_z \log^-|g|\,dm \\ &= \int P_z \log|g|\,dm,\end{aligned}$$

thus proving (3.12).

Now to complete the proof of this theorem. Let $f = c\,b(\phi_1/\phi_2)\,F$ as in Corollary 3.8 and put $\phi = \phi_1/\phi_2$. Since $|f(w)| = |F(w)|$ a.e. $[m]$ on $\partial\mathbb{D}$, $F \in H^p$. Also $|\phi| = 1$ a.e. $[m]$ on $\partial\mathbb{D}$. So if it can be shown that $|b(z)\,\phi(z)| \leq 1$ on $\mathbb{D}$, this will show that $b\,\phi$ is an inner function and complete the proof. But for $|z| < 1$,

$$\begin{aligned}\log|F(z)| &= \int P_z(w)\log|F(w)|\,dm(w) \\ &= \int P_z(w)\log|f(w)|\,dm(w) \\ &\geq \log|f(z)|\end{aligned}$$

by (3.12). Therefore $1 \geq \left|\frac{f(z)}{F(z)}\right| = |b(z)\,\phi(z)|$. □

It is now possible to obtain the promised extension to H^1 of Theorems 1.4 and 1.5. The result is stated for all p, though it is only necessary to offer a proof for the case that $p = 1$. The subspace $\mathcal{H}^p$ of L^p is defined as before (1.5).

3.13 Theorem. *If $f \in H^p$, $1 \leq p \leq \infty$, then $g(w) = \lim_{r\to 1-} f(rw)$ defines a function in L^p such that $\hat{g}(n) = 0$ for $n < 0$, $\tilde{g} = f$, and $||g||_p = \lim_{r\to 1-} ||f_r||_p$. Thus the map that takes a function f in H^p onto its boundary values g is an isometric isomorphism of H^p onto the subspace $\mathcal{H}^p$ of L^p. For $p = \infty$, this map is also a weak* homeomorphism.*

Proof. We can assume that $p = 1$. If $f \in H^1$, then $f = hk$ for functions h and k in H^2 (3.9); also, let h and k denote the respective boundary functions in $\mathcal{H}^2$. By Theorem 1.4, $\tilde{h} = h$, $\tilde{k} = k$, $||h_r - h||_2 \to 0$, and $||k_r - k||_2 \to 0$. If $0 < r,\ s < 1$, then

$$\begin{aligned}\int |f(rw) - f(sw)|\,dm(w) &\leq \int |h(rw)[k(rw) - k(sw)]|\,dm(w) + \\ &\qquad \int |k(sw)[h(rw) - h(sw)]|\,dm(w)\end{aligned}$$

$$\leq \ ||h_r||_2||k_r - k_s||_2 + ||k_s||_2||h_r - h_s||_2.$$

If $M = \max\{||h||_2, ||k||_2\}$ and $\varepsilon > 0$, then there is an $r_0 < 1$ such that $||k_r - k_s||_2 < \varepsilon/2M$ and $||h_r - h_s||_2 < \varepsilon/2M$ for $r, s > r_0$. Hence $||f_r - f_s||_1 < \varepsilon$ whenever $r, s > r_0$. That is, $\{f_r\}$ is a Cauchy net in L^1. Let $F \in L^1$ such that $||f_r - F||_1 \to 0$ as $r \to 1-$. By Theorem 3.8, $f = \tilde{F}$. By Fatou's Theorem, $F(w) = \lim_{r\to 1-} f(rw) = g(w)$ a.e. $[m]$ on $\partial\,\mathbb{D}$. But since $f_r \to g$ in L^1 norm, $\hat{f}_r(n) \to \hat{g}(n)$ for all n. Thus $\hat{g}(n) = 0$ for $n < 0$. It is also clear that $||f_r||_1 \to ||g||_1$. □

We will henceforth make no distinction between functions in the Hardy spaces H^p and their boundary values. That is, with no warning we will consider functions in H^p as functions on $\mathbb{D}$ or on $\partial\,\mathbb{D}$ unless there is a distinct expository advantage in making a distinction.

We have seen that if $f \in L^p$, $1 \leq p \leq \infty$, and $\hat{f}(n) = 0$ for all $n < 0$, then f is the boundary function of a function in H^p. We are therefore justified in calling functions in L^p whose negative Fourier coefficients are 0 analytic functions. What are the "analytic measures?" That is, if μ is a regular Borel measure on $\partial\,\mathbb{D}$ and $\hat{\mu}(n) = 0$ for $n < 0$, what can we conclude, if anything, about μ and $\tilde{\mu}$? The answer, contained in the F and M Riesz Theorem below, is that we get nothing new, since such measures must be absolutely continuous with respect to Lebesgue measure and therefore have Radon-Nikodym derivatives equal to a function in the Hardy space H^1.

3.14 The F and M Riesz Theorem. *If $\mu \in M(\partial\,\mathbb{D})$ and $\hat{\mu}(n) = 0$ for $n < 0$, then $\mu \ll m$ and $d\mu/dm \in H^1$.*

Proof. By familiar arguments (see, for example, the proof of Theorem 1.3), if $f = \tilde{\mu}$, then $f(z) = \sum_{n=0}^{\infty} \hat{\mu}(n) z^n$ for $|z| < 1$. Thus f is an analytic function. Also, $||f_r||_1 \leq ||\mu||$ and so $f \in H^1$. If $g(w) = \lim_{r\to 1-} f(rw)$, then Theorem 3.13 implies that $\tilde{g} = f$. Hence $\widetilde{\mu - g} = 0$. But now an easy computation shows that for every integer n, $\widehat{\mu - g}(n) = 0$, and so $\hat{\mu}(n) = \hat{g}(n)$ for all n. Since the trigonometric polynomials are dense in $C(\partial\,\mathbb{D})$, it follows that $\mu = gm$. □

We close this section with a discussion of weak convergence in the H^p spaces. Part of this discussion (the part for $p > 1$) could have been presented earlier, while the consideration of the case where $p = 1$ is dependent on the F and M Riesz Theorem. We begin with the case $p > 1$.

3.15 Proposition. *If $1 < p \leq \infty$, $f \in H^p$, and $\{f_n\}$ is a sequence in H^p, then the following statements are equivalent.*

(a) *$f_n \to f$ weakly in L^p (weak* in L^∞ if $p = \infty$).*

(b) *$\sup_n ||f_n||_p < \infty$ and $f_n(z) \to f(z)$ uniformly on compact subsets of $\mathbb{D}$.*

(c) $\sup_n ||f_n||_p < \infty$ *and* $f_n(z) \to f(z)$ *for all* z *in* $\mathbb{D}$.

(d) $\sup_n ||f_n||_p < \infty$ *and* $f_n^{(k)}(0) \to f^{(k)}(0)$ *for all* $k \geq 0$.

Proof. (a) *implies* (b). By the Principle of Uniform Boundedness, $\sup_n ||f_n|| < \infty$. If K is a compact subset of $\mathbb{D}$, let $s = \sup\{|z| : z \in K\}$; so $s < 1$. Let $s < r < 1$; by Cauchy's Theorem, for any analytic function h on $\mathbb{D}$ and z in K,

$$h(z) = \frac{1}{2\pi i}\int_{|\zeta|=r} \frac{h(\zeta)}{\zeta - z}\, d\zeta = \frac{1}{2\pi}\int_0^{2\pi} \frac{h(re^{i\theta})}{re^{i\theta} - z}\, d\theta.$$

Now if $h \in H^p$, $h_r \to h$ in L^p norm. So letting $r \to 1$ in the preceding equation gives that

$$h(z) = \frac{1}{2\pi}\int_0^{2\pi} \frac{h(e^{i\theta})}{e^{i\theta} - z}\, d\theta$$

for all z in K. By Exercise 6, $\{(e^{i\theta} - z)^{-1} : z \in K\}$ is compact in L^q, where q is the index dual to p. By Exercise 7, $f_n(z) \to f(z)$ uniformly on K.

(c) *implies* (a). Assume that $1 < p < \infty$. The fact that $\{f_n\}$ is norm bounded implies, by the reflexivity of L^p, that there is a function g in L^p such that $f_n \to_{\text{cl}} g$ weakly in L^p. Since L^q is separable, there is a subsequence $\{f_{n_k}\}$ such that $f_{n_k} \to g$ weakly. If $p = \infty$, then the fact that L^1 is separable implies there is a g in L^∞ and a subsequence $\{f_{n_k}\}$ such that $f_{n_k} \to g$ weak*. In either case, for all integers m, $\hat{g}(m) = \lim \hat{f}_{n_k}(m)$. Hence $g \in H_p$. Also, the fact that $(e^{i\theta} - z)^{-1} \in L^q$ implies that $\tilde{g}(z) = \langle g, (e^{i\theta} - z)^{-1}\rangle = \lim_{n_k \to \infty}\langle f_{n_k}, (e^{i\theta} - z)^{-1}\rangle = \lim_{n_k\to\infty} f_{n_k}(z) = f(z)$. Hence f is the unique weak (respectively, weak*) cluster point of $\{f_n\}$ and so $f_n \to f$ weakly (respectively, weak*).

It is clear that (b) implies (c) and the proof that (d) is equivalent to the remaining conditions is left to the reader. □

What happens if $p = 1$? By the F and M Riesz Theorem, H^1 can be identified, isometrically and isomorphicly, with $\mathcal{L} \equiv \{\mu \in M(\partial\,\mathbb{D}) : \hat{\mu}(n) = 0$ for $n < 0\}$. Now it is easy to see that this is a weak* closed subspace of $M = M(\partial\,\mathbb{D})$. So if $\mathcal{L}_\perp = \{f \in C(\partial\,\mathbb{D}) : \int f\,h dm = 0$ for all h in $H^1\}$, then $H^1 \cong (C(\partial\,\mathbb{D})/\mathcal{L}_\perp)^*$. That is, H^1 is the dual of a Banach space and therefore has a weak* topology. In fact, this is precisely the relative weak* topology it inherits via its identification with the subspace $\mathcal{L}$ of M. Thus a sequence $\{f_n\}$ in H^1 converges weak* to f in H^1 if and only if $\int g f_n dm \to \int g f\, dm$ for all g in $C(\partial\,\mathbb{D})$. We therefore have the following analogue of the preceding proposition.

3.16 Proposition. *If* $f \in H^1$ *and* $\{f_n\}$ *is a sequence in* H^1, *then the following statements are equivalent.*

(a) $f_n \to f$ *weak* in* H^1.

(b) $\sup_n \|f_n\|_1 < \infty$ *and* $f_n(z) \to f(z)$ *uniformly on compact subsets of* $\mathbb{D}$.

(c) $\sup_n \|f_n\|_1 < \infty$ *and* $f_n(z) \to f(z)$ *for all* z *in* $\mathbb{D}$.

(d) $\sup_n \|f_n\|_1 < \infty$ *and* $f_n^{(k)}(0) \to (0)$ *for all* $k \geq 0$.

The proof is left to the reader.

Exercises

1. Let ϕ_1 and ϕ_2 be singular inner functions corresponding to the singular measures μ_1 and μ_2. If ϕ is the function ϕ_1/ϕ_2 in the Nevanlinna class, show that ϕ is an inner function if and only if $\mu_1 \geq \mu_2$.

2. Let $\mathcal{F}$ be the collection of all inner functions and observe that $\mathcal{F}$ is a semigroup under multiplication. If $\phi \in \mathcal{F}$, characterize all the divisors of ϕ. Apply this to the singular inner function ϕ corresponding to the measure $\alpha\delta_1$, where δ_1 is the unit point mass at 1 and $\alpha > 0$.

3. If ϕ is an inner function, is it possible for $1/\phi$ to belong to some H^p space? Can $1/\phi$ belong to some L^p space? If f is any H^p function such that $1/f \in H^1$, what can you say about f?

4. If $f \in H^1$ and $\operatorname{Re} f(z) > 0$ for all z, show that f is an outer function. If ϕ is an inner function, show that $1 + \phi$ is outer.

5. If Δ is a measurable subset of $\partial\,\mathbb{D}$ having positive Lebesgue measure and a and b are two positive numbers, show that there is an outer function f in H^∞ with $|f(w)| = a$ a.e. on Δ and $|f(w)| = b$ a.e. on $\partial\,\mathbb{D} \setminus \Delta$. Show that if $h \in L^p$, $1 \leq p \leq \infty$, and $h \geq 0$, then there is a function f in H^p such that $h = |f|$ a.e. on $\partial\,\mathbb{D}$ if and only if $\log h \in L^1$; show that the function f can be chosen to be an outer function.

6. If $1 \leq q < \infty$, the function $z \to (e^{i\theta} - z)^{-1}$ is a continuous function from $\mathbb{D}$ into L^q. Thus for any compact subset K of $\mathbb{D}$, $\{(e^{i\theta} - z)^{-1} : z \in K\}$ is a compact subset of L^q.

7. If $\mathcal{X}$ is a Banach space, $\{x_n\}$ is a sequence in $\mathcal{X}$ such that $x_n \to 0$ weakly, and K is a norm compact subset of $\mathcal{X}^*$, then $\sup\{|x^*(x_n)| : x^* \in K\} \to 0$ as $n \to \infty$.

8. Prove that condition (d) in Proposition 3.15 is equivalent to the remaining ones.

9. Prove Proposition 3.16.

10. Show that for f in L^1, $\int f\phi dm = 0$ for all ϕ in H^∞ if and only if $f \in H^1$ and $f(0) = 0$. Denote this subspace of H^1 by H^1_0. Show that H^∞ is isometrically isomorphic to $(L^1/H^1_0)^*$.

11. Let b_n be the Blaschke product with a zero of multiplicity n at $1 - \frac{1}{n}$. Show that $\{b_n\}$ converges weak* in H^∞. What is its limit?

12. (a) Show that if μ is a complex valued measure on $\partial\mathbb{D}$, (3.3) defines an analytic function ϕ on $\mathbb{D}$ with no zeros. (b) Show that $f(z) = -\int \frac{w+z}{w-z}\, d\mu(w)$ is the unique analytic function on $\mathbb{D}$ such that $\phi = e^f$ and $f(0) = -\mu(\partial\mathbb{D})$. (c) Show that $f^{(n)}(0) = -2n! \int \overline{w^n} d\mu(w)$ for $n \geq 0$. (d) Show that if μ is a singular measure and $\phi \equiv 1$, then $\mu = 0$. (e) Show that if μ is a real-valued measure and $\phi \equiv 1$, then $\mu = 0$. (f) Now assume that μ_1 and μ_2 are two positive singular measures that represent the same singular inner function ϕ. That is, assume that $\exp(-\int \frac{z+w}{w-z}\, d\mu_1(w)) = \exp(-\int \frac{w+z}{w-z} d\mu_2(w)) = \phi(z)$. Show that $\mu_1 = \mu_2$.

13. If ϕ is an inner function, find all the points a on $\partial\mathbb{D}$ such that ϕ has a continuous extension to $\mathbb{D} \cup \{a\}$.

§4 The Disk Algebra

Here we study an algebra of continuous functions on the closed disk (or the unit circle) that is related to the Hardy spaces.

4.1 Definition. The *disk algebra* is the algebra A of all continuous functions f on cl $\mathbb{D}$ that are analytic on $\mathbb{D}$.

It is easy to see that A is a Banach algebra with the supremum norm. In fact, it is a closed subalgebra of H^∞. In light of the recent sections of this book, the proof of the next result should offer little difficulty to the reader. (This was covered in Exercise 1.2.)

4.2 Theorem. *The map $f \to f|\partial\mathbb{D}$ is an isometric isomorphism of A onto the subalgebra $\mathcal{A} = \{g \in C(\partial\mathbb{D}) : \hat{g}(n) = 0$ for $n < 0\}$ of $C(\partial\mathbb{D})$. Furthermore, if $g \in \mathcal{A}$ and $f = \tilde{g}$, then $f|\partial\mathbb{D} = g$ and $f_r \to g$ uniformly on $\partial\mathbb{D}$.*

We will no longer make a distinction between the disk algebra A and the algebra $\mathcal{A}$ consisting of its boundary values. That is, we will often think of A as a subalgebra of $C(\partial\mathbb{D})$.

4.3 Proposition. *The analytic polynomials are uniformly dense in the disk algebra.*

Proof. This is easy to prove if you first observe that the Cesàro means of a function in the disk algebra are analytic polynomials and then apply Theorem 18.7.5. □

4.4 Theorem. *If $\rho : A \to \mathbb{C}$ is a non-zero homomorphism, then there is a point a in $\operatorname{cl}\mathbb{D}$ such that $\rho(f) = f(a)$ for all f in A. Thus the maximal ideal space of A is homeomorphic to $\operatorname{cl}\mathbb{D}$, and under this identification the Gelfand transform is the identity map.*

Proof. If $\rho : A \to \mathbb{C}$ is a non-zero homomorphism, let $a = \rho(z)$. Since $||\rho|| = 1$, $|a| \leq 1$. It follows by algebraic manipulation that $\rho(p) = p(a)$ for all polynomials in z. In light of the preceding proposition, $\rho(f) = f(a)$ for all f in A. Conversely, if $|a| \leq 1$ and $\rho(f) = f(a)$ for all f in A, then ρ is a homomorphism. Thus $\rho \to \rho(z)$ is a one-to-one correspondence between the maximal ideal space and cl $\mathbb{D}$. The proof of the fact that this correspondence is a homeomorphism and the concomitant fact about the Gelfand transform is left to the reader. □

4.5 Proposition. $\{\operatorname{Re} p|\partial\mathbb{D} :$ *p is an analytic polynomial$\}$ is uniformly dense in $C_{\mathbb{R}}(\partial\mathbb{D})$.*

Proof. If $g(w) = \sum_{k=-n}^{n} C_k w^k$, where $C_{-k} = \overline{C}_k$, then $g = \operatorname{Re} p$ for some analytic polynomial. On the other hand, the Cesàro means of any function in $C_{\mathbb{R}}(\partial\mathbb{D})$ are such trigonometric polynomials and the means converge uniformly on $\partial\mathbb{D}$ (18.7.5). □

4.6 Corollary. *If μ is a real-valued measure on $\partial\mathbb{D}$ such that $\int p\, d\mu = 0$ for every analytic polynomial p, then $\mu = 0$.*

4.7 Corollary. *If μ is a real-valued measure on $\partial\mathbb{D}$ such that $\int p\, d\mu = 0$ for every analytic polynomial p with $p(0) = 0$, then there is a real constant c such that $\mu = cm$.*

Proof. Let $c = \int 1 dm = \mu(\partial\mathbb{D})$. It is easy to check that, if $\nu = \mu - cm$, then $\int p d\nu = 0$ for all polynomials p. The result now follows from the preceding corollary. □

This essentially completes the information we will see here about the disk algebra. There is more information available in the references.

Now we turn our attention to some related matters that will be of use later. Note that this puts the finishing touches to the proof of Theorem 14.5.8.

4.8 Theorem. *If $f \in H^1$, the following statements are equivalent.*

(a) *The function $\theta \to f(e^{i\theta})$ is of bounded variation on $[0, 2\pi]$.*

(b) *The function f belongs to the disk algebra and $\theta \to f(e^{i\theta})$ is absolutely continuous.*

(c) *The derivative of f belongs to H^1.*

Proof. (a) *implies* (b). Let $u(\theta) = f(e^{i\theta})$; we are assuming here that u is a function of bounded variation. But since $f \in H^1$, $\frac{1}{2\pi}\int_0^{2\pi} u(\theta)e^{in\theta}d\theta = \int f(w)w^n\, dm(w) = 0$ for $n \geq 1$. Using integration by parts, this implies that, for $n \geq 1$,

$$\begin{aligned} 0 &= \frac{u(\theta)}{in}e^{in\theta}\Big|_0^{2\pi} - \frac{1}{in}\int_0^{2\pi} e^{in\theta}du(\theta) \\ &= -\frac{1}{in}\int_0^{2\pi} e^{in\theta}du(\theta). \end{aligned}$$

The F and M Riesz Theorem now implies that u is an absolutely continuous function whose negative Fourier coefficients vanish. In particular, u is continuous and, since $f = \tilde{u}$, $f \in A$.

(b) *implies* (c). Since $u(\theta) = f(e^{i\theta})$ is absolutely continuous, for $0 < r < 1$ and for all θ,

$$f(re^{i\theta}) = \frac{1}{2\pi}\int_0^{2\pi} P_r(\theta - t)f(e^{it})dt.$$

Differentiating both sides with respect to θ gives

$$ire^{i\theta}f'(re^{i\theta}) = \frac{1}{2\pi}\int_0^{2\pi}\frac{\partial}{\partial\theta}[P_r(\theta - t)]f(e^{it})dt.$$

Since P_r is an even function, this implies

$$ire^{i\theta}f'(re^{i\theta}) = \frac{1}{2\pi}\int_0^{2\pi}\frac{\partial}{\partial t}[P_r(\theta - t)]u(t)\, dt.$$

Since u is absolutely continuous, integration by parts yields

$$ire^{i\theta}f'(re^{i\theta}) = \frac{1}{2\pi}\int_0^{2\pi} P_r(\theta - t)u'(t)\, dt.$$

This implies that $izf'(z)$ is an analytic function on $\mathbb{D}$ that is the Poisson integral of the L^1 function u'. Hence $izf'(z)$ belongs to H^1. But this implies that $f' \in H^1$.

(c) *implies* (b). Let h denote the boundary values of f'. So $h \in L^1$, $\hat{h}(-n) = 0$ for $n > 0$, and f' is the Poisson integral of h. Let $g(\theta) = \int_0^\theta i\, e^{it}\, h(t)\, dt$; so $g(0) = 0$. Also $g(2\pi) = \int_0^{2\pi} i\, e^{it}\, h(t)\, dt = i\int_0^{2\pi} e^{-i(-t)}dt = i\hat{h}(-1) = 0$. Now g is absolutely continuous and $g'(\theta) = ie^{i\theta}h(\theta)$ a.e. For $n < 0$,

$$\hat{g}(n) = \frac{1}{2\pi}\int_0^{2\pi} g(\theta)\, e^{-in\theta}d\theta$$

$$
\begin{aligned}
&= -\frac{1}{in}\,\frac{1}{2\pi}\int_0^{2\pi} g(\theta)\,d(e^{-in\theta}) \\
&= -\frac{1}{2\pi\, n}\,e^{-in\theta} g(\theta)\,|_0^{2\pi} + \frac{1}{2\pi\, in}\int_0^{2\pi} e^{-in\theta} g'(\theta)\,d\theta \\
&= \frac{1}{2\pi\, n}\int_0^{2\pi} h(\theta)\,e^{-i(n-1)\theta} d\theta \\
&= \frac{1}{2\pi\, n}\,\hat{h}(n-1) \\
&= 0.
\end{aligned}
$$

Thus $\tilde{g}$ belongs to the disk algebra and $\tilde{g}|\partial\,\mathbb{D} = g$ is absolutely continuous. Since we have already shown that (b) implies (c), we have that $\tilde{g}' \in H^1$. Moreover an examination of the proof that (b) implies (c) reveals that

$$\frac{d}{d\theta} g(e^{i\theta}) = \lim_{r\to 1-} ie^{i\theta}\tilde{g}(re^{i\theta}).$$

But $\frac{d}{d\theta} g(e^{i\theta}) = ie^{i\theta}h(\theta) = ie^{i\theta}f'(e^{i\theta})$. Thus

$$\lim_{r\to 1-} [f'(re^{i\theta}) - \tilde{g}'(re^{i\theta})] = 0$$

a.e. Since $f' - \tilde{g}' \in H^1$, $f' = \tilde{g}'$. Therefore there is a constant C such that $f = \tilde{g}+C$. This implies that $f \in A$ and $\theta \to f(e^{i\theta})$ is absolutely continuous.

Since it is clear that (b) implies (a), this completes the proof. □

It is worth recording the following fact that surfaced in the preceding proof.

4.9 Corollary. *If f belongs to the disk algebra and $\theta \to f(e^{i\theta})$ is absolutely continuous, then*

$$\frac{d}{d\theta} f(e^{i\theta}) = ie^{i\theta} \lim_{r\to 1-} f'(re^{i\theta}) a.e.$$

4.10 Corollary. *If f belongs to the disk algebra and $\theta \to f(e^{i\theta})$ is absolutely continuous, then the length of the curve $\theta \to f(e^{i\theta})$ is $2\pi\, ||f'||_1$.*

Exercises

1. Show that the only inner functions that belong to A are the finite Blaschke products.

2. Show that the function $f(z) = (1-z)\exp\left[\frac{1+z}{1-z}\right]$ is continuous as a function on $\partial\,\mathbb{D}$, but there is no function in A that equals f on $\partial\,\mathbb{D}$.

3. Let K be a compact subset of $\partial\mathbb{D}$ having zero Lebesgue measure. (a) Show that there is an integrable, continuous function $w : \partial\mathbb{D} \to [-\infty, -1]$ such that $w(z) = -\infty$ if and only if $z \in K$. (b) Prove the result of Fatou that says there is a function f in A with $f(z) = 0$ if and only if $z \in K$.

4. Let γ be a rectifiable Jordan curve and $G = \text{ins }\gamma$; let $\tau : \mathbb{D} \to G$ be a Riemann map and extend τ to be a homeomorphism of $\text{cl }\mathbb{D}$ onto $\text{cl }G$. Generalize Corollary 4.10 by showing that, if Δ is a Borel subset of γ, the arc length measure of Δ is $2\pi \int_{f^{-1}(\Delta)} |\tau'(w)|\, dm(w)$.

5. Keep the notation of the preceding exercise. Show that if E is a subset of $\partial\mathbb{D}$, then $m(E) = 0$ if and only if $\tau(E)$ is a measurable subset of γ with arc length measure 0.

6. Keep the notation of Exercise 4. Let $a = e^{i\alpha}$ and let $\sigma : [0,1] \to \mathbb{D} \cup \{a\}$ be a curve with $\sigma(1) = a$ and $|\sigma(t)| < 1$ for $0 \le t < 1$. Assume that σ has a well defined direction at a that is not tangent to $\partial\mathbb{D}$; that is, assume that $\theta \equiv \lim_{t\to 1-} \arg[\sigma(t) - a]$ exists and $\theta \ne \alpha \pm \pi/2$. If $\frac{d}{dt}\,\tau(e^{it})$ exists at $t = \alpha$, does the angle between $\tau \circ \sigma$ and γ exist at $\tau(a)$ and is it equal to θ? Is this true a.e. on $\partial\mathbb{D}$?

§5 The Invariant Subspaces of H^p

The purpose of this section is to prove the following theorem (Beurling [1949]).

5.1 Beurling's Theorem. *If $1 \le p < \infty$ and $\mathcal{M}$ is a closed linear subspace of H^p such that $z\mathcal{M} \subseteq \mathcal{M}$ and $\mathcal{M} \ne (0)$, then there is a unique inner function ϕ with $\phi(0) \ge 0$ such that $\mathcal{M} = \phi H^p$. If $p = \infty$ and $\mathcal{M}$ is a weak* closed subspace of H^∞ such that $z\mathcal{M} \subseteq \mathcal{M}$ and $\mathcal{M} \ne (0)$, then there is a unique inner function ϕ with $\phi(0) \ge 0$ such that $\mathcal{M} = \phi H^\infty$.*

A subspace $\mathcal{M}$ of H^p such that $z\mathcal{M} \subseteq \mathcal{M}$ is called an *invariant subspace* of H^p. Strictly speaking, such a subspace is an invariant subspace of the operator defined by multiplication by the independent variable. Clearly if $\mathcal{M} = \phi H^p$ for some inner function ϕ, then $\mathcal{M}$ is a closed invariant subspace of H^p. So Beurling's Theorem characterizes the invariant subspaces of H^p.

Beurling's Theorem is one of the most celebrated in functional analysis. It is one of the first results that make a deep connection between operator theory and function theory. For a proof in the case that $p = 2$ that uses only operator theory, see Theorem I.4.12 in Conway [1991]. The proof here will require some additional work.

Begin by introducing an additional class of functions related to the Nevanlinna class. Let N^+ denote those functions f in the class N that have a factorization $f = c\,b\,\phi\,F$, where $c \in \mathbb{C}$ with $|c| = 1$, b is a Blaschke product, ϕ is an inner function, and F is an outer function. Referring to the factorization of functions in the Nevanlinna class (Corollary 3.8), we see that the functions in N^+ are precisely those in N for which no inner function is required in the denominator of this factorization. The first result will be stated but not proved. Its somewhat difficult proof is left to the reader. (Also see Duren [1970], Theorem 2.10.)

5.2 Theorem. *If $f \in N$, then $f \in N^+$ if and only if*

$$\lim_{r\to 1-} \frac{1}{2\pi}\int_0^{2\pi} \log^+ |f(re^{i\theta})|\, d\theta = \frac{1}{2\pi}\int_0^{2\pi} \log^+ |f(e^{i\theta})|\, d\theta.$$

Note that if ϕ is a singular inner function, $1/\phi \in N$ but $1/\phi \notin N^+$. For $f = 1/\phi$ the left hand side of this equality is finite and not 0 while the right hand side is 0.

The next result is an easy consequence of the definition on the class N^+.

5.3 Proposition. *If $f \in N^+$ and $f \in L^p$, $1 \le p \le \infty$, then $f \in H^p$.*

To facilitate the proof of Beurling's Theorem introduce the notation $[f] =$ the closed linear span of $\{f, zf, z^2 f, \ldots\}$ in H^p whenever $f \in H^p$, $1 \le p < \infty$. Note that $[f] = \text{cl}\,\{pf : p \text{ is a polynomial}\}$ and $[f]$ is the smallest invariant subspace of H^p that contains the function f. If $f \in H^\infty$, $[f]$ is the weak* closed linear span of the same set.

5.4 Lemma. *If f is an outer function in H^p, $1 \le p \le \infty$, then $[f] = H^p$.*

Proof. If $[f] \ne H^p$, then there is a continuous linear functional L on H^p (L is weak* continuous if $p = \infty$) such that $L(pf) = 0$ for every polynomial p but $L \ne 0$. Let q be the index conjugate to p and let $g \in L^q$ such that $L(h) = \int hg\, dm$ for every h in H^p. So $\int pfg\, dm = 0$ for every polynomial p, but there is at least one integer $n \ge 0$ with $\int z^n g\, dm \ne 0$. Thus $g \notin H_0^q$, the H^q functions that vanish at 0. On the other hand we do have that $\int z^n fg\, dm = 0$ for all $n \ge 0$, so that $k = fg \in H_0^1$.

Now f has no zeros in $\mathbb{D}$ and so k/f is an analytic function on $\mathbb{D}$. Since $k \in H_0^1$ and $f \in H^p$, $\log|k|$ and $\log|f|$ both belong to L^1. Therefore $\log|k/f| = \log|k| - \log|f| \in L^1$. Using the fact that f is outer and $k \in H^1$, it follows that $k/f \in N^+$. But on $\partial\mathbb{D}$, $k/f = g \in L^q$. Thus Proposition 5.3 implies that $g \in H_0^q$, a contradiction. □

We now fix our attention on the case $p = 2$.

5.5 Lemma. *Beurling's Theorem is true for $p = 2$.*

Proof. First, let g be a function in $\mathcal{M}$ such that the order of its zero at $z = 0$ is the smallest of all the functions in $\mathcal{M}$. It must be that $g \notin z\mathcal{M}$. Indeed, if $g \in z\mathcal{M}$, then $g = zf$ for some f in $\mathcal{M}$; but the order of the zero of f at 0 is one less than the order of the zero of g at 0, a contradiction. Thus $\mathcal{M}$ properly contains $z\mathcal{M}$.

Let $f\phi \in \mathcal{M} \cap (z\mathcal{M})^\perp$ with $||\phi||_2 = 1$. So for all $n \geq 1$, $0 = \langle \phi, \phi z^n \rangle = \int |\phi|^2 \overline{z^n} dm$. By taking complex conjugates we see that all the Fourier coefficients of $|\phi|^2$ are zero except possibly for the coefficient for $n = 0$. Thus $|\phi|^2$ is the constant function. Since $||\phi||_2 = 1$, we have that $|\phi| \equiv 1$, and so ϕ is an inner function.

Claim. $\dim[\mathcal{M} \cap (z\mathcal{M})^\perp] = 1$.

Indeed let ψ be a function in $\mathcal{M} \cap (z\mathcal{M})^\perp$ such that $\psi \perp \phi$ and $||\psi||_2 = 1$. It is left to the reader to show that $\{\overline{z^n}\phi, \overline{z^k}\psi : n, k \geq 0\}$ is an orthonormal set in L^2. Therefore for any $n, k \geq 0$, $0 = \langle \overline{z^n}\phi, \overline{z^k}\psi \rangle = \int \overline{z^n} z^k \phi \overline{\psi} dm$. Thus $0 = \int h\phi\overline{\psi} dm$ for every h in $L^1(\partial\mathbb{D})$ and so $\phi\overline{\psi} = 0$. Since both ϕ and ψ must be inner functions, this is impossible. This contradiction shows that $\dim[\mathcal{M} \cap (z\mathcal{M})^\perp] = 1$ and hence is spanned by ϕ.

Let $\mathcal{N} = [\phi]$; so $\mathcal{N} \leq \mathcal{M}$. Let $h \in \mathcal{M} \cap \mathcal{N}^\perp$. So $h \perp \phi$; since $\dim[\mathcal{M} \cap (z\mathcal{M})^\perp] = 1$, this says that $h \in z\mathcal{M}$. But an easy argument shows that $\dim[z\mathcal{M} \cap (z^2\mathcal{M})^\perp] = 1$. Since $h \perp z\phi$, we also get that $h \in z^2\mathcal{M}$. Continuing this argument we arrive at the fact that $h \in z^n\mathcal{M}$ for all $n \geq 1$. But this says that h is an analytic function with a zero at $z = 0$ of infinite order. Hence $h = 0$ and it must be that $\mathcal{M} = \mathcal{N} = [\phi]$. □

Now we prove Beurling's Theorem for the case $p = \infty$.

5.6 Lemma. *If $\mathcal{M}$ is a weak* closed invariant subspace of H^∞ and $\mathcal{M} \neq (0)$, then there is an inner function ϕ such that $\mathcal{M} = \phi H^\infty$.*

Proof. Let $\mathcal{N}$ be the closure of $\mathcal{M}$ in H^2. It is immediate that $\mathcal{N}$ is an invariant subspace of H^2. By Lemma 5.5 there is an inner function ϕ such that $\mathcal{N} = \phi H^2$. It is claimed that $\mathcal{M} = \phi H^\infty$. Actually it is easy to see that $\mathcal{M} \subseteq \phi H^\infty$ since $\mathcal{M} \subseteq \mathcal{N} \cap H^\infty$.

For the other inclusion, let $\{f_n\} \subseteq \mathcal{M}$ such that $||f_n - \phi||_2 \to 0$. By passing to a subsequence if necessary, we may assume that $f_n \to \phi$ a.e. $[m]$ on $\partial\mathbb{D}$. Since $\mathcal{M} \subseteq \phi H^\infty$, for each n there is a function h_n in H^∞ such that $f_n = \phi h_n$. Define v_n on $\partial\mathbb{D}$ by $v_n(w) = 1$ when $|h_n(w)| \leq 1$ and $v_n(w) = |h_n(w)|$ otherwise. Now $v_n^{-1} \in L^\infty$ and $\log(v_n^{-1}) \in L^1$. Thus there is an outer function g_n in H^∞ such that $|g_n| = v_n^{-1}$ on $\partial\mathbb{D}$ (Proposition 3.6). But $f_n \to \phi$ a.e on $\partial\mathbb{D}$ and so $h_n = \overline{\phi} f_n \to \overline{\phi}\phi = 1$ a.e. on $\partial\mathbb{D}$. This in turn implies that $g_n \to 1$ a.e. on $\partial\mathbb{D}$. But $||g_n||_\infty \leq 1$ for all n and so $g_n \to 1$ weak* in H^∞. By Proposition 3.15, $g_n(z) \to 1$ for all z in $\mathbb{D}$. Thus $g_n(z)f_n(z) \to \phi(z)$ for z in $\mathbb{D}$. But $||g_n f_n||_\infty = ||v_n^{-1} h_n||_\infty \leq 1$. By Proposition 3.15, $g_n f_n \to \phi$ weak* in H^∞. Since $g_n f_n \in \mathcal{M}$, $\phi \in \mathcal{M}$. Therefore $\phi H^\infty \subseteq \mathcal{M}$. □

The proof of the next lemma is immediate from the factorization of functions in H^p. The details are left to the reader.

5.7 Lemma. *If ϕ and ψ are inner functions, then $\phi H^p = \psi H^p$ if and only if there is a scalar α with $|\alpha| = 1$ and $\psi = \alpha\phi$.*

Proof of Beurling's Theorem. Let $\mathcal{M}$ be an invariant subspace of H^p, $1 \leq p < \infty$. Because $L^q \subseteq L^1$, $\mathcal{M} \cap H^\infty$ is weak* closed and invariant in H^∞. If $g \in \mathcal{M}$ and $g = g_0 g_1$ with g_0 outer and g_1 inner, let $\{p_n\}$ be a sequence of polynomials such that $||p_n g_0 - 1||_p \to 0$ (Lemma 5.4). It follows that $p_n g \to g_1$ in L_p and so $g_1 \in \mathcal{M} \cap H^\infty$. That is, the inner factor of every function in $\mathcal{M}$ belongs to $\mathcal{M} \cap H^\infty$. In particular, $\mathcal{M} \cap H^\infty \neq (0)$. Thus (Lemma 5.6) there is an inner function ϕ such that $\mathcal{M} \cap H^\infty = \phi H^\infty$.

We now show that $\mathcal{M} = \phi H^p$. In fact if $g \in \mathcal{M}$ and $g = g_0 g_1$ as in the preceding paragraph, $g_1 \in \phi H^\infty$; let $g_1 = \phi\psi$ for some ψ in H^∞. So $g = \phi\psi g_0 \in \phi H^p$; that is, $\mathcal{M} \subseteq \phi H^p$. On the other hand, $\phi \in \mathcal{M}$ and so $\phi p \in \mathcal{M}$ for every polynomial p. By taking limits we get that $\phi H^p \subseteq \mathcal{M}$.

The uniqueness statement follows immediately from Lemma 5.7. □

Some notes are in order. First, H^∞ is a Banach algebra and the invariant subspaces of H^∞ are precisely the ideals of this algebra. So Beurling's Theorem characterizes the weak* closed ideals of H^∞. In Exercise 1 the weak* continuous homomorphisms from H^∞ into $\mathbb{C}$ are characterized. A discussion of the Banach Algebra H^∞ is a story by itself. (For example, see Garnett [1981].)

Second, if $\mathcal{L}$ = the collection of invariant subspaces of H^p, then $\mathcal{L}$ forms a lattice where the join and meet operations are defined as follows. If $\mathcal{M}$ and $\mathcal{N} \in \mathcal{L}$,

$$\begin{aligned} \mathcal{M} \vee \mathcal{N} &= \text{cl}\,[\text{span}(\mathcal{M} \cup \mathcal{N})], \\ \mathcal{M} \wedge \mathcal{N} &= \mathcal{M} \cap \mathcal{N}. \end{aligned}$$

It is left to the reader to check that $\mathcal{L}$ with these operations satisfies the axioms of a lattice. In the exercises Beurling's Theorem is applied to the study of this lattice. Also see §3.10 in Conway [1991].

Exercises

1. If $\rho : H^\infty \to \mathbb{C}$ is a non-zero weak* continuous homomorphism, show that there is a unique point a in $\mathbb{D}$ such that $\rho(f) = f(a)$ for all f in H^∞.

   Exercises 2 through 14 are interdependent.

2. Let I denote the set of all inner functions ϕ of the form

$$\phi = z^k b(z)\psi(z),$$

where $k \geq 0$, b is a Blaschke product with $b(0) > 0$ (possibly $b \equiv 1$), and ψ is a singular inner function (possibly $\psi \equiv 1$). Call this the *canonical factorization* of a function in I. Show that for an inner function ϕ, $\phi^{(n)}(0) > 0$ for the first positive integer n with $\phi^{(n)}(0) \neq 0$ if and only if $\phi \in I$.

3. Note that I is a semigroup under multiplication and that 1 is the identity of I. Show that this semigroup has no zero divisors; that is, if ϕ and $\psi \in I$ and $\phi\psi = 1$, then $\phi = \psi = 1$. Show that this makes it possible to define the greatest common divisor and least common multiple of two functions in I, and that they are unique when they exist. If ϕ_1 and $\phi_2 \in I$, then $\gcd(\phi_1, \phi_2) =$ the *greatest common divisor* of ϕ_1 and ϕ_2 and $\operatorname{lcm}(\phi_1, \phi_2) =$ the *least common multiple* of ϕ_1 and ϕ_2 when they exist. The next exercise guarantees the existence.

4. If ϕ_1 and $\phi_2 \in I$ and $1 \leq p \leq \infty$, then: (a) $\phi_1 H^p \wedge \phi_2 H^p = \phi H^p$, where $\phi = \operatorname{lcm}(\phi_1, \phi_2)$; (b) $\phi_1 H^p \vee \phi_2 H^p = \phi H_p$, where $\phi = \gcd(\phi_1, \phi_2)$.

