Complex Analysis Topics for Undergraduates and Beginning Researchers:
an Exploration with Unsolved Problems [1]
(Date: June 7, 2011)

Michael Brilleslyper

Michael Dorff

Jane McDougall

James Rolf

Beth Schaubroeck

Rich Stankewitz

Kenneth Stephenson

UNITED STATES AIR FORCE ACADEMY
E-mail address: Mike.Brilleslyper@usafa.edu

BRIGHAM YOUNG UNIVERSITY
E-mail address: mdorff@math.byu.edu

COLORADO COLLEGE
E-mail address: JMcDougall@ColoradoCollege.edu

UNITED STATES AIR FORCE ACADEMY
E-mail address: rolf.jim@gmail.com

UNITED STATES AIR FORCE ACADEMY
E-mail address: Beth.Schaubroeck@usafa.edu

BALL STATE UNIVERSITY
E-mail address: rstankewitz@bsu.edu

UNIVERSITY OF TENNESSEE, KNOXVILLE
E-mail address: kens@math.utk.edu

Contents

KEN STEPHENSON

Introduction

This book is written for undergraduate students who have studied some complex analysis and want to explore some research topics in the field. We envision this book could be used as:

- a supplement for a standard undergraduate complex analysis course allowing students as a group or as individuals to explore a research topic;
- a guide for undergraduate research projects for an advanced student or a small group of students; or
- a resource for senior capstone courses.

The nature of this book is quite different from other mathematics texts. Rather than having a straightforward linear and highly structured introduction to the material, this book is designed to regularly involve the student in a very hands-on way. This book focuses on discovery, self-driven investigation, and creative problem posing. Some of the ideas are part of the standard fare learned in a course which focuses solely on the topic, while others may be new to the field. We want to inspire the reader to investigate, explore, conjecture, and pursue mathematical ideas. Students are taken on a guided tour of the topics and are given many opportunities to stray from the text to pursue their own investigations.

Interlaced in the reading for each chapter are exercises, explorations using computer applets, and projects. These activities are an essential part of the students' learning of the topic. For this reason, most of these activities end with the phrase ***Try it out!*** to remind the student that the activity needs to be done before going on with the reading. The types of activities are as follows:

- **Examples** - The reader should be sure that she/he can follow the arguments and provide small details when needed.
- **Exercises** - These have a well-defined goal and should be done by the reader before going on to the next paragraph in the text. Skipping these would result in the reader missing an important skill or idea that will be fundamental to her/his understanding.
- **Explorations** - These also should be done before going on in the text. Generally, these do not have a well-defined problem that you are trying to "solve". Some may include undirected investigating or "playing" with applets. There is no specific outcome expected, but much will be gained by this activity. Such activities are at the heart of what this book is about since it is getting the student to explore on her/his own and see what she/he can come up with.
- **Small Projects** - These are more involved activities than those listed above and are optional. It may take up to a few weeks to complete a Small Project.
- **Large Projects** - These are similar to a Small Project, but on a larger scale. A Large Project could be a semester long project, a capstone project, or an honors thesis.

- **Additional Exercises** - Additional exercises may include any of the previous activities. However, these appear at the end of the chapter and are optional.

While working on these activities, it is a good habit to consider such questions as "Why was this problem posed?" "Why is it interesting?" "What if I changed the problem slightly, does it make it easier? harder? impossible?" "What does such a problem say about the general theory?" Thinking about such questions is what it means to do mathematics research and investige the unknown. Pausing to constantly ask new questions, investigate these questions, and mathematically describe these questions can make for very slow reading. The reader should measure progress, not by the number of pages read, but by the amount of independent thought given to the material. If the student reads just a few pages of a chapter and then becomes motivated to work on a problem or set of problems she/he devised on their own, the authors of this book would be delighted.

