Elliptic Partial Differential Equations and Quasiconformal Mappings in the Plane

KARI ASTALA
TADEUSZ IWANIEC
GAVEN MARTIN

Elliptic Partial Differential Equations and Quasiconformal Mappings in the Plane

Princeton Mathematical Series

1. The Classical Groups *by Hermann Weyl*
8. Theory of Lie Groups: I *by C. Chevalley*
9. Mathematical Methods of Statistics *by Harald Cramér*
14. The Topology of Fibre Bundles *by Norman Steenrod*
17. Introduction to Mathematical Logic, Vol. I *by Alonzo Church*
19. Homological Algebra *by H. Cartan and S. Eilenberg*
28. Convex Analysis *by R. T. Rockafellar*
30. Singular Integrals and Differentiability Properties of Functions *by E. M. Stein*
32. Introduction to Fourier Analysis on Euclidean Spaces *by E. M. Stein and G. Weiss*
33. Étale Cohomology *by J. S. Milne*
35. Three-Dimensional Geometry and Topology, Volume 1 *by William P. Thurston. Edited by Silvio Levy*
36. Representation Theory of Semisimple Groups: An Overview Based on Examples *by Anthony W. Knapp*
38. Spin Geometry *by H. Blaine Lawson, Jr., and Marie-Louise Michelsohn*
43. Harmonic Analysis: Real Variable Methods, Orthogonality, and Oscillatory Integrals *by Elias M. Stein*
44. Topics in Ergodic Theory *by Ya. G. Sinai*
45. Cohomological Induction and Unitary Representations *by Anthony W. Knapp and David A. Vogan, Jr.*
46. Abelian Varieties with Complex Multiplication and Modular Functions *by Goro Shimura*
47. Real Submanifolds in Complex Space and Their Mappings *by M. Salah Baouendi, Peter Ebenfelt, and Linda Preiss Rothschild*
48. Elliptic Partial Differential Equations and Quasiconformal Mappings in the Plane *by Kari Astala, Tadeusz Iwaniec, and Gaven Martin*

Elliptic Partial Differential Equations and Quasiconformal Mappings in the Plane

Kari Astala, Tadeusz Iwaniec, and Gaven Martin

Princeton University Press
Princeton and Oxford

Published by Princeton University Press
41 William Street, Princeton, New Jersey 08540

In the United Kingdom: Princeton University Press
6 Oxford Street, Woodstock, Oxfordshire OX20 1TW

Library of Congress Cataloging-in-Publication Data

Astala, Kari, 1953-
Elliptic partial differential equations and quasiconformal mappings in the plane / Kari Astala, Tadeusz Iwaniec, and Gaven Martin.
p. cm. – (Princeton mathematical series ; 48)
Includes bibliographical references and index.
ISBN 978-0-691-13777-3 (hardcover : acid-free paper) 1. Differential equations, Elliptic. 2. Quasiconformal mappings. I. Iwaniec, Tadeusz. II. Martin, Gaven. III. Title.
QA377.A836 2009
515'.93–dc22 2008037099

British Library Cataloging-in-Publication Data is available

This book has been composed in LaTeX

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

Printed on acid-free paper. ∞

press.princeton.edu

Printed in the United States of America

10 9 8 7 6 5 4 3 2 1

To our families
Tuulikki, Eero & Eeva
Grażyna & Krystyna
Dianne, Jennifer & Amy

This page intentionally left blank

Contents

Preface

This book presents the most recent developments in the theory of planar quasiconformal mappings and their wide-ranging applications in partial differential equations and nonlinear analysis, conformal geometry, holomorphic dynamical systems, singular integral operators, inverse problems, the geometry of mappings and, more generally, the calculus of variations. It is a simply amazing fact that the mathematics that underpins the geometry, structure and dimension of such concepts as Julia sets and limit sets of Kleinian groups, the spaces of moduli of Riemann surfaces, conformal dynamical systems and so forth is the *very same* as that which underpins existence, regularity, singular set structure and so forth for precisely the most important class of differential equations one meets in physical applications, namely, second-order divergence-type equations. All these subjects are inextricably linked in two dimensions by the theory of quasiconformal mappings.

There have been profound developments in the three or four decades since the publication of Lars Ahlfors' beautiful little book [8] and the classical text of Olli Lehto and Kalle Virtanen [229]. Indeed, whole subjects have blossomed, conformal and holomorphic dynamics, holomorphic motions, nonlinear partial differential equations and connections with the calculus of variations, to name just a few.

This book gives a fairly comprehensive account of the modern theory, but for those planning to present a semester course in the theory of quasiconformal mappings and their applications in modern complex analysis, the contents of Chapters 3 and 5, with selected applications chosen from Chapters 12, 13 and some of the later chapters, should provide ample material at an easy pace. Further, the material in Chapter 4 presents a reasonable and self-contained introduction to harmonic analysis and the theory of singular integral operators in two dimensions.

The latter parts of the book present perhaps the most recent advances in the area. Indeed, more than a few results and proofs in this monograph are new. These chapters also serve to illustrate the wide applicability of the ideas and techniques developed in the earlier part of the book.

It is our pleasure to acknowledge the wide-ranging support we have had from a number of places that has made this book possible. First, we have all been partly supported by the Academy of Finland, the Marsden Fund of New

Zealand and the National Science Foundation of the United States at one time or another. We all shared Research in Peace fellowships at Institute Mittag-Leffler (Sweden) where the first real progress toward a book was made. Of course our home institutions—Massey University (New Zealand), Syracuse University (United States) and the University of Helsinki (Finland)—have all hosted and supported us as a group at various times.

There are many people to thank as well. In particular, Pekka Koskela let us use his notes on quasisymmetric functions (a good part of Chapter 3), Stanislaw Smirnov let us present his unpublished proof of the dimension bounds for quasicircles and Laszlo Lempert communicated Chirka's proof of the λ-lemma to us while we were at Oberwolfach. There are also the people who read various drafts of the book and made substantial and valuable comments. These include Tomasz Adamowicz, Samuel Dillon, Daniel Faraco, Peter Haïssinsky, Jarmo Jääskeläinen, Matti Lassas, Martti Nikunen, Jani Onninen, Lassi Päivärinta, Istvan Prause, Eero Saksman, Carlo Sbordone, Ignacio Uriarte-Tuero and Antti Vähäkangas. We would also like to thank the team at Princeton University Press—Kathleen Cioffi, Carol Dean, Lucy Day Hobor and Vickie Kearn—who skillfully guided us through the production process and whose considerable efforts improved this book.

Finally, during the writing of this book the "quasi-world" was saddened by the premature death of one of its leading figures, Juha Heinonen. We wish to record here the deep respect we have for Juha and the contributions he made. He was an inspiration to all of us.

Kari Astala, Tadeusz Iwaniec, Gaven Martin

Helsinki, Syracuse and Auckland, 2009

Elliptic Partial Differential Equations and Quasiconformal Mappings in the Plane

This page intentionally left blank

Chapter 1

Introduction

This book relates the most modern aspects and most recent developments in the theory of planar quasiconformal mappings and their application in conformal geometry, partial differential equations (PDEs) and nonlinear analysis. There are profound applications in such wide-ranging areas as holomorphic dynamical systems, singular integral operators, inverse problems, the geometry of mappings and, more generally, the calculus of variations—all of which are presented here. It is a simply amazing fact that the mathematics that underpins the geometry, structure and dimension of such concepts as Julia sets and limit sets of Kleinian groups, the spaces of moduli of Riemann surfaces, conformal dynamical systems and so forth is the *very same* as that which underpins existence, regularity, singular set structure and so forth for precisely the most important class of equations one meets in physical (and other) applications, namely, second-order divergence-type equations. All these theories are inextricably linked in two dimensions by the theory of quasiconformal mappings.

Because of these and other compelling applications, there has recently been considerable pressure to extend classical results from conformal geometry to more general settings, for instance, to obtain optimal bounds on the existence, regularity and geometric properties of solutions of quasilinear and general nonlinear systems in the plane both in the classical elliptic setting and now in the degenerate elliptic setting. Here one moves from the established theory of quasiconformal mappings, through the theory of weakly quasiregular mappings, and comes to the more general class of Sobolev mappings of finite distortion. This progression is natural as one seeks greater knowledge about the fine properties of these mappings for implementation. Even for such well-known problems as the nonlinear $\overline{\partial}$-problem, we find that precise L^2-bounds lead to a simple and beautiful proof of the extension theorem for holomorphic motions. In the same vein, we use optimal regularity to prove Pucci's conjecture, as well as related precise results to give a solution to Calderón's problem on impedance tomography and also to Painlevé's problem on the size and structure of removable singular sets for solutions to elliptic and degenerate elliptic equations.

These precise results are in a large part due to a new understanding of the relationship between quasiconformal mappings and holomorphic flows on the one hand, and, on the other, precise results on the L^p-invertibility of classes of singular integral operators called Beltrami operators. However, there have been other recent developments in the theory of quasiconformal mappings—notably in the field of analysis on metric spaces principally established by Heinonen and Koskela. These advances could not leave a book such as ours untouched, for they clarify many of the basic facts and the precise hypotheses necessary to prove them and often provide elementary and clear proofs. Thus the reader will find novelty and simplicity here even for the foundations of the theory, which now go back more than half a century.

Another novelty in the approach of this book is the use of many of the significant advances in harmonic analysis made over the last few decades; these include H^1-BMO duality, maximal function estimates, the theory of nonlinear commutators and integral estimates for Jacobian determinants both above and below their natural Sobolev domain of definition, all crucial for our studies of optimal regularity and nonlinear PDEs, as well as the Painlevé problem on removable singularities. The reader will have ample opportunity to see these powerful modern techniques in diverse applications.

1.1 Calculus of Variations, PDEs and Quasiconformal Mappings

The strong interplay among the calculus of variations, partial differential equations and the geometric theory of mappings (which is what this book is all about) has a long and distinguished history—going back at least to d'Alembert who in 1746 first related the derivatives of the real and imaginary part of a complex function in his work on hydrodynamics [51, p. 497]. These equations came to be known as the Cauchy-Riemann equations.

Conservation laws and equations of motion or state in physics and mathematics are described by divergence-type second-order differential equations. This is no accident. It is a fundamental precept of physics that a system acts so as to minimize some action functional—Hamilton's principle of least action. Hamilton's principle applies quite generally to classical fields such as the electromagnetic, gravitational and even quantum fields. We are therefore naturally led to study the minima of energy functionals, regularity of minimizers and other aspects of the calculus of variations. We give a classical problem a review in the next section. Loosely, minima satisfy an associated Euler-Lagrange equation that appears in divergence form as a result of integration by parts in the derivation of the equation. Similar examples appear in continuum mechanics and materials science.

On the other hand, general conservation laws are described as follows. Suppose the flux density of a scalar quantity e, such as density, concentration, temperature or energy, is $q = \mathcal{A}(z, \nabla e)$, a function of the gradient of e. A basic assumption of continuum physics is that the gain of the physical quantity in a domain Ω corresponds to the loss of this quantity across the boundary $\partial\Omega$. Thus

$$\int_{\partial\Omega} q \cdot \nu = \int_{\Omega} f$$

Here f denotes the source density and ν denotes the outer normal. The above identity is called the conservation law with respect to the flux q and leads (we describe how in Section 16.3) to the differential equation

$$\dot{e} + \mathrm{div}(q) = \dot{e} + \mathrm{div}\mathcal{A}(z, \nabla e) = f$$

This is a conservation law for the physical quantity e. In the steady-state case we obtain a second-order equation in divergence form for q.

With so many compelling applications in hand, it is no wonder that there is considerable interest in the topological and analytic properties of the minimizers of various functionals and also in solutions of second-order equations in divergence form. These topological and analytic properties describe, for instance, the flow lines of the field and the structure and size of any singular set.

Let us explain using an elementary example from the calculus of variations how related first- and second-order equations might arise. Consider deforming the unit disk $\mathbb{D}$ to another domain Ω minimizing energy. This was in fact Riemann's approach to his mapping theorem, and which he called the Dirichlet principle. He obtained the desired conformal mapping as an absolute minimizer of the Dirichlet energy. Weierstrass showed Riemann's argument was not generally valid, however Hilbert later ironed out the details—ultimately requiring some regularity of $\partial\Omega$. As this discussion suggests, the example is quite classical, but it's solution contains many key ideas and provides us with some important lessons.

Problem. *Given a simply connected domain Ω,*
(a) *find the homeomorphism of minimal energy mapping the disk to Ω,*
(b) *find the minimizer subject to prescribed boundary values.*

The energy of a mapping is defined as the Dirichlet integral, so we are asked to find

$$\min_{f:\mathbb{D}\to\Omega} \left\{ \int_{\mathbb{D}} \|Df(z)\|^2 \, dz \right\}, \qquad \|Df\|^2 = |f_x|^2 + |f_y|^2,$$

over all homeomorphisms, with the possible restriction $f|\partial\mathbb{D} = g_o$. In order to solve this problem (if it is possible at all), we should consider the correct function space to start looking for a solution. For the minimum to be finite, we certainly need for there to be some mapping f_0 satisfying the hypotheses (the gradient of f_0 should be square-integrable with correct boundary values). Given this mapping, we can then assume the existence of a sequence tending to the minimum (a *minimizing sequence*). Then comes the difficult problem of proving this sequence has a convergent subsequence whose limit is sufficiently regular to satisfy the hypotheses (thus the need for a priori estimates). For the problem in hand, Hadamard's inequality for matrices $A \in \mathbb{R}^{2\times 2}(\mathbb{C})$ states $\|A\|^2 = \mathrm{tr}\,(A^t A) \geqslant 2\,\det A$ and therefore gives the pointwise almost everywhere estimate

$$\|Df(z)\|^2 \geqslant 2\,J(z,f) = 2\det\,Df(z)$$

(we consider only orientation-preserving homeomorphisms, meaning that the Jacobian $J(z,f) \geqslant 0$ almost everywhere in $\mathbb{D}$.) Then for every homeomorphism of Sobolev class $W^{1,2}(\mathbb{D})$, we have

$$\int_{\mathbb{D}} \|Df\|^2 \geqslant 2\int_{\mathbb{D}} J(z,f) = 2|f(\mathbb{D})| = 2|\Omega|,$$

providing a lower bound on the minimum. Consequently, if there is to be an absolute minimizer f achieving this lower bound we must have it solving the first(!)-order equation for an absolute minimizer

$$\|Df(z)\|^2 = 2\,J(z,f)$$

Some linear algebra (we have equality in Hadamard's estimate) shows this to be equivalent to

$$D^t f(z)\,Df(z) = J(z,f)\,\mathbf{I},$$

where $\mathbf{I}$ is the identity matrix. This is the equation for a conformal mapping, of course—in complex notation this system is the Cauchy-Riemann equations (which points to the virtue of complex notation). Back to our problem, if we prescribe the boundary values and they happen not to be those of a conformal mapping, then a minimizer cannot achieve our a priori lower bound. Another approach is to vary a supposed minimizer f by a parameterized family of homeomorphisms of $\mathbb{D}$ that are the identity near the boundary, say φ_t normalized so $\varphi_0(z) = z$. Since f is a minimizer we must have

$$\frac{d}{dt}\int_{\mathbb{D}} \|D(f\circ\varphi_t)\|^2 \,\Bigg|_{t=0} = 0,$$

leading to the second-order Euler-Lagrange equation for f, $\mathrm{div} Df = \Delta f = 0$. Thus the minimum should be a harmonic mapping with the given boundary values, and the question boils down to whether our prescribed boundary values g_0 have a harmonic homeomorphic extension to $\mathbb{D}$. The Poisson formula gives a harmonic function, and we are left to discuss the topological properties of this solution. A way forward here is to show that the Jacobian is continuous and does not vanish (so local injectivity) and use the monodromy theorem, but the geometry of the domain and the boundary values must come into play. For instance, without some convexity assumption on Ω the mean value of g_0 may lie outside Ω. It is a classical theorem of Choquet, Kneser and Rado that as soon as Ω is convex, one can solve the posed problem with homeomorphic boundary data and the solution is a smooth diffeomorphism.

We may consider the above problem in more general circumstances. For instance, if $H : \Omega \to \mathbb{R}^{2\times 2}$, symmetric and positive definite, is some measurable function describing some material property of Ω, we could seek to minimize the new energy functional

$$\int_{\mathbb{D}} \langle H(f(z))Df(z), Df(z)\rangle$$

We use Hadamard's inequality in the form

$$\langle \sqrt{H}Df, \sqrt{H}Df\rangle \geqslant 2\sqrt{\det H}\,\det(Df)$$

and, as before, an absolute minimizer must satisfy the nonlinear PDE

$$D^t f H(f) Df = \sqrt{\det H(f)}\, J(z,f)\,\mathbf{I}$$

It is only in two dimensions that such an equation is not overdetermined (this accounts for higher-dimensional rigidity), and we have the possibility of finding a solution in quite reasonable generality. For conformal geometry we are interested in the case $\det H \equiv 1$ yielding the nonlinear Beltrami equation

$$D^t f\, H(f)\, Df = J(z,f)\mathbf{I}$$

If in the above we consider a tensor field $G : \mathbb{D} \to \mathbb{R}^{2\times 2}$, $\det G \equiv 1$, we have

$$D^t f\, Df = J(z,f)\, G,$$

equivalent to a linear (over $\mathbb{C}$) equation called the complex Beltrami equation,

$$\frac{\partial f}{\partial \bar z} = \mu(z)\,\frac{\partial f}{\partial z}\,,$$

which we will spend quite a bit of time discussing. Finally, if we consider a constrained problem and look for the Euler-Lagrange equation we are quickly led to second-order equations in divergence form (arising from the necessary integration by parts) for the real and imaginary parts of $f = u + iv$,

$$\operatorname{div} G^{-1}\,\nabla u = 0, \qquad\qquad \operatorname{div} G^{-1}\,\nabla v = 0,$$

and ultimately to more general second-order equations in divergence form.