5. For ϕ_1 and ϕ_2 in I, say that $\phi_1 \geq \phi_2$ if $\phi_2 | \phi_1$; that is, if ϕ_1 is a multiple of ϕ_2. Henceforth it will always be assumed that I has this ordering. It is customary to define $\phi_1 \vee \phi_2 \equiv \operatorname{lcm}(\phi_1, \phi_2)$ and $\phi_1 \wedge \phi_2 \equiv \gcd(\phi_1, \phi_2)$; with these definitions show that I becomes a lattice and the map $\phi \to \phi H^p$ is a lattice anti-isomorphism from I onto $\mathcal{L}_0$, the lattice of non-zero invariant subspaces of H^p.

6. Let $\phi_1, \phi_2 \in I$ and let $\phi_j = z^{k_j} b_j \psi_j$ be the canonical factorization of $\phi_j, j = 1, 2$. Furthermore, let μ_j be the positive singular measure on $\partial\mathbb{D}$ associated with the singular function ψ_j. Prove that $\phi_1 \geq \phi_2$ if and only if: (i) $k_1 \geq k_2$; (ii) the zeros of b_1 contain the zeros of b_2, counting multiplicities; (iii) $\mu_1 \geq \mu_2$.

7. Let μ_1 and μ_2 be two positive measures on a compact set X and set $\mu = \mu_1 + \mu_2$; set $f_j =$ the Radon-Nikodym derivative of μ_j with respect to $\mu, g = \max(f_1, f_2)$, and $h = \min(f_1, f_2)$. Put $\nu = g\mu$ and $\eta = h\mu$. Prove the following. (a) $\mu_1, \mu_2 \leq \nu$ and if σ is any positive measure on X such that $\mu_1, \mu_2 \leq \sigma$, then $\nu \leq \sigma$. (b) $\mu_1, \mu_2 \geq \eta$ and if σ is any positive measure on X such that $\mu_1, \mu_2 \geq \sigma$, then $\eta \geq \sigma$. For two positive measures μ_1 and μ_2, the measures ν and η will be denoted by $\mu_1 \vee \mu_2$ and $\mu_1 \wedge \mu_2$, respectively.

8. Let ϕ_1 and ϕ_2 be functions in I with canonical factorizations $\phi_j = z^{k_j} b_j \psi_j$, where b_j is a Blaschke product with zeros Z_j, each repeated as often as its multiplicity, and ψ_j is a singular inner function with corresponding singular measure μ_j. Let $k_1 \vee k_2 = \max(k_1, k_2)$ and $k_1 \wedge k_2 = \min(k_1, k_2)$; $b_1 \vee b_2 =$ the Blaschke product with zeros $Z_1 \cup Z_2$ and $b_1 \wedge b_2 =$ the Blaschke product with zeros $Z_1 \cap Z_2$; $\psi_1 \vee \psi_2 =$

the singular inner function with measure $\mu_1 \vee \mu_2$ and $\psi_1 \wedge \psi_2 =$ the singular inner function with measure $\mu_1 \wedge \mu_2$. Prove the following.
(a) $\phi_1 \vee \phi_2 = z^{k_1 \vee k_2}(b_1 \vee b_2)(\psi_1 \vee \psi_2)$ and $\phi_1 H^p \wedge \phi_2 H^p = (\phi_1 \vee \phi_2)H^p$.
(b) $\phi_1 \wedge \phi_2 = z^{k_1 \wedge k_2}(b_1 \wedge b_2)(\psi_1 \wedge \psi_2)$ and $\phi_1 H^p \wedge \phi_2 H^p = (\phi_1 \wedge \phi_2)H^p$.

9. With the notation of the preceding exercise, $\phi_1 H^p \vee \phi_2 H^p = H^p$ if and only if $k_1 = k_2 = 0$, $Z_1 \cap Z_2 = \emptyset$, and $\mu_1 \perp \mu_2$.

10. Show that if $1 \le p < \infty$ and $\mathcal{M}_1$ and $\mathcal{M}_2$ are non-zero invariant subspaces of H^p, then $\mathcal{M}_1 \cap \mathcal{M}_2 \neq (0)$. State and prove the corresponding fact for weak* closed ideals of H^∞.

11. Using the notation of Exercise 8, show that if $k_1 = k_2 = 0$, Z_1 and Z_2 have disjoint closures, and μ_1 and μ_2 have disjoint closed supports, then $\phi_1 H^p + \phi_2 H^p = H^p$. Show that $\phi_1 H^p + \phi_2 H^p = H^p$ if and only if there are functions u_1 and u_2 in H^∞ such that $\phi_1 u_1 + \phi_2 u_2 = 1$. (Remark. This condition is equivalent to the requirement that $\inf\{|\phi_1(z)| + |\phi_2(z)| : |z| < 1\} > 0$. This is a consequence of the Corona Theorem (Carleson [1962])).

12. If $\{\phi_i\} \subseteq I$, describe the inner functions ϕ and ψ such that $\bigvee \phi_i H^p = \phi H^p$ and $\bigwedge \phi_i H^p = \psi H^p$. Give necessary and sufficient conditions that $\bigvee \phi_i H^p = H^p$. That $\bigwedge \phi_i H^p = (0)$.

13. Let ϕ be the singular inner function corresponding to the measure $\alpha\delta_1$, $\alpha > 0$. If $\mathcal{M}_1$ and $\mathcal{M}_2$ are invariant subspaces for H^p, $1 \le p < \infty$, and both contain ϕH^p, show that either $\mathcal{M}_1 \subseteq \mathcal{M}_2$ or $\mathcal{M}_2 \subseteq \mathcal{M}_1$. (See Sarason [1965] for more on this situation.)

14. Let $\mathcal{M}_1$ and $\mathcal{M}_2$ be invariant subspaces of H^p such that $\mathcal{M}_1 \subseteq \mathcal{M}_2$. If ϕ_1 and ϕ_2 are the corresponding functions in I, give a necessary and sufficient condition in terms of ϕ_1 and ϕ_2 that $\dim(\mathcal{M}_2/\mathcal{M}_1) < \infty$. If $\dim(\mathcal{M}_2/\mathcal{M}_1) > 1$, show that there is an invariant subspace $\mathcal{M}$ such that $\mathcal{M}_1 \subseteq \mathcal{M} \subseteq \mathcal{M}_2$ and $\mathcal{M} \neq \mathcal{M}_1$ or $\mathcal{M}_2$.

15. (Conway [1973]) Endow I with the relative weak* topology from H^∞, and for each ϕ in I let P_ϕ be the orthogonal projection of H^2 onto ϕH^2. Show that if a sequence $\{\phi_n\}$ in I converges weak* to ϕ in I, then $P_{\phi_n} \to P_\phi$ in the strong operator topology.

16. Show that the inner functions are weak* dense in the unit ball of H^∞.

§6 Szegö's Theorem

In this section we will study the spaces $P^p(\mu)$, the closure of the polynomials in $L^p(\mu)$ for a measure μ supported on the unit circle. For $p = \infty$, $P^\infty(\mu)$

denotes the weak* closure of the polynomials in $L^\infty(\mu)$. In particular we will prove Szegö's Theorem (6.6).

Let A denote the disk algebra (§4) and let $A_0 = \{f \in A : f(0) = 0\}$; so A_0 is a maximal ideal in the algebra A. If μ is a positive measure on $\partial\mathbb{D}$, let $P_0^p(\mu)$ be the closure of A_0 in $L^p(\mu)$. ($p = \infty$?) Since $P^p(\mu)$ is the closure of A in $L^p(\mu)$, we have that $\dim[P^p(\mu)/P_0^p(\mu)] \leq 1$. If $\mu = m$, Lebesgue measure on $\partial\mathbb{D}$, then this dimension equals 1. Can it be 0? The answer is easily seen to be yes by taking μ to be the unit point mass at 1. However the general answer is the key to this section.

First, a few recollections from measure theory. If μ is a measure and $\mu = \nu + \eta$, where ν and η are mutually singular measures, then for $1 \leq p \leq \infty$ the space $L^p(\mu)$ splits into a direct sum, $L^p(\mu) = L^p(\nu) \oplus L^p(\eta)$. For a function f in $L^p(\mu)$ this decomposition is achieved by restricting f to the two disjoint carriers of the mutually singular measures. Thus for $f = g \oplus h$, $||f||_p = ||g||_p + ||h||_p$ when $1 \leq p < \infty$ and $||f||_\infty = \max\{||g||_\infty, ||h||_\infty\}$. This natural decomposition will be always lurking in the background of the discussion that follows.

6.1 Proposition. *If $1 \leq p \leq \infty$ and μ is a positive measure on $\partial\mathbb{D}$ with $\mu = \mu_a + \mu_s$ the Lebesgue decomposition of μ with respect to m, then*

$$P_0^p(\mu) = P_0^p(\mu_a) \oplus L^p(\mu_s).$$

Proof. It is rudimentary that $P_0^p(\mu) \subseteq P_0^p(\mu_a) \oplus L^p(\mu_s) \subseteq L^p(\mu_a) \oplus L^p(\mu_s)$. If $p^{-1} + q^{-1} = 1$, let $g \in L^q(\mu) = L^q(\mu_a) \oplus L^q(\mu_s)$ such that $\int g f d\mu = 0$ for all f in A_0. Thus $\int z^n g d\mu = 0$ for all $n \geq 1$. By the F and M Riesz Theorem, $\bar{g}\mu \ll m$. Thus $g = 0$ a.e. $[\mu_s]$ and so $g \in L^q(\mu_a)$ and $g \perp L^p(\mu_s)$. Consequently $g \perp P_0^p(\mu_a) \oplus L^p(\mu_s)$. The proposition now follows by the Hahn-Banach Theorem. □

6.2 Corollary. *If $1 \leq p < \infty$ and μ is a positive measure on $\partial\mathbb{D}$ with absolutely continuous part μ_a, then*

$$\inf_{f \in A_0} \int |1 - f|^p d\mu = \inf_{f \in A_0} \int |1 - f|^p d\mu_a.$$

We have seen that if h is a non-negative function in $L^1(m)$, then $\log h$ may fail to be integrable. In fact, $\log h \in L^1$ if and only if there is a function f in H^1 with $h = |f|$ (Proposition 3.6). Because $\log x \leq x - 1$ for $x > 0$, the only way that $\log h$ can fail to be integrable is for $\int \log h dm = -\infty$; that is, the approximating sums for the integral must diverge to $-\infty$. If this is indeed the case, then the expression $\exp\left[\int \log h dm\right]$ that appears in the subsequent text is to be interpreted as 0 (what else?).

6.3 Proposition. *If h is a non-negative function in $L^1(m)$, then*

$$\exp\left[\int \log h\, dm\right] = \inf\left\{\int he^g\, dm : g \in L^1_{\mathbb{R}}(m) \text{ and } \int g\, dm = 0\right\}.$$

Proof. If $g \in L^1_{\mathbb{R}}(m)$ and $\int g\, dm = 0$, then $\int \log(he^g) dm = \int(\log h + g)dm = \int \log h\, dm$. Thus letting $\mu = m$ in Proposition 19.4.13 and replacing h by $\log(he^g)$, we get

$$\exp\left[\int \log h\ dm\right] \leq \inf\left\{\int he^g dm : g \in L^1_{\mathbb{R}}(m) \text{ and } \int g\ dm = 0\right\}.$$

If $\varepsilon > 0$, let $c_\varepsilon = \int \log(h+\varepsilon)dm$ and put $g_\varepsilon = c_\varepsilon - \log(h+\varepsilon)$. Thus $g_\varepsilon L^1_{\mathbb{R}}(m)$ and $\int g_\varepsilon dm = 0$. Also $\int he^{g_\varepsilon} dm = \int he^{c_\varepsilon}(h+\varepsilon)^{-1} dm = e^{c_\varepsilon} \int h(h+\varepsilon)^{-1} dm$. Now $c_\varepsilon \to \int \log h dm$ by monotone convergence. On the other hand, $\int h(h+\varepsilon)^{-1} dm \to 1$. Thus $\int he^{g_\varepsilon} dm \to \exp\left[\int \log h dm\right]$, proving the proposition. □

6.4 Lemma. *If h is a non-negative function in $L^1(m)$ and $g \in L^1_{\mathbb{R}}(m)$ with $\int g dm = 0$, then there exists a sequence of functions $\{g_n\}$ in $L^\infty_{\mathbb{R}}(m)$ such that $\int g_n dm = 0$ for all $n \geq 1$ and $\int he^{g_n} dm \to \int he^g dm$ as $n \to \infty$.*

Proof. Let $f_n = g$ if $|g| \leq n$ and 0 otherwise. Then

$$\int he^{f_n} dm = \int_{|g|\leq n} he^g dm + \int_{|g|>n} h\ dm.$$

Since h and $g \in L^1(m)$, $\int_{|g|>n} h\ dm \to 0$ as $n \to \infty$. By the Monotone Convergence Theorem, $\int_{|g|\leq n} he^g dm \to \int he^g dm$. Therefore $\int he^{f_n} dm \to \int he^g dm$. Let $g_n = f_n - \int f_n$. Then $\int he^{g_n} dm = \exp\left(-\int f_n\right) \int he^{f_n} dm$. Since $\int f_n \to \int g = 0$, $\exp\left(-\int f_n\right) \to 1$ and so $\int he^{g_n} dm \to \int he^g dm$. □

6.5 Proposition. *If $h \in L^1(m)$ and $h \geq 0$, then*

$$\exp\left[\int \log h\ dm\right] = \inf_{f\in A_0}\left[\int he^{\operatorname{Re} f} dm\right].$$

Proof. Let $\alpha \equiv \inf_{f\in A_0}\left[\int he^{\operatorname{Re} f} dm\right]$ and $\beta \equiv \inf\left\{\int he^g dm : g \in L^1_{\mathbb{R}}(m)\right.$ and $\left.\int g\ dm = 0\right\}$. By Proposition 6.3, it must be shown that $\alpha = \beta$. But if $f \in A_0$, then $\int \operatorname{Re} f\ dm = 0$ and so $\alpha \geq \beta$. To obtain the other inequality, we first use the preceding lemma to see that $\beta = \inf\left\{\int he^g dm : g \in L^\infty_{\mathbb{R}}(m)\right.$ and $\left.\int g\ dm = 0\right\}$. Since for any polynomial p, $\int \operatorname{Re} p\ dm = \operatorname{Re} p(0)$, it follows from Proposition 4.5 that $\{\operatorname{Re} f : f \in A_0\}$ is uniformly dense in $\left\{g \in C_{\mathbb{R}}(\partial\mathbb{D}) : \int g\ dm = 0\right\}$. Thus if $g \in L^\infty_{\mathbb{R}}(m)$ with $\int g\ dm = 0$, then there is a sequence of functions $\{f_n\}$ from A_0 such that $\{\operatorname{Re} f_n\}$ is uniformly bounded and $\operatorname{Re} f_n \to g$ a.e. on $\partial\mathbb{D}$. By the Lebesgue Dominated

Convergence Theorem, $\int he^{\mathrm{Re}\, f_n} dm \to \int he^g dm$. Hence $\alpha \leq \int he^g dm$. But g was arbitrary, so $\alpha \leq \beta$. □

6.6 Szegö's Theorem. *If $1 \leq p < \infty$, μ is a positive measure on $\partial \mathbb{D}$, $\mu = \mu_a + \mu_s$ is the Lebesgue decomposition of μ with respect to m, and h is the Radon-Nikodym derivative of μ_a with respect to m, then*

$$\inf_{f \in A_0} \int |1 - f|^p d\mu = \exp\left[\int \log h \, dm\right].$$

Proof. Using the preceding proposition,

$$\exp\left[\int \log h \, dm\right] = \inf_{g \in A_0} \int he^{p\, \mathrm{Re}\, g} dm.$$

But $e^{p\, \mathrm{Re}\, g} = |e^g|^p$ and if $g \in A_0$, then $f = 1 - e^g \in A_0$. Hence $e^{p\, \mathrm{Re}\, g} = |1 - f|^p$ and

6.7
$$\exp\left[\int \log h \, dm\right] \geq \inf_{f \in A_0} \int |1 - f|^p h \, dm.$$

Now let $g \in A_0$ and apply (6.7) to the function $h = |1 - g|^p$. Since $|1 - f - g + fg|^p$ is subharmonic, this yields

$$\begin{aligned} \exp\left[\int \log |1 - g|^p dm\right] &\geq \inf_{f \in A_0} \int |1 - f - g + fg|^p dm \\ &\geq 1. \end{aligned}$$

This says that $\log |1 - g|^p \in L^1$ and $a \equiv \int \log |1 - g|^p dm \geq 0$. Put $k = \log |1 - g|^p - a$ and $c = e^a$. Thus $\int k \, dm = 0$, $c \geq 1$, and $ce^k = |1 - g|^p$. By Proposition 6.3, applied to the original function h, $\exp\left[\int \log h \, dm\right] \leq \int he^k dm \leq c \int he^k dm = \int |1 - g|^p h \, dm$. Combining this with (6.7) we get that

$$\exp\left[\int \log h \, dm\right] = \inf_{g \in A_0} \int |1 - g|^p h \, dm.$$

The theorem now follows from Corollary 6.2. □

Of course the left hand side of the equation in Szegö's Theorem is precisely the distance in $L^p(\mu)$ from the constant function 1 to the space $P_0^p(\mu)$. If this distance is zero, then $1 \in P_0^p(\mu)$. But this implies that, for $n \geq 1$, $\overline{z} z^n = z^{n-1} \in P_0^p(\mu)$. Thus $P_0^p(\mu)$ is invariant for multiplication by $\overline{z}$ as well as z. Therefore $P_0^p(\mu)$ contains all polynomials in z and $\overline{z}$ and must equal $L^p(\mu)$. This proves the next corollary.

6.8 Corollary. *If $1 \leq p < \infty$, μ is a positive measure on $\partial \mathbb{D}$, and $\mu = \mu_a + \mu_s$ is the Lebesgue decomposition of μ with respect to m, then $P_0^p(\mu) =$*

$L^p(\mu)$ if and only if

$$\int \log\left(\frac{d\mu_a}{dm}\right) dm = -\infty.$$

Note that the condition in the preceding corollary is independent of p. The condition for $p = \infty$ is different and less restrictive on μ. The proof of the final proposition of this section is immediate from Proposition 6.1.

6.9 Proposition. *If μ is a positive measure on $\partial\,\mathbb{D}$, then $P_0^\infty(\mu) \neq L^\infty(\mu)$ if and only if $m \ll \mu$.*

Exercises

1. Without using any of the results of this section, show that if $\alpha > 0$ and $\Gamma = \{e^{i\theta} : \pi \geq |\theta| \geq \alpha\}$, then every continuous function on Γ is the uniform limit of analytic polynomials.

2. If $\tau : \mathbb{D} \to \mathbb{D}$ is an analytic function and $f \in H^p$, is $f \circ \tau$ in H^p?

3. Give an example of a measure μ on $\partial\,\mathbb{D}$ such that μ and m are mutually absolutely continuous and $P_o^p(\mu) = L^p(\mu)$. Note that $P_0^\infty(\mu) \neq L^\infty(\mu)$.

4. If $f \in L^2$, show that the closed linear span in L^2 of the functions $\{z^n f : n \geq 0\}$ is L^2 if and only if: (i) f does not vanish on a set of positive measure; (ii) $\log|f|$ is not Lebesgue integrable. What happens for L^p with $p \neq 2$?

5. For $1 \leq p \leq \infty$ and μ a positive measure on $\partial\,\mathbb{D}$, show that the following are equivalent: (a) $P_0^p(\mu) \neq L^p(\mu)$; (b) $P^p(\mu) \neq L^p(\mu)$; (c) $P_0^p(\mu) \neq P^p(\mu)$.

Chapter 21

Potential Theory in the Plane

In this chapter we will continue the study of the Dirichlet problem. The emphasis here will be on finding the limits of the results and employing measure theory to do so.

Potential theory is most fully developed in the literature for n-dimensional Euclidean space with $n \geq 3$. There is an essential difference between the plane and $\mathbb{R}^n$ for $n \geq 3$; the standard potential for $\mathbb{C}$ is the logarithmic potential of §19.5, while for higher dimensions it is a Newtonian potential. Classical works on complex analysis treat the case $n = 2$, but usually from a classical point of view. In this chapter the treatment of the classical results will be from a more modern point of view. Many of the results carry over to higher dimensions and the interested readers are invited to pursue this at their convenience. Good general references for this are Brelot [1959], Carleson [1967], Helms [1975], Landkof [1972], and Wermer [1974].

§1 Harmonic Measure

If G is a hyperbolic open set in the extended plane and u is a real-valued continuous function on $\partial_\infty G$, then u is a solvable function for the Dirichlet problem. This leads to the following elementary result.

1.1 Proposition. *If G is a hyperbolic open set and $a \in G$, the map $u \to \hat{u}(a)$ is a positive linear functional of norm 1 on the space $C_\mathbb{R}(\partial_\infty G)$.*

Proof. Proposition 19.7.10 implies that $u \to \hat{u}$ is linear. If u is a positive function in $C_\mathbb{R}(\partial_\infty G)$, then $0 \in \hat{\mathcal{P}}(u, G)$; hence $\hat{u} \geq 0$. Therefore $u \to \hat{u}(a)$ is a positive linear functional. Since $\hat{1} = 1$, this functional has norm 1. □

According to the preceding proposition, for each a in G there is a unique probability measure ω_a supported on $\partial_\infty G$ such that

1.2
$$\hat{u}(a) = \int_{\partial_\infty G} u \, d\omega_a$$

for each u in $C_\mathbb{R}(\partial_\infty G)$.

1.3 Definition. For any hyperbolic open set G and any point a in G, the unique probability measure ω_a supported on $\partial_\infty G$ and satisfying (1.2) for every u in $C_{\mathbb{R}}(\partial_\infty G)$ is called *harmonic measure* for G at a. To indicate the dependence of the measure on G, the notation ω_a^G will be used. If K is a compact subset of $\mathbb{C}$ and $a \in \operatorname{int} K$, then harmonic measure for K at a is the same as harmonic measure for int K at a.

The point to remember here is that harmonic measure is only defined for hyperbolic sets. If K is compact, then $G = \operatorname{int} K$ is bounded and thus hyperbolic. In this book there will be little need for harmonic measure on a compact set.

The next result is left as an exercise for the reader.

1.4 Proposition. *If G is a hyperbolic set and $a \in G$, then (1.2) holds for all bounded Borel functions u.*

Note that this proposition says that $\omega_a(\Delta) = \hat{\chi}_\Delta(a)$ for any Borel subset of $\partial_\infty G$ and any point a in G.

From the comments following Theorem 15.2.5 we see that, for an analytic Jordan region, harmonic measure at the point a is absolutely continuous with respect to arc length measure on the boundary; and the Radon-Nikodym derivative is $\partial g_a/\partial n$, the derivative of the Green function with respect to the exterior normal to the boundary. An interpretation of Theorem 10.2.4 shows that if $G = \mathbb{D}$ and $a \in \mathbb{D}$, then $d\omega_a = P_a\, dm$, where P_a is the Poisson kernel at a and m is normalized arc length measure on $\partial\mathbb{D}$. Thus in these examples, harmonic measure and arc length measure on the boundary are mutually absolutely continuous. (Recall that two measures μ and ν are said to be mutually absolutely continuous if they have the same sets of measure 0.) This is not an isolated incident.

1.5 Theorem. *If G is a bounded simply connected region such that ∂G is a rectifiable Jordan curve, then harmonic measure for G and arc length measure on ∂G are mutually absolutely continuous.*

Proof. Fix a in G and let $\tau : \mathbb{D} \to G$ be the Riemann map such that $\tau(0) = a$ and $\tau'(0) > 0$. By Theorem 14.5.6 (also see Theorem 20.4.8), τ extends to a homeomorphism $\tau : \operatorname{cl} \mathbb{D} \to \operatorname{cl} G$. Thus $\theta \to \tau(e^{i\theta})$ is a parameterization of ∂G. Since ∂G is a rectifiable curve, Theorem 14.5.8 implies that $\theta \to \tau(e^{i\theta})$ is absolutely continuous and τ' is in the space H^1. According to (20.4.9) the derivative of the function $\theta \to \tau(e^{i\theta})$ is $ie^{i\theta}\tau'(e^{i\theta})$, where $\tau'(e^{i\theta})$ is the radial limit of τ'. So if $u : \partial G \to \mathbb{R}$ is a continuous function, the measure μ defined on ∂G by

$$\int u\, d\mu = \frac{1}{2\pi}\int_0^{2\pi} u(\tau(e^{i\theta}))\, |\tau'(e^{i\theta})|\, d\theta$$

is (non-normalized) arc length measure.

Claim. If E is a Borel subset of $\partial\mathbb{D}$, then $m(E) = 0$ if and only if $\mu(\tau(E)) = 0$.

From the above formula, we know that for any Borel set Δ contained in ∂G, $\mu(\Delta) = \int_{\tau^{-1}(\Delta)} |\tau'|\, dm$. Assume that $m(E) = 0$. Since τ is a homeomorphism, $E = \tau^{-1}(\tau(E))$ and so $\mu(\tau(E)) = \int_E |\tau'|\, dm = 0$. Conversely, if $\mu(\tau(E)) = 0$, then $\tau' = 0$ a.e. $[m]$ on E. Since $\tau' \in H^1$ and is not the zero function, this implies that $m(E) = 0$.

On the other hand, if $u \in C_{\mathbb{R}}(\partial G)$, then $u \circ \tau \in C_{\mathbb{R}}(\partial \mathbb{D})$. Let h be the solution of the Dirichlet problem on $\mathbb{D}$ with boundary values $u \circ \tau$. It is easy to see that $h \circ \tau^{-1}$ is the solution of the Dirichlet problem on G with boundary values u. That is, $h \circ \tau^{-1} = \hat{u}$. Hence

$$\int u\, d\omega_a = \hat{u}(a) = h(0) = \int u \circ \tau\, dm.$$

Since u was arbitrary, this implies that $\omega_a = m \circ \tau^{-1}$. Thus for a Borel subset Δ of ∂G, $\omega_a(\Delta) = 0$ if and only if $m(\tau^{-1}(\Delta)) = 0$. In light of the claim, this proves the theorem. □

The above result has an extension to a finitely connected Jordan region with rectifiable boundary. If G is such a region and its boundary $\Gamma = \partial G$ consists of analytic curves, then this is immediate from Theorem 15.2.5. If Γ only consists of rectifiable curves, then a more careful analysis is needed. Also see Exercise 4.

In Exercise 2 part of a fact that emerged from the preceding proof is extracted for later use. The basic problem that is touched on in that exercise is the following. If $\tau : G \to \Omega$ is a surjective analytic function, can harmonic measure for Ω be expressed in terms of harmonic measure for G? The next proposition settles this for bounded Dirichlet regions.

1.6 Proposition. *Suppose G is a bounded Dirichlet region, $\tau : \mathbb{D} \to G$ is an analytic function with $\tau(0) = \alpha$, and $\tilde{\tau}$ denotes the radial limit function of τ. If $\tilde{\tau}(\partial \mathbb{D}) \subseteq \partial G$, then $\omega_\alpha^G = m \circ \tilde{\tau}^{-1}$.*

Proof. Let $f : \partial G \to \mathbb{R}$ be a continuous function and $\hat{f}$ the solution of the Dirichlet problem on G with boundary values f; so $\hat{f}$ extends to a continuous function on cl G with $\hat{f} = f$ on ∂G. If $a \in \partial \mathbb{D}$ such that τ has a radial limit at a and $r \to 1-$, then $\hat{f} \circ \tau(ra) \to f(\tilde{\tau}(a)) = f \circ \tilde{\tau}(a)$. That is, the bounded harmonic function $\hat{f} \circ \tau$ on $\mathbb{D}$ has radial limits equal to $f \circ \tilde{\tau}$ a.e. on $\partial \mathbb{D}$. Therefore $\hat{f} \circ \tau = \widehat{f \circ \tilde{\tau}}$ and so

$$\int f\, d\omega_\alpha^G = \hat{f}(\tau(0)) = \widehat{f \circ \tilde{\tau}}(0) = \int f \circ \tilde{\tau}\, dm = \int f\, d\, m \circ \tilde{\tau}^{-1}.$$

□

In general, we will not be so concerned with the exact form of harmonic measure but rather with its measure class; that is, with the sets of harmonic measure 0. When two measures μ and ν are mutually absolutely continuous, we can form the two Radon-Nikodym derivatives $d\mu/d\nu$ and $d\nu/d\mu$. Say that μ and ν are *boundedly mutually absolutely continuous* if they are mutually absolutely continuous and the two Radon-Nikodym derivatives are bounded functions.

1.7 Theorem. *If G is a hyperbolic open subset of the plane and a and b belong to the same component of G, then the harmonic measures for G at a and b, ω_a and ω_b, are boundedly mutually absolutely continuous. Moreover, there is a constant $\rho > 0$ (depending only on a, b, and G) such that if H is any hyperbolic open set containing G and μ_a and μ_b are the harmonic measures for H at a and b, then $\rho\mu_a \leq \mu_b \leq \rho^{-1}\mu_a$.*

Proof. This follows from Harnack's Inequality. If $\overline{B}(a;R) \subseteq G$, $|b-a| = r < R$, and u is a positive continuous function on $\partial_\infty G$, then Harnack's Inequality implies that $\rho\hat{u}(a) \leq \hat{u}(b) \leq \rho^{-1}\hat{u}(a)$, where $\rho = (R-r)/(R+r)$. Thus $\rho \int u \, d\omega_a \leq \int u \, d\omega_b \leq \rho^{-1} \int u \, d\omega_a$.

But this implies that $\rho\, \omega_a(\Delta) \leq \omega_b(\Delta) \leq \rho^{-1}\, \omega_a(\Delta)$ for every Borel set Δ contained in $\partial_\infty G$. Thus ω_a and ω_b are boundedly mutually absolutely continuous and

$$\rho \leq \frac{d\omega_b}{d\omega_a} \leq \rho^{-1}.$$

Note that this constant ρ depends on a, b, and G alone. Thus the same inequalities hold for μ_a and μ_b:

$$\rho \leq \frac{d\mu_b}{d\mu_a} \leq \rho^{-1}.$$

If a and b belong to the same component of G, then there are points $a_0, \ldots, a_n$ and positive numbers $R_0, \ldots, R_n$ such that: (i) $a = a_0$, $b = a_n$; (ii) $|a_j - a_{j-1}| < R_j$, $1 \leq j \leq n$; (iii) $\overline{B}(a_j; R_j) \subseteq G$. By the preceding paragraph, for $1 \leq j \leq n$, ω_{a_j} and $\omega_{a_{j-1}}$ are boundedly mutually absolutely continuous with constant ρ_j. Thus ω_a and ω_b are mutually absolutely continuous and

$$\frac{d\omega_a}{d\omega_b} = \frac{d\omega_{a_0}}{d\omega_{a_1}} \frac{d\omega_{a_1}}{d\omega_{a_2}} \cdots \frac{d\omega_{a_{n-1}}}{d\omega_{a_n}};$$

hence

$$\rho \leq \frac{d\omega_b}{d\omega_a} \leq \rho^{-1},$$

where $\rho = \rho_1\rho_2 \ldots \rho_n$. Similarly, the same inequality holds for the measures μ_a and μ_b. $\square$

This result has some interesting consequences. If μ and ν are boundedly mutually absolutely continuous measures, then the identity mapping on bounded Borel functions induces a bounded bijection of $L^p(\mu)$ onto $L^p(\nu)$ for all p. (See Exercise 3.) Thus we can legitimately say that $L^p(\mu) = L^p(\nu)$

in this case. So if a and b belong to the same component of G, $L^p(\omega_a) = L^p(\omega_b)$ for all p. We are led to the following definition for regions only.

1.8 Definition. If G is a region and $1 \le p < \infty$, $L^p(\partial G) \equiv L^p(\omega_a)$, where a is any point in G.

So the preceding discussion says that the definition of this space $L^p(\partial G)$ does not depend on the choice of point a in G, though the value of the norm of a function in the space will depend on the choice of a. We return to the case of an arbitrary (not necessarily connected) hyperbolic open set G.

1.9 Proposition. *If $a \in G$ and G is hyperbolic, then for any Borel set Δ disjoint from the boundary of the component of G that contains a, $\omega_a(\Delta) = 0$. Thus the support of ω_a is contained in $\partial_\infty G_a$, where G_a is the component of G that contains a.*

Proof. Note that if the notation is as in the statement of the theorem and $\phi \in \hat{\mathcal{P}}(\chi_\Delta, G)$, then $\limsup_{z\to\zeta} \phi(z) \le \chi_\Delta(\zeta) = 0$ for ζ in ∂G_a. Thus $\phi_1 = \max\{\phi, 0\} \in \hat{\mathcal{P}}(\chi_\Delta, G)$ and ϕ_1 vanishes on G_a and so we have that $0 = \widehat{\chi_\Delta}(a) = \omega_a(\Delta)$. $\square$

In light of the last few results it is tempting to believe that the harmonic measures for points in distinct components of G are mutually singular. This is not the case, as the next example demonstrates.

1.10 Example. A *string of beads* is a compact set

$$K = [\text{cl } \mathbb{D}] \setminus \cup_{n=1}^\infty \Delta_n,$$

where $\{\Delta_n\}$ is a sequence of open disks in $\mathbb{D}$ having the following properties:

(i) $\text{cl } \Delta_n \cap \text{cl } \Delta_m = \emptyset$ for $n \ne m$;

(ii) the center of each Δ_n lies on the interval $[-1, 1]$;

(iii) $[-1, 1] \setminus \cup_n \Delta_n$ contains no interval;

(iv) $[-1, 1] \cap K$ has positive one-dimensional Lebesgue measure.

To construct K, first construct a Cantor set in $[-1, 1]$ with positive Lebesgue measure and replace each deleted subinterval of $[-1, 1]$ with an open disk Δ_n. The details of the construction of this set are left to the reader.

Notice that $G = \text{int } K$ has two components, each of which is simply connected with a boundary that is a rectifiable Jordan curve; denote these components by Q_+ and Q_-. Let $a_\pm \in Q_\pm$ and let $\omega_\pm =$ harmonic measure for G at the point $a_\pm$. Since $\partial Q_\pm$ is a rectifiable Jordan curve (why?), $\omega_\pm$ and arc length measure on $\partial Q_\pm$ are mutually absolutely continuous (1.5). Thus ω_+ and ω_- are not mutually singular.

We now characterize solvable functions for a hyperbolic open set. But first we will see a lemma.

1.11 Lemma. *If G is a hyperbolic set, $a \in G$, and $f : \partial_\infty G \to [-\infty, \infty)$ is an upper semicontinuous function that is bounded above and satisfies $\hat{f}(a) > -\infty$, then $f \in L^1(\omega_a)$.*

Proof. According to Theorem 19.3.6 there is a decreasing sequence $\{f_n\}$ of continuous functions on $\partial_\infty G$ such that for each ζ in $\partial_\infty G$, $f_n(\zeta) \downarrow f(\zeta)$. By Corollary 19.7.12, $\hat{f}_n(a) \to \hat{f}(a)$. On the other hand, the Monotone Convergence Theorem implies that $\int f_n \, d\omega_a \to \int f \, d\omega_a$. By (1.2), $\hat{f}(a) = \int f \, d\omega_a$. Since f is bounded above and $\int f \, d\omega_a > -\infty$, $f \in L^1(\omega_a)$. □

1.12 Brelot's Theorem. *Let G be a hyperbolic set and let u be an extended real-valued function on $\partial_\infty G$. If u is solvable, then for all z in G, u is integrable with respect to ω_z and*

$$\hat{u}(z) = \int u \, d\omega_z.$$

Conversely, if for each component of G there is a point a in that component such that $u \in L^1(\omega_a)$, then u is solvable.

Proof. By Corollary 19.7.7 it suffices to assume that G is connected. Assume that u is solvable and fix a point z in G. According to Proposition 19.7.8.b there is an increasing sequence of upper semicontinuous functions $\{u_n\}$ on $\partial_\infty G$ such that each u_n is bounded above, $u_n \le u$, and $\hat{u}_n(z) \to \hat{u}(z)$. Similarly, there is a decreasing sequence of lower semicontinuous functions $\{v_n\}$ on $\partial_\infty G$ such that each v_n is bounded below, $v_n \ge u$, and $\hat{v}_n(z) \to \hat{u}(z)$. Because u is solvable, it can be assumed that $\hat{u}_n(z)$ and $\hat{v}_n(z)$ are all finite numbers. By the preceding lemma, each u_n and v_n is integrable with respect to ω_z. But $u_n \le u \le v_n$ and so $u \in L^1(\omega_z)$. Moreover, $\int u_n d\omega_z = \hat{u}_n(z) \le \hat{u}(z) \le \hat{v}_n(z) = \int v_n d\omega_z$. Since $0 = \lim_n \int (v_n - u_n) \, d\omega_z$, we have that $\hat{u}(z) = \int u \, d\omega_z$.

Now assume that there is a point a in G such that $u \in L^1(\omega_a)$. By Theorem 1.7, u is integrable with respect to ω_z for each z in G. It suffices to assume that $u \ge 0$. From measure theory there is a sequence of nonnegative bounded Borel functions $\{u_n\}$ such that $u = \sum_n u_n$. Since each u_n is solvable, for each z in G, $0 \le \sum_n \hat{u}_n(z) = \sum_n \int u_n \, d\omega_z = \int u \, d\omega_z < \infty$. According to Proposition 19.7.11, u is a solvable function. □

The next result will be used later and could have been proved earlier. It was postponed until now because the corollary is the chief objective here and this was meaningless until harmonic measure was introduced. Note that the next proposition generalizes Proposition 19.7.6. (The result is quite useful in spite of the fact that its statement is longer than its proof, usually a sign of mathematical banality.)

1.13 Proposition. *Let G be a hyperbolic open set and let $g : \partial_\infty G \to \mathbb{R}$ be a bounded Borel function; denote by $\hat{g}$ the solution of the Dirichlet Problem on G with boundary values g. If H is an open subset of G and $h : \partial_\infty H \to \mathbb{R}$ is defined by $h(\zeta) = \hat{g}(\zeta)$ for ζ in $\partial_\infty H \cap G$ and $h(\zeta) = g(\zeta)$ for ζ in $\partial_\infty H \cap \partial_\infty G$, then h is a solvable function for H and $\hat{h} = \hat{g}$ on H.*

Proof. In fact, if $\phi \in \hat{\mathcal{P}}(g, G)$, then $\phi|H \in \hat{\mathcal{P}}(h, H)$; hence $\hat{g} \leq \hat{h}$ on H. Similarly, $\check{g} \geq \check{h}$ on H. Since G is hyperbolic, we get that $\check{h} = \hat{h} = \hat{g}$ on H. □

1.14 Corollary. *If G is a hyperbolic open set and H an open subset of G, then for any Borel subset Δ of $\partial_\infty G \cap \partial_\infty H$, $\omega_a^H(\Delta) \leq \omega_a^G(\Delta)$ for all a in H.*

Proof. Note that $z \to \omega_z^G(\Delta)$ is the solution of the Dirichlet problem on G with boundary values $g = \chi_\Delta$. If $h : \partial_\infty H \to \mathbb{R}$ is defined by $h(z) = \omega_z^G(\Delta)$ for z in $\partial_\infty H \cap G$ and $h(z) = g(z)$ for z in $\partial_\infty H \cap \partial_\infty G$, then the preceding proposition implies that for z in H,

$$\omega_z^G(\Delta) = \hat{h}(z) = \int h \, d\omega_z^H \geq \omega_z^H(\Delta)$$

since $h \geq \chi_\Delta$ on $\partial_\infty H$. □

1.15 Corollary. *Let G and Ω be hyperbolic open sets with $G \subseteq \Omega$. If K is a compact subset of $\partial_\infty G$ such that $\omega_a^{\Omega \setminus K}(K) = 0$ for all a in G, then $\omega_a^G(K) = 0$ for all a in G.*

Proof. First observe that the hypothesis implies that K is a subset of $\partial_\infty(\Omega \setminus K) \cap \partial_\infty G$. So the preceding corollary implies $\omega_a^G(K) \leq \omega_a^{\Omega \setminus K}(K)$ for all a in G. □

The section concludes with an application of harmonic measure to obtain a formula for the Green function of a hyperbolic open set. In order to set the stage for this, we must first show that the logarithm is integrable with respect to harmonic measure. Using Theorem 1.12, this will produce a solution of the Dirichlet problem with logarithmic boundary values and lead to the sought formula.

1.16 Lemma. *If Ω is a hyperbolic open subset of $\mathbb{C}_\infty$ such that $\infty \in \Omega$ and $\alpha \in \partial\Omega$, then for every b in Ω, $\omega_b(\{\alpha\}) = 0$ and the function $\zeta \to \log|\zeta - \alpha|$ belongs to $L^1(\omega_b)$.*

Proof. Let $g(z, \infty)$ be the Green function for Ω with singularity at ∞ and put $h(z) = g(z, \infty) - \log|z - \alpha|$. So $h(z) = g(z, \infty) - \log|z| - \log|1 - \alpha/z|$ and h is a harmonic function on Ω. Now $\partial\Omega$ is bounded and so there is a constant C such that $\log|\zeta - \alpha| \leq C$ on $\partial\Omega$.