> "It is not so very important for a person to learn facts. For that he does not really need a college. He can learn them from books. The value of an education in a liberal arts college is not learning of many facts but the training of the mind to think something that cannot be learned from textbooks."–Albert Einstein (in Einstein: His Life and Times by Frank)

This book contains six research topics. Each topic is presented in a self-contained chapter that contains necessary background material, presentation of new material, exercises, explorations, and problems suitable for student projects, and several computer applets that allow the student to explore the topic. Also, each topic is a fairly recent area of research, and there are a lot of new questions to investigate. Here is a brief description of each of the chapters in this book:

(1) Complex Dynamics: This chapter investigates chaos and fractals as they relate to dynamical systems which come from iterating complex valued functions, i.e., given an initial value z_0 we consider the values $z_1 = f(z_0), z_2 = f(z_1) = f(f(z_0))$, $z_3 = f(z_2) = f(f(f(z_0))), \ldots$, and ask what kind of behavior we can have in this sequence z_n. Iteration in this sense arises in Newton's method for approximating roots of complicated functions, and so our chapter begins by asking such questions as: Which initial values will "work" for Newton's method (i.e., converge to a root)? If I change my initial value z_0 slightly, will I get similar or drastically different behavior? Often these questions are pursued computationally, visually, and experimentally with the aid of computer applets. We then extend our discussion by considering the iteration of any complex analytic map, which leads to a pursuit of the mathematics behind the famous Mandelbrot set, and much more.

(2) Soap films, Differential Geometry, and Minimal Surfaces: Minimal surfaces in $\mathbb{R}^3$ are beautiful geometric objects that minimize surface area locally. Visually, they can be thought of as saddle surfaces – at each point, the surface bends

upward in one direction in the same amount as it bends downward in its perpendicular direction. Minimal surfaces are related to soap films that result when a wire frame is dipped in soap solution. In this chapter, we present the necessary background from differential geometry, a field of mathematics in which the ideas and techniques of calculus are applied to geometric shapes, to give an introduction to minimal surfaces. Then we use ideas from complex analysis to present a nice way to describe minimal surfaces and to relate the geometry of the surface with this description. This allows us to begin investigating some of the interesting properties that can be studied with the help of the applets.

(3) **Applications to Flow Problems:** Two dimensional vector fields are used to model and study a wide range of phenomena. Of particular interest are vector fields that are both irrotational and incompressible. Such fields can be used to model the velocity of an ideal fluid flowing in a region or the electric field in a region free of charges. Modeling two dimensional fluid flow is a standard application of the theory of conformal mappings in complex variables. This chapter takes a geometric and visual approach to explore this standard body of work and then extends it to several more applications. Fields of interest typically include various sources or sinks that generate or remove fluid from the flow. Throughout the chapter, examples, theory, and exercises are used to develop methods that allow fields to be modeled that are generated by all types of sources and sinks in a variety of regions. Also, we have provided the applet *FlowTool* that readily displays the streamlines for a field with various sources and sinks. The applet permits real-time dynamical experimentation with the field. Students with an interest in using technology to visualize mathematical objects will find many opportunities to explore their ideas, though these are not explicit exercises.

(4) **Anamorphosis, Mapping Problems, and Harmonic Univalent Functions:** Complex-valued analytic functions have many very nice properties that are not necessarily possessed by differentiable real-valued functions. For example, if you can differentiate such a complex-valued function one time, then you can differentiate it infinitely many times. In addition, complex-valued analytic functions can always be represented as a Taylor series, and they are conformal (that is, they preserve angles) at points where $f' \neq 0$. Why does an analytic function have these properties? If $f = u + iv$ is an analytic function, then its real part $u(x, y)$ and its imaginary part $v(x, y)$ each satisfy Laplace's equation and thus are both harmonic. Also, u and v satisfy the Cauchy-Riemann equations and are therefore harmonic conjugates of each other. In this chapter we discuss some ideas and problems related to a collection of univalent (i.e., one-to-one) complex-valued functions $f = u + iv$, where u and v satisfy Laplace's equation but not necessarily the Cauchy-Riemann equations. This collection of functions are known as *harmonic univalent functions* or mappings, and contain the collection of analytic univalent functions as a subset.