There are a few important points we would like to draw from this discussion regarding minima of variational problems:

- Unconstrained or absolute minimizers of variational functionals are likely to satisfy first-order differential equations.
- Constrained or stationary mappings will likely satisfy a second-order differential equation.
- We may well find stationary solutions that are not minimizers. Indeed, there might not be a minimizer within the class of homeomorphisms.

Of course, in the most general setting of multiple connected domains, one would consider minimizers in a given homotopy class of maps between domains, or more generally, homotopy classes of maps between Riemann surfaces. Moreover we would seek to minimize more general functionals. Here we find clear connections with Teichmüller theory, surface topology and so forth.

A significant portion of this book is given over to the study of the equations like those we have discovered above where we will seek existence, uniqueness and optimal regularity and so forth for their solutions—and also for the counterparts to these equations in other settings. Later we shall discuss recent developments in the study of existence and uniqueness properties for mappings between planar domains whose boundary values are prescribed and have the smallest mean distortion—this will bring the relevance of the first example discussed above back into focus because of a surprising connection with harmonic mappings and other surprises as well. Indeed, the analogy here with Teichmüller theory is quite strong. This theory is partly concerned with extremal quasiconformal mappings in a homotopy class. These mappings minimize the L^∞-norm of the distortion. We investigate what happens when the L^1-norm of the distortion is minimized instead. Further, in these studies we will find many new and unexpected phenomena concerning existence, uniqueness and regularity for these extremal problems where the functionals are polyconvex but typically not convex. These seem to differ markedly from phenomena observed when studying multi-well functionals in the calculus of variations. The phenomena observed concerning mappings between annuli present a case in point.

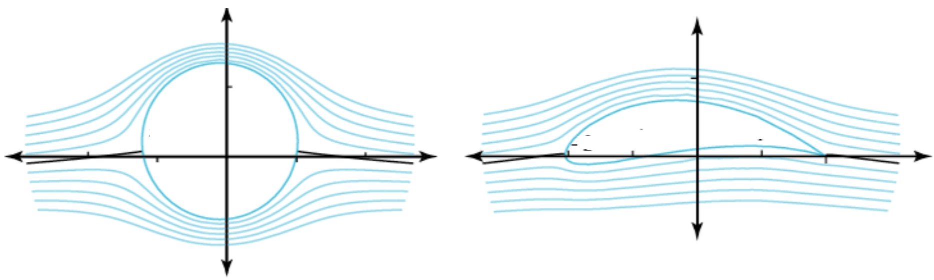
In two dimensions, the methods of complex analysis, conformal geometry and quasiconformal mappings provide powerful techniques, not available in other dimensions, to solve highly nonlinear partial differential equations, especially those in divergence form. Of course the relevance of divergence-type equations to quasiconformal mappings is not new. It has been evident to researchers for at least 70 years, beginning with M.A. Lavrentiev [224], C.B. Morrey [271, 273, 272], R. Caccioppoli [83], L. Bers and L. Nirenberg [54, 56, 57], B. Bojarski [68], Finn [126, 127] and Serrin [325], among many others. In the literature one finds concrete applications in materials science, particularly, nonlinear elasticity, gas flow and fluid flow, and in the calculus of variations going back generations.

One of the primary aims of this book is to give a thorough account of this classical theory from a modern perspective and connect it with the most recent developments.

1.2 Degeneracy

As we have suggested, the equations we consider arise naturally in hydrodynamics, nonlinear elasticity, holomorphic dynamics and several other areas. A good part of this book is concerned with these equations at the extreme limits of regularity and related assumptions on the coefficients. A particular aim is to develop tools to handle these situations where a system of equations might degenerate. Here is a natural example.

In two-dimensional hydrodynamics, the fluid velocity (the gradient of the potential function—see for instance (16.51)—satisfies a Beltrami equation that degenerates as the flow approaches a critical value, the local speed of sound; see (16.54).



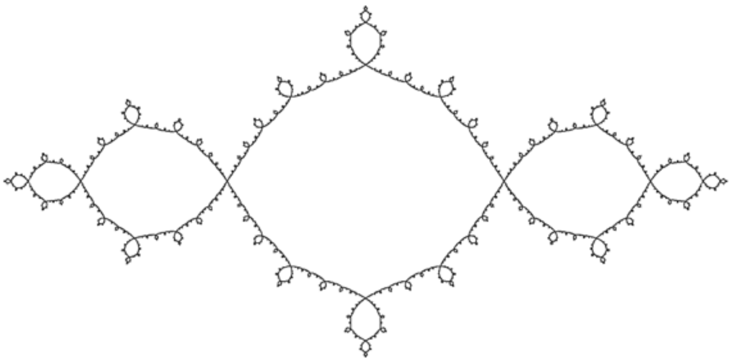
Subsonic fluid flow around a disk and Joukowski aerofoil

What happens as we break the speed of sound? In the 1950s when this was a problem of particular importance, the existence of a shock wave boundary was supposed, presupposing, albeit with good evidence, the particular structure of the singular set; see for instance [100, 144]. On one side of the shock boundary one had an elliptic equation, and on the other a hyperbolic equation. The very early approaches had to assume some degree of analyticity and used various schemes of successive approximations: the Rayleigh-Janzen expansion of the potential function in a power series in the stream Mach number or a modification of this method due to Prandtl and the solution of mixed (transonic) flows by means of power series in the space variables. Perturbative methods were also employed, most based on von Kármán's similarity law for transonic flow [207]. The state of the art as of the late 1950s is described in L. Bers' well-known book [54], and although there has been a great deal of literature on the subject since, most has focused on the study of shock waves in a similar sort of setup (and of course in higher dimensions). In this book we will describe the precise limits of existence and regularity and the structure of the singular set in the degenerate setting—but where there is no shock wave. This allows for isolated points (or even Cantor sets) where one might have degeneracies such as infinite density or pressure. The precise conditions are described in terms of bounded mean oscillation (BMO) bounds on the distortion function of the coefficient - leading to the theory of mappings with exponentially integrable distortion. This was first realized by G. David [103], and here we present substantial sharpening and refinement of these early results. When applied in the setting described above, this theory shows the topological properties of the streamlines and so forth to be the same as those for subsonic incompressible flows (really the Stoilow factorization theorem showing that these mappings are topologically equivalent to analytic mappings).

Thus a significant problem addressed in this book is to see how to relax the classical assumptions on the Beltrami equations making them uniformly elliptic, so as to study the nonuniformly elliptic (that is, degenerate elliptic) setting and yet save as much of the theory as possible.

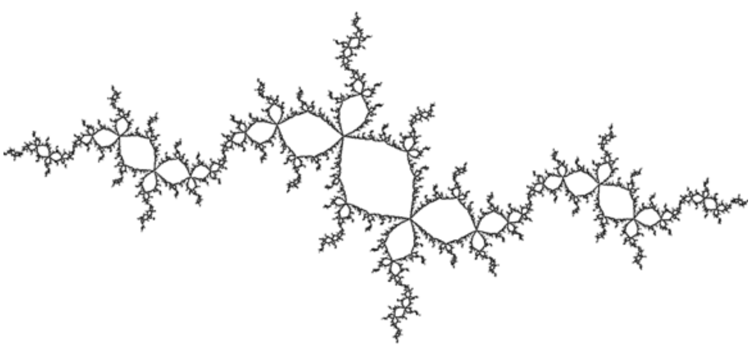
1.3 Holomorphic Dynamical Systems

There are two basic examples of holomorphic dynamical systems. First, is the classical Fatou-Julia theory of iteration of rational mappings of the sphere, [90, 123, 203, 262]. Hardly anyone has not seen the beautiful pictures [293] of the Mandelbrot set and associated Julia sets of quadratic mappings.



The Julia set of a quadratic polynomial.

The theory of quasiconformal maps has played a key role in the study of these conformal dynamical systems ever since D. Sullivan, A. Douady and their coauthors introduced them to the theory [108, 111, 237, 257, 341]. Ideas such as quasiconformal surgery show how one conformal dynamical system can be constructed from another.



The Julia set after quasiconformal surgery: grafted with Douady's rabbit.

A crucial discovery for us, which underpins a good deal of our approach in this book, is the concept of holomorphic motions intoduced by R. Mañé, P. Sad and D. Sullivan [237] and the subsequent conjectures on the extension of these motions by Sullivan and W.P. Thurston [342] and the solution by Z. Słodkowski [329]. This discovery really shows the notions of holomorphically parameterized flows and quasiconformal mappings to be inextricably linked.

In this book we will provide tools that allow one to study the structure, dimension and other properties of Julia sets. As far as the question of degeneracy goes, we will see that the Julia set of $\lambda z + z^2$ is a $\frac{1+|\lambda|}{1-|\lambda|}$-quasicircle if $|\lambda| < 1$ (equivalently for z^2+c when $c = \lambda/2-\lambda^2/4$ lies in the primary component of the Mandelbrot set) and degeneracy occurs as $|\lambda| \to 1$. The dynamical systems obtained are quasiconformally equivalent on hyperbolic components of parameter space. The intriguing question of what happens as $|\lambda| \to 1$ (or generally moves to the boundary of a hyperbolic component) and the uniform bounds in the theory of quasiconformal mappings are lost is still to some measure unresolved. Haïssinsky has shown, using David's work, that for real $\lambda \nearrow 1$ ($c \nearrow \frac{1}{4}$) the sequence of Julia sets converges to a Jordan curve—the cauliflower Julia set—that is the image of the unit circle under a mapping of exponentially integrable distortion.

The second classical example of quasiconformal mappings being applied in conformal dynamical systems is the way they arise naturally in the study of Kleinian groups; through Teichmüller theory and moduli spaces. The modern approach goes back to Bers' seminal work on simultaneous uniformization [53] and Ahlfors' use of quasiconformal mappings in proving geometric finiteness [5]. The key idea again is that in moduli space the Kleinian groups in question are quasiconformally equivalent. What happens as one goes to the boundary and considers, for instance, degenerating sequences of quasi-fuchsian groups? Again we loose the uniform estimates needed in the classical theory of quasiconformal mappings and need to analyze a degenerate situation. We hope that mappings of finite distortion may play a future role in the analytic understanding of these questions.

1.4 Elliptic Operators and the Beurling Transform

The types of first-order equations $\mathcal{L}\, f = 0$ we have seen above have evolved from study of the Cauchy-Riemann operators,

$$\mathcal{L}_1 f = \frac{\partial}{\partial \bar{z}}\, f, \qquad \mathcal{L}_2 f = \frac{\partial}{\partial z}\, f$$

The solutions to $\mathcal{L}_i f = 0$, $i = 1, 2$, represent analytic and anti-analytic functions. A quantatative distinction between these two classes of mappings is that the former are orientation-preserving and the later orientation-reversing (or positive versus negative Jacobian determinant). In fact, this topological dichotomy of solutions applies to *all* first-order elliptic PDEs in the complex plane. The continuous deformation of a general elliptic system $\mathcal{L} f = 0$, perhaps by varying the coefficients, will never change the orientation of solutions unless ellipticity is violated at some moment.

This idea leads to the homotopy classification of all first-order elliptic systems and the corresponding differential operators into the two classes represented by the Cauchy-Riemann equations $\mathcal{L}_1 f = 0$ and its dual $\mathcal{L}_2 f = 0$. The fundamental connection between these classes is made via the Beurling transform, about which we will have much to say. It is a singular integral operator $\mathcal{S}$ of Calderón-Zygmund type bounded in $L^p(\mathbb{C})$, $1 < p < \infty$. It is the remarkable property

$$\mathcal{S} \circ \frac{\partial}{\partial \bar{z}} = \frac{\partial}{\partial z} : C_0^\infty(\mathbb{C}) \to C_0^\infty(\mathbb{C})$$

intertwining the Cauchy-Riemann operators that makes it so important in the L^p-theory of elliptic operators. There are six homotopy classes of second-order elliptic operators in the complex plane, or three equivalence classes of elliptic equations, comprising of combinations of the z- and $\bar{z}$-derivatives. The most important of these is the complex Laplace equation $\frac{\partial^2}{\partial z \partial \bar{z}} f = 0$. Notice that for this equation the "factors" $\frac{\partial}{\partial \bar{z}}$ and $\frac{\partial}{\partial z}$ come from different homotopy classes.

For the other two classes, the first-order factors come from the same homotopy class and this partly explains why the equations and their solutions have significantly different features. For instance, the Fredholm alternative fails; as an example, the equation $f_{\bar{z}\bar{z}} = 0$, $f|\partial\mathbb{D} = 0$, admits the uncountable family of solutions $f(z) = (1 - |z|^2)h(z)$, where h is holomorphic and continuous in the closed unit disk.

A major innovation in this book is the study of second-order PDEs of divergence form in the complex plane,

$$\operatorname{div}\mathcal{A}(z, \nabla u) = 0 \tag{1.1}$$

when $\mathcal{A}$ is only supposed δ-monotone, with *no additional regularity assumption.* Here we are still able to obtain a reduction to a first-order system for the complex gradient $f = u_z$ of a solution and show that it is a quasiregular mapping, if $\mathcal{A}$ is spatially independent. In this way the significant results we obtain for such mappings apply to show that f has good regularity and nice topological properties which the solution u then inherits.

Another important approach is via the duality given by the Hodge * operator. Here we reduce the $\mathcal{A}$-harmonic equation (1.1) to the first-order system

$$-\mathcal{A}(z, \nabla u) = *\nabla v$$

for a function v called the $\mathcal{A}$-harmonic conjugate of u. This approach is particularly useful when $\mathcal{A}(z, \nabla u) = A(z)\nabla u$ for some measurable $A : \Omega \to \mathbb{R}^{2\times 2}$. This leads us to the quasiregular mapping $f = u + iv$ in much the same way as an analytic function is composed of a harmonic function and its harmonic conjugate.

There are also very interesting questions concerning the convergence of *sequences* of operators that we shall address in the book. Here we will meet the notion of G-convergence, which emerges in quite a natural way and exploits the normal family (equicontinuity) properties of quasiregular mappings.

While we give evidence of substantial progress in the theory of elliptic second-order equations in the complex plane, we are sure there remains many interesting phenomena to be discovered and interesting connections to other areas of mathematics to be found.

Chapter 2

A Background in Conformal Geometry

We have mentioned the strong connections between PDEs and geometry. In this chapter we give a gentle introduction to conformal geometry in the plane and describe some of the connections between conformal and Riemannian geometry with PDEs in two dimensions. We shall also try to give a clear account of how the PDEs we shall spend much of our time studying arise from a number of differing perspectives.

In this book the development of the theory of quasiconformal mappings really starts at the beginning of Chapter 3. The material in the present chapter will be quite familiar to those with some experience in geometric function theory. Such readers may therefore skip this chapter and proceed to Chapter 3, returning to the present chapter for background material only when necessary.

2.1 Matrix Fields and Conformal Structures

Let $\mathbf{S}(2)$ denote the space of 2×2 positive definite symmetric matrices with real entries and having determinant equal to 1. A positive definite matrix can be used to define an inner product using the standard (Euclidean) inner product between vectors by the rule

$$\langle \eta, \eta \rangle_G = \langle \eta, G\eta \rangle \tag{2.1}$$

Recall that a Riemannian structure is, roughly, an inner product defined on the tangent bundle of a manifold. Such a structure gives rise to a metric distance defined as the length of the shortest curve between two points. The length of a curve is the sum (integral) of the lengths of its tangent vectors.

Typically Riemannian structures are assumed smooth and second-order invariants such as curvature determine the local isometry type. The equations determining an isometry are nonlinear and overdetermined as they essentially prescribe the Jacobian of a mapping, and this is why compatibility conditions such as curvature are necessary. However, in this book, motivated by applications in PDEs and dynamics, we shall largely consider measurable Riemannian structures. Despite the loss of such higher-order invariants, we shall see that there is a surprisingly rich theory.

For planar domains various differential-geometric constructions are somewhat easier to describe, as we have global coordinates and the tangent bundle trivializes. In particular, an orientable Riemannian metric on a planar domain Ω can simply be viewed as a map $A : \Omega \to \mathbb{R}_+ \cdot \mathbf{S}(2)$, the space of symmetric positive definite matrices. If

$$A = A(z) = \begin{bmatrix} a & b \\ b & c \end{bmatrix},$$

then the length element induced by A is

$$ds = \langle dz, A(z)dz\rangle^{1/2} = \sqrt{a\,dx^2 + 2b\,dxdy + c\,dy^2} \tag{2.2}$$

The most important special case we shall deal with is the hyperbolic metric studied in the next section. We shall henceforth restrict ourselves only to orientable stuctures without saying so every time. This is equivalent to the restriction $\det A(z) > 0$.

If $B = B(z)$ is another Riemannian metric on Ω', a map $f : \Omega \to \Omega'$ is an isometry if and only if it preserves the inner-products. That is, for all $z \in \Omega$, if ξ, ζ are tangent vectors at z, then

$$\langle f_*\xi, f_*\zeta\rangle_B = \langle \xi, \zeta\rangle_A \tag{2.3}$$

Of course, $f_*\xi = Df(z)\xi$, and in view of (2.1) the equation (2.3) is written as

$$\langle Df(z)\xi, B(f(z))Df(z)\zeta\rangle = \langle \xi, A(z)\zeta\rangle \tag{2.4}$$

From here we quickly find that

$$D^t f(z)B(f(z))Df(z) = A(z) \tag{2.5}$$

for all $z \in \Omega$. This is the first-order partial differential equation for a map between the metric structures determined by A and B to be an isometry.

If we simply take determinants of both sides of (2.5) and write $a(z) = \det^{1/2} A$ and $b(z) = \det^{1/2} B$, we obtain the equation

$$J(z, f)b(f(z)) = a(z), \qquad z \in \Omega \tag{2.6}$$

for the unknown mapping f. Solving (2.6) is already a formidable task. For instance, even when $b \equiv 1$, identifying those functions a that are the Jacobians of mappings is an important outstanding problem in analysis. Some results are known once one assumes some smoothness; see for instance [275, 310, 313].