Let $\{u_n\}$ be a sequence of continuous functions on $\partial\Omega$ such that $u_n \geq -C$, $u_n(\zeta) \uparrow \log|\zeta - \alpha|^{-1}$ for ζ in $\partial\Omega$ and $\zeta \neq \alpha$, and $u_n(\alpha) \neq \infty$; put $h_n = \hat{u}_n$. So $\{h_n\}$ is an increasing sequence of harmonic functions on Ω. If $\phi \in \hat{\mathcal{P}}(u_n, \Omega)$ and $\zeta \in \partial\Omega$, then $\limsup_{z\to\zeta}[\phi(z) + \log|z - \alpha|] \leq u_n(\zeta) + \log|\zeta - \alpha| \leq 0 \leq \liminf_{z\to\zeta} g(z, \infty)$. According to the Maximum Principle, $\phi(z) + \log|z - \alpha| \leq g(z, \infty)$ on Ω; hence $\phi \leq h$ on Ω and so $h_n \leq h$. Therefore for each n,

$$\begin{aligned} h(b) &\geq h_n(b) \\ &= \int u_n \, d\omega_b \\ &= u_n(\alpha)\, \omega_b(\{\alpha\}) + \int_{\partial\Omega\setminus\{\alpha\}} u_n \, d\omega_b \\ &\geq u_n(\alpha)\, \omega_b(\{\alpha\}) - C. \end{aligned}$$

So if $\omega_b(\alpha) > 0$, the right hand side converges to ∞ as $n \to \infty$, a contradiction.

To show the integrability of the logarithm, take the limit in the above inequality to get

$$\infty > h(b) \geq -\int \log|\zeta - \alpha| \, d\omega_b \geq -C.$$

Thus $\log|\zeta - \alpha| \in L^1(\omega_b)$. □

Amongst other things, the preceding lemma says that harmonic measure has no atoms for regions of that type. In the above lemma note that if $\alpha \in \mathbb{C} \setminus \partial\Omega$, $\log|\zeta - \alpha|$ is a bounded continuous function on the boundary and is thus integrable. The next result is part of the reason for the preceding lemma and the succeeding proposition is another.

1.17 Proposition. *If G is hyperbolic, then ω_a has no atoms.*

Proof. This is immediate from the preceding lemma since, by the choice of a suitable Möbius transformation, any hyperbolic open set G can be mapped onto a hyperbolic open set Ω such that $\infty \in \Omega$. Once this is done we need only apply Exercise 2 and the preceding lemma. □

1.18 Proposition. *If G is hyperbolic and $a \in \mathbb{C}$, then the function $\zeta \to \log|\zeta - a|$ belongs to $L^1(\omega_b)$ for every b in G.*

Proof. Without loss of generality we may assume that $0 \in G$. Let $\tau(z) = z^{-1}$ and put $\Omega = \tau(G)$. So $\infty \in \Omega$. According to Exercise 2, $\omega^{\Omega}_{1/b} = \omega^G_b \circ \tau$ and $\omega^G_b = \omega^{\Omega}_{1/b} \circ \tau$ (since $\tau = \tau^{-1}$). Fix b in G. By Lemma 1.16, $\log|\zeta - \beta| \in L^1(\omega^{\Omega}_{1/b})$ for every β in $\mathbb{C}$; thus $\log|\zeta^{-1} - \beta| \in L^1(\omega^G_b)$. Taking $\beta = 0$ gives that $\log|\zeta| = -\log|\zeta|^{-1} \in L^1(\omega^G_b)$. Taking $\beta = 1/\alpha$, $\alpha \neq 0$, gives that $\log|\zeta - \alpha| = \log|\zeta^{-1} - \beta| + \log|\zeta| + \log|\alpha| \in L^1(\omega^G_b)$. □

It should be emphasized that in the next theorem the open set is assumed to not contain the point ∞. For a hyperbolic open set that contains ∞ the formula for the Green function is in the succeeding theorem.

1.19 Theorem. *If G is a hyperbolic open set contained in $\mathbb{C}$, $a \in G$, and $g(z,a)$ is the Green function for G with singularity at a, then*

$$g(z,a) = \int \log|\zeta - a| \, d\omega_z(\zeta) - \log|z-a|.$$

Proof. According to the preceding proposition the integral in this formula is well defined. Let $h(z,a)$ denote the right hand side of the above equation and fix a in G. Clearly h is harmonic in $G \setminus \{a\}$. If $\psi \in \check{\mathcal{P}}(\log|\zeta - a|, G)$, then $\psi(z) - \log|z-a|$ is superharmonic on G and, for ζ in $\partial_\infty G$,

$$\begin{aligned}\liminf_{z\to\zeta}[\psi(z) - \log|z-a|] &= \liminf_{z\to\zeta}\psi(z) - \log|\zeta - a| \\ &\geq 0.\end{aligned}$$

But $\int \log|\zeta - a| \, d\omega_z = \inf\{\psi(z) : \psi \in \check{\mathcal{P}}(\log|\zeta - a|, G)\}$. So $h(z,a) \geq 0$ for all z in G. Clearly $h(z,a) + \log|z-a|$ is harmonic near a. By the definition of the Green function, $h(z,a) \geq g(z,a)$.

On the other hand, let ψ be a positive superharmonic function on G such that $\psi(z) + \log|z-a|$ is also superharmonic. It follows that $\liminf_{z\to\zeta}[\psi(z) + \log|z-a|] \geq \log|\zeta - a|$. Thus $\psi(z) + \log|z-a| \geq \int \log|\zeta - a| \, d\omega_z$. Applying Proposition 19.9.10 we get that $g(z,a) \geq h(z,a)$. □

1.20 Theorem. *If Ω is a hyperbolic open set in the extended plane such that $\infty \in \Omega$, $\alpha \in \Omega$, and $g(w,\alpha)$ is the Green function for Ω with singularity at $\alpha \neq \infty$, then for every $w \neq \infty$*

$$g(w,\alpha) = \int_{\partial\Omega} \log|\zeta - \alpha| \, d\omega_w^\Omega(\zeta) - \log|w - \alpha| + g(w,\infty).$$

If β is a point not in Ω, then for every choice of distinct α and w in Ω

$$g(w,\alpha) = \int_{\partial\Omega} \log\left|\left(\frac{\zeta - \alpha}{\zeta - \beta}\right)\left(\frac{w-\beta}{w-\alpha}\right)\right| \, d\omega_w^\Omega(\zeta).$$

In particular, for $w \neq \infty$ in Ω,

$$\begin{aligned} g(w,\infty) &= \int_{\partial\Omega} \log\left|\left(\frac{w-\beta}{\zeta-\beta}\right)\right| \, d\omega_w^\Omega(\zeta) \\ &= -\int_{\partial\Omega} \log|\zeta - \beta| \, d\omega_w^\Omega(\zeta) + \log|w-\beta|. \end{aligned}$$

Proof. In this proof the Green function for Ω will be denoted by g^Ω. Take any point β that does not belong to Ω and let $\tau(w) = (w-\beta)^{-1}$; put

$G = \tau(\Omega)$. So G is hyperbolic, $\infty \notin G$, and $0 = \tau(\infty) \in G$. If g^G is the Green function for G, then for every w and α in Ω

$$\begin{aligned}
g^{\Omega}(w,\alpha) &= g^G(\tau(w),\tau(\alpha)) \\
&= \int_{\partial_\infty G} \log|\xi - \tau(\alpha)|\, d\omega^G_{\tau(w)}(\xi) - \log|\tau(w)-\tau(\alpha)| \\
&= \int_{\partial\Omega} \log\left|\frac{1}{\zeta-\beta} - \frac{1}{\alpha-\beta}\right| d\omega^{\Omega}_w(\zeta) - \log\left|\frac{1}{w-\beta} - \frac{1}{\alpha-\beta}\right| \\
&= \int_{\partial\Omega} \log\left|\left(\frac{\zeta-\alpha}{\zeta-\beta}\right)\left(\frac{w-\beta}{w-\alpha}\right)\right| d\omega^{\Omega}_w(\zeta).
\end{aligned}$$

This gives all but the first of the formulas in the theorem. To obtain the first, assume neither w nor α is ∞ and use the properties of the logarithm and the fact that harmonic measure is a probability measure. This gives

$$\begin{aligned}
g^{\Omega}(w,\alpha) &= \int_{\partial\Omega} \log|\zeta-\alpha|\, d\omega^{\Omega}_w(\zeta) - \log|w-\alpha| \\
&\qquad + \log|w-\beta| - \int_{\partial\Omega} \log|\zeta-\beta|\, d\omega^{\Omega}_w(\zeta) \\
&= \int_{\partial\Omega} \log|\zeta-\alpha|\, d\omega^{\Omega}_w(\zeta) - \log|w-\alpha| + g^{\Omega}(w,\infty).
\end{aligned}$$

□

Exercises

1. Prove Proposition 1.4.

2. If $\tau : G \to \Omega$ is a conformal equivalence that extends to a homeomorphism between the closures in the extended plane, $a \in G$, and $\alpha = \tau(a)$, then $\omega^G_a \circ \tau^{-1} = \omega^{\Omega}_{\alpha}$.

3. Assume that μ and ν are finite measures on the same set X. (a) Show that if μ and ν are mutually absolutely continuous (but not necessarily boundedly so), then the identity mapping on the bounded Borel functions induces an isometric isomorphism of $L^\infty(\mu)$ onto $L^\infty(\nu)$ that is a homeomorphism for the weak* topology. (b) Conversely, show that if $T : L^\infty(\mu) \to L^\infty(\nu)$ is an isometric isomorphism that is the identity on characteristic functions and T is a weak* homeomorphism, then μ and ν are mutually absolutely continuous. (c) If μ and ν are mutually absolutely continuous and $\phi = d\mu/d\nu$, show that for $1 \le p < \infty$, the map $Tf = f\,\phi^{1/p}$ defines an isometric isomorphism of $L^p(\mu)$ onto $L^p(\nu)$ such that $T(uf) = uT(f)$ for all u in $L^\infty(\mu)$ and f in $L^p(\mu)$. (d) Conversely, if there is an isometric isomorphism

$T : L^p(\mu) \to L^p(\nu)$ such that $T(uf) = u\,T(f)$ for all bounded Borel functions u and all f in $L^p(\mu)$, then μ and ν are mutually absolutely continuous.

4. (Spraker [1989]) Suppose γ is a smooth Jordan curve, $G = \text{ins}\,\Gamma$, and $a \in G$. Show that if $\omega_a^G =$ normalized arc length measure on γ, then γ is a circle and a is its center.

5. For $R > 0$, put $G = \mathbb{C}_\infty \setminus \overline{B}(0; R)$ and show that for every continuous function f on ∂G, $\int f\, d\omega_\infty^G = \frac{1}{2\pi}\int_0^{2\pi} f(R\, e^{-i\theta})\, d\theta$.

6. If $G = \mathbb{C}_\infty \setminus [0, 1]$, what is harmonic measure for G at ∞?

§2 * The Sweep of a Measure

If G is a hyperbolic open set and $u \in C(\partial_\infty G)$, let, as usual, $\hat{u}$ be the solution of the Dirichlet problem. Thus if μ is a bounded regular Borel measure carried by G, $\int_G \hat{u}\, d\mu$ is well defined and finite. In fact, $\left|\int_G \hat{u}\, d\mu\right| \le ||\mu||\ ||\hat{u}||_\infty \le ||\mu||\ ||u||_{\partial_\infty G}$. That is, $u \to \int_G \hat{u}\, d\mu$ is a bounded linear functional on $C(\partial_\infty G)$. Thus there is a measure $\hat{\mu}$ supported on $\partial_\infty G$ such that

2.1
$$\int_G \hat{u}\, d\mu = \int_{\partial_\infty G} u\, d\hat{\mu}$$

for every u in $C(\partial_\infty G)$.

2.2 Definition. If G is a hyperbolic open set and $\mu \in M(G)$, the *sweep* of μ is the measure $\hat{\mu}$ defined by (2.1).

Caution here, intrepid reader. The notation for the sweep of a measure is the same as that for the Cauchy transform. This coincidence does not reflect on humanity or provide a commentary on the lack of notational imagination. Though unfortunate, the notation is traditional.

Note that if $a \in G$ and ω_a is harmonic measure for G at a, then for every continuous function u on $\partial_\infty G$, $\hat{u}(a) = \int_{\partial_\infty G} u\, d\omega_a$. Equivalently, $\int_G \hat{u}\, d\delta_a = \int_{\partial_\infty G} u\, d\omega_a$. Thus the sweep of δ_a is ω_a.

There is often in the literature the desire to discuss the sweep of a measure μ carried by a compact subset K of $\mathbb{C}$. In this case the sweep of μ is $\mu|\partial K + \widehat{\mu|(\text{int}\, K)}$.

The next proof is left to the reader.

*This section can be skipped if desired, as the remainder of the book does not depend on it.

2.3 Proposition. *The map $\mu \to \hat{\mu}$ is a contractive linear map of $M(G)$ into $M(\partial_\infty G)$.*

Often we are not so interested in harmonic measure itself but rather with its measure class (that is, the collection of measures that have the same sets of measure 0 as harmonic measure). This applies to non-connected hyperbolic open sets as well as regions. With this in mind, let us introduce the following idea.

2.4 Definition. Let G be a hyperbolic open set with components $G_1, G_2, \ldots$ and for each $n \geq 1$ pick a point a_n in G_n. If ω_n is harmonic measure for G at a_n and $\omega = \sum_n 2^{-n}\omega_n$, then *harmonic measure* for G is any measure on $\partial_\infty G$ that has the same sets of measure zero as ω.

Thus harmonic measure for a hyperbolic open set is actually a measure class $[\omega]$, rather than a specific measure. This means it is impossible to define the L^p space of the class if $1 < p < \infty$. Indeed if the points a_n in Definition 2.4 are replaced by points a_n' with ω_n' the corresponding harmonic measure, the spaces $L^p(\omega_n)$ and $L^p(\omega_n')$ are isomorphic via the identity map but not isometrically isomorphic. Thus, in the presence of an infinite number of components, the spaces $L^p(\omega)$ and $L^p(\omega')$ may not even be isomorphic under the identity map. We can, however, define the L^1 space and the L^∞ space since the definitions of these spaces actually only depend on the collection of sets of measure 0.

2.5 Definition. If G is a hyperbolic open set and ω is harmonic measure for G, define

$$\begin{aligned} L^\infty(\partial_\infty G) &= \{f : f \text{ is } \omega - \text{ essentially bounded}\}, \\ L^1(\partial_\infty G) &= \{\mu \in M(\partial_\infty G) : \mu \ll \omega\}. \end{aligned}$$

As usual, functions in $L^\infty(\partial_\infty G)$ are identified if they agree a.e. $[\omega]$. Thus $L^\infty(\partial_\infty G) = L^\infty(\omega)$ and the definition of the norm on $L^\infty(\partial_\infty G)$ is independent of which form of harmonic measure for G we choose. The same is not quite true for $L^1(\partial_\infty G)$. If ω and ω' are two harmonic measures for G that are not boundedly mutually absolutely continuous, then $L^1(\omega)$ and $L^1(\omega')$ can be significantly different if we define these spaces as spaces of integrable functions. In fact, these spaces may not be equal as sets, let alone isometric as Banach spaces. But the definition of $L^1(\partial_\infty G)$ given above as a subspace of $M(\partial_\infty G)$ with the total variation norm removes this ambiguity.

2.6 Proposition. *If $\mu \in M(G)$, then $\hat{\mu}$, the sweep of μ, belongs to $L^1(\partial_\infty G)$*

Proof. Adopt the notation of Definition 2.4 and let Δ be a compact subset of $\partial_\infty G$ with $\omega(\Delta) = 0$. It suffices to assume that $\mu \geq 0$ and show that $\hat{\mu}(\Delta) = 0$. Let $\{u_n\}$ be a sequence of continuous functions on $\partial_\infty G$ such that $\chi_\Delta \leq u_n \leq 1$ for all $n \geq 1$ and $\{u_n(z)\}$ decreases monotonically to $\chi_\Delta(z)$ for each z in $\partial_\infty G$. If $a \in G$, then $\hat{u}_n(a) = \int u_n \, d\omega_a \to \omega_a(\Delta) = 0$. Since $0 \leq \hat{u}_n(a) \leq 1$ for all $n \geq 1$ and for all a in G, $\hat{\mu}(\Delta) = \lim \int u_n \, d\hat{\mu} = \lim \int_G \hat{u}_n \, d\mu = 0$. □

Exercises

1. Suppose that $\operatorname{int} K$ is a Dirichlet set and show that $u \to \hat{u}$ is a bounded linear map of $C(\partial K)$ into $C(K)$.

   What is the dual of this map?

2. If μ is a positive measure carried by $\mathbb{D}$, define $R_\mu : \partial\mathbb{D} \to \mathbb{R}$ by $R_\mu(\zeta) = \int P_\zeta d\mu$, where P_ζ is the Poisson kernel. Let m be normalized arc length measure on $\partial\mathbb{D}$. (a) Show that R_μ is a non-negative function in $L^1(m)$. (b) If $f \in H^\infty$, show that $\int_{\mathbb{D}} |f|^2 d\mu \leq \int |f|^2 R_\mu dm$. (c) Prove that $\hat{\mu} = R_\mu m$.

§3 The Robin Constant

Recall that, for a compact subset K of $\mathbb{C}$, the Green function for $\mathbb{C}_\infty \setminus K$ near the point at infinity is precisely the Green function for $\mathbb{C}_\infty \setminus \hat{K}$. Keep this in mind while reading the definition of the Robin constant below.

3.1 Definition. If K is any compact subset of $\mathbb{C}$ such that $\mathbb{C}_\infty \setminus K$ is hyperbolic, then the *Robin constant* for K is the number $\operatorname{rob}(K)$ defined by

$$\operatorname{rob}(K) \equiv \lim_{z \to \infty} [g(z, \infty) - \log |z|],$$

where g is the Green function for $\mathbb{C}_\infty \setminus K$. If $\mathbb{C}_\infty \setminus K$ is parabolic, define $\operatorname{rob}(K) = \infty$.

First note that when $\mathbb{C}_\infty \setminus K$ is hyperbolic, $\operatorname{rob}(K) < \infty$ from the definition of the Green function. So $\operatorname{rob}(K) = \infty$ if and only if $\mathbb{C}_\infty \setminus K$ is parabolic. Next, the remarks preceding the definition show that $\operatorname{rob}(K) = \operatorname{rob}(\partial K) = \operatorname{rob}(\hat{K})$ if $\mathbb{C}_\infty \setminus K$ is hyperbolic; Proposition 19.9.5 implies the same thing in the parabolic case. Another observation that will be useful in the sequel is that for any complex number a

$$\operatorname{rob}(K) = \lim_{z \to \infty} [g(z, \infty) - \log |z - a|].$$

Indeed, $\log |z| - \log |z - a| = \log |1 - a/z| \to 0$ as $z \to \infty$.

The following results often make it easier to compute the Robin constant of a compact connected set.

3.2 Proposition. *Let K be a compact connected subset of $\mathbb{C}$ and let G be the component of $\mathbb{C}_\infty \setminus K$ that contains ∞. If $\tau : G \to \mathbb{D}$ is the Riemann map with $\tau(\infty) = 0$ and $\rho = \tau'(\infty) > 0$, then* $\text{rob}(K) = -\log \rho$. *If $f = \tau^{-1} : \mathbb{D} \to G$, then $\rho = \lim_{w\to 0} w\, f(w) = \text{Res}(f; 0)$.*

Proof. Note that $\rho = \tau'(\infty) = \lim_{z\to\infty} z\,\tau(z) = \lim_{w\to 0} w\, f(w)$, which is the residue of f at the simple pole at 0. Thus it suffices to prove the first statement. To this end, we observe that $-\log|\tau(z)| = g(z,\infty)$, the Green function for G with singularity at ∞ (19.9.2). Thus

$$\begin{aligned}
\text{rob}(K) &= \lim_{z\to\infty} [-\log|\tau(z)| - \log|z|] \\
&= \lim_{z\to\infty} -\log|z\,\tau(z)| \\
&= -\log \rho.
\end{aligned}$$

□

3.3 Corollary. *If K is a closed disk of radius R,* $\text{rob}(K) = -\log R$.

For the next corollary see Example 14.1.4.

3.4 Corollary. *If K is a straight line segment of length L,* $\text{rob}(K) = -\log(L/4)$.

We will see the Robin constant again in §10.

3.5 Proposition. *Let K be a compact subset of $\mathbb{C}$ such that $\mathbb{C}_\infty \setminus K$ is hyperbolic, let $\gamma = \text{rob}(K)$, and let G be the component of $\mathbb{C}_\infty \setminus K$ that contains ∞. If ω is harmonic measure for G at ∞ and L_ω is its logarithmic potential, then*

$$L_\omega(a) = \begin{cases} \gamma - g(a,\infty) & \text{if } a \in G \\ \gamma & \text{if } a \notin \text{cl}\, G \end{cases}$$

and $L_\omega(z) \le \gamma$ for all z in $\mathbb{C}$.

Proof. It can be assumed without loss of generality that $K = \hat{K}$. Theorem 1.20 implies that $g(z,a) = \int \log|z-a|\, d\omega_z(\zeta) - \log|z-a| + g(z,\infty)$ for all finite a and z in G, $z \ne a$. But $g(z,a) = g(a,z)$ and so

$$\begin{aligned}
L_{\omega_z}(a) &= \int \log|\zeta - a|^{-1}\, d\omega_z(\zeta) \\
&= [g(z,\infty) - \log|z-a|] - g(a,z).
\end{aligned}$$

Letting $z \to \infty$ and using the fact $L_\omega(a) \to L_\omega(a)$ (Exercise 1), we get that $L_\omega(a) = \gamma - g(a,\infty)$ for any finite point a in G.

Now suppose that $a \notin \text{cl}\ G$. Once again we invoke Theorem 1.20 to get that

$$g(z,\infty) - \log|z-a| = \int \log|z-a|^{-1} d\omega_z(\zeta).$$

Letting $z \to \infty$ shows that $L_\omega(a) = \gamma$ for $a \notin \text{cl } G$.

Finally, the fact that the Green function is positive implies that $L_\omega(z) \le \gamma$ for all z in $\mathbb{C} \setminus \partial G$. But L_ω is lsc and so for each ζ in ∂G, $L_\omega(\zeta) \le \liminf_{z \to \zeta} L_\omega(z) \le \gamma$. □

Exercises

1. If G is any hyperbolic region and $\{z_n\}$ is a sequence in G that converges to z in G, then $\omega_{z_n} \to \omega_z$ in the weak* topology on $M(\partial_\infty G)$.

2. If K is a closed arc on a circle of radius R that has length $\theta\, R$, show that $\text{rob}(K) = -\log[R\ \sin(\theta/4)]$.

3. If K is the ellipse $x^2/a^2 + y^2/b^2 = 1$, then $\text{rob}(K) = -\log[(a+b)/2]$.

4. Let $p(z) = z_n + a_1 z^{n-1} + \ldots + a_n$ and put $K = \{z \in \mathbb{C} : |p(z)| \le R\}$. Show that $\text{rob}(K) = -n^{-1} \log R$. (See Exercise 19.9.3.)

5. If $R > 0$, show that $(2\pi)^{-1} \int_0^{2\pi} \log |z - Re^{i\theta}|^{-1} d\theta$ equals $\log R^{-1}$ if $|z| < R$ and $\log |z|^{-1}$ if $|z| > R$.

6. Let $R > 0$ and let μ be the restriction of area measure to $B(a; r)$. Show that $L_\mu(z) = \pi R^2 \log |z|^{-1}$ if $|z| \ge R$ and $L_\mu(z) = \pi R^2 [\log R^{-1} + 1/2] - \pi |z|^2/2$ for $|z| \le R$.

7. Let μ be a positive measure on the plane with compact support contained in the disk $D = B(a; R)$. Show that $\frac{1}{2\pi} \int_0^{2\pi} L_\mu(a + Re^{i\theta}) d\theta = ||\mu|| \log R^{-1}$.

§4 The Green Potential

In this section G will always be a hyperbolic open set and g its Green function. We will define a potential associated with the Green function g and a positive measure μ carried by G. To justify the definition, we first prove the following.

4.1 Proposition. *If μ is a finite positive measure on G and*

4.2
$$G_\mu(z) = \int g(z, w)\ d\mu(w),$$

then G_μ defines a positive superharmonic function on G that is not identically infinite on any component of G.

Proof. First note that because both the Green function and the measure are positive, the function G_μ is well defined, though it may be that $G_\mu(z) =$

∞ for some z. For $n \geq 1$, let $g_n(z,w) = \min\{g(z,w), n\}$ and put $f_n(z) = \int g_n(z,w)\, d\mu(w)$. By Proposition 19.4.6, g_n is superharmonic. Since μ is positive, each function f_n is superharmonic. Another application of (19.4.6) shows that G_μ is superharmonic.

Suppose there is a disk $B = B(a;r)$ contained in G with $\mu(B) = 0$. Let $C \geq g(z,a)$ for all z in $G \backslash B$. Then $G_\mu(a) = \int_{G\backslash B} g(a,w)\, d_\mu(w) \leq C\, ||\mu|| < \infty$.

Now return to the case that μ is arbitrary and let B be any disk contained in G. Let $\nu = \mu|B$ and $\eta = \mu|(G \setminus B)$. Clearly $G_\mu = G_\nu + G_\eta$ and, by the preceding paragraph, neither G_ν nor G_η are identically ∞ on the component containing B. By Proposition 19.4.11, G_μ is not identically ∞ on this arbitrary component. $\square$

Because the Green function is positive, the function (4.2) is well defined even if the measure μ is not finite (though it must be assumed to be a positive measure). In case μ is not finite, the Green potential may be identically ∞ on some components of G. There may also be some positive measures μ that are infinite but such that the function (4.2) is finite valued. See Example 4.4 below.

4.3 Definition. If G is a hyperbolic open set and μ is a positive (extended real-valued) measure on G, the *Green potential* of μ is the function G_μ defined in (4.2).

4.4 Example. If G is the unit disk $\mathbb{D}$, the Green function is given by $g(z,w) = -\log(|z-w|/|1-z\overline{w}|)$. Thus for any positive measure μ on $\mathbb{D}$,

$$G_\mu(z) = L_\mu(z) - \int \log|1 - z\overline{w}|^{-1} d\mu(w).$$

If $\{a_n\}$ is any sequence of points in $\mathbb{D}$ and $\mu = \sum_n \delta_{a_n}$, then

$$G_\mu(z) = \sum_{n=1}^{\infty} \log\left|\frac{z-a_n}{1-z\bar{a}_n}\right|,$$

and this function is finite-valued for $z \neq a_n$ if $\sum_n (1-|a_n|) < \infty$ (see Exercise 8.5.4).

Shortly we will see a connection between the Green potential and the logarithmic potential of a compactly supported measure similar to the one exhibited for the disk in the preceding example. Each potential has its advantages. The Green potential is always positive and this has important consequences. In addition, the Green potential readily extends to higher dimensional spaces with important applications similar to the ones we will see below. This is not the case for the logarithmic potential. On the other hand, the definition of G_μ depends on the choice of the open set G and is only defined there, while the definition of L_μ is universal.

When we discuss the Green potential, it will always be understood that there is an underlying hyperbolic open set G whose Green function is used to define this potential.

4.5 Theorem. *If G is a hyperbolic open set, ω_z is harmonic measure for G at z, and μ is a positive measure with compact support contained in G, then*

$$G_\mu(z) = L_\mu(z) - \int_{\partial G} L_\mu(\zeta)\, d\omega_z(\zeta)$$

for all z in G.

Proof. Recall (1.16) that for any z, w in G the function $\zeta \to \log|\zeta - w|$ belongs to $L^1(\omega_z)$. Since μ has compact support, it follows that $\log|\zeta - w| \in L^1(\omega_z \times \mu)$. Indeed, for z fixed the function $w \to \int \log|\zeta - w|\, d\omega_z(\zeta)$ is harmonic on G and thus bounded on supp μ. Therefore Fubini's Theorem applies and we get that

$$\begin{aligned}\int_G \int_{\partial G} \log|\zeta - w|^{-1}\, d\omega_z(\zeta)\, d\mu(w) &= \int_{\partial G} \int_G \log|\zeta - w|^{-1} d\mu(w)\, d\omega_z(\zeta) \\ &= \int_{\partial G} L_\mu(\zeta)\, d\omega_z(\zeta).\end{aligned}$$

From Theorem 1.19 it follows that

$$g(z, w) = -\int_{\partial G} \log|\zeta - w|^{-1}\, d\omega_z(\zeta) + \log|z - w|^{-1}.$$

Integrating both sides with respect to μ and using the preceding equation gives the formula for G_μ in the theorem. □

The subsequent corollary follows from the observation that as a function of z the integral in (4.5) is harmonic on G.

4.6 Corollary. *If μ is a positive measure on G with compact support, then $G_\mu - L_\mu$ is a harmonic function on G.*

4.7 Corollary. *If μ is a positive measure on G with compact support, then, considering G_μ as a distribution on G,*

$$\Delta G_\mu = -2\pi\, \mu.$$

Consequently if H is an open subset of G such that G_μ equals a harmonic function a.e. [Area] *on H, $\mu(H) = 0$.*

4.8 Corollary. *If μ and ν are two positive measures on G with compact support and there is a harmonic function h on an open subset H of G with $G_\mu - G_\nu = h$ on H, then $\mu|H = \nu|H$.*

As a positive superharmonic function G_μ has a least harmonic minorant (19.8.2), which must be non-negative. From Proposition 19.9.2 we also know that for any point a in G the greatest harmonic minorant of the function $g_a(z) = g(z, a)$ is the constantly 0 function. This carries over to the Green potential of a positive measure.

4.9 Proposition. *If μ is a positive measure on G, then the greatest harmonic minorant of the function G_μ is 0.*

Proof. Let $\{G_n\}$ be a sequence of open subsets of G whose union is G such that $\text{cl}\, G_n \subseteq G_{n+1}$ and ∂G_n is a finite system of Jordan curves. For z in G_n let ω_z^n be harmonic measure for G_n at z. If

$$h_n(z) = \int_{\partial G_n} G_\mu(\zeta)\, d\omega_z^n(\zeta)$$

then $\{h_n\}$ is an increasing sequence and $h(z) = \lim_n h_n(z)$ for all z is the greatest harmonic minorant of G_μ (19.8.3). In a similar way define

$$g_n(z, w) = \int_{\partial G_n} g(w, \zeta)\, d\omega_z^n(\zeta)$$

where g is the Green function for G. Extend the definition of g_n by letting $g_n(z, w) = g(z, w)$ for z in $G \setminus G_n$; so for each w, $z \to g_n(z, w)$ is superharmonic on G. Since the greatest harmonic minorant of g is 0, Corollary 19.8.3 implies that $\lim_n g_n(z, w) = 0$ for all z and w. But the Green function is positive, so we can apply Fubini's Theorem to get that $h_n(z) = \int g_n(z, w)\, d\mu(w)$. On the other hand, the Maximum Principle implies $g_n(z, w) \le g(z, w)$ for all n. So Lebesgue's Dominated Convergence Theorem implies that $h_n(z) \to 0$; that is, $h = 0$. □

We are now in a position to prove another Riesz Decomposition Theorem for subharmonic functions. (See Theorem 19.5.6.) After reading the statement of the theorem, the reader should look at Exercise 2.

4.10 Riesz Decomposition Theorem. *If u is a subharmonic function on G that is not identically $-\infty$ on any component, then there is a positive measure μ (possibly infinite-valued) such that for every bounded open subset H of G with $\text{cl}\, H \subseteq G$ there is a harmonic function h on H with $u(z) = h(z) - G_{\mu|H}(z)$ for all z in H. The function h is the least harmonic majorant of u on H. If u is negative and h is the least harmonic majorant of u on G, then $u = h - G_\mu$ on G.*

Proof. As in the proof of Theorem 19.5.6, $\mu = \Delta u$ in the sense of distributions; and for any such open set H, if $\nu = \mu|H$, the harmonic function h on H exists such that $u = h - G_\nu$ on H. Since the Green potential is positive, $u \le h$ on H so that h is a harmonic majorant. If k is another harmonic function on H with $u \le k$, then $G_\nu = h - u \ge h - k$. Since the greatest harmonic minorant of G_ν is 0, we have that, on H, $0 \geq h - k$ or $k \geq h$. Therefore h is the least harmonic majorant of u on H.

Now assume that $u \leq 0$ on G. Write $G = \cup_n H_n$, where each H_n is a finitely connected Jordan region with cl $H_n \subseteq H_{n+1}$; let g_n be the Green function for H_n. From the Maximum Principle we have that $g_n \leq g_{n+1}$ so that if $\mu_n = \mu|H_n$, $G_{\mu_n} \leq G_{\mu_{n+1}}$ on G_n. In fact by monotone convergence, $G_{\mu_n}(z) \to G_\mu(z)$ for all z in G. Let h_n be the least harmonic majorant of u on H_n so that $u = h_n - G_{\mu_n}$ there. Since $u \leq 0$, $h_n \leq 0$. Clearly $\{h_n\}$ is increasing. Therefore $h(z) = \lim_n h_n(z)$ is a harmonic function on G and $u = h - G_\mu$. It is easy to see that h is the least harmonic majorant of u. □

4.11 Corollary. *If u is a positive superharmonic function on G that is not identically infinite on any component of G, then u is the Green potential of a positive measure on G if and only if the greatest harmonic minorant of u is 0.*

The next proposition is proved like Proposition 19.5.11.

4.12 Proposition. *If μ is a positive measure with compact support K in G and $G_\mu|K$ is continuous at a point a of K, then G_μ is continuous at a.*

Exercises

1. What is the connection between the formula in Theorem 4.5 relating the Green potential and the logarithmic potential of a positive measure with compact support, and the formula obtained in Example 4.4 for G_μ in the case that $G = \mathbb{D}$.

2. Let G be a hyperbolic open set and let H be an open subset of G. Suppose μ is a positive measure with compact support contained in H and define the Green potentials G_μ^G and G_μ^H of μ using the Green functions for G and H, respectively. Show that there is a positive harmonic function h on H such that $G_\mu^G(z) = G_\mu^H(z) + h(z)$ for all z in G.

3. Prove versions of (4.7), (4.8), and (4.9) for the case that μ is any positive measure on G for which G_μ is finite-valued.

4. Let μ be a positive measure with compact support K in G and assume that $G_\mu(z) < \infty$ for every point z in K. Show that for every $\varepsilon > 0$ there is a compact subset A of K with $\mu(K \setminus A) < \varepsilon$ such that $G_{\mu|A}$ is continuous on G.

5. Show that if f is a positive Borel function on G that is integrable with respect to area measure and $\mu = f \cdot \mathcal{A}|G$, then G_μ is continuous on G.

6. If G is a hyperbolic open set, K is a compact subset of G, and $\{\mu_n\}$ is a sequence of positive measures with supp $\mu_n \subseteq K$ for all $n \geq 1$ such that $\mu_n \to \mu$ weak* in $M(K)$, then $G_\mu(z) \leq \liminf_n G_{\mu_n}(z)$ for all z in G.

7. If μ is a positive measure with compact support contained in the open disk D, show that $G_\mu^D(z) \to 0$ as z approaches any point of ∂D.

§5 Polar Sets

In this section we begin a discussion that will evolve to occupy a significant portion of our attention. The basic idea is to examine sets that in the sense of harmonic functions are small.

Almost every mathematical theory has a concept of smallness or negligibility. An example already seen by the reader is the set of zeros of an analytic function: such a set must be discrete, which is small by the standards of all save the finitely oriented amongst us. Another example is a set of area 0. In measure theory we are conditioned to regard this as a kind of ultimate smallness. In the setting of the theory of analytic or harmonic functions, this turns out to be quite large and certainly not negligible. The appropriate idea of a small compact set in the sense we want is one whose complement in the extended plane is a parabolic region.

That said, we first examine results concerning subsets of the boundary of a hyperbolic set G that have *harmonic measure zero*; that is, subsets Δ of $\partial_\infty G$ such that for every a in G, Δ is ω_a- measurable and $\omega_a(\Delta) = 0$. There is a temptation to aim at a greater degree of generality by considering an arbitrary open set G, not just the hyperbolic ones, and discussing subsets Δ of $\partial_\infty G$ that have inner and outer harmonic measure zero. For example, a set Δ might be said to have outer harmonic measure zero if $\check{\chi}_\Delta(z) = 0$ for all z in G. Similarly, we would say that Δ has inner harmonic measure zero if $\hat{\chi}_\Delta(z) = 0$ for all z in G. It is an illusion, however, that this is added generality.

Indeed, suppose G is not hyperbolic and Δ is any subset of $\partial_\infty G$. A function in $\check{\mathcal{P}}(\chi_\Delta, G)$ is a positive superharmonic function. Since G is not hyperbolic, it is parabolic. Thus the only functions in $\check{\mathcal{P}}(\chi_\Delta, G)$ are constants; in fact, these constant functions must be at least 1. Thus every subset of $\partial_\infty G$ has outer harmonic measure 1. Similarly, if G is not hyperbolic, then the functions in $\hat{\mathcal{P}}(\chi_\Delta, G)$ must be negative constants and so every subset of $\partial_\infty G$ has inner harmonic measure zero.

So we will restrict our attention to hyperbolic sets. Here one could define the inner and outer harmonic measure of arbitrary subsets of $\partial_\infty G$, not just the Borel sets. A set is then harmonically measurable when its outer and inner harmonic measure agree. That is, a set Δ is harmonically measurable if and only if χ_Δ is a solvable function. In fact, Theorem 1.12 says that such sets are precisely those that are measurable with respect to harmonic measure. Attention usually will be restricted to Borel sets, though we will examine arbitrary sets of harmonic measure zero. Recall from measure theory that any subset of a set of measure zero is a measurable set having measure zero.

There are many examples of sets that have harmonic measure zero.

For example, we have already seen that if G is the inside of a rectifiable Jordan curve (or system of curves), then harmonic measure on ∂G and arc length measure are mutually absolutely continuous (Theorem 1.5). So subsets of arc length zero have harmonic measure zero. We begin by characterizing sets of harmonic measure zero in terms of subharmonic functions. To simplify matters, agree to say that for a hyperbolic set G a subset Δ of $\partial_\infty G$ is *harmonically measurable* if for every a in G, Δ is ω_a-measurable.

5.1 Theorem. *If G is a hyperbolic set and Δ is a harmonically measurable subset of $\partial_\infty G$, the following are equivalent.*

(a) $\omega_a(\Delta) = 0$ *for all a in G.*

(b) *If ϕ is a subharmonic function on G that is bounded above and satisfies*
$$\limsup_{z\to\zeta} \phi(z) \leq 0$$
for every ζ in $\partial_\infty G \setminus \Delta$, then $\phi \leq 0$.

(c) *There is a negative subharmonic function ϕ on G that is not identically $-\infty$ on any component of G and satisfies*
$$\lim_{z\to\zeta} \phi(z) = -\infty$$
for all ζ in Δ.

Proof. (a) *implies* (b). Assume $\omega_a(\Delta) = 0$ for all a in G and let ϕ be a subharmonic function as in part (b). Let M be a positive constant such that $\phi(z) \leq M$ for all z in G. Thus $M^{-1}\phi \in \hat{\mathcal{P}}(G, \chi_\Delta)$ and so for every a in G, $0 \leq M^{-1}\phi(a) \leq \hat{\chi}_\Delta(a) = \omega_a(\Delta) = 0$.

(b) *implies* (c). Note that (b) implies that $\phi \leq 0$ for every ϕ in $\hat{\mathcal{P}}(\chi_\Delta, G)$. Thus for all a in G, $0 = \hat{\chi}_\Delta(a) = \check{\chi}_\Delta(a)$. According to Proposition 19.7.8 there is a sequence $\{\phi_n\}$ in $\hat{\mathcal{P}}(-\chi_\Delta, G)$ such that $\phi_n(z) \to 0$ for every z in G. Since $\phi_n \in \hat{\mathcal{P}}(-\chi_\Delta, G)$, $\phi_n \leq 0$. Fix a point a in G; by passing to a subsequence if necessary, it can be assumed that $|\phi_n(a)| < 2^{-n}$ for all $n \geq 1$. Let $\phi = \sum_n \phi_n$; it follows by Proposition 19.4.6 that ϕ is subharmonic. Temporarily assume that G is connected. The restriction on the value of ϕ_n at the point a shows that ϕ is not identically $-\infty$ on G and clearly $\phi \leq 0$. If $\zeta \in \Delta$, then fix $N \geq 1$ and choose $\delta > 0$ such that for $1 \leq n \leq N$
$$\sup\{\phi_n(z) : z \in G,\ |z - \zeta| < \delta\} < -1 - \frac{1}{N}.$$

Again letting $G(\zeta;\delta) = B(\zeta;\delta)\cap G$ we have that

$$\begin{aligned}\sup\{\phi(z) : z\in G(\zeta;\delta)\} &\le \sum_{n=1}^{N} \sup\{\phi_n(z) : z\in G(\zeta;\delta)\}\\ &\quad + \sum_{n=N+1}^{\infty} \sup\{\phi_n(z) : z\in G(\zeta;\delta)\}\\ &\le -N+1.\end{aligned}$$

Hence

$$\limsup_{z\to\zeta} \phi(z) = -\infty.$$

In case G is not connected, let $G_1, G_2, \ldots$ be its components and for each $k\ge 1$ use the preceding paragraph to find a negative subharmonic function ϕ_k that is not identically $-\infty$ on G_k and such that $\lim_{z\to\zeta}\phi_k(z) = -\infty$ for all ζ in $\Delta\cap\partial G_k$. Define ϕ on G by letting it equal $\phi_k + k$ on G_k. This works.