(5) Mappings to Polygonal Domains: A rich source of problems in analysis is determining when, and how, one can create a univalent (one-to-one) function from one region onto another. In this chapter, we consider the problem of mapping the unit disk onto a polygonal domain by two different classes of functions. First for analytic functions we give an overview of the well established Schwarz-Christoffel transformation. This method leads to some very rich mathematics, the study of special functions, so we give a brief primer of a few special functions. We then diverge from analytic function theory and consider the Poisson Integral Formula to find harmonic functions that will serve as mapping functions onto polygonal domains. Proving that these harmonic functions are univalent requires us to explore some less known theory of harmonic functions and some relatively new techniques.

(6) Circle Packing: *Circle packings* are configurations of circles with prescribed patterns of tangency. They exist in quite amazing and often visually stunning variety, but what are they doing in a book on complex analysis? Well, the fact is that complex analysis is at its heart a *geometric* topic. The reader will see this in the global geometry on display throughout Chapters 1–5, but the foundation lies down at the local level where, as the saying goes, "analytic functions map infinitesimal circles to infinitesimal circles." In Chapter 6 this geometry will come to life in the theory of *discrete analytic functions* based on circle packing. Using the Java application *CirclePack*, we will create, manipulate, and display maps between circle packings which are the discrete analogues of familiar functions between plane domains, including some of those encountered earlier in the book. Direct access to the underlying geometry gives new insight into fundamental topics like harmonic measure, extremal length, and branching. Moreover, we will see that our discrete functions not only mimic their classical counterparts, but actually converge to them under refinement. In short, Chapter 6 is about *quantum* complex analysis.

Acknowledgements

out to be well beyond what any of us initially thought we would need or want from him, yet he adapted and produced very professional applets tailored to our specific desires. Since active use of these applets is critical to this text, we especially thank Jim for his tremendous efforts which were most critical to improving the overall project.

Chapter 1: Complex Dynamics As a text which hopes to inspire students by showing them the beauty of complex variables research, I would like to dedicate this chapter to the professors who have inspired me in this fashion, namely, Juha Heinonen, Joe Miles, and Aimo Hinkkanen.

Also, I sincerely thank Neal Coleman, Stephanie Edwards, Dan Lithio, Kevin Pilgrim, and Irina Popovici, who provided very useful feedback after reading through early drafts of this chapter. I also thank Bob Devaney and Phil Rippon for helpful conversations where they shared ideas and insights, and gave encouragement.

<div align="right">Rich Stankewitz</div>

Chapter 2: Soap films, Differential Geometry, and Minimal Surfaces I would like to thank my colleagues Casey Douglas and Michelle Hackman who read through the chapter and provided me feedback. Also, I would like to thank my 2009 and 2010 REU students who worked through the chapters and gave me feedback: Valmir Bucaj, Sarah Cannon, Amanda Curtis, Sam Ferguson, Laura Graham, Jamal Lawson, Rachel Messick, Jessica Spicer, Ryan Viertel, and Melissa Yeung. In addition, I would like to thank the BYU students in my complex analysis course who read through the chapter. Finally, I would like to thank Lonette Stoddard who used Adobe Acrobat to draw many of the images in the chapter.

<div align="right">Michael Dorff</div>

Chapter 3: Applications to Flow Problems I am particularly indebted to Harrison Potter who used ideas from this material in writing his undergraduate honors thesis at Marietta College. He inspired several of the ideas and approaches in this chapter. I would also like to acknowledge Professor Tristan Needham and his beautiful text Visual Complex Analysis for giving me a much greater appreciation of the subject of complex variables. I offer a special thank you to my colleagues Beth Schaubroeck and Jim Rolf, who are often my partners in crime for many of my crazy ideas. Finally, I wish to thank my co-authors for sticking with a project that lasted far longer than anyone thought it would.