Two Riemannian structures $\langle\cdot,\cdot\rangle_A$ and $\langle\cdot,\cdot\rangle_B$ on a domain Ω are said to be *conformally equivalent* if there is a positive function $\phi:\Omega\to\mathbb{R}_+$ such that

$$\langle\cdot,\cdot\rangle_{A(z)} = \phi(z)\langle\cdot,\cdot\rangle_{B(z)},$$

or equivalently, $A(z)=\phi(z)B(z)$. Locally, ϕ effects a change of scale, but all angles are preserved, as

$$\frac{\langle\xi,\zeta\rangle_{\phi A}}{\sqrt{\langle\xi,\xi\rangle_{\phi A}\langle\zeta,\zeta\rangle_{\phi A}}} = \frac{\langle\xi,\phi A\zeta\rangle}{\sqrt{\langle\xi,\phi A\xi\rangle\langle\zeta,\phi A\zeta\rangle}} = \frac{\langle\xi,\zeta\rangle_A}{\sqrt{\langle\xi,\xi\rangle_A\langle\zeta,\zeta\rangle_A}}$$

Measurable Structures and Conformal Mappings

Every Riemannian structure A on Ω is conformally equivalent to another, say

$$G:\Omega\to\mathbf{S}(2), \tag{2.7}$$

for which the determinant is identically equal to 1, namely, $G=(\det A)^{-1/2}A$. We say that G is bounded if the set of matrices $\{G(z): z\in\Omega\}$ is a bounded subset of the 2×2 matrices or, equivalently, bounded as a subset of $\mathbb{R}^4$.

Definition 2.1.1. *If $G:\Omega\to\mathbf{S}(2)$ is bounded and measurable, we call G a* measurable conformal structure *on Ω. If $H:\Omega'\to\mathbf{S}(2)$ is a conformal structure on Ω', a homeomorphic mapping $f:\Omega\to\Omega'$ is said to be* conformal *from (Ω,G) to (Ω',H) if f preserves angles. This means that, for $z\in\Omega$ and unit vectors ξ,ζ*

$$\langle f_*\xi, f_*\zeta\rangle_H = \phi\,\langle\xi,\zeta\rangle_G \tag{2.8}$$

Here ϕ is an unspecified real-valued function. As G and H are only assumed measurable, such an equation is only supposed to hold almost everywhere and accordingly, ϕ is only assumed to be measurable.

However, we shall often have cause to abuse terminology. A homeomorphism (of sufficient Sobolev regularity) $f:\Omega\to\Omega'$ is holomorphic, or complex analytic, if and only if it preserves the standard conformal structure, that is, f satisfies (2.8) with $G=H\equiv\mathbf{I}$. Thus, when there are no obvious measurable conformal structures in sight, we shall use the term "conformal", as one usually does in complex analysis, to mean a holomorphic injection, an injective mapping having a complex derivative at each point of its domain. There will be very little opportunity for confusion here.

As a differential equation, (2.8) reads as

$$D^t f(z) H(f(z)) Df(z) = \phi(z)G(z) \tag{2.9}$$

Since $\det(G)=\det(H)=1$, we must in fact have $\phi(z)=J(z,f)$, and the resulting equation, called the *Beltrami system*, becomes

$$D^t f(z) H(f(z)) Df(z) = J(z,f)G(z) \tag{2.10}$$

and will be the focus of much of our study. A first glance suggests this equation is overdetermined and nonlinear, however, we shall soon see that it is well determined and, while not always linear, it satisfies the nice property of being quasilinear, as does the related nonlinear equation

$$D^t f(z)H(z,f)Df(z) = J(z,f)G(z,f) \tag{2.11}$$

It is difficult at first to appreciate the importance of solving such equations as (2.10). Let us point out simply that if $G = H \equiv \mathbf{I}$, then a solution f satisfies the Cauchy-Riemann equations and therefore represents an analytic equivalence between domains. As a further example, if $\Omega = \Omega' = \mathbb{D}$, the unit disk, then the existence of solutions shows that all Riemannian metrics on $\mathbb{D}$ are equivalent by a conformal change of variables. However, solutions do not exist in complete generality; one needs to place restrictions on G and H, such as boundedness, to guarantee ellipticity. Further, there are topological restrictions on the domains Ω and Ω', and even when there are no topological obstructions, in multiply connected domains there are many "conformal invariants" represented by Teichmüller spaces or moduli spaces.

At this point the reader may ask why we have chosen to speak of measurable structures and not of smooth structures. There are a number of reasons for this. First, in many applications, such as those in elasticity theory, the matrices G and H describe the properties of various media and are seldom smooth. In other applications, such as in holomorphic dynamics, we must construct conformal structures G and H by various infinite processes that will cause smoothness to be lost. From the point of view of mapping theory or the calculus of variations, one seeks the minima of a certain problems. Seldom will these minima be smooth; therefore the equations these minima satisfy should not have smooth coefficients. Yet it is from these equations that we deduce properties, such as continuity or the existence of partial derivatives, of minima. In Teichmüller theory the distance between topologically equivalent surfaces is measured in terms of the mapping of smallest distortion. Such extrema are almost never smooth and yet have very nice structure reflecting the geometry and topology of the surfaces in question. Also, from other points of view, particularly those of PDEs, it is the families of equations and the properties of their solutions that are important. Compactness requirements, that these families of equations should be closed, lead naturally to consideration of those equations with measurable coefficients.

2.2 The Hyperbolic Metric

One of the more useful tools in complex analysis is the hyperbolic metric of a planar domain. We discuss here the hyperbolic metric of the unit disk, sometimes refered to as the Poincaré plane or disk. Later we shall use this metric of $\mathbb{D}$, together with the uniformization theorem, to define the hyperbolic metric of an arbitrary planar domain (other than the full plane or the punctured plane). This is quite a technical and deep fact from complex analysis, in fact, one of the most important results and themes of 19^{th} century mathematics, and we shall not offer proofs here. Mostly, we shall need only the hyperbolic metric on the disk and the triply punctured sphere.

For each $a \in \mathbb{D}$ and $\theta \in \mathbb{R}$, the linear fractional transformation

$$\phi_a(z) = e^{i\theta}\,\frac{z-a}{1-\overline{a}z} \tag{2.12}$$

defines a holomorphic self-homeomorphism of the disk $\mathbb{D}$ to itself. We compute

$$\phi'_a(z) = e^{i\theta}\,\frac{1-|a|^2}{(1-\overline{a}z)^2},$$

from which we have the identity

$$\frac{|\phi'_a(z)|}{1-|\phi_a(z)|^2} = \frac{1}{1-|z|^2} \tag{2.13}$$

Equation (2.13) expresses the fact that each $\phi_a : \mathbb{D} \to \mathbb{D}$ is an isometry of the Riemannian metric

$$ds_{hyp}(z) = \frac{2|dz|}{1-|z|^2}, \quad z \in \mathbb{D}$$

This corresponds to the choice of matrix

$$A = \begin{bmatrix} \frac{4}{(1-|z|^2)^2} & 0 \\ 0 & \frac{4}{(1-|z|^2)^2} \end{bmatrix}$$

in (2.2). Integrating this metric provides the hyperbolic metric $\rho_{\mathbb{D}}$ of the unit disk $\mathbb{D}$,

$$\rho_{\mathbb{D}}(z,w) = \inf_{\gamma} \int_{\gamma} ds_{hyp} \tag{2.14}$$

where the infimum is over all rectifiable curves γ joining z to w in $\mathbb{D}$. From the definition at (2.14) the triangle inequality is clear, and so $\rho_{\mathbb{D}}(z,w)$ is a metric.

Next, symmetry considerations and an integration quickly reveal that

$$\rho_{\mathbb{D}}(0,z) = \log\frac{1+|z|}{1-|z|},$$

while using the transitivity of the group of linear fractional transformations of the disk gives us the more general formula

$$\rho_{\mathbb{D}}(z,w) = \log\frac{|1-\bar{z}w|+|z-w|}{|1-\bar{z}w|-|z-w|}$$

As a consequence, we can make the useful observation that, for all $z \in \mathbb{D}$,

$$|z| = \tanh\frac{1}{2}\,\rho_{\mathbb{D}}(0,z) \tag{2.15}$$

Actually the group of all linear fractional transformations of $\mathbb{D}$, described by (2.12), is isomorphic to the group $PSL(2,\mathbb{R})$, the projective group of 2×2 matrices with real entries and determinant 1. To see this, note that the map

$$\Phi : z \mapsto i\,\frac{1-z}{1+z}$$

is conformal from $\mathbb{D}$ onto the upper half-space $\mathbb{H} = \{z : \Im m(z) > 0\}$. In fact, Φ is an isometry from the hyperbolic metric of the disk to the metric

$$ds = \frac{|dz|}{\Im m(z)}, \qquad z \in \mathbb{H},$$

giving us another model for the hyperbolic plane. The reflection principle quickly identifies the conformal (and hence isometric!) transformations of $\mathbb{H}$ as the linear fractional transformations

$$z \mapsto \frac{az+b}{cz+d}, \qquad a,b,c,d \in \mathbb{R}, \quad ad-bc \neq 0$$

Since we may multiply the numerator and denominator of this fractional transformation by any nonzero constant without affecting the transformation, we may normalize so that $ad - bc = 1$. The reader may wish to verify that this estabishes a topological isomorphism between the group of orientation-preserving isometries of $\mathbb{H}$ and $PSL(2,\mathbb{R})$.

Discrete subgroups of linear fractional transformations are called Fuchsian groups, and it is a fact that every Riemann surface, other than the sphere, torus, plane and punctured plane, admits a complete hyperbolic metric. Further, each such surface can be identified with the orbit space (or quotient) of a Fuchsian group. Thus hyperbolic geometry plays a central role in the theory of surfaces. More of the basic facts concerning hyperbolic geometry, hyperbolic trigonometry, discrete groups and Riemann surfaces can be found in the well-known texts of Beardon [48], Ahlfors-Sario [12] and Farkas-Kra [122], among many others.

2.3 The Space S(2)

There is a natural differential geometric structure on the space $\mathbf{S}(2)$ of symmetric positive definite 2×2 matrices with determinant equal to 1. This structure is induced in turn on the space of measurable conformal structures on a domain, as for instance in (2.7). It is the purpose of this section to take time to recount this important geometric fact.

As we shall see later, the measurable conformal structure induced by a quasiconformal mapping may be described either from the point of view of real analysis and the space $\mathbf{S}(2)$ or in the terminology of complex analysis and the hyperbolic plane $\mathbb{D}$. Both aspects are important for our study, and this dichotomy of real and complex analysis is typical for many topics considered in this monograph.

Reflecting this phenomenon, we shall show that in fact the space $\mathbf{S}(2)$ is isometric to the hyperbolic plane $\mathbb{D}$. This isometry is effected by the correspondences $G \to \mu$, where

$$G = \begin{bmatrix} g_{11} & g_{12} \\ g_{12} & g_{22} \end{bmatrix} \quad \mapsto \quad \mu = \frac{g_{11} - g_{22} - 2ig_{12}}{g_{11} + g_{22} + 2} \tag{2.16}$$

In fact, this correspondence goes much deeper than just this one-to-one isomorphism, and as we shall see it partly reflects the correspondence between the real and complex distortion functions of quasiconformal mappings. We shall see it appearing in a number of different forms.

The space $\mathbf{S}(2)$ is clearly two-dimensional, and any $A \in \mathbf{S}(2)$ can be written in the form

$$A = O^t \begin{bmatrix} \lambda & 0 \\ 0 & 1/\lambda \end{bmatrix} O$$

for some $O \in SO(2,\mathbb{R})$ and $\lambda > 0$. The general linear group $GL(2,\mathbb{R})$ acts transitively on the right on $\mathbf{S}(2)$ via the rule

$$X[G] = |\det X|^{-1} X^t G X \qquad X \in GL(2,\mathbb{R});\ G \in \mathbf{S}(2) \tag{2.17}$$

The Riemannian metric

$$ds^2 = \frac{1}{2}\,\mathrm{tr}\,(Y^{-1}dY)^2 \tag{2.18}$$

on $\mathbf{S}(2)$ gives rise to a metric distance, which we denote by $\rho(G,H)$, for $G,H \in \mathbf{S}(2)$. This metric is invariant under the right action of $GL(2,\mathbb{R})$ and makes $\mathbf{S}(2)$ isometric to the hyperbolic plane $\mathbb{D}$. In particular, $\mathbf{S}(2)$ is simply connected and complete. See [162, p. 518] for this computation in all dimensions.

Here we sketch in two dimensions an elementary argument showing that $\mathbf{S}(2)$ and $\mathbb{D}$ are isometric. We have already indicated how a given matrix in $\mathbf{S}(2)$ should correspond to a point in $\mathbb{D}$, namely,

$$G = \begin{bmatrix} g_{11} & g_{12} \\ g_{12} & g_{22} \end{bmatrix} \mapsto \mu = \frac{g_{11} - g_{22} - 2ig_{12}}{g_{11} + g_{22} + 2} \tag{2.19}$$

Clearly this map is continuous. We next compute that

$$\begin{aligned} (\mathrm{tr}(G)+2)^2|\mu|^2 &= (g_{11}-g_{22})^2 + 4g_{12}^2 = (g_{11}+g_{22})^2 - 4 \\ &= \mathrm{tr}^2(G) - 4 \end{aligned}$$

Thus $|\mu| < 1$, and two matrices G and H have the same image only if

$$\frac{\mathrm{tr}(G)-2}{\mathrm{tr}(G)+2} = \frac{\mathrm{tr}(H)-2}{\mathrm{tr}(H)+2},$$

which implies that $\mathrm{tr}(G) = \mathrm{tr}(H)$ as both $\mathrm{tr}(G), \mathrm{tr}(H) \geqslant 2$. It is now quite clear in view of (2.19) that the map $G \mapsto \mu$ is an injection. Next, the elementary observation that $\mathrm{tr}(G) = |G| + |G|^{-1}$ implies

$$|\mu| = \frac{|G|-1}{|G|+1}, \qquad |G| = \frac{1+|\mu|}{1-|\mu|}$$

where we continue to use $|\cdot|$ to denote the operator norm (that is, the largest singular value) of a matrix.

On the diagonal matrices

$$\Lambda = \begin{bmatrix} \lambda & 0 \\ 0 & 1/\lambda \end{bmatrix},$$

we can compute the metric (2.18) directly,

$$\rho(\mathbf{I}, \Lambda) = \big| \log \lambda \big|$$

Because any $G \in \mathbf{S}(2)$ can be diagonalized and because of the invariance of the metric under the action in (2.17) we have

$$\rho(\mathbf{I}, G) = \rho(\mathbf{I}, O^t \Lambda O) = \rho(\mathbf{I}, \Lambda) = \log |G| \tag{2.20}$$

As the hyperbolic metric $\rho_{\mathbb{D}}$ on the unit disk $\mathbb{D}$ is given by the Riemannian metric $ds = 2|dz|/(1-|z|^2)$, we have the formula

$$\rho_{\mathbb{D}}(0, \mu) = \log \frac{1+|\mu|}{1-|\mu|} = \log |G| \tag{2.21}$$

The two formulas (2.20) and (2.21) show that the map $G \mapsto \mu$ is an isometry on geodesic rays from the identity to geodesic lines passing through the origin and is therefore onto (it is not too difficult to invert this map, however, we shall do it later in a more general setting).

To show that (2.16) provides a global isometry between $\mathbf{S}(2)$ and $\mathbb{D}$, we need the following lemma.

Lemma 2.3.1. *Suppose that under the correspondence (2.16) we have $G \mapsto \mu$ and $H \mapsto \nu$, where $G, H \in \mathbf{S}(2)$. Then*

$$X \mapsto \frac{\mu - \nu}{1 - \overline{\mu}\nu},$$

where $X = \sqrt{G} H^{-1} \sqrt{G} = \sqrt{G}[H^{-1}]$.