(c) *implies* (a). If such a subharmonic function ϕ exists, let $\phi_n = \max\{-1, n^{-1}\phi\}$. It is easy to check that the sequence $\{\phi_n\}\subseteq\hat{\mathcal{P}}(-\chi_\Delta, G)$. If $b\in G$ and $\phi(b) > -\infty$, then $\phi_n(b)\to 0$ as $n\to\infty$. Thus $\omega_b(\Delta) = 0$. Since each component of G contains such a point b, $\omega_a(\Delta) = 0$ for all a in G (Theorem 1.7). □

Note that condition (b) in the preceding theorem is a generalized maximum principle. This also gives a proof of Exercise 13.5.7.

5.2 Corollary. *If G is a hyperbolic set, h is a bounded harmonic function on G, and there is a subset Δ of $\partial_\infty G$ having harmonic measure zero such that $h(z)\to 0$ as z approaches any point of $\partial_\infty G\setminus\Delta$, then $h\equiv 0$.*

Proof. Part (b) of the preceding theorem applied to h and $-h$ implies that both these functions are negative. □

The next proposition is one like several others below that show that certain sets are removable singularities for various classes of functions. In this case we see that sets of harmonic measure zero are removable singularities for bounded harmonic functions.

5.3 Proposition. *Let F be a relatively closed subset of G such that $\partial_\infty F\cap G$ has harmonic measure zero with respect to the set $G\setminus F$. If $h : G\setminus F\to\mathbb{R}$ is a bounded harmonic function, then h has a harmonic extension to G.*

Proof. By the use of an appropriate Möbius transformation, it can be assumed that F is a bounded set. Let $G = \cup_n G_n$, where each G_n is an open set whose boundary consists of a finite number of smooth Jordan curves and $\text{cl}\, G_n\subseteq G_{n+1}$. Fix n for the moment and let $\phi\in\hat{\mathcal{P}}(\chi_{\partial F\cap \text{cl}\, G_n}; G_n\setminus F)$ with

$\phi \geq 0$. Extend ϕ to all of $G \setminus F$ by setting $\phi = \chi_{\partial F \cap \text{cl } G_n}$ on $\partial(G_n \setminus F) = (\partial G_n \setminus F) \cup (\partial F \cap \text{cl } G_n)$ and $\phi = 0$ off cl $(G_n \setminus F)$. The reader can check that this extension of ϕ is usc on $G \setminus F$. By Proposition 19.4.17, ϕ is subharmonic on $G \setminus F$. It can be checked that $\phi \in \hat{\mathcal{P}}(\chi_{\partial F \cap G}, G \setminus F)$ and of course $\phi \geq 0$. From the hypothesis we have that $\phi \equiv 0$. Thus $\omega_z^{G_n \setminus F}(\partial F \cap \text{cl } G_n) = 0$ for all z in G_n.

Now define $g : \partial G_n \to \mathbb{R}$ by $g = h$ on $\partial G_n \setminus F$ and $g = 0$ on $\partial G_n \cap F$; let $\hat{g}$ be the corresponding solution of the Dirichlet problem on G_n (remember that G_n is a Jordan region). Thus $h - \hat{g}$ is a bounded harmonic function on $G_n \setminus F$. If $\zeta \in \partial G_n \setminus F$, $h(z) - \hat{g}(z) \to 0$ as $z \to \zeta$ with z in $G_n \setminus F$. By Corollary 5.2, $h - \hat{g} \equiv 0$ on $G_n \setminus F$. Therefore $\hat{g}$ is a harmonic extension of h to G_n. Since n was chosen arbitrarily, this shows that h can be harmonically extended to all of G. □

Subsets of the boundary of a hyperbolic set that have harmonic measure zero can be considered as sets that are locally small; their smallness is defined relative to the particular open set. We now turn to an examination of sets that are universally small relative to the study of harmonic functions.

5.4 Definition. A set Z is a *polar set* if there is a non-constant subharmonic function u on $\mathbb{C}$ such that $Z \subseteq \{z : u(z) = -\infty\}$.

Consideration of the function $\log |z - a|$ shows that a singleton is polar. A slight improvement shows that finite sets are polar and an application of Lemma 5.6 below shows that all countable sets are polar. From properties of subharmonic functions it follows that every polar set must be Lebesgue measurable with $\mathcal{A}(Z) = 0$. Indeed, if u is as in the definition and $E = \{z : u(z) = -\infty\}$, then $E = \cap_n \{z : u(z) < -n\}$ so that E is a G_δ set. Since u is not identically $-\infty$, $\mathcal{A}(E) = 0$ (19.4.11). Since $Z \subseteq E$, Z is Lebesgue measurable and $\mathcal{A}(Z) = 0$. In particular, polar sets have no interior.

The next result says that the property of being a polar set is locally verifiable.

5.5 Proposition. *A set Z is polar if and only if there is an open set G that contains Z and a subharmonic function v on G that is not identically $-\infty$ on any component of G and such that $Z \subseteq \{z \in G : v(z) = -\infty\}$.*

Proof. Assume there is an open set G that contains Z for which there is a subharmonic function v on G that is not identically $-\infty$ on any component of G and with $Z \subseteq \{z \in G : v(z) = -\infty\}$. Replacing G by $\{z : v(z) < 0\}$, we may assume that $v < 0$ on G. Let $G = \cup_n G_n$, where cl G_n is a compact subset of G_{n+1}. Let $d_n = \operatorname{diam} G_n$. Let μ be the positive measure on G that defines the positive distribution Δv (19.5.6 and 18.4.9). Note that $\mu(G_n) < \infty$ for all n. According to Theorem 4.10 there is a harmonic function h_n on G_n such that if $\mu_n = \mu | G_n$, then $v = h_n - G_{\mu_n}$ on G_n. Also

$G_{\mu_n} = L_{\mu_n} - f_n$, where f_n is harmonic on G_n. If we define

$$u_n(z) = -\int_{G_n} \log \frac{d_n}{|z-w|}\, d\mu(w),$$

then we arrive at the equation $v(z) = u_n(z) + k_n(z)$ for all z in G_n, where $k_n = f_n + h_n - d_n\mu(G_n)$ is harmonic on G_n. It is clear that u_n is a subharmonic function on $\mathbb{C}$ with $u_n(z) \le 0$ for z in G_n and $u_n(z) = -\infty$ on $Z \cap G_n$. Since $\mu|G_n$ has compact support, u_n is not identically $-\infty$. Thus $u_n(z) > -\infty$ a.e. [Area] (19.4.11). Pick a point a with $u_n(a) > -\infty$ for all $n \ge 1$ and choose constants $c_n > 0$ such that $\sum_n c_n u_n(a) > -\infty$. Put $u = \sum_n c_n u_n$. If $m \ge 1$ is fixed and $n \ge m$, $u_n(z) \le 0$ for z in G_m. It follows from Proposition 19.4.6 that u is subharmonic. Since $u(a) > -\infty$, u is not identically $-\infty$. It is routine to see that $Z \subseteq \{z : u(z) = -\infty\}$.

The converse is obvious. □

5.6 Lemma. *If $\{Z_n\}$ is a sequence of polar sets, then $\cup_n Z_n$ is polar.*

Proof. According to Proposition 5.5, there is a subharmonic function u_n on $\mathbb{C}$ that is not identically $-\infty$ such that $Z_n \subseteq \{z : u_n(z) = -\infty\}$. Let $a \in \mathbb{C}$ such that $u_n(a) > -\infty$ for all n. For each $n \ge 1$ there is a constant k_n such that $u_n(z) \le k_n < \infty$ for z in $B(0; n)$. Choose constants $c_n > 0$ such that $\sum_n c_n[k_n - u_n(a)] < \infty$. By (19.4.6), $u = \sum_n c_n[u_n - k_n]$ is a subharmonic function on $\mathbb{C}$, $u(a) > -\infty$, and $Z \subseteq \{z : u(z) = -\infty\}$. □

Using this lemma we see that to show a set is polar it suffices to show that every bounded subset of it is polar.

5.7 Proposition. *If K is a non-trivial compact connected set, then K is not a polar set.*

Proof. Suppose u is a subharmonic function on $\mathbb{C}$ such that $u(z) = -\infty$ on K; it will be shown that u is identically $-\infty$. Let G be a bounded open subset of $\{z : u(z) < 0\}$ that contains K. Now use the Riemann Mapping Theorem to get a conformal equivalence $\tau : \mathbb{D} \to \mathbb{C}_\infty \setminus K$ with $\tau(0) = \infty$. Note that if $\{z_n\}$ is a sequence in $\mathbb{D}$ such that $|z_n| \to 1$, all the limit points of $\{\tau(z_n)\}$ lie in K. Thus $u \circ \tau$ is a subharmonic function on $\mathbb{D}$ and $u \circ \tau(z) \to -\infty$ as z approaches any point on the unit circle. By the Maximum Principle, $u \circ \tau$ is identically $-\infty$ and thus so is u. □

5.8 Corollary. *If Z is a polar set, then every compact subset of Z is totally disconnected.*

5.9 Proposition. *If Z is a polar set and G is a bounded open set, then, for every point a in G, $Z \cap \partial G$ is ω_a^G-measurable and $\omega_a^G(Z \cap \partial G) = 0$.*

Proof. Without loss of generality we may assume that there is a non-constant subharmonic function v on $\mathbb{C}$ such that $Z = \{z : v(z) = -\infty\}$. By a previous argument, $Z \cap \partial G$ is a G_δ set and therefore ω_a^G-measurable. To show that $\omega_a^G(Z \cap \partial G) = 0$, it suffices to show that $\omega_a^G(K) = 0$ for every compact subset K of $Z \cap \partial G$. Let U be a bounded open set that contains K and put $H = G \cup U$. Since cl H is compact, there is a finite constant M such that $v(z) \leq M$ for all z in cl H. Thus $v - M$ is a negative non-constant subharmonic function on H that is $-\infty$ on K. By Theorem 5.1, $\omega_a^{H \setminus K}(K) = 0$ for all a in $H \setminus K$. But $G \subseteq H \setminus K$ and $K \subseteq \partial G \cap \partial(H \setminus K)$. By Corollary 1.14, $\omega_a^G(K) \leq \omega_a^{H \setminus K}(K) = 0$ for all a in G. □

The preceding proposition will be extended to all hyperbolic open sets G (Theorem 8.4, below) but some additional theory will be required.

We will rejoin the examination of arbitrary polar sets later (Theorem 7.5), but now we turn our attention to compact polar sets and their relation with some recently acquired friends. There are compact sets K contained in the boundary of an open set G such that $\omega_a^G(K) = 0$ for all a in G and K is not polar. For example, if G is the unit square (with vertices $0, 1, 1+i$, and i) and K is the usual Cantor ternary set, then ∂G is a rectifiable Jordan curve and so $\omega_a^G(K) = 0$ for all a in G. But K is not polar (8.16). If, however, K has harmonic measure zero for every open set whose boundary contains it, then K is polar. This as well as the equivalence of additional conditions will be seen in the next theorem. Remember (5.8) that compact polar sets must be totally disconnected and thus without interior.

5.10 Theorem. *For a compact totally disconnected set K, the following are equivalent.*

(a) *K is a polar set.*

(b) *If G is any bounded region with $K \subseteq \partial G$, then $\omega_a^G(K) = 0$ for every point a in G.*

(c) *If G is a bounded region that contains K and u is a bounded harmonic function on $G \setminus K$, then u admits a harmonic extension to G.*

(d) *If γ is a Jordan curve such that $K \subseteq G \equiv \operatorname{ins}\gamma$, then $\omega_a^{G \setminus K}(K) = 0$ for every point a in $G \setminus K$.*

(e) *There is a bounded region G that contains K and there is a point a in $G \setminus K$ such that $\omega_a^{G \setminus K}(K) = 0$.*

(f) *If G is any bounded region containing K, then $\omega_a^{G \setminus K}(K) = 0$ for every point a in $G \setminus K$.*

(g) *$\mathbb{C} \setminus K$ is parabolic.*

(h) *The only bounded harmonic functions on $\mathbb{C} \setminus K$ are the constant functions.*

(i) *There is no positive non-zero measure μ on K such that L_μ is bounded above.*

As might be expected, the proof of this theorem requires a lemma. It is perhaps surprising that it only requires one.

5.11 Lemma. *If G is a hyperbolic set and Δ is a measurable subset of $\partial_\infty G$ such that $0 < \omega_a(\Delta) < 1$ for some point a in G, then*

$$\inf_{z\in G} \omega_z(\Delta) = 0 \text{ and } \sup_{z\in G} \omega_z(\Delta) = 1.$$

Proof. Put $M = \sup\{\omega_z(\Delta) : z \in G\}$; so $0 < M \leq 1$. If $\phi \in \hat{\mathcal{P}}(\chi_\Delta, G)$, then $\phi \leq \hat{\chi}_\Delta \leq M$ and so $\phi/M \leq 1$. Hence $\phi/M \in \hat{\mathcal{P}}(\chi_\Delta, G)$, so that $\phi/M \leq \hat{\chi}_\Delta$; equivalently, $\phi \leq M\hat{\chi}_\Delta$. Taking the supremum over all such ϕ gives that $\hat{\chi}_\Delta \leq M\hat{\chi}_\Delta$. From the hypothesis we have that $M = 1$.

The proof of the statement about the infimum is similar. □

Proof of Theorem 5.10. That (a) implies (b) is immediate from Proposition 5.9.

(b) *implies* (c). This is a consequence of Proposition 5.3.

(c) *implies* (d). Let γ be a Jordan curve such that $K \subseteq G \equiv \text{ins } \gamma$ and put $u(z) = \omega_z^{G\setminus K}(K)$. So u is the solution of the Dirichlet problem on $G \setminus K$ with boundary values χ_K. According to (c) u has an extension to a harmonic function $h : G \to [0,1]$. But χ_K is continuous at points of γ, so u, and thus h, extends continuously to cl G with $h(z) = \chi_K(z) = 0$ for z in $\gamma = \partial G$. But then the Maximum Principle implies $h \equiv 0$. This implies (d).

(d) *implies* (e). This is trivial.

(e) *implies* (g). Suppose (g) does not hold. That is, assume that $\mathbb{C} \setminus K$ is hyperbolic; so there is a subharmonic function ϕ on $\mathbb{C} \setminus K$ such that $\phi \leq 0$ and ϕ is not constant. By Exercise 19.4.7 we can assume that ϕ is subharmonic on $\mathbb{C}_\infty \setminus K$. Let G be any bounded region in the plane that contains K. Put $m = \max\{\phi(z) : z \in \partial G\}$. Note that because ϕ is subharmonic at ∞, $\phi(z) \leq m$ for z in $\mathbb{C}\setminus G$. Let $M = \sup\{\phi(z) : z \in G\setminus K\}$ and observe that $m \leq M$.

Claim. There is a point a in $G \setminus K$ with $\phi(a) > m$ and so $m < M$.

Otherwise we would have that $\phi \leq m$ on $\mathbb{C}\setminus K$ and ϕ attains its maximum value at an interior point (in ∂G), thus contradicting the assumption that ϕ is not constant.

Define $\phi_1 = (\phi - m)/(M - m)$. It is left to the reader to check that for every ζ in $\partial(G \setminus K)$, $\limsup_{z\to\zeta} \phi_1(z) \leq \chi_K(\zeta)$. Hence $\phi_1 \in \hat{\mathcal{P}}(\chi_K, G \setminus K)$.

The claim shows that there is a point a in $G \setminus K$ with $\phi_1(a) > 0$. By definition this implies $\omega_a^{G\setminus K}(K) > 0$. Now $G \setminus K$ is connected and so this implies that $\omega_z^{G\setminus K}(K) > 0$ for all z in $G \setminus K$. Since G was arbitrary, this says that condition (e) does not hold.

(g) *implies* (f). Suppose (f) does not hold; so there is a bounded region G containing K and a point a in $G \setminus K$ with $\omega_a^{G\setminus K}(K) > 0$. Thus there is a function ϕ in $\hat{\mathcal{P}}(\chi_K, G \setminus K)$ with $\phi(a) > 0$. By replacing ϕ with $\max\{\phi, 0\}$ we may assume that $\phi \geq 0$ on $G \setminus K$. But then for any ζ in ∂G, $0 \leq \limsup_{z\to\zeta} \phi(z) \leq \chi_K(\zeta) = 0$. Thus $\phi(z) \to 0$ as z approaches any point of ∂G. If we extend ϕ to be defined on all of $\mathbb{C} \setminus K$ by setting $\phi(z) = 0$ for $z \notin G$, then it follows that ϕ is usc. By Proposition 19.4.17, ϕ is subharmonic on $\mathbb{C} \setminus K$. Clearly $\phi \leq 1$ and not constant. Therefore $\mathbb{C} \setminus K$ is hyperbolic.

(f) *implies* (a). Let G be a bounded region that contains K so that $\omega_z^{G\setminus K}(K) = 0$ for all z in $G \setminus K$. According to Theorem 5.1 there is a subharmonic function ϕ on $G \setminus K$ such that $\phi \leq 0$, ϕ is not identically $-\infty$, and $\phi(z) \to -\infty$ as z approaches any point of K. If ϕ is extended to be defined on G by letting $\phi(z) = -\infty$ for z in K, then ϕ is subharmonic on G and this shows that K is polar.

(c) *implies* (h). Let h be a bounded harmonic function on $\mathbb{C} \setminus K$. If G is any bounded region that contains K, then $h|(G \setminus K)$ admits a continuation to G by condition (c). Thus h has a continuation to a bounded harmonic function on $\mathbb{C}$ and must therefore be constant.

(h) *implies* (g). Suppose (g) is not true; so $G = \mathbb{C} \setminus K$ is hyperbolic and not the whole plane. This implies that K has more than one point. Let $a \in G$. Since ω_a has no atoms, there is a Borel set $\Delta \subseteq K = \partial G$ such that $0 < \omega_a(\Delta) < 1$. It follows from Lemma 5.11 that $h(z) = \omega_z(\Delta)$ is a non-constant bounded harmonic function on G and so (h) is not true.

(c) *implies* (i). Assume that μ is a positive measure supported on K such that there is a constant M with $L_\mu \leq M$; it will be shown that $\mu = 0$. Let G be a bounded open set containing K. If $d = \operatorname{diam} G$, then $L_\mu(z) \geq d^{-1}||\mu||$ for all z in $\operatorname{cl} G$. Thus L_μ is a bounded harmonic function on $G \setminus K$. By (c), L_μ has a harmonic extension to G. By Corollary 19.5.5, $\mu = 0$.

(i) *implies* (g). Assume that $\mathbb{C} \setminus K$ is hyperbolic. If ω is harmonic measure for $\mathbb{C}_\infty \setminus K$ at ∞, Proposition 3.5 implies L_ω is bounded above by $\operatorname{rob}(K)$. □

Condition (i) will be seen in the future. This particular characterization of polar sets will resurface in §7 when the notion of logarithmic capacity is encountered.

Exercises

1. If G is an open set with components $\{H_n\}$ and $\Delta = \partial_\infty G \setminus \cup_n \partial_\infty H_n$, then Δ has harmonic measure zero.

2. Show that if K is a compact subset of $\mathbb{C}$, then $\mathbb{C} \setminus K$ is parabolic if and only if $\mathbb{C}_\infty \setminus K$ is parabolic.

§6 More on Regular Points

In this section the results obtained on polar sets will be applied to obtain a more accurate and complete picture of the sets of regular and irregular points of a hyperbolic open set. The first result along this line generalizes the fact that isolated points in the boundary of a region are irregular (19.10.8).

6.1 Proposition. *If K is a polar set contained in the hyperbolic set G, then no point of K is a regular point for $G \setminus K$.*

Proof. Let $a \in K$ and put $\Omega = G \setminus K$; fix $r > 0$ such that $B = B(a; r)$ has cl $B \subseteq G$. If h is the solution of the Dirichlet problem on $B \cap \Omega$ with boundary values $|\zeta - a|$, then $h(z) = r$ on $\partial B \cap \Omega$ since ∂B is a connected subset of the boundary of $B \cap \Omega$. We want to show that $\limsup_{z \to a} h(z) > 0$.

Recall that K is totally disconnected and let L be a compact subset of $K \cap B$ such that $a \in L$ and L is relatively open in K. There is an open set U in $\mathbb{C}$ such that $L = K \cap U$; without loss of generality it can be assumed that $U \subseteq B$. Now L is a polar set and h is a bounded harmonic function on $U \setminus L$; thus h has a harmonic extension to U. Denote this extension by h_1. So $\limsup_{z \to a} h(z) = \lim_{z \to a} h_1(z) = h_1(a)$. But $h_1 \geq 0$, so, by the Maximum Principle, $h_1(a) > 0$. □

The next theorem gives an additional list of conditions that can be added to those in Theorem 19.10.1 that are equivalent to the regularity of a point. These are given in terms of the Green function.

6.2 Theorem. *If G is a hyperbolic region and $a \in \partial_\infty G$, then the following are equivalent.*

(a) *a is a regular point of G.*

(b) *There is a point w in G such that if $g(z, w)$ is the Green function for G with pole at w, then $g(z, w) \to 0$ as $z \to a$.*

(c) *For every w in G, if $g(z, w)$ is the Green function for G with pole at w, then $g(z, w) \to 0$ as $z \to a$.*

Proof. It is trivial that (c) implies (b) and the fact that (b) implies (a) is immediate from Theorem 19.10.1 and the fact that the Green function is a positive superharmonic function. It remains to prove that (a) implies (c). Assume that a is a regular point for G and fix an arbitrary point w in G. Without loss of generality we can assume $w = \infty$. Let $r > 0$ and put $B = B(a; r)$. Let u be the solution of the Dirichlet problem on $G(a; r) = B \cap G$ with boundary values $\chi_{\partial B}$. According to Proposition 19.10.5, a is a regular point for the open set $G(a; r)$; thus $u(z) \to 0$ as $z \to a$ in $G(a; r)$.

Let $0 < r_1 < r$ and put $B_1 = B(a; r_1)$. If $m = \sup\{u(z) : z \in \partial B_1\}$, then $0 < m < 1$. Let $h(z) = 1 - (1 - m)[\log(|z - a|/r)/\log(r_1/r)]$ for $r_1 \le |z - a| \le r$. Note that $h(z) = 1$ for $|z - a| = r$ and $h(z) = m$ for $|z - a| = r_1$.

Put $H = G(a; r) \setminus \text{cl}\, B_1$ and define $v : \partial H \to \mathbb{R}$ by letting $v(z) = u(z)$ for z in $\partial H \cap G(a; r) = \partial B_1 \cap G(a; r)$ and $v(z) = \chi_{\partial B}(z)$ for z in $\partial H \cap \partial G(a; r)$. If $\hat{v}$ is the solution of the Dirichlet problem on H with boundary values v, then Proposition 1.13 implies $\hat{v}(z) = u(z)$ for z in H. It is left as an exercise for the reader to verify that $h \in \check{\mathcal{P}}(v, H)$. Hence $h \ge \hat{v} = u$ on H.

Now let $C = (1 - m)^{-1} \log(r/r_1) > 0$. So $Ch(z) = C + \log(r^{-1}|z - a|)$ and hence

6.3 $$C\, u(z) \le C + \log(r^{-1}|z - a|) \text{ on } H.$$

Define the function ψ on G by setting $\psi(z) = C + \log(r^{-1}|z - a|)$ for z in $G \setminus B(a; r)$ and $\psi(z) = C\, u(z)$ for z in $G(a; r)$. Since u extends continuously to $G \cap \partial B$ and is equal to 1 there, ψ is a continuous function on G. It is left to the reader to verify that (6.3) implies that ψ is superharmonic. Since ψ is clearly positive and $\psi(z) - \log|z|$ is also superharmonic near ∞, $\psi \ge g(z, \infty)$ by Proposition 19.9.10. Therefore on $G(a; r)$, $0 \le g(z, \infty) \le C\, u(z) \to 0$ as $z \to a$. $\square$

The next lemma is stated and proved for bounded regions. It is also true for unbounded regions; see Exercise 1.

6.4 Lemma. *If G is a bounded region, $a \in G$, and $G_n \equiv \{z \in G : g^G(z, a) > n^{-1}\}$, then G_n is connected, $g^{G_n}(z, a) = g^G(z, a) - n^{-1}$, and $\omega_a^{G_n}(\partial G \cap \partial Gn) = 0$.*

Proof. If G_n is not connected, there is a component H of G_n that does not contain a. Define ψ on G by letting $\psi(z) = n^{-1}$ for z in H and $\psi(z) = g^G(z, a)$ elsewhere. It is left as an exercise to show that ψ is superharmonic, $\psi \ge 0$, and $\psi(z) + \log|z - a|$ is also superharmonic. By Proposition 19.9.10, $\psi(z) \ge g^G(z, a)$. But on H, $\psi(z) = n^{-1} < g^G(z, a)$, a contradiction. Therefore G_n is connected.

Now fix $n \ge 1$ and define h on G_n by $h(z) = g^G(z, a) - n^{-1}$. It is easily checked that $h \ge 0$, h is harmonic on $G_n \setminus \{a\}$, and $h(z) + \log|z - a|$ is harmonic near a. By definition of the Green function, $h(z) \ge g^{G_n}(z, a)$ on G_n. On the other hand, $h(z) - g^{G_n}(z, a)$ is harmonic on G_n and, for any ζ in ∂G_n, $\limsup_{z\to\zeta}[h(z) - g^{G_n}(z, a)] \le \limsup_{z\to\zeta} h(z) - \liminf_{z\to\zeta} g^{G_n}(z, a) \le 0$. Thus $h(z) \le g^{G_n}(z, a)$ and so equality holds.

Finally put $K_n = \partial G \cap \partial G_n$ and let $\phi \in \hat{\mathcal{P}}(\chi_{K_n}, G_n)$; without loss of generality we may assume that $\phi \ge 0$ (19.7.8). So for any ζ in $\partial G_n \cap G$, $0 \le \liminf_{z\to\zeta} \phi(z) \le \limsup_{z\to\zeta} \phi(z) \le \chi_{K_n}(\zeta) = 0$; that is, for all ζ in $\partial G_n \cap G$, $\phi(z) \to 0$ as $z \to \zeta$ with z in G_n. So if ψ is defined on G by $\psi(z) = g^G(z, a) - n^{-1}\phi(z)$ for z in G_n and $\psi(z) = g^G(z, a)$ otherwise, ψ

is a positive superharmonic function on G such that $\psi(z) + \log|z-a|$ is also superharmonic (verify!). Thus $\psi(z) \geq g^G(z,a)$ by (19.9.10). But this implies $\phi \leq 0$ so that $\phi \equiv 0$. Therefore $\omega_a^{G_n}(K_n) = 0$. □

6.5 Theorem. *If G is hyperbolic and E is the set of irregular points for G, then E is a polar set that is the union of a sequence of compact subsets of $\partial_\infty G$.*

Proof. In light of Corollary 19.10.10 it suffices to assume that G is connected. First consider the case that G is bounded, so that the preceding lemma applies; adopt its notation. Let $K_n = \partial G \cap \partial G_n$; by Theorem 6.2, $E = \cup_n K_n$ is the set of irregular points of G. It only remains to show that each set K_n is polar. This will be done by using Theorem 5.10.i.

Note that for each $n \geq 1$,

$$\begin{aligned} g^{G_{n+1}}(z,a) &= g^{G_{n+1}}(a,z) \\ &= \int \log|z-\zeta|\, d\omega_a^{G_{n+1}}(\zeta) - \log|z-a|. \end{aligned}$$

The right hand side of this equation is upper semicontinuous throughout $\mathbb{C}$. So for any point w in K_n

$$\begin{aligned} &\int \log|w-\zeta|\, d\omega_a^{G_{n+1}}(\zeta) - \log|w-a| \\ &\qquad \geq \limsup_{z\to w}\{g^{G_{n+1}}(z,a) : z \in G_n\} \\ &\qquad \geq \limsup_{z\to w}\{g^{G}(z,a) - \frac{1}{n+1} : z \in G_n\} \\ &\qquad \geq \frac{1}{n} - \frac{1}{n+1}. \end{aligned}$$

Hence

$$\int \log|w-\zeta|\, d\omega_a^{G_{n+1}}(\zeta) > \log|w-a| + \frac{1}{n(n+1)}$$

for all w in K_n. If it were the case that K_n is not a polar set, then there is a probability measure μ supported on K_n and a finite constant M such that $L_\mu(z) \leq M$ for all z. But

$$\begin{aligned} L_\mu(a) &= \int \log|w-a|^{-1}\, d_\mu(w) \\ &> \int\int \log|w-\zeta|^{-1}\, d\omega_a^{G_{n+1}}(\zeta)\, d\mu(w) \\ &= \int L_\mu(\zeta)\, d\omega_a^{G_{n+1}}(\zeta) \end{aligned}$$

6.6
$$= u(a),$$

where u is the solution of the Dirichlet problem on G_{n+1} with boundary values L_μ. Now $L_\mu - u$ is harmonic on G_{n+1} and bounded since L_μ is bounded below on compact sets (Exercise 19.5.1). But for each ζ in $\partial G_{n+1} \cap G$, as $z \to \zeta$ with z in G_{n+1} it holds that

$$g^{G_{n+1}}(z,a) = g^G(z,a) - \frac{1}{n+1} \to g^G(\zeta,a) - \frac{1}{n+1} = 0.$$

By Theorem 6.2 this implies that points in $\partial G_{n+1} \cap G$ are regular for G_{n+1}. Hence $L_\mu(z) - u(z) \to 0$ as $z \to \zeta$ in $\partial G_{n+1} \cap G$. Also $\partial G_{n+1} \cap \partial G = K_{n+1}$ and $\omega_a^{G_{n+1}}(K_{n+1}) = 0$ by the preceding lemma. Therefore Corollary 5.2 implies $L_\mu - u \equiv 0$, contradicting (6.6). Thus no such measure μ exists and K_n is polar.

Now let G be a not necessarily bounded hyperbolic open set and put $G_k = G \cap B(0;k)$. If E_k is the set of irregular points of G_k, then the fact that each component of $\partial B(0;k) \cap G$ is a non-trivial arc of a circle implies that $E_k \subseteq \partial G$. By Proposition 19.10.5 this implies that $\cup_k E_k \subseteq E$. The use of barriers shows that $E = \cup_k E_k$. The general result now follows from the proof of the bounded case. □

6.7 Corollary. *If G is a bounded open set, the collection of irregular points has harmonic measure zero.*

Proof. This is immediate from the preceding theorem and Proposition 5.9. □

The preceding corollary will be extended to hyperbolic open sets G in Corollary 8.4 below.

Exercises

1. Use Proposition 19.9.8 to prove Lemma 6.4 for unbounded regions.

2. If G is hyperbolic, K is a compact subset of G, and ζ is a regular boundary point of G, show that as $z \to \zeta$, $g(w,z) \to 0$ uniformly for w in K.

§7 Logarithmic Capacity: Part 1

Recall that, for a compact set K, $M(K)$ is the set of regular Borel measures on K. For a non-compact set E, let $M^c(E)$ be all the measures that belong to $M(K)$ for some compact subset K of E. (The superscript "c" here stands for "compact.") That is, $M^c(E)$ is the set of regular Borel measures whose support is compact and contained in E. $M^c_+(E)$ will denote the positive measures belonging to $M^c(E)$ and $M^c_{\mathbb{R}}(E)$ is the set of real-valued measures belonging to $M^c(E)$. If E is any set, let

$$M^c_I(E) \equiv \{\mu : \mu \in M^c_{\mathbb{R}}(E) \text{ and } L_\mu \in L_1(|\mu|)\}.$$

It may occur that $M^c_I(E) = \emptyset$ (for example, if E is a single point). If $\mu \in M^c_I(E)$, let

$$I(\mu) = \int\int \log \frac{1}{|z-w|}\, d\mu(z)\, d\mu(w).$$

For such measures μ, $I(\mu)$ is a well defined finite number. Indeed $I(\mu) = \int L_\mu d\mu$. The number $I(\mu)$ is called the *energy integral* of μ. The use of this term, as well as the term logarithmic potential, is in analogy with the terminology for electrostatic potentials in three dimensions.

7.1 Lemma. *If $\mu \in M_+(E)$, then either $\mu \in M^c_I(E)$ or*

$$\lim_{n\to\infty} \int \min\{L_\mu, n\}\, d\mu = +\infty.$$

Proof. This is a direct consequence of the fact that the logarithmic potential of a positive measure is bounded below on compact sets. □

In light of this lemma we can define $I(\mu) = \int L_\mu d\mu$ for all positive measures, where we admit the possibility that $I(\mu) = \infty$. Another piece of notation is that $M^c_1(E)$ will denote all the probability measures in $M^c(E)$. That is, $M^c_1(E) = \{\mu \in M^c_+(E) : ||\mu|| \equiv \mu(\operatorname{supp}\mu) = 1\}$.

7.2 Definition. If E is any set such that $M^c_I(E) \neq \emptyset$, define

$$v(E) \equiv \inf\{I(\mu) : \mu \in M^c_1(E)\}.$$

The *logarithmic capacity* of such a set E is defined by

$$c(E) = e^{-v(E)}.$$

If E is such that $M^c_I(E) = \emptyset$, define $v(E) = \infty$ and $c(E) = 0$. A property that holds at all points except for a set of capacity zero is said to hold *quasi-everywhere*. This is abbreviated "q.e."

The term capacity is used in analysis in a variety of ways. The common thread here is that it is a way of associating with sets a number that measures the smallness of the set relative to the theory under discussion. Another role of capacities is to assist in making estimates. That is, a set having small capacity will imply the existence of certain functions having rather precise technical properties. An instance of this occurs in §14 below, where Wiener's criterion for the regularity of a boundary point is established. In the present case, it will shortly be proved that a compact set K has zero capacity if and only if it is polar (Theorem 7.5, below). In $\mathbb{R}^d$ for $d \geq 3$, there is an analogous notion of capacity where compact sets having zero capacity are removable sets of singularities for bounded harmonic functions. (See Landkof [1972], p 133.) In the study of bounded analytic functions there is also a notion of analytic capacity with similar properties. (See Conway [1991], p 217.) There is also a general theory of capacity that originated in Choquet [1955]. (Also see Carleson [1967].) It is becoming common to refer to a capacity that fits into this general theory as a "true" or Choquet capacity.

Unfortunately, the logarithmic capacity defined above is not a true capacity (nor is the analytic capacity used in the study of analytic functions).

A modification of the logarithmic capacity can be made that produces a true capacity. The extra effort to do this is modest and it is therefore presented in conjunction with the development of logarithmic capacity. This will only be defined for subsets of the disk $r\mathbb{D}$.

If $\mu \in M_+^c(r\mathbb{D})$, define

$$L_\mu^r(z) = \int \log \frac{2r}{|z-w|}\, d\mu(w)$$

for all z in $r\mathbb{D}$. Note that $L_\mu^r(z) \geq 0$ on $r\mathbb{D}$, though it may be infinite valued. Similarly define

$$I_r(\mu) = \int\int \log \frac{2r}{|z-w|}\, d\mu(w)\, d\mu(z)$$

for μ in $M_+^c(r\mathbb{D})$.

7.3 Definition. For any subset E of $r\mathbb{D}$, define

$$v_r(E) \equiv \inf\{I_r(\mu) : \mu \in M_i^c(E)\}.$$

The *r-logarithmic capacity* of E is defined by

$$c_r(E) = \frac{1}{v_r(E)}.$$

We note the equations $L_\mu^r = ||\mu|| \log 2r + L_\mu$ and $I_r(\mu) = ||\mu||^2 \log 2r + I(\mu)$, from which it follows that $v_r(E) = \log 2r + v(E)$ for all subsets E of $r\mathbb{D}$. Thus a subset E of $r\mathbb{D}$ has logarithmic capacity zero if and only if it has r-logarithmic capacity zero; in fact, for such sets, $c_r(E) = [\log(2r/c(E))]^{-1}$.

The r-logarithmic capacity is a true or Choquet capacity defined on the subsets of $r\mathbb{D}$. However logarithmic capacity as defined in (7.2) is more closely related to the geometric properties of analytic and harmonic functions on the plane. For example, it will be shown that, for any compact set K, $c(K) = e^{-\gamma}$, where γ is the Robin constant for K (Theorem 10.2 below). The r-logarithmic capacity has also been heavily used in the study of analytic and harmonic functions; see Beurling [1939], Carleson [1967], and Richter, Ross, and Sundberg [1994].

It is also the case that the Green potential can be used to define a Green capacity (for subsets of the parent set G) that is a Choquet capacity. This will not be done here. The interested reader can look at Helms [1975]. Also see Landkof [1972] and Brelot [1959]. A distinct advantage of the Green capacity is that it generalizes to higher dimensional spaces. A disadvantage is that it is restricted to and dependent on the chosen set G.

In what follows, results will only be stated for logarithmic capacity unless there is a difficulty with the corresponding fact for r-logarithmic capacity or a particular emphasis is called for. Of course, exact formulas or numerical estimates for logarithmic capacity will not carry over directly to r-logarithmic capacity, though some modification will. The first result is a collection of elementary facts. The proofs are left to the reader.

7.4 Proposition.

(a) *If $E_1 \subseteq E_2$, then $c(E_1) \leq c(E_2)$.*

(b) *For any set E, $c(E) = \sup\{c(K) : K$ is a compact subset of $E\}$.*

(c) *If $T(z) = az + b$, $a \neq 0$, then $v(T(E)) = v(E) - \log|a|$ and $c(T(E)) = |a|\, c(E)$. Thus $c(E) = 0$ if and only if $c(T(E)) = 0$.*

7.5 Theorem. *If E is a Borel set, the following are equivalent.*

(a) *E has positive capacity.*

(b) *There is a non-zero measure μ in $M_+(E)$ such that L_μ is bounded above.*

(c) *There is a compact subset of E that is not polar.*

(d) *There is a positive measure μ with $I(\mu) < \infty$ and $\mu(E) > 0$.*

If K is a compact set that has positive capacity, then $c(K) \geq e^{-\gamma}$, where γ is the Robin constant for K.

Proof. In light of part (b) of the preceding proposition, it suffices to assume that E is a compact set K. By part (c) of the preceding proposition, with an appropriate choice of constants it can be assumed that $K \subseteq B(0; 1/2)$. (This is a typical use of (7.4.c) and will be seen again in the course of this development.) The virtue of this additional assumption is that $\log|z - w|^{-1} \geq 0$ for all z, w in K.

(a) *implies* (b). Let $\nu \in M_1(K)$ with $I(\nu) < \infty$. Since $K \subseteq B(0; 1/2)$, $L_\nu(z) \geq 0$ on K. Since $L_\nu \in L^1(\nu)$, there is a constant M such that if $F = \{z \in \text{supp}\ \nu : L_\nu(z) \leq M\}$, then $\nu(F) > 0$. Now L_ν is a lower semicontinuous function: thus F is compact. If $\mu = \nu|F$, then $L_\mu(z) \leq L_\nu(z) \leq M$ for all z in F and hence throughout $\mathbb{C}$ (19.5.10).

(b) *implies* (c). This is immediate from Theorem 5.10.

(c) *implies* (d). By Theorem 5.10 there is a μ in $M_+(K)$ such that $L_\mu \leq M$; hence $I(\mu) = \int L_\mu d\mu < \infty$.

(d) *implies* (a). If μ is a positive measure with $\mu(K) > 0$ and $I(\mu) < \infty$, then by the definition of $v(K)$, $v(K) < \infty$ and so $c(K) > 0$. This completes the proof that statements (a) through (d) are equivalent.

If $c(K) > 0$, G is the component of $\mathbb{C}_\infty \setminus K$ that contains ∞, and ω is harmonic measure for G at ∞, then Proposition 3.5 states that $L_\omega \leq \gamma$ on the plane. Thus $v(K) \leq I(\omega) \leq \int L_\omega d\omega \leq \gamma$ and so $c(K) \geq e^{-\gamma}$. □

Think of the first of the following two corollaries as a result about absolute continuity of measures with respect to logarithmic capacity, even though logarithmic capacity is not a measure.

7.6 Corollary. *If K is a compact set with $c(K) > 0$ and $\mu \in M_I^c(K)$, then $|\mu|(\Delta) = 0$ for every Borel set Δ with $c(\Delta) = 0$.*

The next corollary is immediate from Corollary 5.8.