<div align="right">Michael Brilleslyper</div>

Chapter 4: Anamorphosis, Mapping Problems, and Harmonic Univalent Functions I would like to thank my 2009 and 2010 REU students who worked through the chapters and gave me feedback: Valmir Bucaj, Sarah Cannon, Amanda Curtis, Sam Ferguson, Laura Graham, Jamal Lawson, Rachel Messick, Jessica Spicer, Ryan Viertel,

A note on using Java applets

Essential to the exposition in this text is the use of Java applets to explore the nature of complex functions. One applet, *ComplexTool*, will be demonstrated here. Many of the features of *ComplexTool* are also used in other applets, and *ComplexTool* is used in most of the chapters in this text.

We can graph the image of almost any domain set under almost any function using *ComplexTool*. Let's begin by examining how the image of the the unit disk $\mathbb{D}$ under the function $f(z) = (1+z)^2$ is drawn. First, open *ComplexTool* (see Figure 0.1).
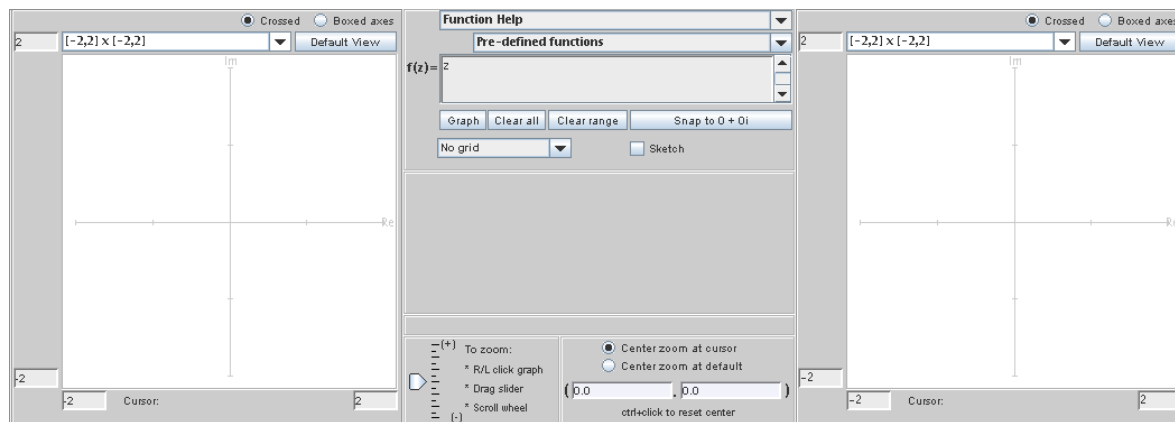
FIGURE 0.1. The applet *ComplexTool*

In the middle section near the top there is a box that has **f(z)=** before it. In this box, enter (1+z)$\wedge$ 2. (In general, you will enter functions into *ComplexTool* using the same syntax you would in a standard computer algebra system.) Below this, there is a window that reads `No grid`. Click on the down arrow ▼ and choose the option `Circular grid`; an image of a circular grid should appear on the left to show the domain being graphed. Next, click on the button `Graph` which is in the middle section below the function you entered earlier. The image of the circular grid under the action of the function should appear on the right. To change the size of the image window, click on the down arrow ▼ above the image and chose a different size, such as `Re: [-3,3] Im: [-3,3]`. You can also zoom in and out by left- or right-clicking on your mouse, by dragging the slider in the bottom left of the center panel, or by turning the scroll wheel on the mouse. Also, you can move the axes so that the image is centered by positioning the cursor over the image, clicking on the mouse button, and dragging the image to the left (see Figure 0.2). To go back to the original region you graphed, push the `Default View` button above either the domain or range. If you want to change the domain that is being graphed, you can do so manually by changing `Center`, `Interior circles`, `Rays`, the boxes surrounding θ, and the boxes surrounding `radius`. Alternatively, you can check the box that says `Vary` θ or `Vary`