Proof. First, we find the square root of the positive definite matrix G. This is a positive definite matrix $\sqrt{G}$ such that $\sqrt{G} \cdot \sqrt{G} = G$. By direct calculation

$$\sqrt{G} = \frac{1}{\sqrt{\mathrm{tr}(G)+2}} \begin{bmatrix} g_{11}+1 & g_{12} \\ g_{12} & g_{22}+1 \end{bmatrix} = \frac{1}{\sqrt{\mathrm{tr}(G)+2}} \left(\begin{bmatrix} g_{11} & g_{12} \\ g_{12} & g_{22} \end{bmatrix} + \mathbf{I} \right)$$

Next, given the matrices $G, H \in \mathbf{S}(2)$, we see that

$$X = \sqrt{G}H^{-1}\sqrt{G} = \frac{1}{\mathrm{tr}(G)+2}\left(GH^{-1}G + GH^{-1} + H^{-1}G + H^{-1}\right)$$

Therefore

$$\begin{aligned}
(\mathrm{tr}(G)+2)x_{11} &= g_{12}^2 h_{11} + (1+g_{11})(h_{22} + g_{11}h_{22} - 2g_{12}h_{12}) \\
(\mathrm{tr}(G)+2)x_{22} &= (1+g_{22})^2 h_{11} + g_{12}^2 h_{22} - 2(1+g_{22})g_{12}h_{12} \\
(\mathrm{tr}(G)+2)x_{12} &= g_{12}(h_{11} + g_{22}h_{11} + h_{22} + g_{11}h_{22}) - h_{12}(g_{11} + g_{22} + 2g_{11}g_{22})
\end{aligned}$$

Moreover,

$$\mathrm{tr}(X) + 2 = g_{11}h_{22} + g_{22}h_{11} - 2g_{12}h_{12} + 2$$

$$\frac{x_{11} - x_{22}}{x_{11} + x_{22} + 2} =$$

$$\frac{(2 + g_{11}(2 + g_{11} - g_{22}))h_{22} - (2 + g_{22}(2 - g_{11} + g_{22}))h_{11} + 2(g_{22} - g_{11})g_{12}h_{12}}{(g_{11} + g_{22} + 2)(g_{11}h_{22} + g_{22}h_{11} - 2g_{12}h_{12} + 2)}$$

and

$$\frac{2x_{12}}{x_{11} + x_{22} + 2} =$$

$$\frac{2g_{12}(h_{11} + g_{22}h_{11} + h_{22} + g_{11}h_{22}) - 2h_{12}(g_{11} + g_{22} + 2g_{11}g_{22})}{(g_{11} + g_{22} + 2)(g_{11}h_{22} + g_{22}h_{11} - 2g_{12}h_{12} + 2)}$$

We have identified G with μ and similarly H with ν. We may hence compute $\frac{\mu-\nu}{1-\overline{\mu}\nu}$ using (2.19). On the other hand, we calculated the corresponding expressions for the matrix $X = \sqrt{G}H^{-1}\sqrt{G}$ above. Comparing the results, we find that indeed

$$X \mapsto \frac{\mu - \nu}{1 - \overline{\mu}\nu}$$

under our proposed isometry. □

Of course, the map $z \mapsto (\mu - z)/(1 - \overline{\mu}z)$ is an isometry of the hyperbolic plane. Thus the previous lemma gives

$$\begin{aligned}
\rho_{\mathbb{D}}(\mu, \nu) &= \rho_{\mathbb{D}}\left(0, \frac{\mu - \nu}{1 - \overline{\mu}\nu}\right) \\
&= \rho(\mathbf{I}, X) = \rho(\mathbf{I}, \sqrt{G}[H^{-1}]) \\
&= \rho(\mathbf{I}, \sqrt{G}^{-1}[H]) \\
&= \rho(\sqrt{G}[\mathbf{I}], H) = \rho(G, H)
\end{aligned}$$

The second-to-last equality holds by virtue of the isometric $GL(2, \mathbb{R})$ action and the last equality as $\sqrt{G}[\mathbf{I}] = G$. We have proved the following theorem.

Theorem 2.3.2. *The map $G \mapsto \mu$ induces an isometry between the space $\mathbf{S}(2)$ with the metric $ds^2 = \frac{1}{2}(Y^{-1}dY)^2$ and the hyperbolic plane $(\mathbb{D}, \rho_{\mathbb{D}})$.*

2.4 The Linear Distortion

Given a homeomorphism $f : \Omega \to \Omega'$, we may introduce a quantity that measures the deviation from f to a conformal mapping.

Definition 2.4.1. *The* linear distortion *of f is the measurable function defined by*

$$H(z,f) = \limsup_{r\to 0} \frac{\max_{|\zeta|=r} |f(z+\zeta)-f(z)|}{\min_{|\zeta|=r} |f(z+\zeta)-f(z)|} \tag{2.22}$$

The first example to consider is that of the linear mapping, which we write conveniently in the complex notation

$$f(z) = az + b\bar{z}$$

Indeed, we have decomposed the $\mathbb{R}$-linear mapping f as the sum of a complex linear operator $f_+(z) = az$ and a complex antilinear operator $f_-(z) = b\bar{z}$. A moment's thought gives

$$\begin{aligned} H(z,f) &= \frac{|a|+|b|}{|a|-|b|} \\ J(z,f) &= |a|^2 - |b|^2 \end{aligned}$$

In the first identity we have implicitly assumed that f is orientation-preserving, for example, that $J(z,f) > 0$ or that $|a| > |b|$.

More generally then, if f has a derivative $Df(z)$ at z, we may use the complex differential operators

$$\frac{\partial f}{\partial z} = f_z = \frac{1}{2}(f_x - i\, f_y), \qquad \frac{\partial f}{\partial \bar{z}} = f_{\bar{z}} = \frac{1}{2}(f_x + i\, f_y)$$

and write the derivative as

$$Df(z)\, h = \frac{\partial f}{\partial z}(z)\, h + \frac{\partial f}{\partial \bar{z}}(z)\, \bar{h}, \qquad h \in \mathbb{C} = \mathbb{R}^2$$

In particular, $h \mapsto f_z\,(z)\, h$ is the $\mathbb{C}$-linear part of $Df(z)$, and $h \mapsto f_{\bar{z}}\,(z)\, \bar{h}$ is its antilinear part. Then the norm of the derivative

$$|Df(z)| = \sup\{|Df(z)\, h| : |h| = 1\} = \left|\frac{\partial f}{\partial z}(z)\right| + \left|\frac{\partial f}{\partial \bar{z}}(z)\right| = |f_z| + |f_{\bar{z}}| \tag{2.23}$$

and the Jacobian

$$J(z,f) = \left|\frac{\partial f}{\partial z}(z)\right|^2 - \left|\frac{\partial f}{\partial \bar{z}}(z)\right|^2 = |f_z|^2 - |f_{\bar{z}}|^2 \tag{2.24}$$

In terms of the complex derivatives,

$$\begin{aligned} H(z,f) &= \frac{\max_{|\zeta|=1} |(D_\zeta f)(z)|}{\min_{|\zeta|=1} |(D_\zeta f)(z)|} = \frac{|f_z(z)| + |f_{\bar{z}}(z)|}{|f_z(z)| - |f_{\bar{z}}(z)|} \\ &= |Df(z)|^2 J(z,f)^{-1} \end{aligned} \tag{2.25}$$

at points where $J(z,f) > 0$. In particular, if f has a complex derivative at z, then $H(z,f) = 1$. Thus if f is *conformal* (that is, holomorphic and injective), we have $H(z,f) \equiv 1$. The reader may wish to attempt to establish the converse to this statement, assuming differentiability.

Suggested by (2.11), if f is differentiable almost everywhere, we may write

$$G_f(z) = J(z,f)^{-1} D^t f(z) Df(z)$$

whenever $J(z,f) > 0$. The quantity $G_f(z)$ is called the *distortion tensor* of the mapping f.

When $Df(z) = 0$, the mapping satisfies the condition (2.8), with $\phi = 0$, and thus at those points we may set $G_f(z) = \mathbf{I}$. However, where $Df(z)$ is nonzero but degenerate, that is, $J(z,f) = 0$ but $Df(z) \neq 0$, there is no meaningful definition for $G_f(z)$.

Assuming the set of all such degenerate points has measure zero, then $G_f : \Omega \to \mathbf{S}(2)$ is a measurable map and, in particular, f is a conformal mapping from (Ω, G_f) to $(\Omega', \mathbf{I})$. Moreover,

$$H(z,f) = |G_f(z)| = e^{\rho(G_f(z),\mathbf{I})} \tag{2.26}$$

Thus the deviation (2.22) from f to a conformal map is simply another measure of the distance between the conformal structure induced by f on Ω (namely, G_f) and the Euclidean structure $\mathbf{I}$.

Under the isometry $\mathbf{S}(2) \to \mathbb{D}$, the mapping f induces a mapping $G_f(z) \mapsto \mu_f(z)$. Now $\mu_f : \Omega \to \mathbb{D}$ will be a measurable map, and evidently,

$$H(z,f) = \frac{1 + |\mu_f(z)|}{1 - |\mu_f(z)|} \tag{2.27}$$

It is important to realize that equations (2.25)–(2.27) hold only where f has a derivative. However, the formula for the linear distortion makes sense in much more generality; for instance, one may compute that $z \mapsto z|z|^{-1/2}$ has linear distortion 1 at the origin where it certainly does not have a derivative. Nevertheless, one expects the boundedness of the linear distortion of a mapping to imply some differentiability and geometric properties of a mapping.

It is perhaps a little surprising that the theory of mappings of bounded linear distortion (or quasiconformal mappings) is so rich.

2.5 Quasiconformal Mappings

Quasiconformal mappings are principally mappings of “bounded distortion”. However, there are in fact many ways to measure the distortion of a mapping, and we shall consider a number of them. We consider first the geometric definition through the linear distortion function in (2.22). For clarity of exposition we use here the term mapping of bounded distortion, even if later these turn out to be precisely the quasiconformal mappings.

Mappings of Bounded Distortion

Definition 2.5.1. *A homeomorphism $f : \Omega \to \Omega'$ is called a* mapping of bounded distortion *if it is orientation-preserving and if its linear distortion is uniformly bounded,*

$$\sup_{z\in\Omega} H(z,f) < \infty \tag{2.28}$$

For a more detailed discussion on the concept of orientation-preserving mappings, see Section 2.8. It is important to note that in (2.28) uniform boundedness, and requiring this at every point instead of up to measure zero, is necessary to build a useful theory. For instance, requiring only that $H(z,f) < \infty$ for all $z \in \Omega$ does not give one the desirable compactness or regularity properties. It does motivate the notion of mappings of finite distortion, as discussed later in Chapter 20, when additional integrability conditions are imposed on $H(z,f)$, but without such extra properties not much can be said. Indeed, simply note here that for a diffeomorphism of a domain Ω we have $H(z,f) < \infty$ everywhere, while the space of diffeomorphisms of a domain has very few compactness properties.

It is a little more difficult to see that

$$\text{ess } \sup_{z\in\Omega} H(z,f) = \|H(z,f)\|_\infty \leqslant K < \infty \tag{2.29}$$

alone also does not imply good regularity for homeomorphisms. The problem here is basically with the differentiability properties of the mappings. Equation (2.28) implies that the homeomorphism f is differentiable almost everywhere, and in fact it can be shown that if f is assumed a priori to be in the Sobolev space $W^{1,1}(\Omega)$ of integrable functions whose first distributional derivatives are integrable, then conversely (2.29) implies $H(z,f)$ is uniformly bounded.

To see the problems that arise when we assume only (2.29), consider the following example. Let $u(x)$ denote the Cantor function defined on the real line. Thus $u(x)$ is continuous and increasing with $u'(x) = 0$ almost everywhere. Set

$$f(z) = x + u(x) + iy, \qquad z = x + iy$$

Then f is a homeomorphism $\mathbb{C} \to \mathbb{C}$ and almost everywhere f is differentiable with

$$Df(z) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

In particular, the Jacobian determinant $J(z,f) = 1$ almost everywhere. We compute that $H(z,f) = 1$ almost everywhere, so

$$\operatorname*{ess\,sup}_{z\in\mathbb{C}} H(z,f) = 1$$

Further, the distortion tensor $G_f(z) = \mathbf{I}$ almost everywhere. However, f is not conformal; it is not even of Sobolev class. The change-of-variables formula does not hold (f does not preserve the measure of sets even though $J(z,f) = 1$ almost everywhere). The reader may wish to discover other nasty properties of this homeomorphism. This example reinforces the distinction between functions that are merely differentiable almost everywhere and Sobolev functions. (The reader who is unfamiliar with Sobolev space theory will find the requisite definitions and so forth in the Appendix.)

While the definition (2.28) of mappings of bounded distortion is aesthetically pleasing, in practice it is very difficult to work with. Also, it requires the a priori knowledge that one is working with an injective mapping (at least locally). It is for such reasons that one turns to the analytic definition.

Here we return to the identities (2.25). At points z where f is differentiable with positive Jacobian, we see that the linear distortion $H(z,f)$ is precisely the ratio of the largest and the smallest directional derivatives at z. This interpretation leads to the following definition of a quasiconformal mapping.

Definition of a Quasiconformal Map

Definition 2.5.2. *A homeomorphism $f : \Omega \to \Omega'$ is called K-*quasiconformal *if it is orientation-preserving, if*

$$f \in W^{1,2}_{loc}(\Omega), \tag{2.30}$$

and if the directional derivatives satisfy

$$\max_{\alpha} |\partial_\alpha f(z)| \leqslant K \min_{\alpha} |\partial_\alpha f(z)| \tag{2.31}$$

for almost every $z \in \Omega$.

Often it is convenient to formulate (2.31) as the distortion inequality

$$|Df(z)|^2 \leqslant KJ(z,f) \quad \text{for almost every } z \in \Omega$$

The smallest constant $K = K(f)$ for which (2.31) holds almost everywhere is called the *distortion* of the mapping f.

The above formulation of a quasiconformal mapping in terms of Sobolev spaces is the most useful and flexible, allowing estimates, in taking limits and so on. There are, however, a few subtleties we need to discuss. The requirement $f \in W^{1,2}_{loc}(\Omega)$ implies that f has partial derivatives f_x and f_y almost everywhere (see the Appendix), but being a Sobolev function is not enough for f to be differentiable almost everywhere. For (2.31) to be meaningful one need only set

$$\partial_\alpha f\,(z) = \cos(\alpha)\, f_x\,(z) + \sin(\alpha)\, f_y\,(z), \qquad \alpha \in [0, 2\pi], \tag{2.32}$$

then for almost all $z \in \Omega$, condition (2.31) is well defined as written.

The condition ties together the partial derivatives of f and hence provides geometric information on the mapping. As such it is enough to start the development of the theory. The reader should note, though, that in the next chapter we will show that all *homeomorphic* Sobolev functions are differentiable almost everywhere; with this result the directional derivatives retain their usual meaning,

$$\partial_\alpha f\,(z) = \lim_{r\to 0} \frac{f(z + r\,e^{i\alpha}) - f(z)}{r}$$

Definition of a Quasiregular Map

Now we also want to give up the hypothesis that f is injective, so as to be able to define a wider and more flexible class of mappings particularly useful in the study of planar elliptic PDEs. This is quite analogous to moving to analytic functions from conformal mappings.

We shall require only that the mapping $f \in W^{1,2}_{loc}(\Omega)$, that it is orientation-preserving, so $J(z,f) = |f_z|^2 - |f_{\bar z}|^2 \geqslant 0$ almost everywhere, and that it satisfies the conditions (2.31) and (2.32) on partial derivatives. We shall call these mappings K-*quasiregular.*

The reader may wonder about the choice of the Sobolev regularity $W^{1,2}_{loc}$ in the definitions of quasiconformal and quasiregular mappings. For the homeomorphic $W^{1,1}_{loc}(\Omega)$-maps, this regularity follows automatically from the distortion inequality (2.31) since, by Corollary 3.3.6 from the next chapter, for any compact subset $A \subset \Omega$,

$$\int_A |Df|^2 \leqslant K(f) \int_A J(z,f) \leqslant K(f)|f(A)| < \infty \tag{2.33}$$

However, the last estimate fails for general nonhomeomorphic Sobolev mappings, and it is for these that the precise regularity $f \in W^{1,2}_{loc}(\Omega)$ is necessary. This condition will guarantee the local integrability of the Jacobian, the key property in the geometric study of mappings.

Before dwelling further on these notions, let us make it clear that quasiconformal mappings are precisely the mappings of bounded distortion. In the literature the conditions (2.30) and (2.31) taken together are called the *analytic definition* of a quasiconformal mapping while (2.28) is termed the *geometric definition.*

For a diffeomorphism the equivalence of these two definitions is clear, as the reader may quickly verify. The general case lies much deeper. That every quasiconformal mapping satisfies (2.28) will be shown in Theorem 3.6.2. For the converse direction, in fact much less than (2.28) is required. The recent surprising result of Heinonen and Koskela [160] allows the lim sup condition in the definition of the linear distortion to be replaced by a lim inf condition. It is only through their deep studies of analysis on metric spaces [161] that these and other properties of quasisymmetric mappings have come to light.

The lim inf theorem

Theorem 2.5.3. *Let $f : \Omega \to \Omega'$ be an orientation-preserving homeomorphism between planar domains. If there is $H < \infty$ such that for every $z \in \Omega$*

$$\liminf_{r\to 0} \frac{\max_{|\zeta|=r} |f(z+\zeta) - f(z)|}{\min_{|\zeta|=r} |f(z+\zeta) - f(z)|} \leqslant H, \tag{2.34}$$

then f is quasiconformal.

The proof of this result will take us rather too far from the themes of this book as it uses ideas from the theory of quasisymmetric mappings of metric spaces and the moduli of curve families. The theorem was further refined in [205] where exceptional sets of zero length are allowed. A key point here is that in applications, such as in conformal dynamics (see for instance [303]), at each point one need only find estimates or control the geometry on some sequence of radii tending to 0 as opposed to every sequence tending to 0. It would be very good to have an analytic proof for this result at hand.

Problem. Give an analytic proof for Theorem 2.5.3.

The Beltrami Equation

The previous section relates in (2.25) the linear distortion to the complex derivatives at points where the mappings are differentiable with positive Jacobian. A similar analysis can be made for the analytic definition (2.30) and (2.31) of a quasiconformal mapping. These complex analytic reductions have far-reaching consequences.

A local analysis again leads to consideration of the linear (over $\mathbb{R}$) mappings $f : z \mapsto az + b\bar{z}$. That f is orientation-preserving implies $|b| \leqslant |a|$. We leave it to the reader to verify that $\max_\alpha |\partial_\alpha f(z)| = |a|+|b| = |Df(z)|$ and $\min_\alpha |\partial_\alpha f(z)| = |a| - |b| = J(z, f)/|Df(z)|$ when $|b| \leqslant |a|$, as this calculation is more or less a repetition of that given to find (2.25). As a consequence, under the interpretation (2.32), the distortion inequality (2.31) achieves the form

$$\left|\frac{\partial f}{\partial z}\right| + \left|\frac{\partial f}{\partial \bar{z}}\right| \leqslant K \left(\left|\frac{\partial f}{\partial z}\right| - \left|\frac{\partial f}{\partial \bar{z}}\right| \right),$$

an inequality equivalent to

$$\left|\frac{\partial f}{\partial \bar{z}}\right| \leqslant \frac{K-1}{K+1} \left|\frac{\partial f}{\partial z}\right| \tag{2.35}$$

Furthermore, write $\mu(z) = f_{\bar{z}}(z)/f_z(z)$ when $f_z(z) \neq 0$ and, say, $\mu(z) = 0$ otherwise. This expresses the inequality (2.35) as a linear partial differential equation. We have thus shown the following theorem.