7.7 Corollary. *If K is a compact set with logarithmic capacity 0, then K is a totally disconnected set.*

We know that polar sets have area zero; the next proposition refines this statement into a numerical lower bound for capacity in terms of area.

7.8 Proposition. *If E is a Borel set, then*

$$c(E) \geq \sqrt{\frac{\text{Area}(E)}{\pi\, e}}.$$

Proof. The proof is reminiscent of the proof of Proposition 18.5.3. By virtue of Proposition 7.4.b, we may assume that E is a compact set K. Without loss of generality we may also assume that the area of K is positive. If u is the logarithmic potential of the restriction of area measure to K, then u is a continuous function on the plane; by (19.5.10) it attains its maximum value on K. By translation, we may assume that $0 \in K$ and $u(z) \leq u(0) = \int_K \log|w|^{-1} d\mathcal{A}(w)$. If R is the radius with $\pi R^2 = |K| = \text{Area}(K)$ and $D = \overline{B}(0; R)$, then $\text{Area}(D) = \text{Area}(K)$ and $\text{Area}(K \setminus D) = \text{Area}(D \setminus K)$. Therefore

$$\begin{aligned} u(0) &= \int_{K\cap D} \log|w|^{-1} d\mathcal{A}(w) + \int_{K\setminus D} \log|w|^{-1} d\mathcal{A}(w) \\ &\leq \int_{K\cap D} \log|w|^{-1} d\mathcal{A}(w) + \mathcal{A}(D\setminus K)\log R \end{aligned}$$

$$\begin{aligned}
&\leq \int_{K\cap D} \log|w|^{-1} d\mathcal{A}(w) + \int_{D\setminus K} \log|w|^{-1} d\mathcal{A}(w) \\
&= \int_D \log|w|^{-1} d\mathcal{A}(w) \\
&= 2\pi \int_0^R r\log r^{-1} dr \\
&= \pi R^2[\frac{1}{2} + \log R^{-1}] \\
&= \mathcal{A}(K)\log\sqrt{\frac{\pi\, e}{\mathcal{A}(K)}}.
\end{aligned}$$

Thus

$$\int_K \int_K \log|z-w|^{-1} d\mathcal{A}(z)\, d\mathcal{A}(w) \leq \mathcal{A}(K)^2 \log\sqrt{\frac{\pi\, e}{\mathcal{A}(K)}}.$$

If $\mu = \mathcal{A}(K)^{-1}\mathcal{A}|K$, then $v(K) \leq I(\mu) \leq \log\sqrt{\pi\, e/\mathcal{A}(K)}$, whence the result. □

7.9 Theorem. *If K is a compact set with positive capacity, then there is a probability measure μ with support contained in K such that $I(\mu) = v(K)$.*

Proof. Let $\{\mu_n\}$ be a sequence in $M_1(K)$ such that $I(\mu_n) \to v(K)$.

But $M_1(K)$ is a compact metric space when it has the weak* topology. So by passing to a subsequence if necessary, it can be assumed that there is a measure μ in $M_1(K)$ such that $\mu_n \to \mu$ weak*. By Exercise 1, $I(\mu) = v(K)$. □

7.10 Definition. If K is a compact set and $\mu \in M_1(K)$ such that $I(\mu) = v(K)$, then μ is called an *equilibrium measure* for K. The corresponding logarithmic potential L_μ is called a *conductor* or *equilibrium potential* of E.

Later (10.2) it will be shown that there is only one equilibrium measure and we can speak of *the* equilibrium measure for a compact set. In the companion development of the r-logarithmic capacity, we must make a point explicit. The proof is straightforward.

7.11 Proposition. *If K is a compact subset of $r\mathbb{D}$ and μ is an equilibrium measure for K, then $I_r(\mu) = v_r(K)$. Conversely, if μ is a probability measure on K such that $I_r(\mu) = v_r(K)$, then μ is an equilibrium measure for K.*

7.12 Theorem. (Frostman [1935]) *If K is a compact set and μ is an equilibrium measure, then $L_\mu \leq v(K)$ on $\mathbb{C}$ and $L_\mu = v(K)$ everywhere on K except for an F_σ set with capacity zero.*

Proof. Let $v = v(K)$ and put $E = \{z \in K : L_\mu(z) < v\}$. So $E = \cup_n E_n$, where $E_n = \{z \in K : L_\mu(z) \leq v - n^{-1}\}$. Because L_μ is lsc, each set E_n is closed and so E is an F_σ set. It will be shown that $c(E_n) = 0$ for each $n \geq 1$. If there is an $n \geq 1$ such that $c(E_n) > 0$, let $\nu \in M_1(E_n)$ such that $I(\nu) < \infty$. If $0 < \delta < 1$, $\mu_\delta = (1-\delta)\mu + \delta\nu \in M_1(K)$ and

$$\begin{aligned} I(\mu_\delta) &= (1-\delta)^2 I(\mu) + \delta^2 I(\nu) + 2\delta(1-\delta)\int L_\mu d\nu \\ &= v - 2\delta v + 2\delta \int L_\mu d\nu + \delta^2 A, \end{aligned}$$

where $A = -v + I(\nu) - 2\int L_\mu d\nu$. Hence

$$\begin{aligned} I(\mu_\delta) &\leq v - 2\delta v + 2\delta(v - n^{-1}) + \delta^2 A \\ &= v + \delta[-2n^{-1} + \delta A] \\ &< v \end{aligned}$$

for a suitably small δ. Since $\mu_\delta \in M_1(K)$, this contradicts the definition of v. Hence $c(E_n) = 0$ for all $n \geq 1$. By Lemma 5.6 and Theorem 7.5, the countable union of sets of capacity zero has capacity zero and so $L_\mu \geq v$ q.e.

Now we show that $L_\mu \leq v$ everywhere. Otherwise the Maximum Principle for the logarithmic potential implies there is a point a in $F = \operatorname{supp}\mu$ such that $L_\mu(a) > v$. Since L_μ is lsc, there is an open neighborhood U of a with $L_\mu > v$ in U; because $a \in F$, $\mu(U) > 0$. On $F \setminus U$, $L_\mu \geq v$ q.e by the first part of the proof. By (7.6), $L_\mu \geq v$ a.e. $[\mu]$. Hence $v = I(\mu) = \int_U L_\mu d\mu + \int_{F\setminus U} L_\mu d\mu > v\mu(U) + v\mu(F \setminus U) = v$, a contradiction. □

7.13 Corollary. *If K is compact with $c(K) > 0$ and μ is an equilibrium measure, then L_μ is continuous at each point a where $L_\mu(a) = v(K)$.*

Proof. By Proposition 19.5.11, it suffices to show that $L_\mu|K$ is continuous at a. Since L_μ is lsc, $v(K) = L_\mu(a) \leq \liminf_{z\to a} L_\mu(z) \leq \limsup_{z\to a} L_\mu(z) \leq v(K)$. □

7.14 Proposition. *Let K be a compact set with $c(K) > 0$ and let μ be an equilibrium measure. For every $\varepsilon > 0$ there is a compact subset K_1 of K with $\mu(K \setminus K_1) < \varepsilon$ and such that, if $\mu_1 = \mu|K_1$, then L_{μ_1} is a continuous finite-valued function on $\mathbb{C}$.*

Proof. Put $c = c(K)$ and $v = v(K)$. By Corollary 7.6 and Theorem 7.12, $L_\mu = v$ a.e. $[\mu]$. Let K_1 be a compact subset of K with $\mu(K \setminus K_1) < \varepsilon$ and $L_\mu = v$ on K_1. By the preceding corollary, L_μ is continuous at each point of K_1. Let μ_1 be as in the statement of the proposition and put $\mu_2 = \mu - \mu_1 = \mu|(K \setminus K_1)$. Now $L_\mu = L_{\mu_1} + L_{\mu_2} \leq v$ on $\mathbb{C}$. Since both L_{μ_1} and L_{μ_2} are bounded below on compact sets, it follows that both functions are bounded on compact subsets of the plane. In particular, both functions are finite-valued. Also $L_{\mu_1} = L_\mu - L_{\mu_2}$ and thus L_{μ_1} is usc; since this function is also lsc, it follows that it is continuous. □

We conclude this section with a result that will be used later.

7.15 Proposition. *If $\{K_n\}$ is a sequence of compact sets such that $K_n \supseteq K_{n+1}$ for all n and $\bigcap_n K_n = K$, then $c(K_n) \to c(K)$. If $c(K) > 0$ and if μ_n is an equilibrium measure for K_n, then every weak* cluster point of the sequence $\{\mu_n\}$ is an equilibrium measure for K.*

Proof. A rudimentary argument shows that if U is any open set containing K, then $K_n \subseteq U$ for all sufficiently large n. In particular, there is no loss in generality in assuming that the sequence $\{K_n\}$ is uniformly bounded. By Proposition 7.4.c we may assume that all the sets $K_n \subseteq \{z : |z| \leq 1/2\}$.

From elementary considerations we know that $c(K) \leq c(K_{n+1}) \leq c(K_n)$ so that $\lim_n c(K_n)$ exists and is at least $c(K)$. Equivalently, $v(K) \geq \lim_n v(K_n)$.

Let μ_n be the equilibrium measure for K_n so that $I(\mu_n) = v(K_n)$. There is a subsequence $\{\mu_{n_k}\}$ that converges to a probability measure μ in $M(\mathbb{C}_\infty)$. It is left to the reader to show that supp $\mu \subseteq K$. Since $\log |z-w|^{-1} \geq 0$ on K_1, Fatou's Lemma implies that

$$\begin{aligned} v(K) &\leq I(\mu) \\ &\leq \liminf_k \int\int \log|z-w|^{-1} d\mu_{n_k}(z)\, d\mu_{n_k}(w) \\ &= \liminf_k v(K_{n_k}) \\ &= \lim_k v(K_{n_k}) \\ &\leq v(K). \end{aligned}$$

□

7.16 Corollary. *If K is a compact set and $\{U_n\}$ is a sequence of open sets such that $U_n \supseteq U_{n+1}$ for all n and $\bigcap_n U_n = K$, then $c(U_n) \to c(K)$. In particular, if $\varepsilon > 0$, there is an open set U that contains K with $c(U) < c(K) + \varepsilon$.*

Proof. For each n let V_n be an open neighborhood of K that is bounded with $K_n = \text{cl } V_n \subseteq U_n$. So $c(K_n) \leq c(U_n)$ and the result follows from the proposition. □

Exercises

1. If $\mu, \mu_1, \mu_2, \ldots$ are positive measures whose supports are contained in the compact set K and $\mu_n \to \mu$ weak* in $M(K)$, show that $I(\mu) \leq$

$\liminf_n I(\mu_n)$. Use this to show that $I : M_+(K) \to (-\infty, \infty]$ is lower semicontinuous.

2. Prove the result for r-logarithmic capacity corresponding to Theorem 7.12.

3. Let G be a hyperbolic open set, let K be a compact subset of G with $c(K) > 0$, and let μ be an equilibrium measure for K. Combine Proposition 7.14 with Theorem 4.5 to show that for every $\varepsilon > 0$ there is a compact subset K_1 with $\mu(K \setminus K_1) < \varepsilon$ such that if $\mu_1 = \mu|K_1$, then G_{μ_1} is a continuous finite-valued function on G.

4. If K is a compact set and μ is a probability measure on K, show that $\inf\{L_\mu(z) : z \in K\} \leq v(K)$.

5. Suppose K is a compact subset of some disk of radius $1/2$ and $K = K_1 \cup \cdots \cup K_n$, where each K_j is compact and $K_i \cap K_j = \emptyset$ for $i \neq j$. Show that $v(K) \leq v(K_1) + \ldots + v(K_n)$.

§8 Some Applications and Examples of Logarithmic Capacity

Here we will give a few applications of the preceding section. Some of these applications will tie together loose ends that exist in earlier sections; many will be used to push our study of potential theory further. Later we will see some examples of sets with zero or positive capacity.

We begin with an easy application of Corollary 7.6 that has important implications.

8.1 Theorem. *If G is a hyperbolic open set and E is a Borel subset of $\partial_\infty G$ with $c(E) = 0$, then $\omega_z^G(E) = 0$ for all z in G.*

Proof. Fix z in G. Using a Möbius transformation, there is no loss of generality in assuming $z = \infty$. So $K = \partial_\infty G = \partial G$ has $c(K) > 0$. If $\gamma = \text{rob}(K)$ and $\omega = \omega_\infty^G$, then $L_\omega \leq \gamma$ and so $I(\omega) \leq \gamma$; thus $\omega \in M_I^c(K)$. If E is a Borel set with $c(E) = 0$ and $E \subseteq K$, then Corollary 7.6 implies $\omega(E) = 0$. □

In light of the preceding theorem, the next few results are immediate from previous results in this chapter. Reference to the earlier versions of the results is given at the end of the statement.

8.2 The Maximum Principle. *If G is a hyperbolic open set, E is a Borel subset of $\partial_\infty G$ with $c(E) = 0$, and ϕ is a subharmonic function on the hyperbolic set G that is bounded above and satisfies*

$$\lim_{z \to \zeta} \phi(z) \leq 0$$

for every ζ *in* $\partial_\infty G \setminus E$, *then* $\phi \leq 0$. (See Theorem 5.1.)

8.3 Corollary. *If* G *is a hyperbolic set,* h *is a bounded harmonic function on* G, *and there is a Borel subset* E *of* $\partial_\infty G$ *with* $c(E) = 0$ *such that* $h(z) \to 0$ *as* z *approaches any point of* $\partial_\infty G \setminus \Delta$, *then* $h \equiv 0$.

8.4 Theorem. *If* Z *is a polar set and* G *is a hyperbolic set, then for every point* a *in* G, $Z \cap \partial G$ *is* ω_a^G*-measurable and* $\omega_a^G(Z \cap \partial G) = 0$. (See Proposition 5.9.)

8.5 Corollary. *If* G *is a hyperbolic open set, then the collection of irregular points for* G *is an* F_σ *set with harmonic measure zero.* (See Corollary 6.7.)

Now to produce a few examples. We will start with a sufficient condition for a compact set to have capacity zero. Let K be a compact set with diameter less than or equal to 1. This assumption implies that $\log|z - w|^{-1} \geq 0$ for all z, w in K. For $r > 0$ let $N(r)$ be the smallest number of open disks of radius r that cover K. Note that if K has k elements, $N(r) \leq k$ for all r. So the idea is that if $N(r)$ does not grow too fast, K is a small set.

8.6 Lemma. *With* K *and* $N(r)$ *as above,* $\int_0^1 [r\, N(r)]^{-1} dr < \infty$ *if and only if* $\sum_1^\infty N(2^{-n})^{-1} < \infty$.

Proof. $N(r)$ is increasing so for $1/2^{n+1} \leq r \leq 1/2^n$, $[r\, N(2^{-n-1})]^{-1} \leq [r\, N(r)]^{-1} \leq [r\, N(2^{-n})]^{-1}$. Hence

$$\frac{\log 2}{N(2^{-n-1})} \leq \int_{2^{-n-1}}^{2^{-n}} \frac{1}{r\, N(r)}\, dr \leq \frac{\log 2}{N(2^{-n})}.$$

□

8.7 Proposition. *With* K *and* $N(r)$ *as above,* $c(K) = 0$ *if*

$$\int_0^1 \frac{1}{r\, N(r)}\, dr = \infty.$$

Proof. Assume that $c(K) > 0$ and let μ be a probability measure on K with $I(\mu) < \infty$. For each z in K, let u_z be the increasing, right-continuous function on $[0, 1]$ defined by $u_z(r) = \mu(\overline{B}(z; r))$. Using the change of variables formula,

$$I(\mu) = \int \left[\int_0^1 \log r^{-1} du_z(r)\right] d\mu(z).$$

Using integration by parts we get

$$\int_0^1 \log r^{-1} du_z(r) = u_z(r) \log r^{-1}|_0^1 + \int_0^1 r^{-1} u_z(r)\, dr.$$

Now for $0 < r < 1$,

$$\begin{aligned} u_z(r)\log r^{-1} &= \int_0^r \log t^{-1} du_z(t) \\ &\leq \int_{\overline{B}(z:r)} \log|z-w|^{-1} d\mu(w) \end{aligned}$$

and this converges to 0 as $r \to 0$. Therefore

$$I(\mu) = \int \left[\int_0^1 r^{-1} u_z(r)\, dr \right] d\mu(z).$$

Since u_z is increasing,

$$\begin{aligned} I(\mu) &\geq \int \left[\sum_{n=0}^{\infty} \int_{2^{-n-1}}^{2^{-n}} \frac{1}{r} u_z(2^{-n-1}) dr \right] d\mu(z) \\ &= \log 2 \sum_{n=0}^{\infty} \int \mu(\overline{B}(z;2^{-n-1}))\, d\mu(z). \end{aligned}$$

Put $N_n = N(2^{-n})$ and let $B_k^{(n)} = B(z_k^{(n)}, 2^{-n})$, $1 \leq k \leq N_n$, be disks that cover K. Now any disk of radius r can be covered by 16 disks of radius $r/2$. Hence for a fixed value of n, no point of K can belong to more than 16 of the disks $B_k^{(n+1)}$. Indeed, if $z \in B_k^{(n+1)}$, then $B_k^{(n+1)} \subseteq B(z;2^{-n})$. So if z belonged to more than 16 of the disks $B_k^{(n+1)}$, we could replace each of these by the 16 disks of radius 2^{-n-1} that cover $B(z;2^{-n})$ and reduce the size of N_{n+1}, contradicting its definition. Thus

$$\int \mu(\overline{B}(z;2^{-n-1}))\, d\mu(z) \geq \frac{1}{16} \sum_{k=1}^{N_{n+2}} \int_{B_k^{(n+2)}} \mu(\overline{B}(z;2^{-n-1}))\, d\mu(z).$$

Now $B_k^{(n+2)} \subseteq \overline{B}(z;2^{-n-1})$ whenever $z \in B_k^{(n+2)}$. Therefore

$$\begin{aligned} I(\mu) &\geq \frac{\log 2}{16} \sum_{n=0}^{\infty} \sum_{k=1}^{N_{n+2}} \mu(B_k^{(n+2)})^2 \\ &= \frac{\log 2}{16} \sum_{n=1}^{\infty} \sum_{k=1}^{N_{n+2}} \mu(B_k^{(n+1)})^2. \end{aligned}$$

Using the Cauchy-Schwarz Inequality,

$$\begin{aligned} 1 &= \mu(K)^2 \\ &\leq \left[\sum_{k=1}^{N_{n+1}} \mu(B_k^{(n+1)}) \right]^2 \\ &\leq N_{n+1} \sum_{k=1}^{N_{n+1}} \mu(B_k^{(n+1)})^2. \end{aligned}$$

Therefore

$$I(\mu) \geq \frac{\log 2}{16} \sum_{n=1}^{\infty} \frac{1}{N_{n+1}}$$

and the proposition is proved. □

Now form a Cantor set contained in $[0,1]$ as follows. Let $\{b_n\}$ be a strictly decreasing sequence of positive numbers less than 1 such that $b_1 + 2b_2 + 2^2 b_3 + \cdots \leq 1$. Delete from $[0,1]$ the open subinterval of length b_1 centered in $[0,1]$. Let K_1 be the union of the two closed intervals that remain; each has the same length, a_1. So $1 = 2a_1 + b_1$. Now from each of the two component intervals that make up K_1, delete the open middle interval of length b_2. Let K_2 be the union of the 2^2 closed intervals that remain; each of these closed interval has length a_2. So $a_1 = 2a_2 + b_2$. Continue in this way to get a decreasing sequence of compact sets $\{K_n\}$ satisfying the following for all $n \geq 1$:

8.8
(i) K_n has 2^n components, each of which has length a_n;
(ii) $a_{n-1} = 2a_n + b_n$.

It is left to the reader to prove the next proposition, which is standard measure theory.

8.9 Proposition. *If $\{K_n\}$ satisfies (8.8), then $K = \bigcap_n K_n$ is a totally disconnected set having Lebesgue measure $\lim_n 2^n a_n = 1 - (b_1 + 2b_2 + 2^2 b_3 + \cdots)$.*

8.10 Theorem. *If $\{K_n\}$ is the sequence of compact sets satisfying (8.8) and $K = \bigcap_n K_n$, then $c(K) > 0$ if and only if*

8.11
$$\sum_{n=1}^{\infty} \frac{1}{2^n} \log a_n^{-1} < \infty.$$

If $c(K) > 0$, then

8.12
$$c(K) \geq \exp\left[-2 \sum_{n=1}^{\infty} \frac{1}{2^n} \log a_n^{-1}\right].$$

Proof. Assume (8.11) holds. For each $n \geq 1$ let μ_n be the probability measure $\mu_n = (2^n a_n)^{-1} m | K_n$, where m is Lebesgue measure on the line. The strategy here will be to show that there is a constant M such that $I(\mu_n) \leq M$ for all n. This implies that $v(K_n) \leq M$ so that $c(K_n) \geq \exp(-M)$. The result will then follow from Proposition 7.15 and the estimate (8.12) will be obtained by giving the appropriate value of M.

Fix $n \geq 1$ and a point x in K_n. For every $p \geq 0$, let $L_p = K_n \cap \{y : a_{p+1} \leq |y-x| \leq a_p\}$. (We take a_0 to be 1.) So

8.13 $$\int \log|y-x|^{-1} d\mu_n(y) = \sum_{p=0}^{\infty} \frac{1}{2^n\, a_n} \int_{L_p} \log|y-x|^{-1} dy.$$

Note that for any $p \geq 0$ and y in L_p, $\log|y-x|^{-1} \leq \log a_{p+1}^{-1}$. By drawing pictures it can be seen that there are at most $5 = 2 \cdot 2 + 1$ of the component intervals of K_n within a distance a_{n-1} of x. Continuing, there are at most $2 \cdot 2^j + 1$ of the component intervals of K_n within a distance a_{n-j} of x for $1 \leq j \leq n$. Thus for $0 \leq p \leq n-1$ and y in L_p,

$$\begin{aligned}\frac{1}{2^n a_n} \int_{L_p} \log|y-x|^{-1} dy &\leq \frac{1}{2^n a_n} (2 \cdot 2^{n-p} + 1)(\log a_{p+1}^{-1}) a_n \\ &\leq \frac{4}{2^p} \log a_{p+1}^{-1}. \end{aligned} \quad \textbf{8.14}$$

For $p \geq n$,

$$\begin{aligned}\frac{1}{2^n a_n} \int_{L_p} \log|y-x|^{-1} dy &\leq \frac{2}{2^n a_n} \log a_{p+1}^{-1} [a_p - a_{p+1}] \\ &= \frac{2}{2^n} \log a_{p+1}^{-1} \left[\frac{a_p - a_{p+1}}{a_n}\right].\end{aligned}$$

Now from (8.8.ii) $a_p \geq 2a_{p+1}$ for all p; so $a_n \geq 2^{p-n} a_p$ for all $p \geq n$. Hence $[a_p - a_{p+1}] \leq a_p \leq 2^{n-p} a_n$ for all $p \geq n$. This gives that

8.15 $$\frac{1}{2^n a_n} \int_{L_p} \log|y-x|^{-1} dy \leq \frac{2}{2^p} \log a_{p+1}^{-1}$$

for all $p \geq n$.

Combining (8.14) and (8.15) with (8.13), we get

$$\int \log|y-x|^{-1} d\mu_n(y) \leq \sum_{p=0}^{\infty} \frac{4}{2^p} \log a_{p+1}^{-1} \equiv M,$$

which is a bound independent of n. Since x was an arbitrary point of K_n, $I(\mu_n) \leq M$ for all n and this proves half the theorem. The estimate for the capacity (8.12) follows from the preceding inequality.

For the converse, assume $c(K) > 0$. We will use Proposition 8.7, so adopt the notation from there. Using the fact that $a_0 = 1$ and the telescoping of the second series below, we get that

$$\sum_{n=1}^{\infty} \frac{1}{2^{n+1}} \log a_n^{-1} = \sum_{n=0}^{\infty} \frac{1}{2^{n+1}} [\log a_{n+1}^{-1} - \log a_n^{-1}].$$

Observe that $N(a_{n+1}) \leq 2^{n+2}$. Hence

$$\begin{aligned}\sum_{n=1}^{\infty} \frac{1}{2^n} \log a_n^{-1} &\leq 2\sum_{n=0}^{\infty} \frac{1}{N(a_{n+1})} \int_{a_{n+1}}^{a_n} \frac{1}{r}\, dr \\ &\leq 2\sum_{n=0}^{\infty} \int_{a_{n+1}}^{a_n} \frac{1}{r\, N(r)}\, dr \\ &= 2\int_0^1 \frac{1}{r\, N(r)}\, dr\end{aligned}$$

and this is finite by Proposition 8.7. □

8.16 Example. If K is the usual Cantor ternary set, then K has Lebesgue measure zero and positive capacity. In fact, in this case $a_n = 3^{-n}$ and so

$$\sum_{n=1}^{\infty} \frac{1}{2^n} \log a_n^{-1} = \sum_{n=1}^{\infty} \frac{1}{2^n} \log 3^n = \log 3 \sum_{n=1}^{\infty} \frac{n}{2^n}.$$

In fact, by evaluating this last sum we get that $c(K) \geq 3^{-4}$.

8.17 Example. If K is the Cantor set as in (8.8) with $a_n = \exp(-2^n)$, then $c(K) = 0$ and so K is an uncountable set that is polar.

There is a strong connection between logarithmic capacity and Hausdorff measure. See Carleson [1967] and Tsuji [1975]. Also Landkof [1972] computes the logarithmic capacity as well as the Green capacity of several planar sets.

§9* Removable Singularities for Functions in the Bergman Space

In this section we will use logarithmic capacity to characterize the removable singularities for functions in the Bergman space.

9.1 Definition. If G is an open set and $a \in \partial_\infty G$, then a is a *removable singularity* for $L_a^2(G)$ if there is a neighborhood U of a such that each function f in $L_a^2(G)$ has an analytic continuation to $G \cup U$. Let $\text{rem}(G)$ denote the points in $\partial_\infty G$ that are removable singularities for $L_a^2(G)$.

There is, of course, a concept of removable singularity for $L_a^p(G)$. Some of the results below carry over in a straightforward manner to such points. This will not be done here as a key result (Theorem 9.5) for the case $p = 2$

*This section can be skipped if desired, as the remainder of the book does not depend on it.

is not true for arbitrary p. There will be more said about this after that proof.

Before giving some elementary properties of rem(G) and exhibiting some examples, let's prove a basic result that will be useful in these discussions.

9.2 Lemma. *If $H = \{z : |z| > R\}$ and $f(z) = \sum_{n=0}^{\infty} a_n z^{-n}$ is analytic on H, then $f \in L_a^2(H)$ if and only if $a_0 = f(\infty) = 0$, $a_1 = f'(\infty) = 0$, and $\sum_{n=2}^{\infty} |a_n|^2 R^{-(2n-2)} < \infty$.*

If f is a bounded analytic function on H, $f \in L_a^2(H)$ if and only if $f(\infty) = f'(\infty) = 0$.

Proof. A calculation shows that

$$\int_{|z|>R} |f|^2 d\mathcal{A} = 2\pi \sum_{n=0}^{\infty} |a_n|^2 \int_R^{\infty} \frac{1}{r^{2n-1}}\, dr.$$

The first statement is now immediate. For the second statement note that by increasing R we may assume that f is analytic in a neighborhood of cl H. If, in addition, $f(\infty) = f'(\infty) = 0$, then $f(z) = z^{-2}g(z)$, where g is a bounded analytic function on H. □

If f is analytic in a set H as in the preceding lemma, the condition that f and its derivative vanish at infinity is that $0 = \lim_{z\to\infty} f(z) = \lim_{z\to\infty} zf(z)$.

9.3 Proposition. *Fix an open set G and a point a in $\partial_\infty G$.*

(a) *If a is an isolated point of $\partial_\infty G$, then a is removable for $L_a^2(G)$.*

(b) *If $B(a;\delta) \cap \partial_\infty G$ has positive area for every $\delta > 0$, then a is not a removable singularity for $L_a^2(G)$.*

(c) *The set $G \cup$ rem(G) is an open subset of* int$[\mathrm{cl}_\infty G]$.

(d) Area[rem(G)] $= 0$.

Proof. (a) If ∞ is an isolated point of $\partial_\infty G$, then the result follows by Lemma 9.2. So assume that a is an isolated point of ∂G; without loss of generality we may assume that $a = 0$. Since 0 is isolated, f has a Laurent expansion, $f(z) = \sum_{n=-\infty}^{\infty} a_n z^n$. If $R > 0$ such that $z \in G$ when $0 < |z| \le R$, then, as in the proof of Lemma 9.2,

$$\infty > \int_{|z|<R} |f(z)|^2 d\mathcal{A}(z) = \sum_{n=-\infty}^{\infty} |a_n|^2 \int_0^R r^{2n+1} dr.$$

From here we see that $a_n = 0$ for $n < 0$, and so 0 is a removable singularity.

(b) Suppose $a \in$ rem(G); we will only treat the case that a is a finite point. Let $\delta > 0$ and assume each f in $L_a^2(G)$ has an analytic extension to

$G \cup B(a;\delta)$. If $\text{Area}(B(a;\delta) \cap \partial G) > 0$, there is a compact set K contained in $B(a;\delta) \cap \partial G$ with positive area. Write K as the disjoint union of two Borel sets E_1 and E_2 having equal area; let $\mu_j = \text{Area}|E_j$ and put $\mu = \mu_1 - \mu_2$. If f is the Cauchy transform of μ, then f is analytic off K and $f(\infty) = f'(\infty) = 0$ (see 18.5.3). So if $K \subseteq B(0;R)$, Lemma 9.2 implies $f \in L^2_a(\{z : |z| > R\})$; since f is bounded, $f \in L^2_a(G)$.

(c) From the definition of a removable singularity, $G \cup \text{rem}(G)$ is open. For the second part of (c) it suffices to show that $\text{rem}(G) \cap \partial[\text{cl}_\infty G] = \emptyset$. If $a \in \partial[\text{cl}_\infty G]$ and U is any neighborhood of a, let $b \in U \setminus [\text{cl}_\infty G]$; put $f(z) = (z-b)^{-2}$. According to Lemma 9.2, $f \in L^2_a(\{z : |z| > R\})$ for any $R > |b|$. It follows that $f \in L^2_a(G)$ and so $a \notin \text{rem}(G)$.

(d) If it were the case that $\text{rem}(G)$ had positive area, then we could find disjoint compact subsets K_1 and K_2 of $\text{rem}(G)$ with $\text{Area}(K_1) = \text{Area}(K_2) > 0$. Let $\mu = \mathcal{A}|K_1 - \mathcal{A}|K_2$. Let $f = \hat{\mu}$, the Cauchy transform of μ. By (18.5.2) f is a bounded analytic function on $\mathbb{C} \setminus (K_1 \cup K_2)$ and $0 = f(\infty) = f'(\infty)$. Thus $f \in L^2_a(G)$. But $K_1 \cup K_2 \subseteq \text{rem}(G)$ and so f can be extended to a bounded entire function; thus f is constant and so $f = 0$. But this implies that $\mathcal{A}|K_1 = \mathcal{A}|K_2$, a contradiction. □

So the typical case where $\text{rem}(G) \neq \emptyset$ occurs when $G = V \setminus K$, where V is an open set and K is a compact subset of V with zero area. If we want to get K contained in $\text{rem}(G)$, we must have that K is totally disconnected (Exercise 1). But more is required, as we will see in Theorem 9.5 below. First we need an elementary result about analytic functions.

9.4 Lemma. *If V is an open subset of $\mathbb{C}$, K is a compact subset of V, and $f : V \setminus K \to \mathbb{C}$ is an analytic function, then there are unique analytic functions f_0: $V \to \mathbb{C}$ and $f_\infty : \mathbb{C}_\infty \setminus K \to \mathbb{C}$ such that $f_\infty(\infty) = 0$ and $f(z) = f_0(z) + f_\infty(z)$ for z in $V \setminus K$.*

Proof. If $z \in V$, let Γ_0 be a smooth Jordan system in $V \setminus K$ such that $K \cup \{z\}$ is included in the inside of Γ_0. Let

$$f_0(z) = \frac{1}{2\pi i} \int_{\Gamma_0} \frac{f(w)}{w-z}\, dw.$$

Cauchy's Theorem implies that the definition of $f_0(z)$ is independent of the Jordan system Γ_0. Also, it is easy to see that f_0 is an analytic function on V.

Similarly, if $z \in \mathbb{C} \setminus K$, let Γ_∞ be a smooth Jordan system in $V \setminus K$ that contains K in its inside and has the point z in its outside. Let

$$f_\infty(z) = -\frac{1}{2\pi i} \int_{\Gamma_\infty} \frac{f(w)}{w-z}\, dw.$$

Once again the definition of $f_\infty(z)$ is independent of the choice of Jordan system Γ_∞ and $f_\infty : \mathbb{C} \setminus K \to \mathbb{C}$ is an analytic function. It is also easy to see that $f_\infty(z) \to 0$ as $z \to \infty$, so that ∞ is a removable singularity. If $z \in V \setminus K$, then Γ_0 and Γ_∞ can be chosen above so that $\Gamma \equiv \Gamma_0 - \Gamma_\infty$ is also a Jordan system with winding number about the point z equal to 0. Thus Cauchy's Integral Formula implies that $f(z) = f_0(z) + f_\infty(z)$.

To see that f_0 and f_∞ are unique, suppose g_0 and g_∞ are another pair of such functions. So $f_0(z) + f_\infty(z) = g_0(z) + g_\infty(z)$ on $V \setminus K$, so that $f_0(z) - g_0(z) = g_\infty(z) - f_\infty(z)$ there. This says that if h is defined on $\mathbb{C}$ by $h(z) = f_0(z) - g_0(z)$ for z in V and $h(z) = g_\infty(z) - f_\infty(z)$ for z in $\mathbb{C} \setminus K$, h is a well-defined entire function. Since $h(\infty) = 0$, $h \equiv 0$. $\square$

Suppose V is an open set, K is a compact subset of V, and $G = V \setminus K$. As was pointed out, if $K \subseteq \text{rem}(G)$, Area$(K) = 0$. Thus when $K \subseteq \text{rem}(G)$, the restriction map $f \to f|G$ is an isometric isomorphism of $L^2_a(V)$ onto $L^2_a(G)$. Equivalently, if $f \in L^2_a(V)$ and $f = f_0 + f_\infty$ as in the preceding lemma, then it must be that $f_\infty = 0$.

We are now in a position to state and prove the main result of this section.

9.5 Theorem. *If K is a compact subset of $\mathbb{C}$, then the following are equivalent.*

(a) *K is a polar set.*

(b) $L^2_a(\mathbb{C} \setminus K) = (0)$.

(c) *If V is any open set containing K, $K \subseteq$ rem$(V \setminus K)$.*

Proof. First we do the easy part of the proof and show that (b) and (c) are equivalent. If (c) is true and $f \in L^2_a(\mathbb{C} \setminus K)$, then taking $V = \mathbb{C}$ in (c) shows that f has a continuation to an entire function. But $f(\infty) = 0$ and so $f \equiv 0$.

Now assume that (b) holds. Note that (b) says that $K \subseteq \text{rem}(\mathbb{C} \setminus K)$. (Actually, this is an equivalent formulation of (b).) By (9.3.d) Area $(K) = 0$. Let V be any open set containing K. Fix a function f in $L^2_a(V \setminus K)$ and write $f = f_0 + f_\infty$ as in the preceding lemma. To prove (c) it must be shown that $f_\infty = 0$. This will be done by showing that $f_\infty \in L^2_a(\mathbb{C} \setminus K)$.

Observe that if $a \in \mathbb{C} \setminus K$, then

$$\lim_{z \to \infty} z \, \frac{f_\infty(z) - f_\infty(a)}{z - a} = -f_\infty(a).$$

Assume $f_\infty \neq 0$ and choose points a and b in $\mathbb{C} \setminus K$ such that $f_\infty(a) \neq 0 \neq f_\infty(b)$. Put

9.6 $$g(z) = \frac{1}{f_\infty(a)} \left[\frac{f_\infty(z) - f_\infty(a)}{z - a} \right] - \frac{1}{f_\infty(b)} \left[\frac{f_\infty(z) - f_\infty(b)}{z - b} \right].$$

Clearly g is analytic in $\mathbb{C} \setminus K$ and from the prior observation $0 = g(\infty) = g'(\infty)$. Thus if $R > \max\{|z| : z \in K\}$ and $H = \{z : |z| > R\}$, $g \in L^2_a(H)$.

On the other hand, if W is an open set with $K \subseteq W \subseteq \text{cl } W \subseteq V$, $f_\infty = f - f_0 \in L^\infty_a(W \setminus K)$. If W is further restricted so that a, $b \notin \text{cl } W$, then $g \in L^2_a(W \setminus K)$. Therefore $g \in L^2_a(\mathbb{C} \setminus K)$. By (b), $g \equiv 0$. This allows us to use (9.6) to solve for $f_\infty(z)$. Doing this we see that f_∞ is a rational function with precisely one pole; denote this pole by c. If $c \notin V$, then f has an analytic continuation to V. If $c \in V$, then f is analytic on $V \setminus \{c\}$. Now $f \in L^2_a(V \setminus K)$, so $c \in K$.

Since Area$(K) = 0$, $f \in L^2_a(V \setminus \{c\})$. By Proposition 9.3.a, c is a removable singularity for f. Thus f has a continuation to V.

(b) *implies* (a). Without loss of generality we can assume that Area$(K) = 0$ and diam $K < 1$. Assume K is not polar; by Theorem 7.5, $c(K) > 0$. We will exhibit a non-zero function that belongs to $L^2_a(\mathbb{C} \setminus K)$. Let K_1 and K_2 be disjoint compact subsets of K, each of which has positive capacity. For $j = 1, 2$ let μ_j be a probability measure on K_j with logarithmic potential that is bounded above and put $\mu = \mu_1 - \mu_2$. So μ is a non-zero measure carried by K and $\mu(K) = 0$. If $f(z) = \hat{\mu}(z)$, the Cauchy transform of μ, f is analytic in $\mathbb{C}_\infty \setminus K$ with $0 = f(\infty) = f'(\infty)$ (18.5.2). Choose $R > 1$ such that $K \subseteq B(0; R)$ and put $H = \{z : |z| > R\}$. By Lemma 9.2, $f \in L^2_a(H)$.

To show that f is square integrable over $D = B(0; R)$, we first find an estimate. Let $z, w \in K$ with $z \neq w$ and for $0 < \varepsilon < |z - w|/2$ put $B_\varepsilon = B(w; \varepsilon) \cup B(z; \varepsilon)$. Note two things. Because $z \neq w$, the function $\zeta \to [(\zeta - w)(\overline{\zeta} - \overline{z})]^{-1}$ is locally integrable with respect to area measure. Thus

$$\int_D \frac{d\mathcal{A}(\zeta)}{(\zeta - w)(\overline{\zeta} - \overline{z})} = \lim_{\varepsilon \to 0} \int_{D \setminus B_\varepsilon} \frac{d\mathcal{A}(\zeta)}{(\zeta - w)(\overline{\zeta} - \overline{z})}.$$

Also note that $(\overline{\zeta} - \overline{z})^{-1} = 2\overline{\partial}_\zeta[\log|\zeta - z|]$. Thus Green's Theorem implies that

$$\begin{aligned}
\int_{D \setminus B_\varepsilon} \frac{d\mathcal{A}(\zeta)}{(\zeta - w)(\overline{\zeta} - \overline{z})} &= 2\int_{D \setminus B_\varepsilon} \overline{\partial}\left[\frac{\log|\zeta - z|}{\zeta - w}\right] d\mathcal{A}(\zeta) \\
&= \frac{1}{i}\int_{|\zeta| = R} \frac{\log|\zeta - z|}{\zeta - w}\, d\zeta \\
&\quad - \frac{1}{i}\int_{|\zeta - w| = \varepsilon} \frac{\log|\zeta - z|}{\zeta - w}\, d\zeta \\
&\quad - \frac{1}{i}\int_{|\zeta - z| = \varepsilon} \frac{\log|\zeta - z|}{\zeta - w}\, d\zeta \\
&= I - I_\varepsilon - J_\varepsilon.
\end{aligned}$$

Now

$$\begin{aligned}
I_\varepsilon &= \frac{1}{i}\int_0^{2\pi} \log|w - z + \varepsilon e^{i\theta}|\; i\,\varepsilon\, e^{i\theta}\, d\theta \\
&= 2\pi\, \log|z - w|.
\end{aligned}$$

Also

$$J_\varepsilon = \frac{1}{i}\int_0^{2\pi} \frac{\log|\varepsilon\, e^{i\theta}|}{(z - w + \varepsilon\, e^{i\theta})}\, i\,\varepsilon\, e^{i\theta}\, d\theta.$$

So

$$|J_\varepsilon| \le 2\pi\, \frac{\varepsilon\ \log \varepsilon^{-1}}{|z-w| - \varepsilon}.$$

Finally

$$I = \int_0^{2\pi} \frac{\log|Re^{i\theta} - z|}{Re^{i\theta} - w}\, R\, e^{i\theta}\, d\theta.$$

Put $d = \sup\{|z| : z \in K\}$. So $d < R$ and $|R\, e^{i\theta} - w| \ge R - d$. Also $\log|R\, e^{i\theta} - z| \le \log 2R$. Thus there is a constant C_1 depending only on R and K such that $|I| \le C_1 \log 2R$. Therefore

$$\begin{aligned} \left| \int_{D\setminus B_\varepsilon} \frac{dA(\zeta)}{(\zeta - w)(\overline{\zeta} - \overline{z})} \right| &\le |I| + |I_\varepsilon| - |J_\varepsilon| \\ &\le C_1\ \log 2R + 2\pi\ \log|z-w|^{-1} \\ &\qquad + 2\pi\, \frac{\varepsilon\ \log\varepsilon^{-1}}{|z-w| - \varepsilon}. \end{aligned}$$

Letting $\varepsilon \to 0$ we see that there is a constant C that depends only on R and K such that

9.7 $$\left| \int_D \frac{dA(\zeta)}{(\zeta - w)(\overline{\zeta} - \overline{z})} \right| \le C\ \log \frac{2R}{|z-w|}$$

whenever $w \ne z$.