radius to see sliders to dynamically change the θ-range and radius range. (You may have to push the **Graph** button to regraph after you make changes.) You can also position the cursor over the domain and "drag" it around while watching the range dynamically change–this will not change the size of the domain, but will change its center. To go back to where you had originally centered the object, you can push the **Snap to** button.
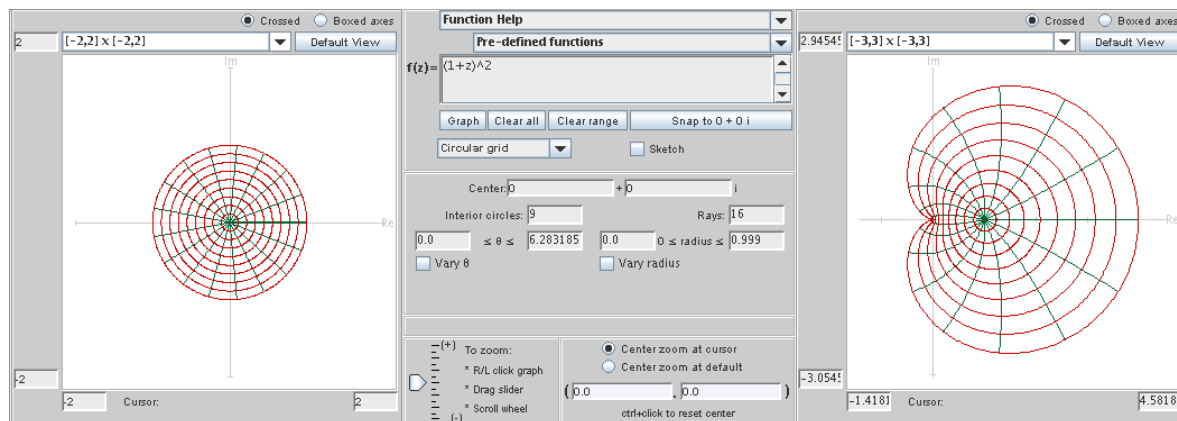
FIGURE 0.2. The image of the unit disk under the map $f(z) = (1 + z)^2$.

To figure out exactly which points in the domain correspond to points on the range, you can check the box in front of **Sketch**. Then you can "scribble" in the domain window and see its image appear in the range. Note that when the **Sketch** function is turned on, you cannot dynamically move the domain, as a mouse-click in the domain window is interpreted as sketching instead.

EXPLORATION 0.1. Use *ComplexTool* to graph several familiar functions. Some functions you might want to try are listed below.

(a) Graph the image of $\mathbb{D}$ under the function $f(z) = \frac{z}{1-z}$.

(b) Consider the two analytic functions $f_1(z) = 2z^2 - z$ and $f_2(z) = \frac{1}{2}z^2 - z$ each defined on $\mathbb{D}$. One of these functions is one-to-one while the other is not. Graph the image of $\mathbb{D}$ under the function $f_1(z)$ and under $f_2(z)$. Determine how you can tell by looking at the applet which the function is one-to-one.

(c) Determine the smallest set of values that you can use for **theta** in a circular wedge domain to get an image of a complete disk under $f(z) = z^3$. (The **Vary** θ option is good to use for this exercise.)

(d) For the function $f(z) = e^z$, plot the images of vertical and horizontal lines. Explain your result mathematically.

(e) Determine if there exists a domain D so that the image of D under $f(z) = \log(z)$ covers the plane.

(f) Demonstrate the periodicity of $\sin(z)$ using *ComplexTool*. [Hint: Find a region to graph (either circular or rectangular), and drag it horizontally to watch its image change.]

Try it out!