Theorem 2.5.4. *Suppose $f : \Omega \to \Omega'$ is a homeomorphic $W^{1,2}_{loc}$-mapping. Then f is K-quasiconformal if and only if*

$$\frac{\partial f}{\partial \bar{z}}(z) = \mu(z)\frac{\partial f}{\partial z}(z) \quad \textit{for almost every } z \in \Omega, \tag{2.36}$$

where μ, called the Beltrami coefficient of f, is a bounded measurable function satisfying

$$\|\mu\|_\infty \leqslant \frac{K-1}{K+1} < 1$$

Often in the literature the term *complex dilatation of f* is used for the Beltrami coefficient μ. We shall later see for quasiconformal mappings that $f_z \neq 0$ almost everywhere. Thus the Beltrami coefficient is uniquely defined up to a set of measure zero. The reader should also note the relations

$$\|\mu\|_\infty = \frac{K(f)-1}{K(f)+1} \quad \text{and} \quad K(f) = \frac{1+\|\mu\|_\infty}{1-\|\mu\|_\infty}$$

From Theorem 2.5.4 and Weyl's Lemma A.6.10 we have a quick proof showing that for a quasiconformal mapping f the distortion $K(f) = 1$ if and only if f is conformal, that is, holomorphic and injective. If we need to give up the hypothesis that f is injective, then we see that f is 1-quasiregular if and only if it is an analytic function.

The differential equation (2.36) is called the *Beltrami equation*. It is this equation that provides the connections from the geometric theory of quasiconformal mappings to complex analysis and to elliptic PDEs. Understanding and utilizing these relations underlie many of the themes of this monograph.

It is in the setting described above that planar quasiconformal mappings were first studied around 1928 by Grötzsch. The term "quasiconformal" was coined by Ahlfors in 1935 [1, 2] when this class of mappings proved to be an integral tool in his geometric development of Nevanlinna theory based on the "length-area" method. Teichmüller found a fundamental connection between quasiconformal mappings and quadratic differentials in his studies on extremal mappings between Riemann surfaces [350] around 1939. Developments of the length-area method led to the definition of quasiconformal mappings in terms of the distortion of the modulus of curve families by Pfluger [295]. These were systematically studied in their own right by Ahlfors from 1953 [3].

The class of quasiconformal diffeomorphisms is not closed under uniform limits. Thus the generalization to Sobolev spaces is absolutely necessary if one is to solve various extremal problems by taking limits. After making this generalization we will find that the limit of a bounded sequence of quasiconformal mappings is either quasiconformal or constant. Therefore it is in this setting that the class of quasiconformal mappings becomes more flexible and has a greater range of applications.

The equivalence between the geometric definition and the analytic definition in the Sobolev setting was shown by Gehring and Lehto in 1959 [137]. This is a relatively deep fact that we shall explore to various extents in this book, giving new and quite different proofs from a modern perspective through the notions of quasisymmetry and also through the theory of holomorphic motions. The connection among quasiconformal mappings, Teichmüller theory and quadratic differentials has been intensively investigated by Ahlfors-Bers, Reich-Strebel and Lehto and others; see [339] and [228] and the references therein.

There are two further routes to the theory of planar quasiconformal mappings. These are via the conformal modulus, the approach taken in Lehto-Virtanen's classic text [229], and the modern approach via "holomorphic motions" due to Sullivan [237]. We shall discuss this last approach quite extensively later.

2.6 Radial Stretchings

There is a class of examples that it is important to have at hand as they typically provide extremal examples. These are the radial stretchings, mappings $f : \mathbb{D}(0,R) \to \mathbb{C}$ of the form

$$f(z) = \frac{z}{|z|}\,\rho(|z|)\,, \qquad f(0) = 0 \tag{2.37}$$

Here the function $t \mapsto \rho(t) > 0$, $0 \leqslant t < R$, is assumed to be continuous and strictly increasing. For $\rho(0) = 0$ the mapping f is continuous at the origin.

Basic examples include $\rho(t) = t^K$ and $\rho(t) = t^{1/K}$ giving rise to the mappings

$$f_1(z) = z|z|^{K-1}$$

and

$$f_2(z) = z|z|^{\frac{1}{K}-1},$$

respectively. These are the standard radial stretchings we shall see frequently, for they arise as extremals for the problems on Hölder continuity, and integrability of the differential, for quasiconformal mappings. For future reference note here that $f_1 = f_2^{-1}$.

We may calculate the differential and distortions of a radial stretching f directly from the definition at points where the derivative $\dot{\rho}$ exists. In complex notation, using the simple identity $\partial_{\bar z}|z| = z\,(2|z|)^{-1}$, we have

$$\frac{\partial f}{\partial z}(z) = \frac{1}{2}\left[\dot{\rho}(|z|) + \frac{\rho(|z|)}{|z|}\right] \tag{2.38}$$

$$\frac{\partial f}{\partial \bar z}(z) = \frac{1}{2}\,\frac{z}{\bar z}\left[\dot{\rho}(|z|) - \frac{\rho(|z|)}{|z|}\right] \tag{2.39}$$

Thus we obtain

$$|Df(z)| = \left|\frac{\partial f}{\partial z}\right| + \left|\frac{\partial f}{\partial \bar z}\right| = \max\left\{\dot{\rho}(|z|), \frac{\rho(|z|)}{|z|}\right\} \tag{2.40}$$

$$J(z,f) \;=\; \left|\frac{\partial f}{\partial z}\right|^2 - \left|\frac{\partial f}{\partial \bar z}\right|^2 \;=\; \frac{\rho(|z|)\dot\rho(|z|)}{|z|} \tag{2.41}$$

It is easy to see that for any radial stretching $H(0,f)=1$. Away from 0 we see that

$$H(z,f) \;=\; |Df(z)|^2 J(z,f)^{-1} \;=\; \max\left\{\frac{|z|\dot\rho(|z|)}{\rho(|z|)}, \frac{\rho(|z|)}{|z|\dot\rho(|z|)}\right\}$$

In particular, for $f(z)=z|z|^{K-1}$, $K>0$, we have

$$H(z,f)=\max\{|K|,|K|^{-1}\}$$

Furthermore, we find that f satisfies the complex Beltrami equation

$$f_{\bar z}=\mu(z)f_z,$$

where the *Beltrami coefficient* μ given by

$$\mu(z)=\frac{z}{\bar z}\,\frac{|z|\dot\rho(|z|)-\rho(|z|)}{|z|\dot\rho(|z|)+\rho(|z|)} \tag{2.42}$$

Hence

$$K(z)=\frac{1+|\mu(z)|}{1-|\mu(z)|}=\max\left\{\frac{|z|\dot\rho(|z|)}{\rho(|z|)}, \frac{\rho(|z|)}{|z|\dot\rho(|z|)}\right\} \tag{2.43}$$

Again, if $f(z)=z|z|^{K-1}$, we have

$$J(z,f)=K|z|^{2(K-1)}, \qquad |Df(z)|=K|z|^{K-1} \tag{2.44}$$

and

$$\mu_f(z)=\frac{K-1}{K+1}\,\frac{z}{\bar z}$$

More generally, if f is a radial stretching, then $J(z,f)$ is always locally integrable. Indeed,

$$\int_{|z|\leqslant\lambda} J(z,f)=2\pi\int_0^\lambda \rho(t)\dot\rho(t)dt=\pi\rho^2(\lambda),$$

the last term being the area of the disk of radius $\rho(\lambda)=f(\lambda)$.

Naturally, the above expressions can be found in real notation, too. For instance, the full Jacobian matrix has the form

$$Df(z)=\frac{\rho(|z|)}{|z|}\mathbf{I}+\Big(\dot\rho(|z|)-\frac{\rho(|z|)}{|z|}\Big)\frac{z\otimes z}{|z|^2},$$

where we have used (and will use elsewhere) the shorthand notation

$$z\otimes z=\begin{bmatrix} x^2 & xy \\ xy & y^2\end{bmatrix}$$

for $z=x+iy$. For further details and identities in this notation see, e.g., [191].

2.7 Hausdorff Dimension

In this section we give the basic definition and properties of Hausdorff dimension. Hausdorff dimension and its close relatives play an important role in planar complex analysis, providing a way of measuring the size of quite general sets. A good reference is Mattila's book [251] on geometric measure theory. The basic definitions and foundational results for the theory are due to Carathéodory [88] and Hausdorff [158] in the second decade of the 20^{th} century.

We shall often meet fractal sets with nonintegral dimension in the theory of quasiconformal mappings. Particularly, they often occur as dynamically defined invariant subsets of the plane (for instance, limit sets of Kleinian groups and Julia sets of rational maps). Self-similarity is often a natural property of dynamically defined sets. Quasiconformal mappings are one of the natural geometric tools to study sets of nonintegral dimension, as it is under quasiconformal deformations of conformal dynamical systems that the invariant sets change their dimensions. Thus a substantial part of our later studies will be aimed at trying to estimate just how much a quasiconformal mapping can change the dimension of a planar set. In fact, we shall be able to present optimal bounds later; see Theorem 13.2.10. It may come as a bit of a surprise to the reader unfamiliar with such things, but distortion of dimension is intimately connected with integrability properties of the derivatives of a mapping.

We begin with a fairly general, but not the most general, construction.

Let $\mathcal{F}$ be the family of all Borel subsets of $\mathbb{C}$ and $\eta : [0,\infty] \to [0,\infty]$ an increasing homeomorphism. For any $0 < \delta \leqslant \infty$ and $A \subset \mathbb{C}$ we set

$$\mathcal{H}_{\eta,\delta}(A) = \inf \left\{ \sum_{i=1}^{\infty} \eta(\operatorname{diam}(E_i)) \ : \ A \subset \bigcup_{i=1}^{\infty} E_i, \quad E_i \in \mathcal{F}, \quad \operatorname{diam}(E_i) < \delta \right\}$$

The function $\mathcal{H}_{\eta,\delta}$ is monotonic and subadditive. Notice that for $A \subset \mathbb{C}$ we have $\mathcal{H}_{\eta,\delta_1}(A) \leqslant \mathcal{H}_{\eta,\delta_2}(A)$ if $\delta_2 < \delta_1$. Hence we can define $\mathcal{H}_\eta$ on subsets A of $\mathbb{C}$ by

$$\mathcal{H}_\eta(A) = \lim_{\delta \searrow 0} \mathcal{H}_{\eta,\delta}(A)$$

The following is an exercise for the reader, see [251] for a proof.

Theorem 2.7.1. *The function $\mathcal{H}_\eta$ defines a Borel regular measure on $\mathbb{C}$.*

s-Dimensional Hausdorff Measure

If we put $\eta(t) = t^s$, the resulting Borel measure is called the s-dimensional Hausdorff measure, denoted $\mathcal{H}^s$. This measure is not locally finite on $\mathbb{C}$ unless $s \geqslant 2$ and is therefore not a Radon measure. However, if $A \subset \mathbb{C}$ is such that for some $s \leqslant 2$ $\mathcal{H}^s(A) < \infty$, then upon restricting to A we do obtain such a measure. That is, $\mathcal{H}^s|A$ is a Radon measure.

Integral-dimensional Hausdorff measures of course play a central role. $\mathcal{H}^0(A)$ is just the cardinality of A. When $s = 1$, the measure $\mathcal{H}^1$ is a generalized length measure. For a rectifiable curve α we can show that $\mathcal{H}^1(\alpha)$ is simply the length of α, while for $s = 2$ we obtain (up to a constant scaling factor) the usual Lebesgue measure. Note that for any $s > 2$ and $A \subset \mathbb{C}$, $\mathcal{H}^s(A) \equiv 0$.

There is no great dependence of Hausdorff measure on the type of covering sets in $\mathcal{F}$. For instance, one may restrict oneself only to covers by convex sets without affecting the measure. If, however, one asks for optimal coverings by sets (say disks) of the *same* diameter, then one is led to the study of Minkowski measures and dimensions; see for instance [251].

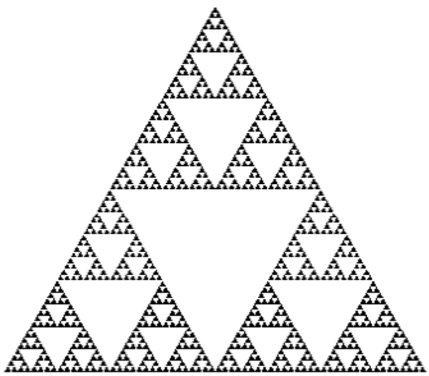
The next result we want to quote, [251, Theorem 4.7], quickly leads to the central concept of Hausdorff dimension.

Theorem 2.7.2. *Let* $0 \leqslant s < t < \infty$ *and* $A \subset \mathbb{C}$. *Then*

- $\mathcal{H}^s(A) < \infty$ *implies* $\mathcal{H}^t(A) = 0$, *and*
- $\mathcal{H}^t(A) > 0$ *implies* $\mathcal{H}^s(A) = \infty$.

We may make use of Theorem 2.7.2 to define Hausdorff dimension,

Hausdorff Dimension



A Sierpiński gasket of dimension $\dim_{\mathcal{H}} = \frac{\log 3}{\log 2}$

The Hausdorff dimension of a set $A \subset \mathbb{C}$ is defined as

$$\dim_{\mathcal{H}}(A) = \sup\{s : \mathcal{H}^s(A) > 0\} = \inf\{t : \mathcal{H}^t(A) < \infty\}$$

Of course, the Hausdorff dimension of a finite set is 0, of a rectifiable curve it is 1 and for an open set it is 2. The topology of a set has implications for its dimension. For instance, any connected nondegenerate subset of $\mathbb{C}$ has dimension at least 1. However, totally disconnected closed sets, usually called Cantor sets, can have any dimension from 0 to 2. Readers may wish to construct such examples for themselves.

Next we record a few simple but useful properties:

- If $A \subset B \subset \mathbb{C}$, then $\dim_{\mathcal{H}}(A) \leqslant \dim_{\mathcal{H}}(B) \leqslant 2$
- For $A_1, A_2, \ldots \subset \mathbb{C}$, $\dim_{\mathcal{H}}(\cup_i A_i) = \sup_i \dim_{\mathcal{H}}(A_i)$.
- If $s < \dim_{\mathcal{H}}(A)$, then $\mathcal{H}^s(A) = \infty$.
- If $t > \dim_{\mathcal{H}}(A)$, then $\mathcal{H}^t(A) = 0$.

Note also that Lipschitz maps, those mappings $f : \Omega \to \mathbb{C}$ satisfying the estimate

$$|f(z) - f(w)| \leqslant L|z - w| \tag{2.45}$$

for some $L > 0$ and all $z, w \in \Omega$, do not increase the dimension. In particular, Hausdorff dimension will be preserved under bilipschitz mappings—those mappings for which both f and f^{-1} are Lipschitz.

2.8 Degree and Jacobian

There are two basic approaches to the notion of local degree for a mapping, the algebraic (see for instance Dold [107]) and the analytic (see for instance Lloyd [234]). Both these notions try to capture the idea of counting the number of solutions $z \in \Omega$ to the equation $f(z) = w$. The reader will be familiar with the winding number of a piecewise smooth closed curve $\gamma : [0, 1] \to \mathbb{C}$,

$$W_\gamma(a) = \frac{1}{2\pi i} \int_\gamma \frac{dz}{z - a}$$

If Ω is bounded by a piecewise smooth Jordan curve γ and $w \notin f(\partial\Omega)$, $a \in \Omega$, then the degree of f at w is simply

$$\deg_\Omega(f, w) = W_\gamma(a) \cdot W_{f\circ\gamma}(w),$$

where $W_\gamma(a) = \pm 1$ accounting for the two possible orientations of the boundary and is constant for $a \in \Omega$. In fact, the winding number is constant on components of $\mathbb{C} \setminus f(\partial\Omega)$. Analytically the degree can be defined as

$$\deg_\Omega(f, w) = \sum_{z \in f^{-1}(w)} \operatorname{sgn} J(z, f)$$

Remark. For a continuous map $f : \Omega \to \mathbb{C}$ and $p \in f(\Omega)$ with $f^{-1}(p)$ compact, the local topological degree is then generally defined as the unique integer representing the homomorphism $H_2(\hat{\mathbb{C}}) \to H_2(\hat{\mathbb{C}}) \cong \mathbb{Z}$ induced on the second homology by the excision isomorphism; see [107, Definition 5.1]. Any map $g : \hat{\mathbb{C}} \to \hat{\mathbb{C}}$ has a well-defined degree independent of a chosen point, so we simply write $\deg(g)$ for such maps.

These two notions are equivalent for sufficiently regular maps and are more generally proven to be equivalent via approximation. The topological definition has the virtue of clearly defining a homotopy invariant notion, while the analytic definition is computable. We recall the basic facts concerning degree.

Theorem 2.8.1. *Let $f : \Omega \to \mathbb{C}$ be continuous, $p \in \mathbb{C} \setminus f(\partial\Omega)$ and $f^{-1}(p)$ compact.*

- *The degree $\deg_\Omega(f, p)$ is constant on components of $\mathbb{C} \setminus f(\partial\Omega)$. That is, if p_1 and p_2 lie in the same component of $\mathbb{C} \setminus f(\partial\Omega)$, then $\deg_\Omega(f, p_1) = \deg_\Omega(f, p_2)$* [234, Theorem 2.1.3].
- *If $H : [0, 1] \times \Omega \to \mathbb{C}$ is a homotopy, Ω a bounded domain, and $w_t = H(t, z_0) \notin H(t, \partial\Omega)$ for $0 \leqslant t \leqslant 1$, then $\deg_\Omega(H_t, w_t)$ is independent of t* [234, Theorem 2.2.4].
- *If f is a homeomorphism, then $\deg_\Omega(f, z) = \pm 1$ for all $z \in \Omega$* [107, IV.5].
- *Degree is multiplicative. Suppose $g : \hat{\mathbb{C}} \to \hat{\mathbb{C}}$ is continuous, then* [107, IV.5]

$$\deg_{g^{-1}(\Omega)}(f \circ g, g(p)) = \deg_\Omega(f, p) \deg(g)$$

We say that a homeomorphism f is *orientation-preserving* if $\deg_\Omega(f, p) \equiv 1$.