Now the fact that L_{μ_1} and L_{μ_2} are bounded above implies that the measure μ can have no atoms. Hence $|\mu \times \mu|(\{(z,z) : z \in K\}) = 0$ by Fubini's Theorem. So (9.7) holds a.e. $[|\mu \times \mu|]$ on $K \times K$ and

$$\begin{aligned} &\int_{K\times K} \left| \frac{dA(\zeta)}{(\zeta - w)(\overline{\zeta}\overline{z})} \right|\ d|\mu \times \mu|(z,w) \\ &\le C \int_K \int_K \log\frac{2R}{|z-w|}\ d|\mu|(z)\ d|\mu|(w) \\ &= 4C\ \log 2R + I(|\mu|) \\ &< \infty. \end{aligned}$$

From Fubini's Theorem we get that

$$\begin{aligned} \int_D |f|^2\, dA &= \int_D \left[\int_K \frac{d\mu(w)}{\zeta - w}\right] \left[\int_K \frac{d\mu(z)}{\overline{\zeta} - \overline{z}}\right]\ dA(\zeta) \\ &\le \int_{K\times K} \left| \int_D \frac{dA(\zeta)}{(\zeta - w)(\overline{\zeta} - \overline{z})} \right|\ d|\mu \times \mu|(z,w) \\ &< \infty. \end{aligned}$$

Thus $f \in L^2_a((\mathbb{C} \setminus K)$.

(a) *implies* (b). Now assume that K is polar; so K is totally disconnected and $\mathbb{C} \setminus K$ is connected. If $f \in L^2_a(\mathbb{C} \setminus K)$, then f has an expansion $f(z) = \sum_{n=2}^{\infty} a_n\, z^{-n}$ (9.2). To show that each such f is the zero function, we need only show that $a_n = 0$ for all $n \geq 2$. Note that this is equivalent to showing that for all $n \geq 1$, $\int_\Gamma z^n\, f(z)\, dz = 0$, where Γ is any smooth Jordan system surrounding K.

Without loss of generality it can be assumed that $K \subseteq \mathbb{D}$. Let Γ be a finite collection of pairwise disjoint smooth positively oriented Jordan curves in $\mathbb{D} \setminus K$ such that $K \subseteq \text{ins}\ \Gamma$; put $\gamma = \text{rob}(\Gamma)$, the Robin constant of the point set Γ.

Put $W = \mathbb{C}_\infty \setminus [\Gamma \cup \text{ins}\ \Gamma] = \text{out}\ \Gamma$, let ω be harmonic measure for W evaluated at ∞, and let g be the Green function for W. If $u = \gamma^{-1} L_\omega$, then u is harmonic on W and, by (3.5), $u(z) = 1 - \gamma^{-1} g(z, \infty)$ for all z in W, and $u(z) = 1$ for all z not in W (because $\partial W = \Gamma$ is a system of curves and so each point of Γ is a regular point).

For $\varepsilon > 0$, set $\Gamma_\varepsilon = \{z : u(z) = 1 - \varepsilon\}$ and $E_\varepsilon = \Gamma_\varepsilon \cup \text{ins}\ \Gamma_\varepsilon$. The value of ε can be chosen as small as desired with Γ_ε a smooth positive Jordan system that contains Γ in its inside. If $U = \{z : u(z) > 0\} \setminus E_\varepsilon$, U is a bounded open set that contains E_ε and has a smooth boundary. Note that if $|z| > 2$, $\text{dist}(z, \Gamma) > 1$ and so $|z - w|^{-1} < 1$ for all w in Γ; thus $u(z) < 0$. Hence $U \subseteq 2\,\mathbb{D}$.

If g is any function analytic in a neighborhood of $\text{cl}[\{z : u(z) > 0\} \setminus \text{cl}[\text{ins}\ \Gamma]]$, then

$$\int_\Gamma g(z)\, dz = \int_{\Gamma_\varepsilon} g(z)\, dz = \frac{1}{1-\varepsilon} \int_{\partial U} g(z)\, u(z)\, dz.$$

Thus applying Green's Theorem we get

9.8 $$\int_\Gamma g(z)\, dz = \frac{2}{1-\varepsilon} \int_U \bar{\partial}[g\, u]\, d\mathcal{A} = \frac{2}{1-\varepsilon} \int_U g\, \bar{\partial} u\, d\mathcal{A}.$$

Taking $g = \partial u$ in (9.8) and using the fact that u is real-valued, we get

$$\int_{\Gamma_\varepsilon} \partial u(z)\, dz = \frac{2}{1-\varepsilon} \int_U |\partial\, u|^2\, d\mathcal{A}.$$

Now fix $n \geq 1$, let $f \in L^2_a(\mathbb{C} \setminus K)$, and let $g = z^n f$ in (9.8); this yields

$$\begin{aligned} \left| \int_\Gamma z^n f(z)\, dz \right|^2 &= \frac{4}{(1-\varepsilon)^2} \left| \int_U z^n f \bar{\partial} u\, d\mathcal{A} \right|^2 \\ &\leq \frac{4}{(1-\varepsilon)^2} \left[\int_U |z^n\, f|^2 d\mathcal{A} \right] \left[\int_U |\partial u|^2 d\mathcal{A} \right] \\ &\leq \frac{2^{n+1}}{1-\varepsilon}\, ||f||^2 \int_{\Gamma_\varepsilon} \partial u(z)\, dz. \end{aligned}$$

According to (19.5.2), $\bar\partial u = -(2\gamma)^{-1}\hat\omega$, the Cauchy transform of ω. Thus

$$\begin{aligned}\int_{\Gamma_\varepsilon} \bar\partial u(z)\, dz &= -\frac{1}{2\gamma}\int_{\Gamma_\varepsilon}\int_{\Gamma}\left[\frac{1}{\zeta - z}\, d\omega(\zeta)\right] dz \\ &= \frac{1}{2\gamma}\int_{\Gamma}\int_{\Gamma_\varepsilon}\left[\frac{1}{z-\zeta}\, dz\right] d\omega(\zeta) \\ &= \frac{1}{2\gamma},\end{aligned}$$

since Γ is in the inside of Γ_ε. This combined with the fact that ε was arbitrary gives that

$$\left|\int_\Gamma z^n f(z)\, dz\right|^2 \leq \frac{2^n\, ||f||^2}{\operatorname{rob}(\Gamma)}$$

for any such system Γ. Since $c(K) = 0$, we can get a sequence of such curve systems $\{\Gamma_k\}$ that squeeze down to K. Thus $\operatorname{rob}(\Gamma_k) \to \infty$; but $\int_{\Gamma_k} z^n f$ remains constant. Thus this integral must be zero and so $f = 0$. □

The first reference for the preceding result that the author is aware of is Carleson [1967], page 73. The proof above is based on Hedberg [1972a]. In this paper and its cousin, Hedberg [1972b], various capacities are introduced that are related to logarithmic capacity and Green capacity, and connections are made with removable singularities of certain spaces of analytic and harmonic functions. In particular, a q-capacity is defined, $1 < q \leq 2$, and a compact set K has q-capacity 0 if and only if $L^p_a(\mathbb{C}\setminus K) = (0)$, where p and q are conjugate exponents. When $2 < q < \infty$ $(1 < p < 2)$, the story is simpler. See Exercise 4. The reader can consult these references for details.

9.9 Theorem. *If G is any open set and $a \in \partial_\infty G$, then $a \in \operatorname{rem}(G)$ if and only if there is a neighborhood U of a such that $c(\operatorname{cl} U \setminus G) = 0$.*

Proof. Suppose U is a neighborhood of a and $c(\operatorname{cl} U \setminus G) = 0$. We assume that a is a finite point, the case that $a = \infty$ being obtained from the finite case by applying a suitable Möbius transformation. Thus $\operatorname{cl} U \setminus G$ is totally disconnected and there is another neighborhood V of a such that $V \cap (\operatorname{cl} U \setminus G)$ is compact. Put $\Omega = G \cup V$ and $K = V \cap (\operatorname{cl} U \setminus G)$. So $K \subseteq \Omega$ and $\Omega \setminus K = G$. By Theorem 9.5 each function in $L^2_a(G)$ has an analytic continuation to Ω and so $a \in \operatorname{rem}(G)$.

If $c(\operatorname{cl} U \setminus G) > 0$ for every neighborhood U of a, then for every such U there is a non-zero function f in $L^2_a(\mathbb{C}\setminus(\operatorname{cl} U \setminus G))$. Thus $f|G \in L^2_a(G)$ and f cannot be extended to U. □

Further results on spaces $L^p_a(G\setminus K)$ can be found in Axler, Conway, and McDonald [1982] and Aleman, Richter, and Ross [preprint].

Exercises

1. If $a \in \partial_\infty G$ and the component of $\partial_\infty G$ that contains a is not trivial, then $a \notin \operatorname{rem}(G)$.

2. If $a \in \partial_\infty G$ and for each f in $L^2_a(G)$ there is a neighborhood U of a such that f has an analytic continuation to $G \cup U$, then $a \in \operatorname{rem}(G)$.

3. In Lemma 9.4, show that if the function f is bounded, then so are f_0 and f_∞.

4. If K is a compact subset of the bounded open set G, $1 \leq p < 2$, and $G \setminus K) = L^p_a(G)$, then $K = \emptyset$.

§10 Logarithmic Capacity: Part 2

We resume the study begun in §7.

10.1 Lemma. *If K is a compact set with positive capacity and μ is an equilibrium measure, then $L_\mu(z) = v(K)$ for all z in* int K.

Proof. Let $a \in \operatorname{int} K$ and let $B = B(a; r)$ such that cl $B \subseteq \operatorname{int} K$. Since the logarithmic potential is superharmonic, $(\pi r^2)^{-1} \int_B L_\mu dA \leq L_\mu(a) \leq v$. But $L_\mu = v = v(K)$ q.e. in K and so $L_\mu = v$ a.e. [Area] (7.12). Thus this integral equals v and we get $L_\mu(a) = v$. □

10.2 Theorem. *If K is a compact set with positive capacity, then:*

(a) *$v(K)$ is the Robin constant γ for K;*

(b) *The equilibrium measure is unique. In fact, if G is the component of $\mathbb{C}_\infty \setminus K$ that contains ∞, then the equilibrium measure for K is harmonic measure for G at ∞.*

(c) *The Robin constant of K is also given by the formulas*

$$\begin{aligned} \operatorname{rob}(K)^{-1} &= \sup\{\mu(K) : \mu \in M_+(K) : L_\mu \leq 1 \text{ on } K\} \\ &= \inf\{\mu(K) : \mu \in M_+(K) : L_\mu \geq 1 \text{ q.e. on } K\} \end{aligned}$$

Moreover if $c(K) > 0$, both the supremum and the infimum are attained for the measure $\gamma^{-1}\omega$, where ω is harmonic measure for G at ∞. The measure $\gamma^{-1}\omega$ is the only measure at which the supremum is attained.

Proof. Adopt the notation in the statement of the theorem and put $v = v(K)$. Let ω be harmonic measure for G at ∞, let μ be an equilibrium measure for K, and let g be the Green function for G. By Proposition 3.5. $L_\omega(z) = \gamma - g(z, \infty)$ for all z in G and $L_\omega \leq \gamma$ on $\mathbb{C}$.

Now $h(z) = v - L_\mu(z)$ is a positive harmonic function on G except for the point ∞. From (19.5.3), $L_\mu(z) + \log|z| \to 0$ as $z \to \infty$. Hence $h(z) - \log|z| \to v$ as $z \to \infty$.

By the definition of the Green function, $\gamma - L_\omega(z) = g(z,\infty) \le h(z) = v - L_\mu(z)$. Letting $z \to \infty$, we get that $\gamma \le v$. Since we already know that $v \le \gamma$ (7.5), this proves (a).

Note that $I(\omega) = \int L_\omega d\omega \le \gamma = v$. By the definition of $v(K)$, $I(\omega) = v$ and ω is an equilibrium measure. Put $\nu = \omega - \mu$. From the preceding paragraph we have that $L_\nu \ge 0$. We also have that $L_\nu \le \gamma - L_\mu$ and is therefore bounded above on compact sets. But $\nu(K) = 0$ implies that $L_\nu(z) \to 0$ as $z \to \infty$; therefore L_ν is a bounded harmonic function on $\mathbb{C}_\infty \setminus K$. Now Theorem 7.12 implies there is an F_σ set E with $c(E) = 0$ such that $L_\omega(z) = L_\mu(z) = \gamma$ for z in $K \setminus E$. By Corollary 7.13, $L_\nu(z) = L_\omega(z) - L_\mu(z) \to 0$ as z approaches any point of $\partial[\mathbb{C} \setminus K] \setminus E$. By the Maximum Principle, $L_\nu = 0$ in $\mathbb{C} \setminus K$. But we also have that $L_\nu(z) = 0$ on int K from the preceding lemma. Hence L_ν vanishes q.e. on $\mathbb{C}$ (off the set E) and thus $L_\nu(z) = 0$ a.e. [Area]. By Theorem 19.5.3, $\omega - \mu = \nu = 0$.

To prove (c), let $\alpha = \sup\{\mu(K) : \mu \in M_+(K) : L_\mu \le 1 \text{ on } K\}$. Taking $\mu = \gamma^{-1}\omega$ in this supremum shows that $\alpha \ge \gamma^{-1}$. On the other hand, if $\mu \in M_+(K)$ with $L_\mu \le 1$ on K, then $\mu_1 = \mu(K)^{-1}\mu \in M_1(K)$ and $I(\mu_1) \le \mu(K)^{-1}$. Hence $\gamma \le \inf\{\mu(K)^{-1} : \mu \in M_+(K) \text{ with } L_\mu \le 1 \text{ on } K\} = \alpha^{-1}$ so that $\alpha \le \gamma^{-1}$. This shows that $\alpha = \gamma^{-1}$ and the supremum is attained for the measure $\gamma^{-1}\omega$. If μ is any positive measure supported by K such that $L_\mu \le 1$ and $\mu(K) = \gamma^{-1}$, put $\mu_1 = \gamma\mu$. So μ_1 is a probability measure and $I(\mu_1) \le \gamma$. Hence $\mu_1 = \gamma\omega$ by part (b).

Now let β denote the infimum in part (c). If μ is any measure in $M_+(K)$ with $L_\mu \ge 1$ q.e. on K, then $L_\mu \ge 1$ a.e. $[\omega]$. Therefore $\gamma^{-1} \le \gamma^{-1} \int L_\mu d\omega = \gamma^{-1} \int L_\omega d\mu \le \mu(K)$. Thus $\gamma^{-1} \le \beta$. On the other hand, $L_\mu \ge 1$ q.e. on K for $\mu = \gamma^{-1}\omega$ so that $\gamma^{-1} = \beta$. □

So you noticed there is no uniqueness statement for the infimum expression for $\text{rob}(K)^{-1}$ in part (c) in the preceding theorem. This is because there is no uniqueness statement! Consider the following example. Put $K = \{z : |z| \le 1/2\}$. So $\text{rob}(K) = \log 2$ (3.3). If δ_0 is the unit point mass at $z = 0$ and m is normalized arc length measure on ∂K, then it is left to the reader to check that for both $\mu = (\log 2)^{-1}\delta_0$ and $\mu = (\log 2)^{-1}m$, $L_\mu \ge 1$ q.e. on K. Where does the proof of uniqueness for the expression for $\text{rob}(K)^{-1}$ as a supremum break down when applied to the expression as an infimum?

10.3 Corollary. *If K is a compact set with positive capacity, G is the component of $\mathbb{C}_\infty \setminus K$ that contains ∞, and ω is harmonic measure for G at ∞, then $c(\partial\hat{K} \setminus \text{supp}\,\omega) = 0$ and $c(K) = c(\hat{K}) = c(\partial\hat{K})$.*

Proof. Let $F = \partial\hat{K} \setminus \text{supp}\,\omega$. For each z in F, L_ω is harmonic in a neighborhood of z. Since $L_\omega \le \gamma$, the Maximum Principle implies $L_\mu(z) <$

γ for all z in F. But Theorem 7.12 implies $L_\mu(z) = \gamma$ q.e. on K, which includes $\partial\hat{K}$. Hence $c(F) = 0$. The rest of this corollary follows from the definition of the Robin constant. □

The next three corollaries follow from computations of the Robin constant in §3.

10.4 Corollary. *If K is a closed disk of radius R, $c(K) = R$.*

10.5 Corollary. *If K is a closed line segment of length L, $c(K) = L/4$.*

10.6 Corollary. *Let K be a compact connected subset of $\mathbb{C}$ and let G be the component of $\mathbb{C}_\infty \setminus K$ that contains ∞. If $\tau : G \to \mathbb{D}$ is the Riemann map with $\tau(\infty) = 0$ and $\rho = \tau'(\infty) > 0$, then $c(K) = \rho$.*

If we combine this last corollary with the proof of the Riemann Mapping Theorem, we have that for a compact connected set K

$$c(K) = \sup\{|f'(\infty)| : f \text{ is analytic on } \mathbb{C}\setminus\hat{K}, f(\infty) = 0, \text{ and } |f| \le 1\}.$$

Thus in the case of compact connected sets, $c(K)$ is the same as analytic capacity (Conway [1991], p 217).

Once again we record pertinent facts about the r-logarithmic capacity. The proof is left as an exercise for the reader.

10.7 Theorem. *If K is a compact subset of $r\mathbb{D}$, then*

$$\begin{aligned} c_r(K) &= \sup\{\mu(K) : \mu \in M_+(K) : L_\mu^r \le 1 \textit{ on } K\} \\ &= \inf\{\mu(K) : \mu \in M_+(K) : L_\mu^r \ge 1 \textit{ q.e. on } K\}. \end{aligned}$$

If $c_r(K) > 0$, both the supremum and the infimum are attained for the measure $c_r(K)^{-1}\omega$, where ω is harmonic measure for $G = \mathbb{C}_\infty \setminus \hat{K}$ at ∞. The measure $c_r(K)^{-1}\omega$ is the only measure at which the supremum is attained.

10.8 Proposition. *Let K be a compact set with positive capacity and let G be the component of $\mathbb{C}_\infty \setminus K$ that contains ∞. If ω is harmonic measure for G at ∞ and $a \in \partial G$, then a is a regular point for G if and only if $L_\omega(a) = v(K)$.*

Proof. If g is the Green function for G, then a is a regular point if and only if $g(z,\infty) \to 0$ as $z \to a$ (5.2). On the other hand, $L_\omega(z) = \gamma - g(z,\infty)$, where $\gamma = v(K)$. So if $L_\omega(a) = \gamma$, then the fact that L_ω is lsc implies that $\gamma = L_\omega(a) \le \liminf_{z\to a} L_\omega(z) \le \limsup_{z\to a} L_\omega(z) \le \gamma$ (3.5). Thus $L_\omega(z) \to \gamma$ as $z \to a$ and so $g(z,\infty) \to 0$ as $z \to a$, showing that a is regular.

Now assume that a is regular. Combining Theorem 10.2 with (7.12) we have that $L_\omega = \gamma$ q.e. on $\mathbb{C}\setminus G$; so, in particular, this equality holds a.e.

[Area]. If $\varepsilon > 0$, the regularity of the point a implies there is a $\delta > 0$ such that $g(z,\infty) < \varepsilon$ for z in $G(a;\delta)$. Thus

$$\begin{aligned}
\gamma \geq L_\omega(a) &\geq \frac{1}{\pi\,\delta^2}\int_{B(a:\delta)} L_\omega d\mathcal{A} \\
&= \frac{1}{\pi\,\delta^2}\int_{B(a:\delta)\setminus G} L_\omega d\mathcal{A} + \frac{1}{\pi\,\delta^2}\int_{G(a:\delta)} L_\omega d\mathcal{A} \\
&\geq \frac{\gamma}{\pi\,\delta^2}\,\mathrm{Area}(B(a;\delta)) + \frac{\gamma-\varepsilon}{\pi\,\delta^2}\,\mathrm{Area}(G(a;\delta) \\
&\geq \gamma - \varepsilon. \quad \square
\end{aligned}$$

We turn now to an application of the uniqueness and identification of the equilibrium measure for a compact set.

10.9 Theorem. *Let G be a hyperbolic region in $\mathbb{C}_\infty$ with $G = \cup_n G_n$ and assume that $a \in G_n \subseteq G_{n+1}$ for all n. If ω and ω_n are the harmonic measures for G and G_n at a, then $\omega_n \to \omega$ weak* in $M(\mathbb{C}_\infty)$.*

Proof. There is no loss in generality in assuming that $a = \infty$. If K and K_n are the complements of G and G_n, then $K = \cap_n K_n$ and ω and ω_n are the equilibrium measures for K and K_n. Recall Proposition 7.15, where it is shown that every weak* cluster point of $\{\omega_n\}$ is an equilibrium measure for K. Since the equilibrium measure is unique and these measures lie in a compact metric space, $\omega_n \to \omega$ weak*. $\square$

Exercises

1. Show that if E is any set and F is a set with $c(F) = 0$, then $c(E) = c(E \cup F) = c(E \setminus F)$.

2. If K is a closed arc on a circle of radius R that has length $\theta\, R$, show that $c(K) = R\sin(\theta/4)$.

3. If K is the ellipse $x^2/a^2 + y^2/b^2 = 1$, then $c(K) = (a+b)/2$.

4. Let $p(z) = z^n + a_1 z^{n-1} + \ldots + a_n$ and put $K = \{z \in \mathbb{C} : |p(z)| \leq R\}$. Show that $c(K) = \sqrt[n]{R}$. (See Exercise 19.9.3.)

§11 The Transfinite Diameter and Logarithmic Capacity

In this section we will identify the logarithmic capacity with another constant associated with compact sets: the transfinite diameter. This identification shows an intimate connection between the logarithmic capacity and the geometry of the plane.

Let K be a fixed compact set and for each integer $n \geq 2$ let $K^{(n)} = \{\overline{z} = (z_1, \ldots, z_n) \in \mathbb{C}^n : z_j \in K \text{ for } 1 \leq j \leq n\}$. Let $c_n = n(n-1)/2 = \binom{n}{2}$, the number of ways of choosing 2 things from n things. Define the constant

11.1 $$\delta_n(K) \equiv \max\left\{\left[\prod_{1\leq j<k\leq n} |z_j - z_k|\right]^{1/c_n} : \overline{z} \in K^{(n)}\right\}.$$

Let's note a few facts. First, because of the compactness of $K^{(n)}$, this is a maximum and not just a supremum. Second, if K has N points, then $\delta_n(K) = 0$ for $n > N$. So for the immediate future assume that K is infinite. In particular, $\delta_n(K) > 0$ for an infinite K. Also note that if $z_j = z_k$ for some $j < k$, the product in (11.1) is zero. So the maximum can be taken over $\overline{z}$ in $K^{(n)}$ with $z_j \neq z_k$ for $1 \leq j < k \leq n$. Next observe that $\delta_2(K) = \operatorname{diam} K$. Finally, if $\overline{z} \in K^{(n)}$, then the *Vandermonde* of $\overline{z}$, $V(\overline{z})$, is defined as the determinant

$$V(\overline{z}) = \det |1, z_j, z_j^2, \ldots, z_j^{n-1}|_{1\leq j\leq n}.$$

It follows that

$$V(\overline{z}) = \prod_{1\leq j<k\leq n} (z_k - z_j).$$

So $\delta_n(K) = \max\{|V(\overline{z})|^{1/c_n} : \overline{z} \in K^{(n)}\}$.

11.2 Proposition. *For any compact set K the sequence $\{\delta_n(K)\}$ is decreasing.*

Proof. Let $\overline{z} \in K^{(n+1)}$ such that $\delta_{n+1} = \delta_{n+1}(K) = |V(\overline{z})|^{1/c_{n+1}}$. Thus

$$\begin{aligned} \delta_{n+1}^{c_{n+1}} &= |z_{n+1} - z_1| \cdots |z_{n+1} - z_n| \prod_{1\leq j<k\leq n} |z_k - z_j| \\ &\leq |z_{n+1} - z_1| \cdots |z_{n+1} - z_n| \delta_n^{c_n}. \end{aligned}$$

Similarly, for each $k = 1, \ldots, n+1$

$$\delta_{n+1}^{c_{n+1}} \leq \delta_n^{c_n} \prod_{j\neq k} |z_k - z_j|.$$

Taking the product of these $n+1$ inequalities gives that

$$\begin{aligned} \delta_{n+1}^{(n+1)c_{n+1}} &\leq \left[\prod_{k=1}^{n+1} \prod_{j\neq k} |z_k - z_j|\right] \delta_n^{(n+1)c_n} \\ &= \delta_{n+1}^{n(n+1)} \delta_n^{n(n-1)(n+1)/2}. \end{aligned}$$

Performing the algebraic simplifications and taking the appropriate roots, we discover that $\delta_{n+1} \leq \delta_n$. □

11.3 Definition. For a compact set K the number

$$\delta_\infty(K) = \lim_{n\to\infty} \delta_n(K)$$

is called the *transfinite diameter* of K.

The existence of this limit is of course guaranteed by the preceding proposition and it is clear why the terminology is used. For finite sets the transfinite diameter is zero, but there are examples of compact sets that are infinite and have zero transfinite diameter. Indeed the next theorem combined with Example 8.17 furnishes such an example.

11.4 Theorem. *If K is a compact set, then the transfinite diameter of K equals its logarithmic capacity.*

Proof. If K is finite, then both $\delta_\infty(K)$ and $c(K)$ are zero. So assume that K is infinite. Let $\delta_n = \delta_n(K)$ and $\delta_\infty = \delta_\infty(K)$.

Let ω be the equilibrium measure for K and let $z_1, \ldots, z_n$ be any points in K. Observe that

$$\sum_{1\le j<k\le n} \log|z_k - z_j|^{-1} \ge c_n \log \delta_n^{-1}$$

even if these points are not distinct. If L is the function defined on $K^{(n)}$ by $L(z_1, \ldots, z_n) =$ the left hand side of the preceding inequality and $\nu = \omega \times \cdots \times \omega$ (n times), then ν is a probability measure and so

$$\begin{aligned}
c_n \log \delta_n^{-1} &\le \int L(z_1, \ldots, z_n) d\nu \\
&= \int \sum_{j=1}^{n-1} \sum_{k=j+1}^{n} \log|z_k - z_j|^{-1} d\omega(z_1) \cdots d\omega(z_n) \\
&= \sum_{j=1}^{n-1} \sum_{k=j+1}^{n} \int\int \log|z_k - z_j|^{-1} d\omega(z_j) d\omega(z_k) \\
&= c_n I(\omega) \\
&= c_n \mathrm{rob}(K).
\end{aligned}$$

Thus $\delta_\infty \ge c(K)$.

For the reverse inequality, choose an arbitrary $\varepsilon > 0$. By Corollary 7.16 there is an open set U that contains K and satisfies $c(U) < c(K) + \varepsilon$. Let n be sufficiently large that $1/\sqrt{n\pi} < \mathrm{dist}(K, \partial U)$ and let $z_1, \ldots, z_n$ be points from K such that $\delta_n = |V(\overline{z})|^{1/c_n}$. Put $B_j = B(z_j; 1/\sqrt{n\pi})$. (So Area$(B_j) = n^{-1}$.) Define $\tau(z)$ to be the number of disks B_j that contain z. Note that τ is a Borel function with $0 \le \tau \le n$. If μ is the measure $\tau\mathcal{A}$, then $\mu > 0$ and $||\mu|| = \sum_j \text{Area}(B_j) = 1$. So $\mu \in M_1(U)$. Therefore

$$\begin{aligned} v(U) &\leq \int\int \log|z-w|^{-1}\tau(z)\,\tau(w)\,d\mathcal{A}(z)\,d\mathcal{A}(w) \\ &= \sum_{j=1}^{n}\int_{B_j}\left[\sum_{k=1}^{n}\int_{B_k}\log|z-w|^{-1}d\mathcal{A}(z)\right]\,d\mathcal{A}(w). \end{aligned}$$

Now for any w, $z \to \log|z-w|^{-1}$ is superharmonic. Thus $\int_{B_k}\log|z-w|^{-1}d\mathcal{A}(z) \leq n^{-1}\log|z_k-w|^{-1}$. Thus

$$v(U) \leq \frac{1}{n}\sum_{j=1}^{n}\int_{B_j}\left[\sum_{k=1}^{n}\log|z_k-w|^{-1}\right]\,d\mathcal{A}(w).$$

Similarly for $j \neq k$, $\int_{B_j}\log|z_k-w|^{-1}d\mathcal{A}(w) \leq n^{-1}\log|z_k-z_j|^{-1}$. If $j=k$, then an evaluation of the integral using polar coordinates gives

$$\int_{B_k}\log|z_k-w|^{-1}d\mathcal{A}(w) = \frac{1}{2n}\,(1+\log n\pi).$$

Hence

$$\begin{aligned} v(U) &\leq \frac{1}{n}\left[\sum_{j\neq k}\frac{1}{n}\log|z_k-z_j|^{-1}+n\,\frac{1}{2n}\,(1+\log n\pi)\right] \\ &= \frac{2}{n^2}\sum_{1\leq j<k\leq n}\log|z_k-z_j|^{-1}+\frac{1}{2n}(1+\log n\pi) \\ &= \frac{2}{n^2}\,\frac{n(n-1)}{2}\,\log\delta_n^{-1}+\frac{1}{2n}\,(1+\log n\pi). \end{aligned}$$

Letting $n \to \infty$ we get that $v(U) \leq \log\delta_\infty^{-1}$ so that $c(K) > c(U)-\varepsilon \geq \delta_\infty - \varepsilon$. Hence $c(K) \geq \delta_\infty(K)$. □

The advantage of the transfinite diameter over logarithmic capacity is that the transfinite diameter is more geometric in its definition. After all, the definition of the logarithmic capacity of a set is given in terms of measures, while the definition of the transfinite diameter is in terms of distances. This is amplified in the next corollary, which would be more difficult to prove without the preceding theorem.

11.5 Corollary. *If K is a compact set and $f : K \to L$ is a surjective function such that there is a constant M with $|f(z)-f(w)| \leq M|z-w|$ for all z, w in K, then $c(L) \leq c(K)$.*

Proof. Let $\zeta_1,\ldots,\zeta_n \in L$ such that $\delta_n^{c_n}(L) = \prod\{|\zeta_k-\zeta_j| : 1 \leq j < k \leq n\}$ and pick $z_1,\ldots,z_n$ in K with $f(z_j)=\zeta_j$. So $\delta_n^{c_n}(L) = \prod\{|f(z_k)-f(z_j)| :$

$1 \le j < k \le n\} \le M\delta_n^{c_n}(K)$; equivalently, $\delta_n(L) \le M^{1/c_n}\delta_n(K)$. Letting $n \to \infty$ gives the conclusion. □

11.6 Corollary. *If γ is a rectifiable Jordan arc or curve and K is a compact subset of γ, then $c(K) \ge |K|/4$, where $|K|$ denotes the arc length of K.*

Proof. Assume that γ is parametrized by $\gamma : [0,1] \to \mathbb{C}$ with $a = \gamma(0)$ a point in K. If $z \in K$, $z \ne a$, there is a unique t in $(0,1]$ such that $\gamma(t) = z$. Define $f : K \to [0, |K|]$ by $f(z) = |K \cap \gamma([0,t])|$. It is clear that f is surjective. If $z, w \in K$ and $z = \gamma(t)$, $w = \gamma(s)$ with $s < t$, then $|f(z) - f(w)| = |K \cap \gamma((s,t])| \le |\gamma([s,t])| \le |z - w|$ since the shortest distance between two points is a straight line. According to the preceding corollary, $c(K) \ge c([0, |K|]) = |K|/4$ by Corollary 10.5. □

11.7 Corollary. *If K is a compact connected set and $d = \operatorname{diam} K$, then $c(K) \ge d/4$.*

Proof. Let $a, b \in K$ such that $|a - b| = d$. By a rotation and translation of K, which does not change the capacity, we may assume that $a = 0$ and $b = d \in \mathbb{R}$. By the choice of a and b, $0 \le \operatorname{Re} z \le d$ for all z in K. Since K is connected, $\operatorname{Re} : K \to [0, d]$ is surjective. According to Corollary 11.5, $c(K) \ge c([0, d]) = d/4$ (10.5). □

There is an equivalent expression for the transfinite diameter and hence the logarithmic capacity of a compact set connected to polynomial approximation. This development is sketched below without proof. The interested reader can see Chapter VII of Goluzin [1969] and §16.2 of Hille [1962].

Let K be an infinite compact set and for $n \ge 1$ let $\mathcal{P}_n$ be the vector space of all polynomials of degree at most n. If $||p||_K \equiv \max\{|p(z)| : z \in K\}$, then $|| \cdot ||_K$ defines a norm on $\mathcal{P}_n$. Let $\mathcal{M}_n$ be the collection of all monic polynomials of degree n. That is, $\mathcal{M}_n$ consists of all polynomials of the form $p(z) = z^n + a_1 z^{n-1} + \cdots + a_n$. Define the constant

$$M_n \equiv \inf\{||p||_K : p \in \mathcal{M}_n\}.$$

Note that $0 \le M_n < \infty$. It can be shown that there is a unique polynomial T_n in $\mathcal{M}_n$ with $||T_n||_K = M_n$. The existence of T_n is easy but the uniqueness is not. The polynomial T_n is called the *Tchebycheff polynomial* for K of order n. The *Tchebycheff constant* for K is defined by $\operatorname{tch}(K) \equiv \lim_{n\to\infty} \sqrt[n]{M_n(K)}$, which always exists and is finite. In fact, it is easy to see that $\operatorname{tch}(K) \le \operatorname{diam} K$. It turns out that the Tchebycheff constant and the transfinite diameter are the same.

Exercise

1. Let K be a compact set and suppose $\overline{z} = (z_1, \dots, z_n) \in K^{(n)}$ such

that $\delta_n^{c_n} = \prod_{1\le j<k\le n} |z_k - z_j|$. Show that $z_1, \ldots, z_n$ belong to $\partial \hat{K}$ (the outer boundary of K).

§12 The Refinement of a Subharmonic Function

Because of the intimate connection of subharmonic functions with so many of the properties of capacity and the solution of the Dirichlet problem, it is no surprise that the ability to manufacture these functions will extend the power of the theory and increase the depth of the results. In this section a new technique is developed that increases our proficiency at constructing subharmonic functions. In the next two sections this augmented skill will be put to good use as we prove Wiener's criterion for regularity.

Let G be a hyperbolic open subset of $\mathbb{C}$ and suppose E is an arbitrary subset of G. For a negative subharmonic function u on G define

$$\begin{aligned} I_E^u(z) \equiv\ & \sup\{\phi(z) : \phi \text{ is a negative subharmonic function on } G \\ & \text{that is not identically } -\infty \text{ on any component of } G \\ & \text{and } \phi \le u \text{ on } E\}. \end{aligned}$$

The function I_E^u is called the *increased function* of u relative to E. Be aware that in the notation for the increased function the role of G is suppressed. Also note that I_E^u may fail to be usc and thus may not be subharmonic. To correct for this, define

$$\hat{I}_E^u(z) = \limsup_{w\to z} I_E^u(w)$$

for all z in G. The function $\hat{I}_E^u$ is called the *refinement* of u relative to E. The term for this function most often seen in the literature is the French word "balayage." This word means "sweeping" or "brushing." The term and concept go back to Poincaré and the idea is that the subharmonic function u is modified and polished (or brushed) to produce a better behaved function that still resembles u on the set E. That the function $\hat{I}_E^u$ accomplishes this will be seen shortly. We are avoiding the word "sweep" in this context as it was used in §2 in a different way.

Here are some of the properties of the increased function and the refinement. Let Φ_E^u be the set of negative subharmonic functions used to define I_E^u.

12.1 Proposition. *If G is a hyperbolic open set, E and F are subsets of G, and u and v are negative subharmonic functions on G, then:*

(a) *$\hat{I}_E^u$ is subharmonic on G;*

(b) *$u \le I_E^u \le \hat{I}_E^u \le 0$ on G;*

(c) $I_E^u = u$ *on* E;

(d) $I_E^u = \hat{I}_E^u$ *on* int E;

(e) $I_\emptyset^u \equiv 0$;

(f) *If* $E \subseteq F$, $\hat{I}_E^u \geq \hat{I}_F^u$;

(g) *If* $u \leq v, \hat{I}_E^u \leq \hat{I}_E^v$;

(h) *If* $t > 0$, $\hat{I}_E^{tu} = t\, \hat{I}_E^U$;

(i) $\hat{I}_E^{u+v} \geq \hat{I}_E^u + \hat{I}_E^v$;

(j) $I_E^u = \hat{I}_E^u$ *on* $G \setminus \text{cl}\, E$ *and this function is harmonic there.*

Proof. Part (a) is a direct application of (19.4.6). Parts (b) through (h) are easily deduced from the definitions, while part (i) follows from the relation $\Phi_E^u + \Phi_E^v \subseteq \Phi_E^{u+v}$. To prove part (j) we will use Lemma 19.7.4. First note that if ϕ_1 and $\phi_2 \in \Phi_E^u$, then so does $\phi_1 \vee \phi_2$. Second, if $\phi \in \Phi_E^u$, D is a closed disk contained in $G \setminus \text{cl}\ E$, and ϕ_1 is the harmonic modification of ϕ on D, then $\phi_1 \in \Phi_E^u$. Hence I_E^u is harmonic on $G \setminus \text{cl}\, E$. Since I_E^u is continuous there, $I_E^u = \hat{I}_E^u$. □

It may be that strict inequality holds in part (b). See Exercise 1. In fact, this exercise is solved by looking at the proof that the origin is not a regular point for the punctured disk and this is no accident. It will be shown later in this section that $I_E^u = \hat{I}_E^u$ except on a polar set (Corollary 12.10).

Now for one of the more useful applications of the refinement of a subharmonic function.

12.2 Proposition. *If* u *is a negative subharmonic function on* G *and* K *is a compact subset of* G*, then there is a positive measure* μ *supported on* K *such that* $\hat{I}_K^u = -G_\mu$ *(the Green potential of* μ*).*

Proof. By Corollary 4.11 it suffices to show that the least harmonic majorant of $\hat{I}_K^u$ is 0. First assume that G is connected and u is bounded on K; so there is a positive constant r such that $u \geq -r$ on K. Fix a point a in G and let $g_a(z) = g(z,a)$ be the Green function for G with pole at a. Since g_a is lsc, there is a positive constant s such that $g_a(z) \geq s$ for all z in K. So for $t = r/s > 0$, $-tg_a \in \Phi_K^u$ and so $-t\ g_a \leq I_E^u$. Now g_a is harmonic on $G \setminus \{a\}$ and so $-t\ g_a \leq \hat{I}_K^u \leq 0$. Thus if h is a harmonic majorant of $\hat{I}_K^u$, $g_a \geq -h/t$. Since the greatest harmonic minorant of g_a is 0, this implies $h \geq 0$ and so the least harmonic majorant of $\hat{I}_K^u$ must be 0.

Now assume that u is bounded on K but G is not connected. Because K is compact, there are at most a finite number of components $G_1, \ldots, G_n$ of G that meet K. Also $K_j \equiv K \cap G_j$ is compact. If H is a component of G different from $G_1, \ldots, G_n$, then the fact that $H \cap K = \emptyset$ yields that $\hat{I}_K^u = 0$ on H (12.1.e). This case now follows from the preceding paragraph applied to each of the components $G_1, \ldots, G_n$.

Now for the arbitrary case. Let h be the least harmonic majorant of $\hat{I}_K^u$. Since $\hat{I}_K^u \leq 0$, $h \leq 0$, it must be shown that $h \geq 0$. By Theorem 4.10 there is a positive measure μ on G such that $\hat{I}_K^u = h - G_\mu$. In fact, $\mu = \Delta \hat{I}_K^u$ in the sense of distributions. Since $\hat{I}_K^u$ is harmonic on $G \setminus K$, $\operatorname{supp}\mu \subseteq K$; in particular, μ is a finite measure. If $v = -G_\mu$, then v is a negative subharmonic function on G and $u = h + v$. By (12.1.i), $\hat{I}_K^u \geq \hat{I}_K^v + \hat{I}_K^h$. Thus h is a harmonic majorant of $\hat{I}_K^h$. But h is bounded on K and so the least harmonic majorant of $\hat{I}_K^h$ is 0. That is, $h \geq 0$. □

12.3 Corollary. *If G is a hyperbolic open set, u a negative subharmonic function on G, and K a compact subset of G, then there is a positive measure μ supported on K such that $u(z) = -G_\mu(z)$ for all z in* int K.