In addition to the action of the applet, you can use the menus across the top to change various settings. Perhaps most helpful is the option to export screen shots of the applet. The `Export Settings` menu allows you to choose which file type to export, and then you can choose to export the domain, range, or the whole screen. As you do the exercises and explorations in this text, you will find it helpful to document the work that you have done by taking careful hand-written or electronic notes.

CHAPTER 1

Complex Dynamics: Chaos, Fractals, the Mandelbrot Set, and More

Rich Stankewitz (text and applet design), Jim Rolf (applet coding and design)

1.1. Introduction

This chapter introduces *complex dynamics*, an area of mathematics that continues to inspire much ongoing research and experimentation. The goal of this chapter is not to give a comprehensive or step-by-step approach to this topic, but rather to get the reader engaged with the general notions, questions, and techniques of the area – but even more so, to encourage the reader to actively pose as well as pursue their own questions. To better understand the nature and purpose of this text, the reader should be sure to read the Introduction before proceeding.

Dynamics, in general, is the study of mathematical "systems" that change over time, i.e., *dynamical* systems. For example, consider a particular Newtonian model for the motion of the planets in our solar system. Here the mathematical "system" is a collection of variables corresponding to the location and velocity of each of the planets relative to the sun, and this system changes over time according to Newton's laws of motion. Thus, as it turns out, one can describe the *process* by which the system evolves (i.e., the rules of how the system changes over time) by differential equations relating all the system variables to each other and to Newton's laws of motion.

Many dynamical systems can be described similarly. Such examples include the population of bacteria in a petri dish, the weather in Muncie, IN (temperature, pressure, and wind velocity, to be more precise), the global temperature of the earth, and the flight of a paper airplane that you might toss across the room. The models for such dynamical systems all have a set of system variables, and some rule or set of equations that describes the process of how these variables change over time. The values of such variables at any particular moment in time t is called the *state* of the system at time t.

Knowing the *initial state* of the system (e.g., today's location and velocity of each planet), we often try to analyze the equations which describe how the system variables change over time in order to the answer such questions as: What will the state of the system be tomorrow? next week? next year? 100 years from now? Will the system in

the long run settle into some sort of equilibrium? Will small changes or errors in our knowledge of the initial state only lead to small changes or errors in the system at some future time, or will such a small initial error lead to huge errors in the future? Such natural questions about one's ability to predict future states of a dynamical system have led to some very useful results of practical importance – and some very beautiful mathematics.

The famous physicist Richard P. Feynman has a quote regarding a common attitude towards the important question of a dynamical system's predictability, or lack thereof ([**15**], p. 9): "Physicists like to think that all you have to do is say, these are the conditions, now what happens next?" Unfortunately (or fortunately, depending on your perspective) there are many systems which we cannot adequately "solve" for the purposes of making useful predictions, and so "what happens next?", a question taken to mean asking what happens well into the future, is simply a question that we cannot answer. Frequently this is due to an inability to find the right pattern, or to "solve" some differential equation. For these systems there is hope that someone in the future (maybe you!) will come along and find a clever way to solve such problems. With such solutions one can then predict the future of these systems and apply their predictions to the real world. But for a large class of systems there is no hope of *ever* being able to predict, with any *useful accuracy*, what such a system will behave like in the future. What's astounding is that, for these systems, it is not a matter of finding the right clever "solution". In fact, we sometimes even have what we thought was a great solution, a formula even, but there is a problem with applying this formula to the real world.

The heart of this problem is that we have only solved a model, an approximation to the real world, and as an approximation we have to accept the fact that there will be some (hopefully small) error built into the setup of our model. The problem creeps in because of two issues that may arise in certain systems: (1) we can never pinpoint the initial conditions *exactly*, and (2) *any* approximation (or error) to the initial conditions leads to errors in our long term predictions so large that we cannot have any confidence at all in our application of our prediction to the real world. Such systems are called *chaotic*, a term you will explore and even be asked to *mathematically define* further into this text. Even though such systems do not allow for precise answers to some of the questions scientist like to ask, much can be gained from studying them.