For a linear mapping $A : \mathbb{C} \to \mathbb{C}$, it is easy to check (from any of the above definitions) that

$$\deg_\Omega(A, p) = \deg(A) = \operatorname{sign} J(p, A) \in \{\pm 1\}, \qquad p \in A(\Omega) \tag{2.46}$$

We note the following corollary.

Corollary 2.8.2. *Let $f : \Omega \to \mathbb{C}$ and suppose that f is differentiable at z_0 with Jacobian determinant $J(z_0, f) \neq 0$. Put $w = f(z_0)$. Then there is $\varepsilon > 0$ so that*

$$\deg_{\mathbb{D}(z_0,r)}(f, w) = \operatorname{sgn} J(z_0, f), \qquad r < \varepsilon$$

Proof. We assume $z_0 = f(z_0) = 0$ and $A = Df(0)$ is a linear map of determinant $J(0, f)$. The Taylor approximation shows that $|f(z) - Az| < o(|z|)$, and so there is $\varepsilon > 0$ such that $|z| < \varepsilon$ implies $|f(z)| > c|z|$, where c depends on $\|A^{-1}\|$. Thus $0 \notin f(\partial\mathbb{D}(0, r))$ and $0 = f^{-1}(f(0))$ for $f|\mathbb{D}(0, r)$, $r < \varepsilon$. Now the homotopy $H(t, z) = \frac{1}{t} f(tz)$ connects the mapping f to the linear transformation A, and the Lipschitz estimate shows $0 \notin H(t, \partial\mathbb{D}(0, r))$. The result now follows by homotopy invariance and (2.46). □

In particular, we see that if a homeomorphism f is differentiable at z_0 with Jacobian determinant $J(z_0, f) \neq 0$, then f is orientation-preserving if and only if $J(z_0, f) > 0$.

2.9 A Background in Complex Analysis

Throughout this book we shall largely be using complex analytic methods. This section is intended to provide, without much proof, some of the basic facts we shall use concerning conformal mappings and so forth.

2.9.1 Analysis with Complex Notation

It will often be convenient to handle with the complex notation also the basic results needed from real analysis. For instance, we have already used the fact that if f is differentiable at a point z, the differential can be written as

$$Df(z)\,h = \frac{\partial f}{\partial z}(z)\,h + \frac{\partial f}{\partial \bar z}(z)\,\overline{h}$$

Furthermore, the operator norm of the derivative takes the form in (2.23) and the Jacobian determinant has the form in (2.24).

Similarly, we may identify the $\mathbb{C}$-linear and the antilinear parts of the derivative $D(f \circ g)(z)$ of a composed function. Consequently, the chain rule takes the form

$$(f \circ g)_z\,(z) = f_w\,(g(z))\; g_z(z) + f_{\overline{w}}\,(g(z))\; \overline{g_{\bar z}\,(z)} \tag{2.47}$$
$$(f \circ g)_{\bar z}\,(z) = f_w\,(g(z))\; g_{\bar z}(z) + f_{\overline{w}}\,(g(z))\; \overline{g_z\,(z)} \tag{2.48}$$

whenever the appropriate pointwise derivatives exist. Moreover, observing that $\partial_z\, z = \partial_{\bar z}\, \bar z = 1$ and $\partial_{\bar z}\, z = \partial_z\, \bar z = 0$, we see that any partial differential operator on $\mathbb{C}$ can be written in terms of ∂_z and $\partial_{\bar z}$ only.

If f is a homeomorphism differentiable at a point z with $Df(z)$ invertible, many times it will be useful to calculate in the complex notation the derivative of the inverse $h = f^{-1}$ at the point $w = f(z)$. Here we start from

$$(h \circ f)(z) = z$$

and compute

$$\begin{aligned} h_w(w)\; f_z(z) + h_{\overline{w}}(w)\; \overline{f_{\bar z}(z)} &= 1 \\ h_w(w)\; f_{\bar z}(z) + h_{\overline{w}}(w)\; \overline{f_z(z)} &= 0 \end{aligned}$$

From these two equations we may eliminate h_w to find $\left(|f_z|^2 - |f_{\bar z}|^2\right) h_{\overline{w}} = -f_{\bar z}$. Hence, wherever $J(z, f) = |f_z|^2 - |f_{\bar z}|^2 \neq 0$, we have at $w = f(z)$ the identities

$$\frac{\partial h}{\partial \overline{w}}(w) = -\frac{f_{\bar z}(z)}{|f_z(z)|^2 - |f_{\bar z}(z)|^2} = -\frac{f_{\bar z}(z)}{J(z, f)} \tag{2.49}$$
$$\overline{\frac{\partial h}{\partial w}}(w) = \frac{f_z(z)}{|f_z(z)|^2 - |f_{\bar z}(z)|^2} = \frac{f_z(z)}{J(z, f)} \tag{2.50}$$

In particular, if f satisfies the equation $f_{\bar z} = \mu(z) f_z$, then

$$-\frac{\partial h}{\partial \overline{w}}(w) = \mu\,(h(w))\, \overline{\frac{\partial h}{\partial w}}(w), \qquad w = f(z) \tag{2.51}$$

Also, the familiar integral formulas can be written in this spirit. For instance, we have the following theorem.

Theorem 2.9.1. (Green's Formula) *Let Ω be a bounded domain with boundary $\partial\Omega$ consisting of a finite number of disjoint and rectifiable Jordan curves.*

Suppose that $f,\ g \in W^{1,1}(\Omega) \cap C(\overline{\Omega})$. Then

$$\int_\Omega \left(\frac{\partial f}{\partial z} + \frac{\partial g}{\partial \bar z} \right) = \frac{i}{2} \int_{\partial\Omega} f\, d\bar z - g\, dz \tag{2.52}$$

For smooth functions this is just the usual Green's formula dressed in complex notation. The general case follows by approximation.

If in Green's formula we choose the functions $g(\tau) = \frac{\phi(\tau)}{z-\tau}$, $f \equiv 0$ and as the domain take $\Omega_\varepsilon = \Omega \setminus \mathbb{D}(z,\varepsilon)$, then letting $\varepsilon \to 0$ gives the classical *generalized Cauchy formula*, the starting point of all complex analysis,

$$\phi(z) = \frac{1}{2\pi i} \int_{\partial\Omega} \frac{\phi(\tau)}{\tau - z} d\tau - \frac{1}{\pi} \int_\Omega \frac{\phi_{\overline{\tau}}(\tau)}{\tau - z}, \qquad z \in \Omega \tag{2.53}$$

Here we have assumed that Ω is bounded, with $\partial\Omega$ a finite union of rectifiable Jordan curves, and

$$\phi \in W^{1,1}(\Omega) \cap C(\overline{\Omega})$$

Another immediate consequence of Green's formula concerns the integral of the Jacobian derivative $J(z,f) = |f_z(z)|^2 - |f_{\bar z}(z)|^2$. The Jacobian has a number of curious properties of fundamental importance in the geometric study of mappings; as an example, the integral of the Jacobian depends only on the boundary values of f. This phenomenon leads naturally to the notion of *null Lagrangians*, to be discussed in more detail in Chapter 19.

Corollary 2.9.2. *Let $f, g \in W^{1,2}(\Omega)$. Suppose f and g have the same boundary values in the sense of Sobolev functions, that is, $f - g \in W_0^{1,2}(\Omega)$, the closure of $C_0^\infty(\Omega)$ in $W^{1,2}(\Omega)$.*

Then

$$\int_\Omega J(z,f) = \int_\Omega J(z,g)$$

Proof. Approximating $f - g$ by functions in $C_0^\infty(\Omega)$ and f by $C^\infty(\Omega)$, we may assume that f and $g = f - (f-g)$ are smooth up to the boundary with $f = g$ on $\partial\Omega$. Consider the algebraic identity

$$f_z\, \overline{f_z} - f_{\bar z}\, \overline{f_{\bar z}} = g_z\, \overline{f_z} - g_{\bar z}\, \overline{f_{\bar z}} + \left[(f-g)\, \overline{f_z}\,\right]_z - \left[(f-g)\, \overline{f_{\bar z}}\,\right]_{\bar z}$$

By Green's formula the integral of the second term on the right is

$$\int_\Omega \left[(f-g)\, \overline{f_z}\,\right]_z = \frac{i}{2} \int_{\partial\Omega} (f-g)\, \overline{f_z}\, d\bar z = 0$$

since $f = g$ on $\partial\Omega$. Similarly the integral of the last term on the right vanishes. Hence we have

$$\int_\Omega J(z,f) = \int_\Omega f_z\,\overline{f_z} - f_{\bar z}\,\overline{f_{\bar z}} = \int_\Omega g_z\,\overline{f_z} - g_{\bar z}\,\overline{f_{\bar z}} = \int_\Omega f_z\,\overline{g_z} - f_{\bar z}\,\overline{g_{\bar z}}$$

since the Jacobian is real-valued. Interchanging the roles of f and g proves the claim. $\square$

2.9.2 Riemann Mapping Theorem and Uniformization

In order to avoid developing any machinery here that we shan't use elsewhere, we take a rather advanced standpoint and collect together basic results concerning conformal mappings and hyperbolic metrics and show how many of them follow from the uniformization theorem. Of course classically many of these results are basic ingredients for the proof of this central result [7]. The reader familiar with the basics of complex analysis may easily skip this material.

We recall that a conformal mapping of a domain $\Omega \subset \mathbb{C}$ is a holomorphic homeomorphism $\phi : \Omega \to \phi(\Omega) = \Omega'$. The first result we wish to present is the Riemann mapping theorem.

Theorem 2.9.3. (Riemann mapping theorem) *Let $\Omega \subset \mathbb{C}$ be a simply connected proper subdomain. Then there is a conformal surjection $\varphi : \mathbb{D} \to \Omega$. Moreover, if ψ is another such mapping, then $\psi^{-1} \circ \varphi : \mathbb{D} \to \mathbb{D}$ is a linear fractional transformation. In particular, given $z_0 \in \Omega$, there exists a unique conformal mapping $\varphi : \mathbb{D} \to \Omega$ with $\varphi(0) = z_0$ and $\varphi'(0) > 0$.*

The uniformization theorem of Koebe and Klein asserts a more general result by identifying the simply connected Riemann surfaces. Recall that a Riemann surface Σ is a two-dimensional topological manifold equipped with a conformal atlas. That is, charts (homeomorphisms onto their images) $\varphi_i : U_i \to \Sigma$, U_i open in $\mathbb{C}$, such that

- $\Sigma = \bigcup_i \varphi_i(U_i)$ and
- $\varphi_i^{-1} \circ \varphi_j : (\varphi_j^{-1} \circ \varphi_i)(U_i) \hookrightarrow \mathbb{C}$ is conformal for all i, j.

The identity chart shows every planar domain to be a Riemann surface.

Theorem 2.9.4. (Uniformization Theorem) *Let Σ be a simply connected Riemann surface. Then Σ is conformally equivalent to exactly one of $\mathbb{C}$, $\hat{\mathbb{C}}$ or $\mathbb{D}$.*

Now it is a basic fact of topology, more precisely covering space theory [147], that any Riemann surface admits a universal cover $\tilde{\Sigma}$, a simply connected Riemann surface, for which the deck transformations (the fundamental group) will act as a discrete group Γ of conformal transformations without fixed points. The discrete groups of conformal transformations of $\mathbb{C}$ and $\hat{\mathbb{C}}$ are easily identified. Any homeomorphism of $\hat{\mathbb{C}}$ has a fixed point. A conformal transformation of $\mathbb{C}$ is a similarity. A similarity without fixed points is a translation. It follows with a little work that $\mathbb{C}$ is the universal cover only for Tori $\approx \mathbb{S} \times \mathbb{S}$ and cylinders $\mathbb{C} \setminus \{0\} \approx \mathbb{S} \times \mathbb{R}$. Every other Riemann surface therefore admits the disk as its universal cover.

We have already identified the conformal transformations of $\mathbb{D}$ as the linear fractional transformations $z \mapsto e^{i\theta}(z-a)/(1-\overline{a}z)$, $|a| < 1$, $\theta \in \mathbb{R}$. These maps act as hyperbolic isometries, as discussed earlier, and so the quotient inherits a hyperbolic metric in a natural way from the covering projection $\mathbb{D} \to \mathbb{D}/\Gamma = \Sigma$. Of importance to us later will be the case of the triply punctured sphere $\hat{\mathbb{C}} \setminus \{0,1,\infty\} = \mathbb{C} \setminus \{0,1\}$. There, the universal covering projection $\mathbb{D} \to \mathbb{C} \setminus \{0,1\}$ is the Weierstrass $\mathcal{P}$-function [6] . Another example is the planar domain $\mathbb{D} \setminus \{0\}$. There the universal covering projection is $\exp\left(\frac{z-1}{z+1}\right)$. If we recall the conformal equivalence between $\mathbb{D}$ and $\mathbb{H}$, one can easily identify the associated group of deck transformations acting on $\mathbb{H}$ as the group $\langle z \mapsto z+k : k \in \mathbb{Z}\rangle$.

In general, for a given domain $\Omega \subset \mathbb{C}$, with $\partial\Omega$ having more than one finite boundary point, there is a universal covering map $\varphi : \mathbb{D} \to \Omega$ given by Theorem 2.9.4. The element of arc for the hyperbolic metric of Ω is given at each $z \in \Omega$ by

$$ds_{hyp} = \frac{2|\psi'(z)||dz|}{1-|\psi(z)|^2}, \tag{2.54}$$

where ψ is a local inverse of φ defined near z. As all local inverses differ by composition with a linear fractional transformation of $\mathbb{D}$, the identity (2.13) shows this metric is well defined. As an example, let the domain $\Omega = \mathbb{D} \setminus \{0\}$ and the covering map $\psi(z) = \exp\left(\frac{z-1}{z+1}\right)$. Using (2.54), a calculation shows that

$$ds_{hyp} = \frac{1}{|z| \log \frac{1}{|z|}}, \qquad z \in \mathbb{D} \setminus \{0\} \tag{2.55}$$

There are many other ways to construct these hyperbolic metrics—for instance through potential theoretic methods or as the solutions to the PDE $\Delta \log \lambda = \lambda^2$, whence the metric $\lambda(z)|dz|$ has constant negative curvature; see [7].

The metric defined in (2.54) is complete and of constant curvature. If Ω is a planar domain with at least two finite points in its boundary, we call Ω a *hyperbolic domain.* Then we define the hyperbolic metric ρ_Ω by integrating the Riemannian metric in (2.54),

$$\rho_\Omega(z,w) = \inf_\gamma \int_\gamma ds_{hyp},$$

where the infimum is over all rectifiable curves γ joining z to w. This infimum is actually attained by a *hyperbolic geodesic.*

2.9.3 Schwarz-Pick Lemma of Ahlfors

Here we shall meet one of the most useful facts concerning holomorphic mappings. The classical Schwarz lemma states that if $\varphi : \mathbb{D} \to \mathbb{D}$ is holomorphic with $\varphi(0) = 0$, then $|\varphi(z)| \leqslant |z|$ and so $|\varphi'(0)| \leqslant 1$. Equality holds only for rotations. The usual proof is to apply the maximum principle to the holomorphic (!) function $\varphi(z)/z$, which does not exceed 1 in modulus on the circle $\partial\mathbb{D}$; see [6]. If $\varphi : \mathbb{D} \to \mathbb{D}$ is holomorphic and $z \in \mathbb{D}$ and we put

$$\Phi_z(\zeta) = \frac{z - \zeta}{1 - \zeta\bar{z}},$$

then

$$\varphi_z(\zeta) = (\Phi_{\varphi(z)} \circ \varphi \circ \Phi_z)(\zeta)$$

has $\varphi_z(\mathbb{D}) \subset \mathbb{D}$ and $\varphi_z(0) = 0$. Therefore, applying the Schwarz lemma to this composition with linear fractional transformations together with a bit of calculation gives the following invariant formulation.

Theorem 2.9.5. *If $\varphi : \mathbb{D} \to \mathbb{D}$ is holomorphic, then*

$$\frac{|\varphi'(z)|}{1 - |\varphi(z)|^2} \leqslant \frac{1}{1 - |z|^2} \tag{2.56}$$

In view of our discussion of the hyperbolic metric of the disk, (2.56) implies that φ is a local (and therefore global) contraction in the hyperbolic metric. If we have a holomorphic map $\varphi : \Omega \to \Omega'$ from one planar domain into another, then this mapping is a local (and hence global) contraction of the hyperbolic metrics. To see this, we simply lift the map via the respective holomorphic coverings to a holomorphic map of the disk $\mathbb{D}$. Locally, the holomorphic covering is an isometry, while the lifted map is a contraction, as we have observed. Thus the original map is a local (and hence global) contraction in view of the definition of the metric in (2.56). We record this as the following theorem.

Theorem 2.9.6. *Let $\varphi : \Omega \to \Omega'$ be a holomorphic mapping between hyperbolic domains. Then φ is a contraction of the hyperbolic metrics,*

$$\rho_{\Omega'}(\varphi(z), \varphi(w)) \leqslant \rho_\Omega(z, w)$$

for all $z, w \in \Omega$.

As a simple application, the identity map is holomorphic and therefore we have the following.

Corollary 2.9.7. *If $\Omega \subset \Omega'$, then for every $z, w \in \Omega$ we have*

$$\rho_{\Omega'}(z, w) \leqslant \rho_\Omega(z, w)$$

Then for $z \in \Omega'$ and $\Omega = \mathbb{D}(z, \operatorname{dist}(z, \partial\Omega'))$ we apply Corollary 2.9.7 to get the following.