Proof. According to Proposition 12.1, $\hat{I}_K^u = u$ on int K. Now use the preceding proposition. □

In the next proposition we will see how the refinement is used to show that subharmonic functions can be extended in some sense.

12.4 Proposition. *If u is a subharmonic function on the disk $B_r = B(a;r)$ and $0 < s < r$, then there is a subharmonic function u_1 on $\mathbb{C}$ such that $u_1 = u$ on B_s and u_1 is finite on $\mathbb{C} \setminus \operatorname{cl} B_s$. In fact, $u_1(z) = \log(|z-a|/r)$ for $|z-a| \geq r$.*

Proof. Clearly, by decreasing r slightly, it can be assumed that u is subharmonic in a neighborhood of cl B_r, and hence bounded above there. There is no loss in generality in assuming that $u \leq 0$ on cl B_r. Put $K =$ cl B_s and consider $\hat{I}_K^u$. From Proposition 12.1 we know that $\hat{I}_K^u = u$ on B_s and is harmonic on $B_r \setminus K$. If g is the Green function for B_r, then Proposition 12.2 implies there is a positive measure μ supported on cl B_s such that $\hat{I}_K^u = -G_\mu$.

But if $\zeta \in \partial B_r$, $g(w,z) \to 0$ uniformly for w in K as $z \to \zeta$ (Exercise 6.2). Thus $G_\mu(z) \to 0$ as $z \to \zeta$ for all ζ in ∂B_r. By Corollary 13.4.13 $\hat{I}_K^u = -G_\mu$ can be extended across the circle ∂B_r. That is, there is a $t > r$ and a function w defined on B_t such that $w = \hat{I}_K^u$ on B_r, w is harmonic on $B_t \setminus$ cl B_s, and $w \leq 0$. Define u_1 on $\mathbb{C}$ by letting $u_1 = \hat{I}_K^u = w$ on B_r and $u_1(z) = \log(|z-a|/r)$ on $\mathbb{C} \setminus B_r$. It follows from Exercise 19.4.9 that u_1 is subharmonic. □

In the next two propositions we will construct subharmonic functions that have specified behavior relative to a polar set. These results are improvements of the definition of a polar set and Proposition 5.5.

12.5 Proposition. *If Z is a polar set and $a \notin Z$, then there is a subharmonic function u on $\mathbb{C}$ such that $u = -\infty$ on Z and $u(a) > -\infty$.*

Proof. By definition there is a subharmonic function v on $\mathbb{C}$ that is not constant with $v(z) = -\infty$ for all z in Z. Let $\{z_n\}$ be a sequence of points in Z and $\{r_n\}$ a sequence of radii with $0 < 2r_n < |z_n - a|$ such that if $B_n = B(z_n; r_n)$, $Z \subseteq \cup_n B_n$. By Proposition 12.4 there is a subharmonic function u_n on $\mathbb{C}$ with $u_n = v$ on B_n and $u_n(z) > -\infty$ for $z \notin \text{cl } B_n$. In particular, $u_n(z) > -\infty$ for $|z-a| \le r_n$. Since u_n is usc, there is a constant α_n such that $-\infty < u_n(z) + \alpha_n \le 0$ for $|z - a| \le r_n$. Now choose positive scalars $\{\beta_n\}$ so that $\sum_n \beta_n[u_n(a)+\alpha_n] > -\infty$. If $u(z) = \sum_n \beta_n[u_n(z)+\alpha_n]$, then it is easy to check that u has the desired properties. □

12.6 Proposition. *If G is a hyperbolic open set, Z is a polar set that is contained in G, and $a \in G \setminus Z$, then there is a negative subharmonic function u on G with $u(z) = -\infty$ for all z in Z and $u(a) > -\infty$.*

Proof. First observe that it suffices to prove the proposition under the additional condition that there is an open disk D that contains Z with $\text{cl } D \subseteq G$. Indeed, if the proposition is proved with this additional condition, then in the arbitrary case let $\{B_n\} = \{B(a_n; r_n)\}$ be a sequence of disks such that $Z \subseteq \cup_n B_n$ and $\text{cl } B_n \subseteq G$ for all n. According to the assumption, there is a negative subharmonic function u_n on G such that $u_n(a) > -\infty$ and $u_n(z) = -\infty$ for z in $Z \cap B_n$. Let $\{\beta_n\}$ be a sequence of positive scalars such that $\sum_n \beta_n u_n(a) > -\infty$ and put $u = \sum_n \beta_n u_n$.

So assume that $D = B(b; s)$ is an open disk with $Z \subseteq D$ and $\text{cl } D \subseteq G$. Let $r > s$ such that $\overline{B}(b; r) \subseteq G$. Using Proposition 12.4 and Proposition 12.5, there is a subharmonic function w on $\mathbb{C}$ such that $w(z) = -\infty$ for all z in Z, $w(a) > -\infty$, and $w(z) = \log(|z - b|/r)$ for z not in $B = B(b; r)$. Let g be the Green function for G and let h be the harmonic function on G such that $g(z, b) = h(z) - \log|z - b|$ for all z in G. So on $G \setminus B$, $w(z) = \log|z - b| - \log r = h(z) - g(z, b) - \log r$.

Define $u(z) = w(z) - h(z) + \log r$ for z in G. So u is a subharmonic function on G and, for z in $G \setminus \text{cl } B$, $u(z) = -g(z, b) \le 0$. Since u is usc, this gives that u is bounded above on all of G. By Theorem 6.5 there is a polar set E contained in $\partial_\infty G$ such that for $\zeta \in \partial_\infty G \setminus E$, $g(z, b) \to 0$ as $z \to \zeta$. Thus for all $\zeta \in \partial_\infty G \setminus E$, $u(z) \to 0$ as $z \to \zeta$. By Theorem 8.2, $u \le 0$ on G. Clearly $u(a) > -\infty$. □

Now for a lemma about upper semicontinuous functions that could have been presented in §19.3 but has not been needed until now. This is the first of two lemmas that are needed for the proof of Theorem 12.9, the main result of this section.

12.7 Lemma. *For an arbitrary family of functions $\{f_i : i \in I\}$ on a separable metric space X and L any subset of I, define $f_L(x) = \sup\{f_i(x) : i \in L\}$. There is a countable subset J of I such that if g is any upper semicontinuous function on X with $g(x) \ge f_J(x)$ for all x in X, then $g(x) \ge f_I(x)$ for all x in X.*

Proof. Before getting into the formal part of the proof, let's recall that an usc function is not allowed to assume the value ∞; $-\infty$ is the only non-finite value it can take on. Thus the countable set J should be chosen so that, if there is a point x with $f_I(x) = \infty$, then $f_J(x) = \infty$. In this way if there is a point where $f_I(x) = \infty$, there can be no usc function g with $g(x) \geq f_J(x)$.

Let $\{U_n\}$ be a countable base for the topology of X and for each $n \geq 1$ let $M_n \equiv \sup\{f_I(x) : x \in U_n\}$. If $M_n < \infty$, there is a point x_1 in U_n with $f_I(x_1) + (2n)^{-1} > M_n$. Thus there is an i_n in I with $f_{i_n}(x_1) + (2n)^{-1} > f_I(x_1)$. Thus for every n with $M_n < \infty$, there is an i_n in I with

$$\sup_{x\in U_n} f_{i_n}(x) + \frac{1}{n} > M_n.$$

If $M_n = \infty$, then an argument similar to the preceding one shows that there is an i_n in I with

$$\sup_{x\in U_n} f_{i_n}(x) > n.$$

Let $J = \{i_1, i_2, ...\}$.

Suppose g is an usc function on X with $g \geq f_J$. Fix an x_0 in X and let $\varepsilon > 0$. Note that if $g(x_0) = -\infty$, there is nothing to verify; so assume that $g(x_0)$ is finite. From the definition of an usc function and the fact that $\{U_n\}$ is a neighborhood base, there is an integer n with $n^{-1} < \varepsilon$ such that $g(x) < g(x_0) + \varepsilon$ for all x in U_n. Now if $M_n = \infty$,

$$\begin{aligned} g(x_0) &= [g(x_0) - \sup_{x\in U_n} g(x)] + [\sup_{x\in U_n} g(x) - \sup_{x\in U_n} f_{i_n}(x)] + \sup_{x\in U_n} f_{i_n}(x) \\ &> -\varepsilon + 0 + n \\ &> -\varepsilon + \frac{1}{\varepsilon}. \end{aligned}$$

Since ε was arbitrary it must be that $g(x_0) = \infty$, an impossibility. Thus $M_n < \infty$ for all n. Therefore

$$\begin{aligned} g(x_0) - \sup_{x\in U_n} f_I(x) &= [g(x_0) - \sup_{x\in U_n} g(x)] + [\sup_{x\in U_n} g(x) - \sup_{x\in U_n} f_{i_n}(x)] \\ &\quad + [\sup_{x\in U_n} f_{i_n}(x) - \sup_{x\in U_n} f_I(x)] \\ &\geq -\varepsilon + 0 - \frac{1}{n} \\ &> -2\varepsilon. \end{aligned}$$

Since ε was arbitrary, $g(x_0) \geq f_I(x_0)$. □

12.8 Lemma. *If $\{\mu_i : i \in I\}$ is a family of positive measures on an open disk D having the following properties:*

(a) supp $\mu_i \subseteq D$ *for all i in I;*

(b) *for any i and j in I there is a k in I with* $G^D_{\mu_k} \leq \min\{G^D_{\mu_i}, G^D_{\mu_j}\}$;

(c) $\sup_i \mu_i(D) < \infty$;

then there is a positive measure μ supported on cl D *such that* $G^D_\mu \leq \inf_i G^D_{\mu_i}$ *and* $G^D_\mu = \inf_i G^D_{\mu_i}$ *q.e.*

Proof. For any measure μ on D put $G_\mu = G^D_\mu$. According to Lemma 12.7 there is a countable subset J of I such that if g is a lsc function with $g \leq \inf_j G_{\mu_j} \equiv \psi_0$, then $g \leq \inf_i G_{\mu_i} \equiv \psi$. Write $\{\mu_j : j \in J\}$ as a sequence $\{\mu_n\}$. By property (b) it can be assumed that $G_{\mu_n} \geq G_{\mu_{n+1}}$ for all $n \geq 1$. By property (c) the sequence $\{\mu_n\}$ can be replaced by a subsequence, so that it can be assumed that there is a positive measure μ on cl D such that $\mu_n \to \mu$ weak* in $M(\text{cl } D)$. By Exercise 4.6, for all z in D, $G_\mu(z) \leq \liminf_n G_{\mu_n}(z) = \lim_n G_{\mu_n}(z) = \psi_0(z)$. Since G_μ is a lsc function, the choice of the subset J implies $G_\mu \leq \psi \leq \psi_0$.

Let $E = \{z \in D : G_\mu(z) < \psi_0(z)\}$; so E is a Borel set. The proof will be complete if it can be shown that $c(E) = 0$. Suppose it is not. Then there is a compact subset K of E with $c(K) > 0$. By Proposition 7.14 there is a positive measure ν on K such that L_ν is continuous and finite-valued on $\mathbb{C}$. But Theorem 4.5 implies that $G_\nu(z) = L_\nu(z) - \int L_\nu(\zeta)\, d\omega^D_z(\zeta)$. Since L_ν is continuous and disks are Dirichlet sets, $G_\nu(z) \to 0$ as z approaches any point of ∂D. Thus defining $G_\nu(z) = 0$ on ∂D makes G_ν a continuous function on cl D. Therefore $\int G_\nu d\mu_n \to \int G_\nu d\mu$. By Fatou's Lemma this implies $\int \psi_0 d\nu \leq \lim_n \int G_{\mu_n} d\nu = \lim_n \int G_\nu d\mu_n = \int G_\nu d\mu = \int G_\mu d\nu$. Thus $\int (G_\mu - \psi_0) d\nu \geq 0$. But $G_\mu - \psi_0 < 0$ on the support of ν, so this is a contradiction and $c(E) = 0$. □

Recall from Proposition 19.4.6 that if $\mathcal{U}$ is a family of subharmonic functions that is locally bounded above, $v = \sup \mathcal{U}$, and $u(z) = \limsup_{w \to z} v(w)$, then u is subharmonic. It was also shown there that if v is usc, then $v = u$. The question arises as to how badly can u and v differ. The next theorem answers this.

12.9 Theorem. *If $\mathcal{U}$ is a set of subharmonic functions on an open set G that is locally bounded above, $v(z) = \sup \mathcal{U}$, and $u(z) = \limsup_{w \to z} v(w)$, then $u = v$ q.e.*

Proof. First enlarge $\mathcal{U}$ to include the functions defined as the maximum of any finite subset of $\mathcal{U}$ and realize that this does not change the value of v and u. Let $\phi_0 \in \mathcal{U}$ and observe that replacing $\mathcal{U}$ by the collection of functions $\{\phi \in \mathcal{U} : \phi \geq \phi_0\}$ does not change the value of v and u. So it can also be assumed that there is a function ϕ_0 in $\mathcal{U}$ such that $\phi \geq \phi_0$ for all ϕ in $\mathcal{U}$.

We first prove a special case of the theorem and then this will be used to prove the theorem in total generality.

Claim. If $G = D$, an open disk, and each function in $\mathcal{U}$ is negative, then there is a Borel subset E of D with $c(E) = 0$ and $u = v$ on $D \setminus E$.

Let K be any closed disk contained in D and for each function ϕ in $\mathcal{U}$ form the refinement $\hat{I}_K^\phi$. So on int K, Proposition 12.1 implies $\hat{I}_K^\phi = I_K^\phi = \phi$. Thus $v = \sup\{\hat{I}_K^\phi : \phi \in \mathcal{U}\}$ on int K. By Proposition 12.2 for each ϕ in $\mathcal{U}$ there is a positive measure μ_ϕ on K such that $\hat{I}_K^\phi = -G_{\mu_\phi}$ on D. Since $\phi_1 \vee \phi_2 \in \mathcal{U}$ whenever both ϕ_1 and ϕ_2 are, it follows that the family of measures $\{\mu_\phi : \phi \in \mathcal{U}\}$ satisfies conditions (a) and (b) of the preceding lemma.

Let B be a closed disk contained in D with $K \subseteq \operatorname{int} B$ and consider the refinement $\hat{I}_B^{-1}$. Once again there is a positive measure η on B such that $\hat{I}_B^{-1} = -G_\eta$ on D. Thus $G_\eta = -\hat{I}_B^{-1} = 1$ on int B. Thus for any ϕ in $\mathcal{U}$, $\mu_\phi(K) = \int G_\eta d\mu_\phi = \int G_{\mu_\phi} d\eta = -\int \hat{I}_K^\phi d\eta \leq -\int \hat{I}_K^{\phi_0} d\eta = \int G_{\mu_{\phi_0}} d\eta = \int G_\eta d\mu_{\phi_0} = \mu_{\phi_0}(K)$. Therefore Lemma 12.8 implies there is a positive measure μ supported on K with $G_\mu \leq G_{\mu_\phi}$ on D and $G_\mu(z) = \inf\{G_{\mu_\phi}(z) : \phi \in \mathcal{U}\}$ except for a Borel subset of D having capacity zero. (Actually, the lemma only gives that the support of μ lies in cl D, but, since each μ_ϕ has its support in K, the proof of the lemma shows that supp$\mu \subseteq K$.) It follows that on int K, $-G_\mu = u$ and $\sup\{-G_{\mu_\phi}(z) : \phi \in \mathcal{U}\} = v(z)$ and there is a Borel subset E_K of int K having zero capacity such that $u = v$ on $(\operatorname{int} K) \setminus E_K$. But D can be written as the union of a sequence of such closed disks. Since the union of a countable number of polar sets is polar, this proves the claim.

Now for the general case. Write G as the union of a sequence of open disks $\{D_n\}$ with cl $D_n \subseteq G$. Since $\mathcal{U}$ is locally bounded above, for each $n \geq 1$ there is a constant M_n such that every ϕ in $\mathcal{U}$ satisfies $\phi \leq M_n$ on D_n. If the claim is applied to the family $\mathcal{U} - M_n = \{\phi - M_n : \phi \in \mathcal{U}\}$, we get that there is a Borel set E_n contained in D_n with $c(E_n) = 0$ such that $u = v$ on $D_n \setminus E_n$. Taking $E = \cup_n E_n$ proves the theorem. □

12.10 Corollary. *If G is a hyperbolic open set, u is a negative subharmonic function, E is a subset of G, and $Z = \{z \in G : I_E^u(z) \neq \hat{I}_E^u(z)\}$, then Z is a Borel set having capacity zero and $Z \subseteq E$.*

Proof. The fact that Z is a Borel set and $c(Z) = 0$ is immediate from the theorem. Now let a be any point in $G \setminus E$; it will be shown that $a \notin Z$. According to Proposition 12.6 there is a negative subharmonic function v on G such that $v(z) = -\infty$ for z in $Z \cap E$ and $v(a) > -\infty$. Thus for any $\varepsilon > 0$, $\hat{I}_E^u + \varepsilon v \leq u$ on E. By definition this implies $\hat{I}_E^u + \varepsilon v \leq I_E^u$ on G. Since $v(a) > -\infty$ and ε is arbitrary, this implies that $\hat{I}_E^u(a) \leq I_E^u(a)$. Since $I_E^u \leq \hat{I}_E^u$, this completes the proof. □

12.11 Corollary. *If G is a hyperbolic open set, u is a negative subharmonic* *function on G, $E \subseteq G$, and Z is a polar subset of E, then $\hat{I}_E^u = \hat{I}_{E\setminus Z}^u$.*

Proof. From Proposition 12.1 we know that $\hat{I}_{E\setminus Z}^u \geq \hat{I}_E^u$. Fix a in $G \setminus Z$ and let w be any negative subharmonic function on G such that $w \leq u$ on $E \setminus Z$. According to Proposition 12.6 there is a negative subharmonic function v on G with $v(z) = -\infty$ for all z in Z and $v(a) > -\infty$. Therefore for any $\varepsilon > 0$, $w+\varepsilon v \leq u$ on all of E. Using the preceding corollary, $\hat{I}_E^u(a) = I_E^u(a) \geq w(a) + \varepsilon v(a)$ for all $\varepsilon > 0$. Letting $\varepsilon \to 0$ gives that $\hat{I}_E^u(a) \geq w(a)$ for all a in $G\setminus Z$. Since Z has area 0, this implies that whenever $\overline{B}(b;r) \subseteq G$, $(\pi r^2)^{-1} \int_{B(b;r)} \hat{I}_E^u dA \geq (\pi r^2)^{-1} \int_{B(b;r)} w\, dA \geq w(a)$. By Proposition 19.4.9 this implies that $\hat{I}_E^u \geq w$ on G. Therefore by definition of $I_{E\setminus Z}^u$, $\hat{I}_E^u \geq I_{E\setminus Z}^u$. Taking lim sup's of both sides and using the observation at the beginning of the proof shows that $\hat{I}_E^u = \hat{I}_{E\setminus Z}^u$. □

Exercises

1. Let $G = \mathbb{D}$ and $E = \{0\}$ and define $u(z) = \log|z|$. Show that $I_E^u(z) = 0$ for z in $\mathbb{D} \setminus \{0\}$ and $I_E^u(0) = -\infty$. Hence $\hat{I}_E^u \equiv 0$ so that $\hat{I}_E^u \neq I_E^u$ on $\mathbb{D}$.

2. Show that there is a subharmonic function u on $\mathbb{C}$ that is finite-valued everywhere but not continuous. (Hint: Let Z be a non-closed polar set, let $a \in \text{cl } Z$ such that $a \notin Z$, and let u be a subharmonic function as in Proposition 12.5. Now massage u.)

3. Let G be a hyperbolic open set, let K be a compact subset of G, and let μ be the positive measure on K such that $\hat{I}_K^1 = -G_\mu$ (12.2). Show that $\mu(K) = \sup\{\nu(K) : \nu$ is a positive measure supported on K such that $G_\nu \leq 1$ on $G\}$. So μ is the "equilibrium measure" for the Green potential. See Helms [1975], p 138, and Brelot [1959], p 52, for more detail.

4. Let G be a hyperbolic open set and W a bounded open set with cl $W \subseteq G$. If u is a negative subharmonic function on G, show that on W, $\hat{I}_{G\setminus W}^u$ is the solution of the Dirichlet problem with boundary values $u|\partial W$.

§13 The Fine Topology

This section introduces the fine topology as a prelude to proving in the next section Wiener's Criterion for a point to be a regular point for the solution of the Dirichlet problem. This topology was introduced by the French school, which proved the basic properties of the topology and its connection with the regularity condition. See Brelot [1959] and the notes there as well as Helms [1975], both of which were used as a source for the preparation for this section and the next.

13.1 Definition. The *fine topology* is the smallest topology on $\mathbb{C}$ that makes each subharmonic function continuous as a function from $\mathbb{C}$ into $[-\infty, \infty)$. This topology will be denoted by $\mathcal{F}$.

Note that the discrete topology is one that makes every function, including the subharmonic ones, continuous. Since the intersection of a collection of topologies is a topology, the fine topology is well defined.

When we refer to a set that belongs to $\mathcal{F}$ we will say that the set is *finely open*. Similarly, we will use expressions such as *finely closed, finely continuous*, etc., when these expressions refer to topological phenomena relative to the fine topology. If a set is called "open" with no modifying adjective, this will refer to the usual topology on the plane. A similar convention applies to other topological terms. For convenience the usual topology on $\mathbb{C}$ will be denoted by $\mathcal{U}$. It is not hard to see that if U is finely open, then so is $a + \alpha U$ for all a in $\mathbb{C}$ and $\alpha > 0$.

13.2 Proposition.

(a) *The fine topology is strictly larger than the usual topology on the plane.*

(b) *If G is open and $\phi : G \to [-\infty, \infty)$ is a subharmonic function, then ϕ is continuous if G has the relative fine topology, $\mathcal{F}_G$.*

(c) *A base for the topology $\mathcal{F}$ consists of all sets of the form*

$$W \cap \bigcap_{k=1}^{n} \{z : \phi_k(z) > c_k\},$$

where $W \in \mathcal{U}$, $\phi_1, \ldots, \phi_n$ are subharmonic functions, and $c_1, \ldots, c_n$ are finite constants.

(d) *The fine topology is a Hausdorff topology.*

Proof. (a) To establish that $\mathcal{U} \subseteq \mathcal{F}$, we need only show that each open disk belongs to $\mathcal{F}$. But the observation that for any a the function $\log |z - a|$ is subharmonic and $B(a; r) = \{z : \log |z-a| < \log r\}$ shows this. To show that this containment is proper, we need only exhibit a subharmonic function that is not continuous.

(b) Suppose $\{z_i\}$ is a net in G that converges finely to a in G. We need to show that $\phi(z_i) \to \phi(a)$. Fix an open disk D whose closure is contained in G and let ϕ_1 be a subharmonic function on $\mathbb{C}$ such that $\phi_1(z) = \phi(z)$ for all z in D (12.4). By definition, $\phi_1(z_i) \to \phi_1(a) = \phi(a)$. But by (a) there is an i_0 such that $z_i \in D$ for $i \geq i_0$. Part (b) now follows.

(c) We observe that $\{z : \phi(z) < c\}$ is open when ϕ is subharmonic, because subharmonic functions are upper semicontinuous. The proof of (c) is now a routine exercise in the definitions that is best left to the reader.

(d) This is immediate from (a). □

13.3 Corollary. *Finely open sets are infinite.*

Proof. This follows immediately from part (c) of the preceding proposition if Proposition 19.4.8 is used to show that, for any subharmonic function ϕ and any real constant c, $\{z : \phi(z) > c\}$ is infinite (or empty). □

Note that the last corollary shows that $\mathcal{F}$ is not the discrete topology.

13.4 Proposition. *Polar sets have no fine limit points.*

Proof. Let Z be a polar set. By replacing Z with $Z\backslash\{a\}$, it can be assumed that $a \notin Z$. By Proposition 12.5 there is a subharmonic function ϕ on $\mathbb{C}$ such that $\phi = -\infty$ on Z and $\phi(a) > \alpha > -\infty$. If $U = \{z : \phi(z) > \alpha\}$, then $U \in \mathcal{F}$ (13.2.a), $a \in U$, and $U \cap Z = \emptyset$. □

13.5 Corollary. *All countable subsets of $\mathbb{C}$ are finely closed and the finely compact sets are finite. Thus $\mathcal{F}$ is not a locally compact topology.*

Proof. Since countable sets are polar, they have no fine limit points by the preceding proposition. Hence they contain all their fine limit points and thus must be finely closed. Since no sequence can have a limit point, finely compact sets must be finite. Now combine this with the fact that finely open sets are infinite and it is clear that the fine topology is not locally compact. □

13.6 Corollary. *Every subset of $\mathbb{C}$ is finely sequentially closed. In particular, the fine topology is not first countable.*

13.7 Definition. A subset E of $\mathbb{C}$ is *thick at* a point a if a is a fine limit point of E. If E is not thick at a, say that E is *thin at* a.

So polar sets are thin at every point. Also since $\mathcal{U} \subseteq \mathcal{F}$, if E is thick at a, a is a limit point (in the usual sense) of E. It is the notion of thickness that is the prime reason for discussing the fine topology. This can be seen by the following argument. Let K be a compact set and let G be the component of $\mathbb{C}_\infty \setminus K$ that contains ∞. Suppose $a \in \partial G$ and K is thick at a. If ω is harmonic measure for G at ∞, then there is a polar set Z contained in K such that for all z in $K \setminus Z$, $L_\omega(z) = \gamma$, the Robin constant for K. Since polar sets are thin at every point and K is thick at a, a topological argument shows that $K \setminus Z$ is thick at a. Thus there is a net $\{z_i\}$ in $K \setminus Z$ such that $z_i \to a$ $(\mathcal{F})$. But then $L_\omega(z_i) \to L_\omega(a)$ since superharmonic functions are finely continuous; thus $L_\omega(a) = \gamma$ and by Proposition 10.8

this implies that a is a regular point for G. So each point of ∂G at which K is thick is a regular point for G. The converse of this will be proved in Theorem 13.16 below.

Here is a bit of notation that will be useful in this study. If E is a subset of G, ϕ is a function defined on G, and a is a limit point of E in the usual topology, define

$$E - \limsup_{z \to a} \phi(z) = \limsup\{\phi(z) : z \in E, z \to a\}.$$

Similarly, define $E - \liminf$ and $E - \lim$. It is perhaps worthwhile to remind the reader here that the definition of the various limits does not use the value of the function at the limit point. Thus $E - \limsup_{z\to a}(z) = \lim_{r\to 0}[\sup\{\phi(z) : z \in E \text{ and } 0 < |z-a| < r\}]$ and $E - \lim_{z\to a}\phi(z) = E - \limsup_{z\to a}\phi(z) = E - \liminf_{z\to a}\phi(z)$ when these two last limits agree.

13.8 Theorem. *If E is any non-empty subset of $\mathbb{C}$ and a is a limit point of E, then following are equivalent.*

(a) *The set E is thin at a.*

(b) *For every $r > 0$, $E \cap B(a;r)$ is thin at a.*

(c) *There is a subharmonic function ϕ defined on $\mathbb{C}$ such that $\phi(a) > E - \limsup_{z\to a}\phi(z) > -\infty$.*

(d) *For any $r > 0$ there is a positive measure μ supported on $B(a;r)$ such that $L_\mu(a) < E - \liminf_{z\to a} L_\mu(z) < \infty$.*

(e) *There is a subharmonic function ϕ defined on $\mathbb{C}$ such that $\phi(a) > E - \lim_{z\to a}\phi(z) = -\infty$.*

(f) *For any $r > 0$ there is a positive measure μ supported on $B(a;r)$ such that $L_\mu(a) < E - \lim_{z\to a} L_\mu(z) = \infty$.*

Proof. First, let's agree that (a) and (b) are equivalent by virtue of basic topology and the fact that $B(a;r) \in \mathcal{F}$ for all $r > 0$. Also observe that it suffices to assume that $a \notin E$.

(a) *implies* (c). Assume that E is thin at a. So there is a set U in $\mathcal{F}$ such that $a \in U$ and $U \cap E = \emptyset$. From Proposition 13.2 there is a neighborhood W of a, subharmonic functions $\phi_1, \dots, \phi_n$ on $\mathbb{C}$, and finite constants $c_1, \dots, c_n$ such that

$$a \in W \cap \bigcap_{k=1}^{n} \{z : \phi_k(z) > c_k\} \subseteq U.$$

Choose $\varepsilon > 0$ such that $\phi_k(a) - \varepsilon > c_k$ for $1 \le k \le n$ and set $\phi = \phi_1 + \dots + \phi_n$. Since each ϕ_k is usc, there is an $r > 0$ such that $B = B(a;r) \subseteq W$ and $\phi_k(z) < \phi_k(a) + \varepsilon/n$ for all z in B. But a is a limit point of E so $B \cap E \neq \emptyset$; if $z \in B \cap E$, then $z \notin U$ and so there is a k, $1 \le k \le n$, with $\phi_k(z) \le c_k$.

Thus

$$
\begin{aligned}
\phi(z) &= \phi_k(z) + \sum_{j\neq k} \phi_j(z) \\
&< c_k + \sum_{j\neq k} [\phi_j(a) + \varepsilon/n] \\
&< \phi_k(a) - \varepsilon + \sum_{j\neq k} \phi_j(a) + \frac{n-1}{n}\varepsilon \\
&= \phi(a) - \varepsilon/n.
\end{aligned}
$$

Hence $E - \limsup_{z\to a} \phi(z) \leq \phi(a) - \varepsilon/n < \phi(a)$. If $-\infty < \alpha < \phi(a)$ and ϕ is replaced by $\max\{\phi, \alpha\}$, all the conditions in (c) are met.

(c) *implies* (d). Let ϕ be as in part (b) and put $D = B(a; r)$ with $r < 1/2$. By Theorem 19.5.6 there is a positive measure μ on D and a harmonic function h on D such that $\phi|D = h - L_\mu$. Since h is continuous at a,

$$
\begin{aligned}
-L_\mu(a) &= \phi(a) - h(a) \\
&> E - \limsup_{z\to a} \phi(z) - h(a) \\
&= E - \limsup_{z\to a} (-L_\mu(z)) \\
&> -\infty
\end{aligned}
$$

and (d) follows.

(d) *implies* (e). Let $D = B(a; 1/2)$; so $L_\mu \geq 0$ on D. Put $C = E - \limsup_{z\to a}[-L_\mu(z)] < -L_\mu(a)$. Let $\{\varepsilon_n\}$ be a sequence of positive numbers such that $\varepsilon_n < 1/2$ and $\varepsilon_n \to 0$ monotonically; let $D_n = B(a; \varepsilon_n)$ and $\mu_n = \mu|D_n$. Since $L_\mu(a)$ is finite, μ does not have an atom at a. By the Monotone Convergence Theorem, $L_{\mu_n}(a) \to 0$. Replacing $\{\varepsilon_n\}$ by a subsequence if necessary, we can assume that $\sum_n L_{\mu_n}(a) < \infty$. If $\phi = -\sum_n L_{\mu_n}$, then ϕ is subharmonic on $\mathbb{C}$ and $\phi(a) > -\infty$.

If $h_n = L_{\mu_n} - L_\mu = -L_{(\mu-\mu_n)}$, then h_n is harmonic on D_n; so $h_n(z) \to h_n(a)$ as $z \to a$. If $0 < \delta < 1/2$ and $E_\delta = \{z \in E : 0 < |z - a| < \delta\}$, then

$$
\sup\{-L_{\mu_n}(z) : z \in E_\delta\} \leq \sup\{-L_\mu(z) : z \in E_\delta\} + \sup\{-h_n(z) : z \in E_\delta\}.
$$

Thus

$$
\begin{aligned}
E - \limsup_{z\to a}\{-L_{\mu_n}(z)\} &\leq C - h_n(a) \\
&= C - L_{\mu_n}(a) + L_\mu(a) \\
&\leq C + L_\mu(a) \\
&< 0.
\end{aligned}
$$

So for any $m \geq 1$,

$$
E - \limsup_{z\to a} \sum_{n=1}^{m} [-L_{\mu_n}(z)] \leq m[C + L_\mu(a)].
$$

Therefore

$$\begin{aligned} E - \limsup_{z \to a} \phi(z) &= E - \limsup_{z \to a} \sum_{n=1}^{\infty} [-L_{\mu_n}(z)] \\ &\le E - \limsup_{z \to a} \sum_{n=1}^{m} [-L_{\mu_n}(z)] \\ &\le m[C + L_\mu(a)] \end{aligned}$$

for all $m \ge 1$. Hence $E - \lim_{z \to a} \phi(z) = -\infty < \phi(a)$.

(e) *implies* (f). This is like the proof that (c) implies (d).

(f) *implies* (a). Suppose μ is as in (f). If E were thick at a, then there would exist a net $\{z_i\}$ in E such that $z_i \to a$ in the fine topology. Since L_μ is $\mathcal{F}$-continuous, $L_\mu(z_i) \to L_\mu(a)$. But this is a contradiction since $L_\mu(a) < \infty$ and $L_\mu(z_i) \to \infty$. □

There are other statements equivalent to those in the preceding theorem that involve capacity; for example, see Meyers [1975].

13.9 Corollary. *If E is a Borel set and*

$$\lim_{r \to 0} \frac{|\partial B(a : r) \cap E|}{r} > 0,$$

where the absolute value signs denote arc length measure on the circle $\partial B(a;r)$, then E is thick at a.

Proof. Suppose E is thin at a. By the preceding theorem there is a subharmonic function ϕ defined on $\mathbb{C}$ such that $\phi(a) > E - \lim_{z \to a} \phi(z) = -\infty$. Since ϕ is usc, there is a constant M such that $M \ge \phi(z)$ for $|z - a| \le 1$. Replacing ϕ by $\phi - M$, it can be assumed that $\phi(z) \le 0$ for $|z - a| \le 1$. Thus for $r < 1$ and $\Delta_r = \{\theta : a + re^{i\theta} \in E\}$,

$$\begin{aligned} -\infty < \phi(a) &\le \frac{1}{2\pi} \int_0^{2\pi} \phi(a + re^{i\theta})\, d\theta \\ &\le \frac{1}{2\pi} \int_{\Delta_r} \phi(a + re^{i\theta})\, d\theta \\ &\le \frac{[\text{measure}(\Delta_r)]}{2\pi} \sup_{\theta \in \Delta_r} \phi(a + re^{i\theta}) \\ &\le \frac{|\partial B(a;r) \cap E|}{2\pi\, r} \sup_{z \in \partial B(a;r)} \phi(z). \end{aligned}$$

But $\sup\{\phi(z) : z \in \partial B(a;r)\} \to -\infty$ as $r \to 0$. Therefore $|\partial B(a;r) \cap E|/r \to 0$ as $r \to 0$, contradicting the assumption. □

If the set E is contained in a hyperbolic open set, then the equivalent formulation of thinness can be improved.

13.10 Proposition. *If G is a hyperbolic open set, $E \subseteq G$, and a is a limit point of E that belongs to G, then the following are equivalent.*

(a) *E is thin at a.*

(b) *There is a negative subharmonic function ϕ on G that is also bounded below and satisfies $\phi(a) > E - \limsup_{z\to a} \phi(z)$.*

(c) *There is a negative subharmonic function ϕ defined on G such that $\phi(a) > E - \lim_{z\to a} \phi(z) = -\infty$.*

Proof. (a) *implies* (b). Let $D = B(a; s)$ be chosen such that cl $D \subseteq G$. By Theorem 13.8 it can be assumed that $E \subseteq D$ and there is a subharmonic function ϕ_1 on $\mathbb{C}$ such that $\phi_1(a) > E - \limsup_{z\to a} \phi_1(z)$. Let $B_r = B(a;r)$ be chosen with $r > s$ such that cl $B_r \subseteq G$. By Proposition 12.4 there is a subharmonic function ϕ_2 on $\mathbb{C}$ with $\phi_2 = \phi_1$ on D and $\phi_2(z) = \log(|z-a|/r)$ for $|z-a| > r$.

Let g be the Green function for G and let h be the harmonic function on G such that $g(z,a) = h(z) - \log|z-a|$ for z in G. Define $\phi_3 : G \to [-\infty, \infty)$ by $\phi_3(z) = \phi_2(z) - h(z) + \log r$. So ϕ_3 is subharmonic on G. For z in $G \setminus$ cl B_r, $\phi_3(z) = -g(z,a) \leq 0$. Since ϕ_3 is usc, there is a constant M with $\phi_3 \leq M$ on cl B_r. Let $\phi_4 = \phi_3 - M$. So ϕ_4 is a negative subharmonic function on G and it is easy to check that since $\phi_2 = \phi_1$ on D, $E - \limsup_{z\to a} \phi_4(z) < \phi_4(a)$. Now let $\phi = \max\{\phi_4, \phi_4(a) - 1\}$ and the proof is complete.

(b) *implies* (c). The proof of this is similar to the corresponding proof in Theorem 13.8 and will not be given.

(c) *implies* (a). Since subharmonic functions are finely continuous, a cannot be a fine limit point of E. □

13.11 Definition. If G is any set and u is an extended real-valued function defined on G, say that u *peaks* at a point a in G if for every neighborhood V of a, $u(a) > \sup\{u(z) : z \in G \setminus V\}$.

Note that if u peaks at a, then u attains its maximum value on G at the point a. This alone, however, will not guarantee that u peaks at a. The object here will be to find superharmonic functions on an open set G that peak at certain points. (Actually, sticking with the fixation of this book on subharmonic functions, we will be concerned with finding subharmonic functions u such that $-u$ peaks at a.) If G is a hyperbolic open set and $a \in G$, then the Green function g is a function such that g_a peaks at a, $g_a \geq 0$, and g_a is superharmonic. The next example will be used later.

13.12 Example. If $R > 0$ and μ is the restriction of area measure to $B(a; R)$, then L_μ peaks at a. In fact, according to Exercise 3.6, $L_\mu(z) = \pi R^2 \log|z-a|^{-1}$ if $|z-a| \geq R$ and $L_\mu(z) = \pi R^2[\log R^{-1} + 1/2] - \pi|z -$

$a|^2/2 = L_\mu(a) - \pi|z-a|^2/2$ for $|z-a| \le R$. It is left to the reader to show that $\sup\{L_\mu(z) : |z-a| > r\} < L_\mu(a)$ for any $r > 0$.

The first task is to relate peaking with thinness.

13.13 Proposition. *Let G be a hyperbolic open set, let $a \in G$, and suppose u is a negative subharmonic function on G such that $-u$ peaks at a. If E is a subset of G, then E is thin at a if and only if $\hat{I}^u_E(a) > u(a)$.*

Proof. Since $\hat{I}^u_E = \hat{I}^u_{E\setminus\{a\}}$ (12.11), we may assume, without loss of generality, that $a \notin E$. But we also know that $\hat{I}^u_E = I^u_E$ on $G \setminus E$ (12.10) so that, in particular, $\hat{I}^u_E(a) = I^u_E(a)$. So we want to show that E is thin at a if and only if $I^u_E(a) > u(a)$.

Assume that $I^u_E(a) > u(a)$. To show that E is thin at a, it is assumed that a is a limit point of E. From the definition of I^u_E, there is a subharmonic function ϕ on G such that $\phi \le u$ on E and $\phi(a) > u(a)$. Thus $E - \limsup_{z\to a} \phi(z) \le E - \limsup_{z\to a} u(z) \le u(a) < \phi(a)$. By Theorem 13.8, E is thin at a.

Now assume that E is thin at a. If a is not a limit point of E, then there is a neighborhood V of a such that $V \cap E = \emptyset$ and cl $V \subseteq G$. Since $-u$ peaks at a, $u(a) < \inf\{u(z) : z \in G\setminus V\} \le \inf\{I^u_E(z) : z \in G\setminus V\}$. But I^u_E is harmonic on V (12.1) and so the Maximum Principle implies $I^u_E(z) > u(a)$ for all z in V; in particular, this holds for $z = a$. So it can be assumed that a is a limit point of E and E is thin at a.