Corollary 2.9.8. *For $z \in \Omega'$*

$$\frac{1}{2\operatorname{dist}(z, \partial\Omega')} \leqslant \delta_{\Omega'}(z) \leqslant \frac{2}{\operatorname{dist}(z, \partial\Omega')},$$

where $ds_{hyp} = \delta_{\Omega'}(z)|dz|$ is the length element of the hyperbolic metric of Ω'.

Here we have found it convenient to include the lower bound, which is a consequence of the Koebe $\frac{1}{4}$-theorem and in particular the estimate of (2.70) that we will come to in a moment.

2.9.4 Normal Families and Montel's Theorem

A family of continuous functions $\mathcal{F} = \{\varphi : \Omega \to \mathbb{C}\}$ is called *normal* if every sequence $\{\varphi_i\}_{i=1}^{\infty} \subset \mathcal{F}$ contains a subsequence converging locally uniformly to a limit function φ.

Two ideas from topology suffice to establish that a family is normal. These are the notion of equicontinuity and the Ascoli theorem.

Equicontinuity

Let (X, d) and (Y, σ) be metric spaces and let $\mathcal{F} = \{\varphi : X \to Y\}$ be a family of functions. We call $\mathcal{F}$ *equicontinuous* at $x_0 \in X$ if for every $\varepsilon > 0$ there is $\delta > 0$ such that $d(x, x_0) < \delta$ implies $\sigma(\varphi(x), \varphi(x_0)) < \varepsilon$ for all $\varphi \in \mathcal{F}$. We say that $\mathcal{F}$ is equicontinuous if it is equicontinuous at each $x_0 \in X$.

Next, the central result linking normal families and equicontinuity is the Arzela-Ascoli theorem.

Theorem 2.9.9. *If (X, d) is a separable metric space and (Y, σ) a compact metric space, then every equicontinuous family of mappings $\mathcal{F} = \{\varphi : X \to Y\}$ is a normal family.*

There are a number of variations of this theorem. One of the most common is to replace the hypothesis that (Y, σ) is a compact metric space by the hypothesis that the family $\mathcal{F}$ is locally bounded and (Y, σ) is complete.

In our applications we are going to take $Y = \hat{\mathbb{C}}$ and σ as the spherical metric arising from the Riemannian metric $ds_{sph} = |dz|/\sqrt{1 + |z|^2}$. Then

$$\sigma(z, w) = \frac{|z - w|}{\sqrt{1 + |z|^2}\sqrt{1 + |w|^2}}$$

with the obvious interpretation if z or w is equal to ∞.

To use the Arzela-Ascoli theorem we need to relate the hyperbolic and spherical metrics. Notice that both these metrics are conformally equivalent to the Euclidean metric. The density of the hyperbolic metric is nowhere vanishing and tends uniformly to ∞ near $\partial\Omega$. Hence it has a positive minimum in Ω. As σ is bounded, we have the following theorem.

Theorem 2.9.10. *Let $\Omega \subset \mathbb{C}$ be a hyperbolic domain. Then there is a constant C_Ω such that*

$$\sigma(z, w) \leqslant C_\Omega \rho_\Omega(z, w)$$

for all $z, w \in \Omega$.

There is no uniform bound independent of domain. If $\Omega = \mathbb{D}(0, R)$, $R > 1$, then $\sigma(0,1) = \frac{1}{\sqrt{2}}$ and $\rho_\Omega(0,1) = \log \frac{R+1}{R-1} \to 0$ as $R \to \infty$.

Now as a consequence, we have Montel's theorem.

Theorem 2.9.11. *Let $z_0, z_1 \in \mathbb{C}$ and let Ω be a planar domain. Suppose that $\mathcal{F} \subset \{\varphi : \Omega \to \mathbb{C} \setminus \{z_0, z_1\}\}$ is a family of holomorphic maps omitting these two values. Then $\mathcal{F}$ is a normal family.*

Proof. $\mathbb{C} \setminus \{z_0, z_1\}$ is a hyperbolic domain, and therefore each element of $\mathcal{F}$ is a contraction of the hyperbolic metrics if Ω is also a hyperbolic domain. In that case to see that the family $\mathcal{F}$ is equicontinuous when viewed as mappings into $\hat{\mathbb{C}}$ we use Theorem 2.9.6 which tells us that for each $z, w \in \Omega$,

$$\sigma(\varphi(z), \varphi(w)) \leqslant C\ \rho_{\mathbb{C}\setminus\{z_0,z_1\}}(\varphi(z), \varphi(w)) \leqslant C\ \rho_\Omega(z, w)$$

with constant C depending only on z_0 and z_1. If Ω is not a hyperbolic domain we simply restrict to a hyperbolic subdomain and the result follows easily. □

2.9.5 Hurwitz's Theorem

Hurwitz's theorem asserts that the limits of nonvanishing holomorphic functions are either nonvanishing or identically 0. It is usually proved as an application of Rouché's Theorem and the maximum principle.

Theorem 2.9.12. *Let $\Omega \subset \mathbb{C}$ be a domain and $\{\varphi_i\}_{i=1}^{\infty}$ be a sequence of nonvanishing holomorphic functions converging locally uniformly to $\varphi : \Omega \to \mathbb{C}$. Then either $\varphi \equiv 0$ or φ does not vanish in Ω*

Using this theorem the reader may easily prove the following important consequence.

Corollary 2.9.13. *Let $\Omega \subset \mathbb{C}$ be a domain and $\{\varphi_i\}_{i=1}^{\infty}$ be a sequence of conformal mappings converging locally uniformly to $\varphi : \Omega \to \mathbb{C}$. Then either φ is constant, or φ is conformal.*

2.9.6 Bloch's Theorem

Bloch's constant $\mathcal{B}_0$ is defined to be the supremum of all numbers δ such that each holomorphic function f in the unit disk $\mathbb{D}$ with $|f'(0)| = 1$ possesses a continuous inverse in some disk of radius δ; see [63]. We have the following theorem of Ahlfors; see for instance [7].

Theorem 2.9.14. $\mathcal{B}_0 \geqslant \sqrt{3}/4$.

For locally univalent mappings (so with the further assumption $f'(z) \neq 0$, $z \in \mathbb{D}$), the similarly defined constant is denoted $\mathcal{B}_\infty$, and it is known that $\mathcal{B}_\infty > \frac{1}{2}$. The exact determination of $\mathcal{B}_0$ remains a famously difficult open problem. The prevailing conjecture, due to Ahlfors and Grunsky (1937), asserts that $\mathcal{B}_0 = \Gamma(\frac{1}{3})\Gamma(\frac{11}{12})/\sqrt{1+\sqrt{3}}\,\Gamma(\frac{1}{4}) = 0.4719\ldots$.

2.9.7 The Argument Principle

We will later need the following formulation of the argument principle, which is essentially a corollary of Theorem 2.8.1. An elementary proof using classical complex analysis can be found in [6].

Theorem 2.9.15. *If f is holomorphic in $\mathbb{D}$, if it is continuous in $\overline{\mathbb{D}}$ and if $f\big|_{\partial\mathbb{D}}$ is homotopic to the identity relative to $\mathbb{C} \setminus \{0\}$, then $f(z_0) = 0$ at precisely one point $z_0 \in \mathbb{D}$.*

2.10 Distortion by Conformal Mapping

From a historical perspective, the notion of a quasiconformal mapping has its roots in the study of the geometric properties of conformal mappings. The philosophy here is that locally a conformal mapping is close to its linearization $z \mapsto f(z_0) + f'(z_0)(z - z_0)$—a similarity mapping of $\mathbb{C}$. Such mappings form a finite-dimensional family and because of this are quite "rigid." One expects that conformal mappings should inherit some of this geometric rigidity. This is a property that we will state in an explicit and quantitative manner in Theorem 2.10.9. In fact, this view leads naturally to the notion of a quasisymmetric mapping, to be studied in the next chapter. For the reader's convenience we describe the basic distortion theorems of conformal mappings.

2.10.1 The Area Formula

We shall first formulate and prove a classical result of Gronwall [148], soon thereafter rediscovered by Bieberbach [60], that is known as the area formula. It will have important applications later. It also leads us to the important Koebe distortion theorems.

Theorem 2.10.1. *Suppose that $f \in W^{1,2}_{loc}(\mathbb{C})$ is analytic outside the disk $\mathbb{D}(0,r)$ and has the expansion*

$$z + b_1 z^{-1} + b_2 z^{-2} + \cdots \tag{2.57}$$

near ∞. Then

$$\int_{\mathbb{D}(0,r)} J(z,f) = \pi \left(r^2 - \sum_{n=1}^{\infty} n|b_n|^2 r^{-2n} \right) \tag{2.58}$$

In particular, if f is orientation-preserving (so $J(z,f) \geqslant 0$ almost everywhere), then

$$\sum_{n=1}^{\infty} n|b_n|^2 r^{-2n} \leqslant r^2 \tag{2.59}$$

Proof. As f is analytic outside $\mathbb{D}(0,r)$, the power series in (2.57), converges uniformly on $\{z : |z| \geqslant s\}$ for any fixed $s > r$. Then we use Green's formula

(2.52) to write

$$\begin{aligned}
\int_{\mathbb{D}(0,s)} J(z,f) &= \int_{\mathbb{D}(0,s)} |f_z|^2 - |f_{\bar z}|^2 = \frac{1}{2i}\int_{|z|=s} \overline{f}\, f_z\, dz \\
&= \frac{1}{2i}\int_{|z|=s} (\bar z + \overline{b_1}\, \bar z^{-1} + \cdots)(1 - b_1 z^{-2} - 2b_2 z^{-3} + \cdots)dz \\
&= \frac{1}{2i}\int_{|z|=s} (s^2 - |b_1|^2 s^{-2} - 2|b_2|^2 s^{-4} - \cdots)\, \frac{dz}{z} \\
&= \pi(s^2 - |b_1|^2 s^{-2} - 2|b_2|^2 s^{-4} - \cdots)
\end{aligned}$$

Now we may let $s \to r$ to obtain (2.58) and of course (2.59) when $J(z,f) \geqslant 0$ almost everywhere. □

We can use the area inequality in (2.58) to estimate certain functions as follows. Suppose that $f \in W^{1,2}_{loc}(\mathbb{C})$ is analytic outside the disk $\mathbb{D}(0,r)$ and has the expansion

$$z + b_1 z^{-1} + b_2 z^{-2} + \cdots \tag{2.60}$$

near ∞. We obtain from the Cauchy-Schwarz inequality the easy estimate

$$\begin{aligned}
|f(z) - z| &\leqslant \frac{|b_1|}{|z|} + \frac{|b_2|}{|z|^2} + \cdots \\
&\leqslant \left(\frac{|b_1|^2}{r^2} + \frac{2|b_2|^2}{r^4} + \frac{3|b_3|^2}{r^6} + \cdots\right)^{1/2} \left(\frac{r^2}{|z|^2} + \frac{r^4}{2|z|^4} + \frac{r^6}{3|z|^6} + \cdots\right)^{1/2} \\
&\leqslant r\left(\log\frac{|z|^2}{|z|^2 - r^2}\right)^{1/2}
\end{aligned}$$

Therefore, in particular, we obtain for $|z| > 1.3\,r$ the estimate

$$|f(z) - z| \leqslant r \tag{2.61}$$

If f is a homeomorphism of $\mathbb{C}$, then (2.59) holds even without assuming the $W^{1,2}_{loc}$-regularity. We therefore have the following corollary.

Corollary 2.10.2. *If $f \in W^{1,1}_{loc}(\mathbb{C})$ is a homeomorphism analytic outside the disk $\mathbb{D}(0,r)$ with $|f(z) - z| = o(1)$ at ∞, then*

$$|f(z)| < |z| + 3r, \quad \textit{for all } z \in \mathbb{C} \tag{2.62}$$

Proof. The only things to observe are that, first, the assumption $|f(z) - z| = o(1)$ at ∞ together with analyticity implies a series expansion as in (2.59) and thus we have (2.61) for $|z| > 1.3r$, and second, as f is supposed a homeomorphism, the estimate clearly follows for $|z| < 1.3r$ as f is an open mapping. □

Another interesting application concerns the solutions to the Beltrami system with the normalization at ∞ as above (we shall later call them principal solutions). Suppose f is a $W^{1,2}_{loc}(\mathbb{C})$-solution to the equation

$$\frac{\partial f}{\partial \bar z} = \mu(z)\,\frac{\partial f}{\partial z}, \tag{2.63}$$

where $|\mu(z)| \leqslant k\,\chi_D$ for $D = \mathbb{D}(0,r)$ and $0 \leqslant k < 1$. Then f is analytic outside the disk D. Suppose we have the further normalisation that

$$|f(z) - z| = o(1), \qquad z \text{ near } \infty \tag{2.64}$$

As we have noticed already, this gives us a series expansion as in (2.58) converging locally uniformly on $|z| > r$. We have the formula (see (2.53))

$$f(z) = z - \frac{1}{\pi}\int_D \frac{f_{\bar z}(\tau)}{\tau - z}\,d\tau$$

and hence

$$b_1 = \lim_{z\to\infty} z(f(z) - z) = \frac{1}{\pi}\int_D f_{\bar z}(\tau)\,d\tau,$$

which gives us the estimate

$$\begin{aligned} |b_1|^2 &\leqslant \left(\frac{1}{\pi}\int_D |f_{\bar z}|\right)^2 \leqslant \frac{r^2}{\pi}\int_D |f_{\bar z}|^2 \\ &\leqslant \frac{k^2 r^2}{\pi(1-k^2)}\int_D (|f_z|^2 - |f_{\bar z}|^2) \\ &= \frac{k^2 r^2}{(1-k^2)}\left(r^2 - \sum_{n=1}^{\infty} n|b_n|^2 r^{-2n}\right) \end{aligned}$$

and hence

$$\begin{aligned} |b_1|^2\left(1 + \frac{k^2}{1-k^2}\right) &\leqslant \frac{k^2 r^2}{(1-k^2)}\left(r^2 - \sum_{n=2}^{\infty} n|b_n|^2 r^{-2n}\right) \\ &\leqslant \frac{k^2 r^4}{(1-k^2)} \end{aligned}$$

This establishes the following corollary.

Corollary 2.10.3. *For a solution of (2.63) normalized by (2.64), we have the estimate*

$$|b_1| \leqslant k r^2 \tag{2.65}$$

We have equality if and only if

$$f(z) = \begin{cases} z + \lambda r^2/z, & |z| \geqslant r \\ z + \lambda \bar z, & |z| \leqslant r \end{cases} \tag{2.66}$$

where λ is any complex number with $|\lambda| = k$.

The only thing not clear in the corollary is the claim (2.66) regarding equality. However, this follows by consideration of the fact that, in order to have equality in (2.65), we must have had equality in every step of our calculation.

2.10.2 Koebe $\frac{1}{4}$-Theorem and Distortion Theorem

The Koebe $\frac{1}{4}$-Theorem is one of the first and also one of the most powerful distortion theorems one meets in complex analysis. With the correct interpretation this result implies universal distortion estimates in hyperbolic disks, satisfied by *all* conformal mappings.

To find these we first consider the analytic function $g(z) = z + b_0 + b_1 z^{-1} + \cdots$, which we suppose is conformal in the exterior of the unit disk and further that $g(z) \neq w$ for $|z| > 1$. Then the branch

$$h(z) = \sqrt{g(z^2) - w} = z + \frac{1}{2}(b_0 - w)z^{-1} + \cdots$$

is well defined and conformal in the exterior disk. Furthermore, for any $r > 1$ its restriction to $\{z : |z| > r\}$ extends to a global mapping $h \in W^{1,2}_{loc}(\mathbb{C})$. Thus Theorem 2.10.1 gives

$$|w - b_0| \leqslant 2 \tag{2.67}$$

Often it is convenient to use this result in the following form.

Theorem 2.10.4. *Suppose $g : \mathbb{C} \to \mathbb{C}$ is a homeomorphism, which is conformal in the exterior of the unit disk. If g has the development $g(z) = z + b_0 + b_1 z^{-1} + \cdots$ for $|z| > 1$, then*

$$g(\mathbb{D}) \subset \mathbb{D}(b_0, 2)$$

The famous Koebe $\frac{1}{4}$-theorem is a quick consequence.

Theorem 2.10.5. *Suppose that $\varphi : \mathbb{D} \to \mathbb{C}$ is conformal and normalized by $\varphi(0) = 0$ and $\varphi'(0) = 1$. Then*

$$\mathbb{D}(0, \frac{1}{4}) \subset \varphi(\mathbb{D})$$

Proof. With our assumptions $\varphi(z) = z + a_2 z^2 + \cdots$ for $z \in \mathbb{D}$. The conjugate

$$g(z) = \frac{1}{\varphi(z^{-1})} = z - a_2 + \mathcal{O}(\frac{1}{z}), \qquad |z| > 1,$$

never vanishes, and thus (2.67) implies the classical bound of Bieberbach,

$$|a_2| \leqslant 2 \tag{2.68}$$

Also if $w \notin \varphi(\mathbb{D})$, the function

$$\varphi_1(z) = \frac{w\,\varphi(z)}{w - \varphi(z)} = z + (a_2 + \frac{1}{w})z^2 + \mathcal{O}(z^2)$$

satisfies the assumptions of the theorem, and hence we have additionally

$$|a_2 + \frac{1}{w}| \leqslant 2 \tag{2.69}$$

Combining the bounds shows that $|w| \geqslant 1/4$ whenever $w \notin \varphi(\mathbb{D})$. □

We may view Theorem 2.10.4 as the counterpart to Koebe's result at ∞. In bounded domains the following form of Koebe's $\frac{1}{4}$-theorem applies in fact to all conformal mappings, independently of their normalization.

Theorem 2.10.6. *Suppose that f is conformal in a domain Ω with $f(\Omega) = \Omega' \subset \mathbb{C}$. Let $z_0 \in \Omega$. Then*

$$\frac{1}{4}\,|f'(z_0)|\text{dist}(z_0, \partial\Omega) \leqslant \text{dist}\,(f(z_0), \partial\Omega') \leqslant |f'(z_0)|\text{dist}(z_0, \partial\Omega) \tag{2.70}$$

The first inequality in (2.70) follows from Koebe's theorem, applied to

$$\varphi(z) = \frac{f(z_0 + z\,d) - f(z_0)}{d\,f'(z_0)}, \qquad d = \text{dist}(z_0, \partial\Omega),$$

while the latter inequality is a consequence of the Schwarz lemma, applied to $f^{-1} : \mathbb{D}\,(f(z_0), d') \to \mathbb{D}(z_0, d)$, where $d' = \text{dist}\,(f(z_0), \partial\Omega')$.

The Bieberbach bound (2.68) also provides us with uniform distortion estimates as soon as we are able to express it in an invariant form. To reveal this we introduce, for each mapping f conformal in $\mathbb{D}$, the *Koebe transform*

$$\varphi(z) = \frac{f(\frac{z+w}{1+\overline{w}z}) - f(w)}{(1-|w|^2)f'(w)}, \qquad z \in \mathbb{D} \tag{2.71}$$

Here $w \in \mathbb{D}$ is arbitrary.

An elementary calculation gives

$$\varphi''(0) = (1-|w|^2)\frac{f''(w)}{f'(w)} - 2\overline{w} \tag{2.72}$$

Since φ is conformal in $\mathbb{D}$ with $\varphi(0) = 0$ and $\varphi'(0) = 1$, Bieberbach's coefficient estimate yields the following theorem.

Theorem 2.10.7. *If f is conformal in the unit disk $\mathbb{D}$, then*

$$(1-|w|^2)\frac{|f''(w)|}{|f'(w)|} \leqslant 6, \qquad w \in \mathbb{D}$$

We are now in a position to prove the first of the Koebe distortion theorems. For our purposes an invariant formulation, such as the following, is the most prefered.

Theorem 2.10.8. *Suppose that f is conformal in the unit disk $\mathbb{D}$ and $z, w \in \mathbb{D}$. Then*

$$e^{-3\rho_{\mathbb{D}}(z,w)} \leqslant \frac{|f'(z)|}{|f'(w)|} \leqslant e^{3\rho_{\mathbb{D}}(z,w)}$$

Proof. Since f is conformal, the function $g(z) = \log f'(z)$ is analytic in $\mathbb{D}$. Theorem 2.10.7 tells us that $|g'(z)| \leqslant 3\, ds_{hyp}(z)$, and the claim follows via an integration,

$$\left|\log \frac{|f'(z)|}{|f'(w)|}\right| \leqslant |g(z) - g(w)| \leqslant 3\,\rho_{\mathbb{D}}(z,w) \qquad \square$$

It is remarkable that each of Theorems 2.10.5 – 2.10.8 is sharp, as the reader may verify using the Koebe function $f_0(z) = \frac{z}{(1-z)^2} = \frac{1}{4}\left(\frac{1+z}{1-z}\right)^2 - \frac{1}{4}$. The function f_0 maps the unit disk $\mathbb{D}$ conformally onto $\mathbb{C} \setminus (-\infty, -\frac{1}{4}]$.

According to Theorem 2.10.8, one may consider the derivative of a conformal mapping as almost constant on hyperbolic disks! It is to be expected that then, with suitable interpretation, the mapping itself should almost be a similarity when restricted to a subdomain bounded in the hyperbolic metric.

This fact turns out to be true and is perhaps most conveniently expressed in the notation of the next theorem. Here note that a homeomorphism f in a domain Ω is a similarity if and only if

$$\frac{|f(z) - f(w)|}{|f(\zeta) - f(w)|} = \frac{|z - w|}{|\zeta - w|} \qquad \text{for all } z, w, \zeta \in \Omega \tag{2.73}$$

Theorem 2.10.9. *Suppose that f is conformal in the unit disk $\mathbb{D}$. Let z_1, z_2 and $w \in \mathbb{D}$ with*

$$\rho_{\mathbb{D}}(z_1, w) + \rho_{\mathbb{D}}(z_2, w) \leqslant M < \infty$$

Then

$$\frac{|f(z_1) - f(w)|}{|f(z_2) - f(w)|} \leqslant e^{4M}\, \frac{|z_1 - w|}{|z_2 - w|} \tag{2.74}$$

Proof. We will use the Koebe transform (2.71) and evaluate $\varphi(\zeta_j)$, where $z_j = (\zeta_j + w)/(1 + \overline{w}\zeta_j)$ and $j = 1, 2$. Then $\zeta_j = (z_j - w)/(1 - \overline{w} z_j)$. Hence

$$\frac{f(z_1) - f(w)}{f(z_2) - f(w)} = \frac{\varphi(\zeta_1)}{\varphi(\zeta_2)} = \frac{z_1 - w}{z_2 - w}\, \frac{\varphi(\zeta_1)}{\zeta_1}\, \frac{\zeta_2}{\varphi(\zeta_2)}\, \frac{1 - \overline{w}\, z_2}{1 - \overline{w}\, z_1}$$

To estimate the last expression we note that

$$\begin{aligned} \log\left|\frac{1 - \overline{w}\, z_2}{1 - \overline{w}\, z_1}\right| &\leqslant \log\left(\frac{1 + \left|\frac{z_1 - w}{1 - \overline{w}\, z_1}\right|}{1 - \left|\frac{z_2 - w}{1 - \overline{w}\, z_2}\right|}\right) \\ &\leqslant \rho_{\mathbb{D}}(z_1, w) + \rho_{\mathbb{D}}(z_2, w) \end{aligned}$$

Since $\rho_{\mathbb{D}}(\zeta_j, 0) = \rho_{\mathbb{D}}(z_j, w)$, it remains to show that

$$e^{-3\,\rho_{\mathbb{D}}(\zeta,0)} \leqslant \frac{|\varphi(\zeta)|}{|\zeta|} \leqslant e^{3\,\rho_{\mathbb{D}}(\zeta,0)}, \qquad \zeta \in \mathbb{D} \tag{2.75}$$

In fact, by Theorem 2.10.8

$$|\varphi(\zeta)| = \Big|\int_0^\zeta \frac{\varphi'(z)}{\varphi'(0)} dz\Big| \leqslant \int_0^{|\zeta|} e^{3\,\rho_{\mathbb{D}}(z,0)}|dz| \leqslant |\zeta| e^{3\,\rho_{\mathbb{D}}(\zeta,0)}$$

For the former of the inequalities in (2.75), note that this is clear if $|\varphi(\zeta)| \geqslant 1/4$. Otherwise, by Koebe $\frac{1}{4}$-theorem, the interval $[\varphi(\zeta), 0] \subset \varphi(\mathbb{D})$. As $\varphi'(z)dz$ has a constant argument on $\varphi^{-1}[\varphi(\zeta), 0]$, we have

$$|\varphi(\zeta)| = \int_0^{|\zeta|} |\varphi'(z)|\,|dz| \geqslant \int_0^{|\zeta|} e^{-3\,\rho_{\mathbb{D}}(z,0)}|dz| \geqslant |\zeta| e^{-3\,\rho_{\mathbb{D}}(\zeta,0)}$$

Combining these estimates gives the inequality (2.74). □

The above theorem is an invariant version of the second Koebe distortion theorem, expressing in a compact and quantitative manner the fact that locally every conformal mapping is close to a similarity. Here, though, no claim is made on the sharpness of (2.74) in terms of the exponent $4M$. On the other hand, an important fact in Theorem 2.10.9 is the conformal invariance; via a change of variables it applies immediately to all mappings f conformal in a simply connected domain Ω.

We note also the following immediate consequence.

Corollary 2.10.10. *Conformal mappings of the plane are similarities.*

Proof. With a scaling, the estimate (2.74) holds in any disk $\mathbb{D}(0, r) = r\mathbb{D}$. If we denote $M_r = \rho_{r\mathbb{D}}(z_1, w) + \rho_{r\mathbb{D}}(z_1, w)$, then (2.74) attains the form

$$\frac{|z_1 - w|}{|z_2 - w|} e^{-4M_r} \leqslant \frac{|f(z_1) - f(w)|}{|f(z_2) - f(w)|} \leqslant \frac{|z_1 - w|}{|z_2 - w|} e^{4M_r} \tag{2.76}$$

Fixing the points z_1, z_2 and w but letting $r \to \infty$ gives $M_r = \rho_{r\mathbb{D}}(z_1, w) + \rho_{r\mathbb{D}}(z_1, w) = \rho_{\mathbb{D}}(z_1/r, w/r) + \rho_{\mathbb{D}}(z_1/r, w/r) \to 0$. Hence f satisfies (2.73). □

Bounds on the distortion of ratios such as in (2.74) quickly yield a large spectrum of various geometric properties. Indeed, the geometric study of mappings requires general notions that allow such conclusions, and for much larger classes than just (the very rigid) conformal mappings. These considerations will lead in a natural manner to the concept of quasisymmetry, which is studied and utilized in the next section. In this terminology, Theorem 2.10.9 tells us that all conformal mappings are uniformly quasisymmetric in subdomains with bounded hyperbolic diameter.

Chapter 3

The Foundations of Quasiconformal Mappings

One of the more interesting and important recent developments in the theory of quasiconformal mappings has been the development of the study of these mappings on more general spaces than those that are locally Euclidean. In particular, the work of Heinonen and Koskela [161] develops the theory in certain metric spaces where some reasonable measure theoretic and geometric properties hold. Part of the motivation behind this work is that often while studying classical problems in geometry one is naturally led to singular spaces as quotients or, for instance, to the boundaries of geometric objects such as groups. In these situations one has no a priori knowledge that the space in question is a manifold. While we do not discuss these developments in this book, this work does have important ramifications. It has led to a new understanding, and in particular new and simplified proofs, of many of the foundational properties of quasiconformal mappings. Here we will present some of this material, partly making use of the lecture notes of P. Koskela [215].

3.1 Basic Properties

Here is our starting point.

Definition 3.1.1. *An orientation-preserving homeomorphism* $f : \Omega \to \Omega'$ *is* K-quasiconformal, $1 \leqslant K < \infty$, *if*

$$f \in W^{1,2}_{loc}(\Omega)$$

and

$$\max_{\alpha} |\partial_\alpha f(z)| \leqslant K \min_{\alpha} |\partial_\alpha f(z)| \tag{3.1}$$

for almost every $z \in \Omega$.

In particular, a mapping f is 1-quasiconformal if and only if it is conformal, that is, a univalent holomorphic mapping. The reader should not forget the precise meaning or background of (3.1) nor the complex interpretation through the Beltrami equation, both discussed in Section 2.5. However, the complex analytic approach will not show its real virtues much before Section 5.

Beginning with the above definition, we will prove in this chapter the first and most fundamental properties of quasiconformal mappings as embodied in the following two theorems.

Theorem 3.1.2. *Let $f : \Omega \to \Omega'$ be a K-quasiconformal mapping from the domain $\Omega \subset \mathbb{C}$ onto $\Omega' \subset \mathbb{C}$ and let $g : \Omega' \to \mathbb{C}$ be a K'-quasiconformal mapping. Then*

- *$f^{-1} : \Omega' \to \Omega$ is K-quasiconformal.*
- *$g \circ f : \Omega \to \mathbb{C}$ is KK'-quasiconformal.*
- *For all measurable sets $E \subset \Omega$, $|E| = 0$ if and only if $|f(E)| = 0$.*
- *The Jacobian determinant $J(z, f) > 0$ almost everywhere in Ω.*

Thus the family of quasiconformal mappings forms a pseudogroup with respect to composition (actually a group if the domain and range are the same). The next theorem addresses local compactness.

Theorem 3.1.3. *Let $f_\nu : \Omega \to \mathbb{C}$, $\nu = 1, 2, \ldots$, be a bounded sequence of K-quasiconformal mappings defined on the domain $\Omega \subset \mathbb{C}$. Then there is a subsequence converging locally uniformly on Ω to a mapping f,*

$$f_{\nu_k} \to f,$$

and f is either a K-quasiconformal mapping or a constant.

The key concept we shall use in this chapter to establish these results is that of quasisymmetry, which we now look at.

3.2 Quasisymmetry

The very definition of a quasiconformal map supposes the function to be defined on an open set. However, from time to time in this book we will be concerned with the restrictions of quasiconformal mappings to smaller sets or mappings perhaps defined on fractal-like sets that might have a quasiconformal extension. This suggests that we seek a notion of (quasi)conformality applicable in general subsets of the plane (or even in arbitrary metric spaces).

In fact, we have already met such a notion, in the invariant formulation of the Koebe distortion estimates in Theorem 2.10.9. This point of view leads us naturally to the very useful concept of quasisymmetry. The notion was introduced by Ahlfors and Beurling [11] on the real line and formulated for general metric spaces by Tukia and Väisälä [359].

Definition of Quasisymmetry

Definition 3.2.1. *Let $\eta : [0,\infty) \to [0,\infty)$ be an increasing homeomorphism, $A \subset \mathbb{C}$ and*

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

a mapping. We say f is η-quasisymmetric if for each triple $z_0, z_1, z_2 \in A$ we have

$$\frac{|f(z_0) - f(z_1)|}{|f(z_0) - f(z_2)|} \leqslant \eta\left(\frac{|z_0 - z_1|}{|z_0 - z_2|}\right) \tag{3.2}$$

Should f be defined on an open set, we will assume that it is orientation-preserving and further, we say f is quasisymmetric if there is some η as above for which f is η-quasisymmetric.

Notice that by definition $\eta(0) = 0$ and $\eta(t) \to \infty$ as $t \to \infty$. A typical example of such a homeomorphism η is $\eta(t) = C\,t^\alpha$, for constants $C \geqslant 1$ and $\alpha > 0$. In fact, as a consequence of what is proved later, η can always be chosen in this form (see Section 3.10). Note also that the condition at (3.2) can be made to make sense even for mappings of metric spaces.

It is an immediate consequence of the definition that a quasisymmetric mapping is continuous and injective, and by Lemma 3.2.2 below we see that it is a homeomorphism onto its image. Furthermore, when A is an open set, $z_0 \in A$ and $r < \operatorname{dist}(z_0, \partial A)$; then, with a suitable choice of z_1 and z_2 on the circle $\{|z - z_0| = r\}$, we have

$$\frac{\max_{|z-z_0|=r} |f(z) - f(z_0)|}{\min_{|z-z_0|=r} |f(z) - f(z_0)|} = \frac{|f(z_1) - f(z_0)|}{|f(z_2) - f(z_0)|} \leqslant \eta(1) \tag{3.3}$$

The right hand side of (3.3) is independent of r. We may take the limit as $r \to 0$ on the left to see that quasisymmetric mappings defined on open subsets are, in particular, mappings of bounded distortion. This suggests a strong relationship between quasiconformality and quasisymmetry, which we shall also explore in this chapter.

It is not at all obvious that the inverse of a quasiconformal homeomorphism is quasiconformal. However, for quasisymmetric mapping this is an elementary observation.

Lemma 3.2.2. *Let $f : \Omega \to \Omega'$ be an η-quasisymmetric mapping onto. Then $f^{-1} : \Omega' \to \Omega$ is σ-quasisymmetric with*

$$\sigma(t) = \frac{1}{\eta^{-1}(1/t)}$$

Proof. As f is injective and onto, given a triple of points a_i, we may put $z_i = f^{-1}(a_i)$, $i = 0, 1, 2$. Then we have, from the fact that f is η-quasisymmetric,

$$\frac{|a_0 - a_2|}{|a_0 - a_1|} \leqslant \eta \left(\frac{|z_0 - z_2|}{|z_0 - z_1|} \right),$$

and hence

$$\frac{|f^{-1}(a_0) - f^{-1}(a_1)|}{|f^{-1}(a_0) - f^{-1}(a_2)|} \leqslant 1/\eta^{-1} \left(\frac{|a_0 - a_2|}{|a_0 - a_1|} \right),$$

which is what we wanted. □

There are no bounded conformal mappings of the plane—similarly, for quasisymmetric mappings.

Lemma 3.2.3. *Every entire quasisymmetric mapping f (that is defined in the whole complex plane $\mathbb{C}$) is a surjection, $f(\mathbb{C}) = \mathbb{C}$.*

Proof. The condition (3.2) implies $|f(z)| \to \infty$ when $|z| \to \infty$ since we may fix z_0, z_1 and let $z_2 \to \infty$. Thus f extends continuously to $\hat{\mathbb{C}}$ with $f(\infty) = \infty$. Thus $f(\hat{\mathbb{C}})$ is open and closed in $\hat{\mathbb{C}}$; that is $f(\hat{\mathbb{C}}) = \hat{\mathbb{C}}$. □

These observations suggest the route we will take to establishing the basic properties of quasiconformal mappings. After auxiliary results we first show that quasisymmetric mappings are quasiconformal. The principal problem here is in proving the function lies in the correct Sobolev class, $W^{1,2}_{loc}(\Omega)$. Next we show that quasiconformal mappings of $\mathbb{C}$ are quasisymmetric. The principal problem in that case is establishing control on global distortion functions. It will then follow, for instance, that the inverse of a K-quasiconformal mapping is quasiconformal. That the inverse is in fact K-quasiconformal will not be too difficult to establish.

Classically, the equivalence of quasiconformality and quasisymmetry for maps of $\mathbb{C}$ was a consequence of the modulus definition of quasiconformality (see for instance [229]) once one has some knowledge of extremal rings and the like. Here we shall offer a couple of alternate routes to this result directly in what follows and somewhat later as a consequence of the theory of holomorphic motions.

3.3 The Gehring-Lehto Theorem

Throughout the theory of quasiconformal mappings we often need quite subtle methods to control the geometry of mappings. In fact, in our later studies we will be basically on the borderline of where a geometric study of mappings is even possible. However, conditions (2.30) and (2.31) for the partial derivatives, or the bound (2.28) for the linear distortion, are rather weak by themselves. To improve the situation we need a few additional basic facts from real analysis to carry the geometric information up from the infinitesimal level.