By Proposition 13.10 there is a negative subharmonic function ϕ on G that is bounded below such that $\phi(a) > E - \limsup_{z\to a} \phi(z)$. Choose a constant M such that $\phi(a) > M > E - \limsup_{z\to a} \phi(z)$ and let D be an open disk about a such that $M > \phi(z)$ for all z in $D \cap E$. For $t > 0$ define the function $w_t = u(a) + t[\phi - M]$. Thus for z in $D \cap E$,

$$w_t(z) < u(a).$$

Now u peaks at a so there is a constant $C > 0$ with $u(z) - u(a) \ge C$ for all z in $G \setminus D$. Because ϕ is a bounded function, the parameter t can be chosen so small that $t|\phi(z) - M| \le C$ for all z in G. Thus for z in $G \setminus D$, $t[\phi(z) - M] \le C \le u(z) - u(a)$. Therefore $w_t(z) = u(a) + t[\phi(z) - M] \le u(z)$ for all z in $G \setminus D$. On the other hand, for z in $D \cap E$, $w_t(z) < u(a) \le u(z)$ since $-u$ peaks at a. Thus $w_t \le u$ on E and so $w_t \le I^u_E$. But $w_t(a) = u(a) + t[\phi(a) - M] > u(a)$, proving the proposition. $\square$

The next result is often called the Fundamental Theorem for thinness. (It is probably best here to avoid the variety of possible minor jokes that such a title affords.)

13.14 Theorem. *If E is any set and $Z = \{a \in E : E$ is thin at $E\}$, then Z is a polar set.*

Proof. Cover E by a sequence of open disks $\{D_n\}$ each having radius less than 1/2 and let $Z_n = Z \cap D_n$. Clearly $Z = \cup_n Z_n$ and if each Z_n is shown to be a polar set, then so is Z. Thus it can be assumed that $E \subseteq D$, an open disk of radius $< 1/2$. Let μ be the restriction of area measure to D and put $u = -L_\mu$. So u is a negative subharmonic function on D. If $a \in Z$, let $r > 0$ such that $\overline{B}(a;r) \subseteq D$, put $\mu_1 = \mu|B(a;r)$ and $\mu_2 = \mu|[D \setminus B(a;r)]$, and let $u_j = -L_{\mu_j}$, $j = 1, 2$. Clearly $u = u_1 + u_2$ and both u_1 and u_2 are negative subharmonic functions on D. By (13.12) $-u_1$ peaks at a and so by the preceding proposition $\hat{I}_E^{u_1}(a) > u(a)$. Thus $\hat{I}_E^{u}(a) \geq \hat{I}_E^{u_1}(a) + \hat{I}_E^{u_2}(a) > u(a) = I_E^u(a)$, since $a \in E$. Thus $Z \subseteq \{z : \hat{I}_E^u(a) > I_E^u(a)\}$ and this is a polar set by Corollary 12.10. □

13.15 Corollary. *A set Z is polar if and only if it is thin at each of its points.*

Proof. The proof of one direction of the corollary is immediate from the theorem. The proof of the other direction was already done in (13.4). □

In Theorem 13.14 let us emphasize that the polar set Z is contained in E. Clearly the set of all points at which E is thin is not polar since it is thin at each point of $\mathbb{C} \setminus \text{cl } E$. But not even the set of points in cl E at which E is thin is necessarily a polar set. For example, if E is the set of all points in $\mathbb{C}$ with rational real and imaginary points, then E is countable and hence polar and hence thin at every point.

13.16 Theorem. *If G is a hyperbolic open set and $a \in \partial_\infty G$, then a is a regular point of G if and only if $\mathbb{C}_\infty \setminus G$ is thick at a.*

Proof. Put $E = \mathbb{C}_\infty \setminus G$. If E is thick at a, the argument given just after the definition of thickness shows that a is a regular point for G.

For the converse, assume that a is a regular point of G. By (19.10.1) there is a non-constant negative subharmonic function u on $D \setminus E = G(a;r)$ such that $(D \setminus E) - \lim_{z \to a} u(z) = 0$. Also assume that E is thin at a and let's work toward a contradiction. According to Proposition 13.10 there is a bounded subharmonic function ϕ on D such that $\phi(a) > E - \limsup_{z \to a} \phi(z)$. By changing scale it may be assumed that $\phi(a) = 1$ and $E - \limsup_{z \to a} \phi(z) < -1$. Choose δ, $0 < \delta < 1$, so that $D_1 = B(a;\delta)$ satisfies cl $D_1 \subseteq D$ and $\phi(z) \leq -1$ for all z in $(\text{cl } D_1) \cap (E \setminus \{a\})$. The contradiction to the assumption that E is thin at a will be obtained by showing that there is a positive scalar t such that $\phi \leq -t\, u$ on $D_1 \setminus E$. Indeed once this is shown it follows that $(D_1 \setminus E) - \limsup_{z \to a} \phi(z) \leq (D_1 \setminus E) - \limsup_{z \to a} [-t\, u(z)] = 0$. But $E - \limsup_{z \to a} \phi(z) < -1$, so we get that $1 = \phi(a) = \limsup_{z \to a} \phi(z) \leq 0$, a contradiction.

It will be shown that there is a positive scalar t such that for any $\varepsilon > 0$

13.17 $$(D_1 \setminus E) - \limsup_{\zeta \to z} [\phi(\zeta) + t\, u(\zeta) + \varepsilon \log|\zeta - \alpha|] \leq 0$$

for all z in $\partial(D_1 \setminus E) = (\partial D_1 \setminus E) \cup (\partial E \cap \text{cl } D_1)$. If this is shown to hold, then the Maximum Principle implies that, for every $\varepsilon > 0$, $\phi + t u + \varepsilon \log|z-a| \leq 0$ on $D_1 \setminus E$. Letting $\varepsilon \to 0$ shows that $\phi \leq -t u$, as desired.

From the properties of ϕ there is an open set W containing $\partial D_1 \cap E$ such that $\phi < 0$ on W. Thus no matter how the positive number t is chosen, (13.17) holds for all z in $W \cap (\partial D_1 \setminus E)$. Also ϕ is bounded and u is negative and usc, so $t > 0$ can be chosen with $\phi + t u < 0$ on $\partial D_1 \setminus W$; fix this choice of t. Combining this with the preceding observation shows that (13.17) holds for all z in $\partial D_1 \setminus E$. If $z \in \partial E \cap \text{cl } D_1$ and $z \neq a$, then $\phi(z) \leq -1$ and so $\phi(\zeta) < 0$ in a neighborhood of z. Thus $(D_1 \setminus E) - \limsup_{\zeta \to z}[\phi(\zeta) + t\, u(\zeta) + \varepsilon \log|\zeta - a|] \leq \varepsilon \log|z-a| \leq \varepsilon \log \delta < 0$, so (13.17) holds for z in $\partial E \cap \text{cl } D_1$ and $z \neq a$. If $z = a$, then (13.17) also holds, since the left hand side is $-\infty$. □

Remarks. The fine topology is fully studied in the literature. In addition to Brelot [1959] and Helms [1975], cited in the introduction of this section, the following sources are useful: Hedberg [1972b], Hedberg [1993], and Landkof [1972].

Exercises

1. Suppose the set E is thin at a and $f : \mathbb{C} \to \mathbb{C}$ is a mapping such that $|f(z) - f(w)| \leq |z - w|$ and $|f(a) - f(z)| = |a - z|$ for all z, w in $\mathbb{C}$. Show that $f(E)$ is thin at a.

2. Is there a set E such that both E and $\mathbb{C} \setminus E$ are thin at a?

§14 Wiener's Criterion for Regular Points

In this section Wiener's Criterion for the regularity of a boundary point will be stated and proved. This statement is in terms of logarithmic capacity.

14.1 Definition. For any subset E in $\mathbb{C}$, define the *outer capacity* of E as the number

$$c^*(E) = \inf\{c(U) : U \text{ is open and contains } E\}.$$

It is clear that

$$c^*(E) = \exp(-v^*(E)),$$

where

$$v^*(E) = \sup\{v(U) : U \text{ is open and contains } E\}.$$

It is a standard fact about capacities that $c^*(E) = c(E)$ for a collection of sets that includes the Borel sets. This is not proved in this book, but we have seen that this is the case when E is compact (Corollary 7.16). The reader interested in this fact can see Carleson [1967].

14.2 Wiener's Criterion. *Let E be a set, $a \in \mathbb{C}$, and $0 < l < 1/2$; for each $n \geq 1$, let $E_n = E \cap \{z : \lambda^{n+1} \leq |z-a| \leq \lambda^n\}$. The set E is thick at a if and only if*

14.3
$$\sum_{n=1}^{\infty} \frac{n}{\log[c^*(E_n)^{-1}]} = \infty.$$

First a few observations and corollaries. Note that the condition (14.3) is equivalent to

14.4
$$\sum_{n=1}^{\infty} \frac{n}{v^*(E_n)} = \infty.$$

The reason for the restriction $\lambda < 1/2$ is to guarantee that diam $E_n < 1$ for all $n \geq 1$ and thus that $c^*(E_n) < 1$ for all n. Thus the series does not diverge by virtue of having one of its terms equal to ∞. This also assures that each term in the series is positive.

To dispose of the trivialities, observe that a is not a limit point if and only if $E_n = \emptyset$ for all n larger than some n_0. In this case $c^*(E_n) = 0$ for $n \geq n_0$ and the series in (14.3) is convergent. The same holds of course if E_n is polar for all large n.

Originally Wiener stated his criterion as a necessary and sufficient condition for a point to be a regular point for the solution of the Dirichlet problem. The following corollary is immediate from Wiener's Criterion and Theorem 13.16. All the corollaries below will be stated for finite boundary points of the hyperbolic open set G. The condition for the regularity of the point at ∞ is obtained by inversion.

14.5 Corollary. *If G is a hyperbolic open set, $a \in \partial G$, $0 < l < 1/2$, and $K_n = \{z \in \mathbb{C} \setminus G : \lambda^{n+1} \leq |z-a| \leq \lambda^n\}$, then a is a regular point if and only if*

$$\sum_{n=1}^{\infty} \frac{n}{\log[c(K_n)^{-1}]} = \infty.$$

14.6 Corollary. *With the notation of the preceding corollary, if μ_n is the equilibrium measure for K_n, then a is a regular point of G if and only if*

$$\sum_{n=1}^{\infty} \frac{L_{\mu_n}(a)}{\log[c(K_n)^{-1}]} = \infty.$$

Proof. Put $v_n = \log[c(K_n)^{-1}]$. For z in K_n, $\lambda^{n+1} \leq |z-a| \leq \lambda^n$ and so $n \leq [\log \lambda^{-1}]^{-1} \log |z-a|^{-1} \leq n+1$. Hence $n \leq [\log \lambda^{-1}]^{-1} L_{\mu_n}(a) \leq n+1$

and so the two series $\sum_n n/v_n$ and $\sum_n L_{\mu_n}(a)/v_n$ simultaneously converge or diverge. □

Now observe that the function $x \log x^{-1}$ is increasing for $0 < x < e^{-1}$. Adopting the notation of the preceding corollary, $K_n \subseteq A_n \equiv \{z : \lambda^{n+1} \leq |z-a| \leq \lambda^n\}$ and so $c(K_n) \leq c(A_n) = \lambda^n$ (10.6). So if $c_n = c(K_n)$, we have that $c_n \log c_n^{-1} \leq \lambda^n \log \lambda^{-n} = n\lambda^n \log \lambda^{-1}$. Thus the following corollary follows from Corollary 14.5.

14.7 Corollary. *With the notation of Corollary* 14.5, *if* $\sum_n c(K_n)/\lambda^n = \infty$, *then* a *is a regular point of* G.

14.8 Corollary. *If* G *is a hyperbolic open set,* $a \in \partial G$, *and, for* $r > 0$, $c(r) \equiv c((\mathbb{C} \setminus G) \cap \overline{B}(a;r))$ *and*

$$\limsup_{r \to 0} \frac{\log r^{-1}}{\log c(r)^{-1}} > 0,$$

then a *is a regular point of* G.

Proof. Assume that $a = 0$ and put $E_r = (\mathbb{C} \setminus G) \cap \overline{B}(a;r)$ and $v(r) = \log c(r)^{-1}$. Fix λ, $0 < \lambda < 1/2$, and let $\{K_n\}$ and $\{\mu_n\}$ be as in Corollary 14.6 above. Put $v_n = \log c(K_n)^{-1}$; so $L_{\mu_n} = v_n^{-1}$ q.e. on K_n.

Suppose a is not a regular point of G and let $\varepsilon > 0$; it will be shown that the lim sup in this corollary is less than this arbitrary ε. According to Corollary 14.6, $\sum_n L_{\mu_n}(a)/v_n < \infty$. Thus there is an integer N such that

$$\sum_{n=m}^{\infty} \frac{L_{\mu_n}(0)}{v_n} < \frac{\varepsilon}{2}$$

for all $m \geq N$. Fix $r < \lambda^N$ and let m be such that $\lambda^{m+1} < r \leq \lambda^m$. Let η be the positive measure supported on E_r such that $L_\eta = 1$ q.e. on E_r and $\eta(E_r) = v(r)^{-1}$. For each n let $\psi_n(w) = \int_{K_n} \log |z-w|^{-1} d\eta(z)$ and $\psi = \sum_{n=m}^{\infty} \psi_n$. Note that with $B_n = (\mathbb{C} \setminus G) \cap \{z : \lambda^{n+1} \leq |z| < \lambda^n\}$

$$\begin{aligned}
\psi(0) &= \sum_{n=m}^{\infty} \int_{K_n} \log |z|^{-1} d\eta(z) \\
&= \int_{K_m} \log |z|^{-1} d\eta(z) + \sum_{n=m+1}^{\infty} \int_{B_n} \log |z|^{-1} d\eta(z) + \\
&\quad + \sum_{n=m+1}^{\infty} \int_{K_n \setminus B_n} \log |z|^{-1} d\eta(z) \\
&= L_\eta(0) + \sum_{n=m+1}^{\infty} \int_{K_n \setminus B_n} \log |z|^{-1} d\eta(z).
\end{aligned}$$

So $L_\eta(0) \le \psi(0) \le 2L_\eta(0)$.

Now $L_{\mu_n} = v_n$ q.e. on K_n and hence $\eta(K_n) = v_n^{-1} \int_{K_n} L_{\mu_n} d\eta = v_n^{-1} \int_{K_n} \psi_n d\mu_n$. But $L_\eta = 1$ q.e. on E_r and $\log|z-w|^{-1} \ge 0$ on E_r, so for μ_n almost all z in K_n, $1 = L_\eta(z) = \int_{E_r} \log|z-w|^{-1} d\eta(w) \ge \psi_n(z)$. Therefore

$$\eta(K_n) \le v_n^{-1}$$

for all n and so

$$\begin{aligned} \psi_n(0) &= \int_{K_n} \log|z|^{-1} d\eta(z) \\ &\le (n+1)(\log\lambda^{-1})\eta(K_n) \\ &\le v_n^{-1} \frac{n+1}{n} n \log\lambda^{-1} \\ &\le 2v_n^{-1} L_{\mu_n}(0). \end{aligned}$$

Combining this with previous inequalities we get

$$\begin{aligned} L_\eta(0) &\le \psi(0) \\ &= \sum_{n=m}^{\infty} \psi_n(0) \\ &\le 2\sum_{n=m}^{\infty} \frac{L_{\mu_n}(0)}{v_n} \\ &< \varepsilon. \end{aligned}$$

But basic inequalities show that $L_\eta(0) \ge v(r)^{-1} \log r^{-1}$ and so we have that $v(r)^{-1} \log r^{-1} \to 0$ as $r \to 0$. $\square$

14.9 Corollary. *With the notation of Corollary* 14.5, *if* $\kappa_n = \text{Area}(K_n)$ *and* $\sum_n \kappa_n/\lambda^{2n} = \infty$, *then a is a regular point.*

Proof. According to Proposition 7.8, $c_n \equiv c(K_n) \ge [\kappa_n/\pi e]^{1/2}$. Since $K_n \subseteq B(a;\lambda^n)$ we also have that $\kappa_n \le \pi\lambda^{2n}$. Hence

$$\frac{\kappa_n}{\lambda^{2n}} = \frac{\sqrt{\kappa_n}}{\lambda^n}\frac{\sqrt{\kappa_n}}{\lambda^n} \le \sqrt{\pi}\sqrt{\pi e}\,\frac{c_n}{\lambda^n}.$$

Thus a is a regular point by Corollary 14.7. $\square$

Now to begin the proof of Wiener's Criterion.

14.10 Lemma. *If U is a bounded open set, then there is a positive measure μ supported on ∂U such that $\mu(\partial U) = 1/v(U)$ and $L_\mu = 1$ on U.*

Proof. Write U as the union of a sequence of compact sets $\{K_n\}$ with $K_n \subseteq \text{int} K_{n+1}$. So $c(K_n) \uparrow c(U)$ and $v_n = v(K_n) \downarrow v(U) = v$. Let ν_n be the equilibrium measure for K_n and put $\mu_n = v_n^{-1}\nu_n$. Now $||\mu_n|| = \mu_n(U) = v_n^{-1} \le v^{-1} < \infty$. So $\{\mu_n\}$ is a uniformly bounded sequence of positive measures in $M(\text{cl } U)$; by passing to a subsequence if needed, it can be assumed that there is a positive measure μ on cl U such that $\mu_n \to \mu$ weak*. Because supp $\mu_n \subseteq \partial K_n$, supp $\mu \subseteq \partial U$. Also $v_n^{-1} = \mu_n(\text{cl } U) = \int 1 d\mu_n \to \mu(\text{cl } U)$; thus $\mu(\partial U) = v^{-1}$. Finally, fix a z in U and pick m with z in int K_m. By (10.1) $L_{\mu_n}(z) = 1$ for all $n \ge m$. But it is easy to use Tiezte's Extension Theorem to show that $\mu_n \to \mu$ weak* in $M(\text{cl } U \setminus \text{int } K_m)$. Since $w \to \log|z-w|^{-1}$ is continuous there, $L_{\mu_n}(z) \to L_\mu(z)$. □

14.11 Lemma. *If* $0 < \lambda < 1/2$, *there is a constant* $C = C(\lambda)$ *such that for* $n \ge 1$, *if* $\lambda^{n+1} \le |z| \le \lambda^n$ *and* w *satisfies either* $|w| \le \lambda^{n+2}$ *or* $\lambda^{n-1} \le |w| \le 1/2$, *then* $\log|z-w|^{-1} \le C \log|w|^{-1}$.

Proof. Put $L = \lambda^{-1}$ and draw circles centered at 0 with radii λ^{n+2}, λ^{n+1}, λ^n, and λ^{n-1}; fix z with $\lambda^{n+1} \le |z| \le \lambda^n$. First, suppose $|w| \le \lambda^{n+2}$ and examine the picture to see that $|w-z| \ge \lambda^{n+1} - \lambda^{n+2} = \lambda^{n+2}(L-1)$. Thus $|w-z|^{-1} \le L^{n+2}(L-1)^{-1} \le |w|^{-1}(L-1)^{-1}$. Thus

$$\begin{aligned}\frac{\log|w-z|^{-1}}{\log|w|^{-1}} &\le 1 + \frac{\log(L-1)^{-1}}{\log|w|^{-1}}\\ &\le 1 + \frac{\log(L-1)^{-1}}{(n+2)\log L}\\ &\le 1 + \frac{\log(L-1)^{-1}}{2\log L}.\end{aligned}$$

Now suppose $\lambda^{n-1} \le |w| \le 1/2$. Another examination of the picture shows that $|w-z| \ge \lambda^{n-1} - \lambda^n = \lambda^{n-1}(1-\lambda)$ and so $|w-z|^{-1} \le L^{n-1}(1-\lambda)^{-1} \le |w|^{-1}(1-\lambda)^{-1}$. Thus

$$\begin{aligned}\frac{\log|w-z|^{-1}}{\log|w|^{-1}} &\le 1 + \frac{\log(1-\lambda)^{-1}}{\log|w|^{-1}}\\ &\le 1 + \frac{\log(1-\lambda)^{-1}}{\log 2}.\end{aligned}$$

□

Proof of Wiener's Criterion. By translation it can be assumed that the point $a = 0$. Also recall (13.8) that a set E is thick at 0 if and only if for every $r > 0$, $E \cap r\mathbb{D}$ is thick at 0. Thus in this proof it suffices to assume that $E \subseteq \frac{1}{2}\mathbb{D}$. Thus $\log|z-w|^{-1} > 0$ for all z, w in E. As pointed out in the discussion following the statement of (14.2), it can be assumed that $v^*(E_n) > 0$ for all n. We will prove the contrapositive of (14.2).

Assume that the series (14.3) converges and choose an arbitrary sequence of positive numbers $\{\varepsilon_n\}$ such that $\sum_n n\varepsilon_n < \infty$. Now for each n choose an open set U_n that contains E_n such that $\operatorname{cl} U_n \cap \{z : |z| \le \lambda^{n+2}\} = \emptyset$ and

$$\frac{1}{v^*(E_n)} \le \frac{1}{v(U_n)} < \frac{1}{v^*(E_n)} + \varepsilon_n.$$

By Lemma 14.10 for each n there is a positive measure μ_n supported on ∂U_n with $\mu_n(\partial U_n) = v(U_n)^{-1}$ and $L_{\mu_n}(z) = 1$ for all z in U_n. For each n define the subharmonic function $\phi_n = -\sum_{k=n}^{\infty} L_{\mu_k}$. Note that ϕ_n is negative on $\frac{1}{2}\mathbb{D}$.

Note that for any $k \ge 1$ and z in ∂U_k, $|z| \ge \lambda^{k+2}$ and so

$$\begin{aligned} L_{\mu_k}(0) &= \int_{\partial U_k} \log |z|^{-1} d\mu_k(z) \\ &\le (\log \lambda^{-(k+2)}) \mu_k(\partial U_k) \\ &\le (k+2) \log \lambda^{-1} \left[\frac{1}{v^*(E_k)} + \varepsilon_k \right]. \end{aligned}$$

Since both $\sum_k k/v^*(E_k)$ and $\sum_k k\varepsilon_k$ converge, this implies there is an n such that $\phi_n(0) > -1$; fix this value of n. If $k \ge n$ and $z \in E_k$, which is contained in U_k, then $L_{\mu_k}(z) = 1$. Since each L_{μ_k} is non-negative on E, this shows that $\phi_n \le -1$ on

$$\bigcup_{k=n}^{\infty} E_k = E \cap \{z : 0 < |z| \le \lambda^n\}.$$

Hence $E - \limsup_{z\to 0} \phi_n(z) \le -1 < \phi_n(0)$. By (13.10), E is thin at 0.

Now suppose that E is thin at 0; it can be assumed that 0 is a limit point of E. So (13.8) there is a positive measure μ with compact support contained in $\frac{1}{2}\mathbb{D}$ such that $L_\mu(0) < E - \lim_{z\to 0} L_\mu(z) = +\infty$. Put $\alpha_n = \inf\{L_\mu(z) : z \in E_n\}$; so $\alpha_n \to \infty$. Let δ be an arbitrary positive number; since L_μ is lsc, $\{z : L_\mu(z) > \alpha_n - \delta\}$ is an open set containing E_n. Put $U_n = \{z : L_\mu(z) > \alpha_n - \delta\} \cap \{z : \lambda^{n+1} \le |z| \le \lambda^n\}$. So U_n is a Borel set that contains E_n; it suffices to show that $\sum_n n/v(U_n) < \infty$.

Let $\{\varepsilon_n\}$ be a sequence of positive numbers such that $\sum_n n\varepsilon_n < \infty$ and for each n choose a compact subset K_n of U_n such that $v(K_n)^{-1} > v(U_n)^{-1} - \varepsilon_n$. It is therefore sufficient to show that

$$\sum_{n=1}^{\infty} \frac{n}{v(K_n)} < \infty.$$

Breaking this sum into 6 separate sums, we see that it suffices to show that

$$\sum_{n=1}^{\infty} \frac{6n+j}{v(K_{6n+j})} < \infty$$

for $j = 0, 1, \ldots, 5$. This will be shown for $j = 0$. (By only considering the series consisting of every sixth term, this will separate the sets K_{6n+j}

sufficiently that estimates such as in Lemma 14.11 can be applied.) Note that $K_{6n} \subseteq U_{6n} \subseteq A_n \equiv \{z : \lambda^{6n+2} < |z| < \lambda^{6n-1}\}$.

Claim. $K_{6k} \cap A_n = \emptyset$ for $k \neq n$.

To see this, first assume that $k > n$; so $k = n + p$, $p \geq 1$. Thus $\lambda^{6k} = \lambda^{6n+6p} < \lambda^{6n+2}$ and so $K_{6k} \cap A_n = \emptyset$. Now assume that $k < n$; so $k = n - p$, $p \geq 1$. Thus $\lambda^{6k+1} = \lambda^{6n-6p+1} > \lambda^{6n-1}$. This proves the claim. In particular, the sets $\{K_{6n}\}$ are pairwise disjoint.

Now $K = \{0\} \cup \bigcup_n K_{6n}$ is a compact subset of $\frac{1}{2}\mathbb{D}$ and $u = -L_\mu$ is a negative subharmonic function there. If $\phi = \hat{I}^u_K$, then Proposition 12.2 says there is a positive measure ν supported on K such that $\phi = -G_\nu = -L_\nu + h$ for some harmonic function h on $\frac{1}{2}\mathbb{D}$. Because $-\infty < u(0) \leq \phi(0)$ and h is harmonic, $L_\nu(0) < \infty$. Thus $\nu(\{0\}) = 0$. Write

$$L_\nu(z) = \int_{K_{6n}} \log|z-w|^{-1} d\nu(w) + \int_{\bigcup_{k\neq n} K_{6k}} \log|z-w|^{-1} d\nu(w).$$

According to Lemma 14.11 there is a constant C depending only on λ such that for z in K_{6n} and w in $\bigcup_{k\neq n} K_{6k}$, $\log|w-z|^{-1} \leq C\log|w|^{-1}$. Thus for all z in K_{6n},

$$\int_{\bigcup_{k\neq n} K_{6k}} \log|z-w|^{-1} d\nu(w) \leq C||\nu||.$$

Thus

14.12 $$L_\nu(z) \leq C||\nu|| + \int_{K_{6n}} \log|z-w|^{-1} d\nu(w).$$

By (12.10) there is a polar set Z contained in K such that $-L_\nu + h = \phi = u = -L_\mu$ on $K \setminus Z$. Let $|h| \leq M$ on K. Therefore on $K_n \setminus Z$, $L_\nu = -\phi + h = L_\mu + h \geq \alpha_n - \delta - M$ and so $(K \setminus Z) - \lim_{z\to 0} L_\nu(z) = \infty$. So if C_1 is any number with $C_1 - C||\nu|| \geq c > 0$, there is an integer m such that $L_\nu(z) \geq C_1$ on $[\cup_{n\geq m} K_{6n}] \setminus Z$.

Hence if $n \geq m$ and $z \in K_{6n}$, (14.12) implies

$$\int_{K_{6n}} \log|z-w|^{-1} d\nu(w) \geq L_\nu(z) - C||\nu|| \geq c.$$

Thus $v(K_{6n})^{-1} \leq c^{-1}\nu(K_{6n})$ by Theorem 10.2. But for z in K_{6n}, $\lambda^{6n+1} \leq |z| \leq \lambda^{6n}$, so that $6n\log\lambda^{-1} \leq \log|z|^{-1} \leq (6n+1)\log\lambda^{-1}$. Therefore

$$\begin{aligned}\sum_{n=1}^{\infty} \frac{6n}{v(K_{6n})} &\leq c^{-1}\sum_{n=1}^{\infty} 6n\,\nu(K_{6n}) \\ &\leq \frac{1}{c\log\lambda^{-1}} \sum_{n=1}^{\infty} \int_{K_{6n}} \log|z|^{-1} d\nu(z)\end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{c\log\lambda^{-1}}\,L_\nu(0) \\
&< \infty.
\end{aligned}
$$

□

References*

Abikoff, W. [1981], *The uniformization theorem*, Amer Math Monthly **88**, 574-592. (123, 129)

Adams, R.A. [1975], *Sobolev Spaces*, Academic Press, New York. (259)

Aharonov, D. [1984], *The De Branges Theorem on Univalent Functions*, Technion, Haifa. (156)

Ahlfors, L.V. [1973], *Conformal Invariants*, McGraw-Hill, New York. (123)

Ahlfors, L.V., and A. Beurling [1950], *Conformal invariants and function theoretic null sets*, Acta Math **83**, 101-129. (193)

Aleman, A., S. Richter, and W.T. Ross, *Bergman spaces on disconnected domains*, (preprint). (263, 351)

Askey, R., and G. Gasper [1976], *Positive Jacobi polynomial sums*, Amer J Math **98**, 709-737. (156, 157)

Axler, S. [1986], *Harmonic functions from a complex analysis viewpoint*, Amer Math Monthly **93**, 246-258. (74)

Axler, S., J.B. Conway, and G. McDonald [1982], *Toeplitz operators on Bergman spaces*, Canadian Math J **34**, 466-483. (351)

Baernstein, A., et al, *The Bieberbach Conjecture*, Amer Math Soc, Providence. (133)

Bagby, T. [1972], *Quasi topologies and rational approximation*, J Funct Analysis **10**, 259-268. (173, 263)

Bell, S.R., and S.G. Krantz [1987], *Smoothness to the boundary of conformal maps*, Rocky Mountain J Math **17**, 23-40. (55)

Bergman, S. [1947], *Sur les fonctions orthogonales de plusiers variables complexes avec les applications à la theorie des fonctions analytiques*, Gauthi Villars, Paris. (169)

Bergman, S. [1950], *The kernel function and conformal mapping*, Math Surveys V, Amer Math Soc, Providence. (169)

Bers, L. [1965], *An approximation theorem*, J Analyse Math **14**, 1-4. (173)

Beurling, A. [1939], *Ensembles exceptionnels*, Acta Math **72**, 1-13. (334)

*The numbers following each reference indicate the page on which the reference is cited.

Beurling, A. [1949], *On two problems concerning linear transformations in Hilbert space*, Acta Math **81**, 239-255. (290)

de Branges, L. [1985], *A proof of the Bieberbach conjecture*, Acta Math **154**, 137-152. (64, 132)

Brelot, M. [1959], *Éléments de la theorie classique du potential*, Centre de Documentation Universitaire, Paris. (234, 301, 364, 367, 376)

Brennan, J.E. [1977], *Approximation in the mean by polynomials on non-Carathédory domains*, Ark Math **15**, 117-168. (173)

Carleson, L. [1962], *Interpolation by bounded analytic functions and the corona problem*, Ann Math (2) **76**, 547-559. (295)

Carleson, L. [1966], *On the convergence and growth of partial sums of Fourier series*, Acta Math **116**, 135-157. (199)

Carleson, L. [1967], *Selected Problems on Exceptional Sets*, Van Nostrand, Princeton. (234, 301, 333, 334, 344, 351, 377)

Choquet, G. [1955], *Theory of capacities*, Ann Inst Fourier **5**, 131-295. (234, 333)

Cima, J., and Matheson [1985], *Approximation in the mean by polynomials*, Rocky Mountain J Math **15**, 729-738. (173)

Collingwood, E.F., and A.J. Lohwater [1966], *The Theory of Cluster Sets*, Cambridge University Press, Cambridge. (44)

Conway, J.B. [1986], *Functions of One Complex Variable*, Springer-Verlag, New York.

Conway, J.B. [1973], *A complete Boolean algebra of subspaces which is not reflexive*, Bull Amer Math Soc **79**, 720-722. (295)

Conway, J.B. [1990], *A Course in Functional Analysis*, Springer-Verlag, New York. (169, 174, 185)

Conway, J.B. [1991], *The Theory of Subnormal Operators*, Amer Math Soc Surveys **36**, Providence. (197, 290, 293, 333, 354)

Duren, P.L. [1970], *Theory of H^p Spaces*, Academic Press, New York. (269, 291)

Evans, L.C., and R.F. Gariepy [1992], *Measure Theory and Fine Properties of Functions*, CRC Press, Boca Raton. (259)

Farrell, O.J. [1934], *On approximation to an analytic function by polynomials*, Bull Amer Math Soc **40**, 908-914. (172)

Fisher, S.D. [1983], *Function Theory on Planar Domains*, Wiley, New York. (129)

Fisher, Y., J.H. Hubbard, and B.S. Wittner [1988], *A proof of the uniformization theorem for arbitrary plane domains*, Proc Amer Math Soc **104**, 413-418. (129)

Fitzgerald, C.H., and C. Pommerenke [1985], *The de Branges Theorem on univalent functions*, Trans Amer Math Soc **290**, 683-690. (133, 160)

Ford, L.R. [1972], *Automorphic Functions*, Chelsea Publ. Co., New York. (123)

Frostman, O. [1935], *Potential d' équilibre et capacité des ensembles*, Lund Univerisity Sem **3**. (234, 336)

Gamelin, T.W. [1969], *Uniform Algebras*, Prentice Hall, Englewood Cliffs, NJ. (197)

Gamelin, T.W., and D. Khavinson [1989], *The isoperimetric inequality and rational approximation*, Amer Math Monthly **96**, 18-30. (193)

Garnett, J. [1981], *Bounded Analytic Functions*, Academic Press, New York. (293)

Goluzin, G.M. [1969], *Geometric Theory of Functions of a Complex Variable*, Amer Math Soc Translations **26**, Providence. (90, 134, 359)

Hayman, W.K., and P.B. Kennedy [1976], *Subharmonic Functions*, Academic Press, New York. (234)

Hedberg, L.I. [1972a], *Approximation in the mean by analytic functions*, Trans Amer Math Soc **163**, 157-171. (171, 173, 351)

Hedberg, L.I. [1972b], *Non-linear potentials and approximation in the mean by analytic functions*, Math Z **129**, 299-319. (173, 234, 351, 376)

Hedberg, L.I. [1993], *Approximation by harmonic functions and stability of the Dirichlet problem*, Expo Math **11**, 193-259. (173, 376)

Heins, M. [1946], *On the number of 1-1 directly conformal maps which a multiply-connected plane region of finite connectivity p* (> 2) *admits onto itself*, Bull Amer Math Soc **52**, 454-457. (97)

Helms, L.L. [1975], *Introduction to Potential Theory*, Robert E. Kreiger Publ. Co., Huntington, NY. (234, 301, 334, 367, 368, 376)

Hille, E. [1962], *Analytic Function Theory*, Ginn and Co., Boston. (359)

Hocking, J.G., and G.S. Young [1961], *Topology*, Addison-Wesley, Reading MA. (48)

Hoffman, K. [1962], *Banach Spaces of Analytic Functions*, Prentice-Hall, Englewood Cliffs, NJ. (269)

Hunt, R.A. [1967], *On the convergence of Fourier series, orthogonal expan-*

sions and their continuous analogies, Proc Conference at Edwardsville, Ill, Southern Illinois Univ. Press, 235-255, Carbondale, IL. (199)

Jenkins, J.A. [1991], *On analytic paths*, Constantine Carathéodory: An International Tribute, edited by Th. M. Rassias, World Scientific Publ. Co., 548-553. (21)

Katznelson, Y. [1976], *An Introduction to Harmonic Analysis*, Dover, New York. (199, 203)

Koosis, P. [1980], *Introduction to H_p spaces*, London Math Soc Lecture Notes **40**, Cambridge Univ. Press, Cambridge. (269)

Landkof, N.S. [1972], *Foundations of Modern Potential Theory*, Springer-Verlag, Heidelberg. (234, 301, 333, 334, 344, 376)

Lang, S. [1985], $SL_2(R)$, Springer-Verlag, New York. (123)

Lindberg, P. [1982], *A constructive method for L^p-approximation by analytic functions*, Ark Math **20**, 61-68. (173)

Markusevic, A.I. [1934], *Conformal mapping of regions with variable boundary and applications to the approximation of analytic functions by polynomials*, Dissertation, Moskow. (172)

Massey, W.S. [1967], *Algebraic Topology: An Introduction*, Harcourt, Brace, & World, New York. (110)

Mergeljan, S.N. [1953], *On the completeness of systems of analytic functions*, Uspeki Math Nauk **8**, 3-63. Also, Amer Math Soc Translations **19** (1962), 109-166. (173)

Meyers, N. [1975], *Continuity properties of potentials*, Duke Math J **42**, 157-166. (372)

Minda, C.D. [1977], *Regular analytic arcs and curves*, Colloq Math **38**, 73-82. (21)

Mozzochi, C.J. [1971], *On the Pointwise Convergence of Fourier series*, Springer-Verlag Lecture Notes **199**, Berlin. (199)

Nehari, Z. [1975], *Conformal Mapping*, Dover, New York. (134)

Newman, M.H.A. [1964], *Elements of the topology of plane sets of points*, Cambridge University Press, Cambridge. (6)

Ohtsuka, M. [1967], *Dirichlet Problem, Extremal Length and Prime Ends*, Van Nostrand Reinhold, New York. (44)

Pfluger, A. [1969], *Lectures on Conformal Mapping*, Lecture Notes from Indiana University. (129)

Pommerenke, C. [1975], *Univalent Functions*, Vandenhoeck & Ruprecht, Göttingen. (33, 49, 146)

Radó, T. [1922], *Zur Theorie der mehrdeutigen konformen Abbildungen*, Acta Math Sci Szeged **1**, 55-64. (107)

Richter, S., W.T. Ross, and C. Sundberg [1994], *Hyperinvariant subspaces of the harmonic Dirichlet space*, J Reine Angew Math **448**, 1-26. (334)

Rubel, L.A., and A.L. Shields [1964], *Bounded approximation by polynomials*, Acta Math **112**, 145-162. (173)

Sarason, D. [1965], *A remark on the Volterra operator*, J Math Anal Appl **12**, 244-246. (295)

Spraker, J.S. [1989], *Note on arc length and harmonic measure*, Proc Amer Math Soc **105**, 664-665. (311)

Stout. E.L. [1971], *The Theory of Uniform Algebras*, Bogden and Quigley, Tarrytown. (197)

Szegö, G. [1959], *Orthogonal Polynomials*, Amer Math Soc, Providence. (156, 157)

Tsuji, M. [1975], *Potential Theory in Modern Function Theory*, Chelsea, New York. (344)

Veech, W.A. [1967], *A Second Course in Complex Analysis*, W.A. Benjamin, New York. (129)

Weinstein, L. [1991], *The Bieberbach conjecture*, International Research J (Duke Math J) **5**, 61-64. (133)

Wermer, J. [1974], *Potential Theory*, Springer-Verlag Lecture Notes **408**, Berlin. (234, 301)

Whyburn, G.T. [1964], *Topological Analysis*, Princeton Univ. Press, Princeton. (3)

List of Symbols

Index

Graduate Texts in Mathematics

continued from page ii

121 LANG. Cyclotomic Fields I and II. Combined 2nd ed.
122 REMMERT. Theory of Complex Functions. *Readings in Mathematics*
123 EBBINGHAUS/HERMES et al. Numbers. *Readings in Mathematics*
124 DUBROVIN/FOMENKO/NOVIKOV. Modern Geometry—Methods and Applications. Part III.
125 BERENSTEIN/GAY. Complex Variables: An Introduction.
126 BOREL. Linear Algebraic Groups.
127 MASSEY. A Basic Course in Algebraic Topology.
128 RAUCH. Partial Differential Equations.
129 FULTON/HARRIS. Representation Theory: A First Course. *Readings in Mathematics*
130 DODSON/POSTON. Tensor Geometry.
131 LAM. A First Course in Noncommutative Rings.
132 BEARDON. Iteration of Rational Functions.
133 HARRIS. Algebraic Geometry: A First Course.
134 ROMAN. Coding and Information Theory.
135 ROMAN. Advanced Linear Algebra.
136 ADKINS/WEINTRAUB. Algebra: An Approach via Module Theory.
137 AXLER/BOURDON/RAMEY. Harmonic Function Theory.
138 COHEN. A Course in Computational Algebraic Number Theory.
139 BREDON. Topology and Geometry.
140 AUBIN. Optima and Equilibria. An Introduction to Nonlinear Analysis.
141 BECKER/WEISPFENNING/KREDEL. Gröbner Bases. A Computational Approach to Commutative Algebra.
142 LANG. Real and Functional Analysis. 3rd
143 DOOB. Measure Theory.
144 DENNIS/FARB. Noncommutative Algebra.
145 VICK. Homology Theory. An Introduction to Algebraic Topology. 2nd ed.
146 BRIDGES. Computability: A Mathematical Sketchbook.
147 ROSENBERG. Algebraic K-Theory and Its Applications.
148 ROTMAN. An Introduction to the Theory of Groups. 4th ed.
149 RATCLIFFE. Foundations of Hyperbolic Manifolds.
150 EISENBUD. Commutative Algebra with a View Toward Algebraic Geometry.
151 SILVERMAN. Advanced Topics in the Arithmetic of Elliptic Curves.
152 ZIEGLER. Lectures on Polytopes.
153 FULTON. Algebraic Topology: A First Course.
154 BROWN/PEARCY. An Introduction to Analysis.
155 KASSEL. Quantum Groups.
156 KECHRIS. Classical Descriptive Set Theory.
157 MALLIAVIN. Integration and Probability.
158 ROMAN. Field Theory.
159 CONWAY. Functions of One Complex Variable II.

John B. Conway is Professor of Mathematics at the University of Tennessee, Knoxville. He received his Ph.D. in mathematics from Louisiana State University in 1965.

Professor Conway has published several papers in functional analysis including some which have applied the theory of complex variables. He has also published **A Course in Functional Analysis** in this same series and the first volume of the present work entitled **Functions of One Complex Variable I**.

ISBN 0-387-94460-5

EAN

9 780387 944609

ISBN 0-387-**94460**-5