In this chapter we will study certain so-called "simple" chaotic systems that can be analyzed using the tools of complex analysis. We do this because such systems are of interest in their own right, but also because it will lead to understanding fundamental principles of incredibly complicated systems, like those mentioned above. The dynamical systems we focus on are *discrete iterative* systems, which are in some sense the easiest kind of dynamical systems. For these, time is represented by a natural number n, and there is no need to solve a differential equation to determine the system's state down the road - one simply needs to repeatedly apply a function. Furthermore, the states of our systems are described not by a large number of variables (as are needed to

represent all the positions and velocities of the all the planets), but just a single complex variable. The system is then called iterative because the state z_n of the system at time n will evolve according to the rule $z_{n+1} = f(z_n)$ where f is a complex valued function. Thus, given an initial state, computing the future state z_n is just a simple matter of computing the values $z_1 = f(z_0), z_2 = f(z_1) = f(f(z_0)), z_3 = f(z_2) = f(f(f(z_0)))$, and so on. But don't be fooled, *predicting* the behavior of the sequence of states z_n (e.g., deciding if this sequence z_n converges or not), an altogether different problem, is by no means simple!

In this chapter we investigate several types of discrete iterative systems which make use of tools from complex function theory, including Newton's method, polynomial iteration, exponential iteration, and trigonometric iteration. We also consider what happens when such systems are perturbed by changing a parameter, thus leading us into what is called bifurcation theory. There are, however, many ways to perturb a system and in the concluding section we describe two such ways: perturbation with a pole and random dynamics. We begin, however, with perhaps the most familiar discrete dynamical system, Newton's method.

Although we are not studying the above mentioned "real world" dynamical systems directly, we should keep in mind that these systems exhibit many of the same behaviors as the systems we do study. In fact, much of the same phenomena we encounter here have direct analogs in flavor, if not in a strict mathematical sense, in the millions of dynamical systems you may encounter everyday.

How to use this chapter

The sections of this chapter can, of course, be worked through in the order presented. However, one wishing to proceed to Section 1.3 more quickly may do so by skipping all of Section 1.2 with the exceptions of Sections 1.2.3 and 1.2.6. Also, Section 1.6 may be skipped by anyone not wishing to investigate the dynamics of transcendental entire maps. Reading the entire chapter and working on several additional exercises as well as large/small projects would suit a 3 credit hour 15 week semester reading course. However, one looking for a 2-3 week group or individual project for the end of the semester in an undergraduate complex variables course could pursue either A) Section 1.2 or B) Sections 1.2.3 and 1.2.6 together with Section 1.3.

Appendix A on page 406 contains a notation page, as well as a brief exposition of results from a standard undergraduate complex variables course. The reader should use it for reference as needed. Appendix B on page 419 contains a review, and perhaps, a brief introduction to a few new concepts related to the Riemann sphere. These concepts are very necessary for this chapter, and so the reader should expect to work through all of Appendix B; however, it is not necessary that it be completed in full before beginning this chapter. There are also three chapter appendices which begin on page 96 providing added information relevant to just this chapter.

This chapter will utilize the following applets which can be accessed at http://rstankewitz.iweb.bsu.edu/DynamicsApplets.htm.

1. *Real Newton Method Applet* - is used to visualize the real-valued Newton's method.

2. *Complex Newton Method Applet* - is used to visualize the complex-valued Newton's method.

3. *Real Function Iterator Applet* - is used for iterating any real function, and seeing the orbit displayed as a numerical list and as points plotted on a number line.

4. *Complex Function Iterator Applet* - is used for iterating any complex function, and seeing the orbit displayed as a numerical list and as points plotted in the complex plane.

5. *Cubic Polynomial Complex Newton Method Applet* - is used for exploring the attracting basins for Newton's method applied to the family of cubic polynomials $p_\rho(z) = z(z-1)(z-\rho)$.

6. *Global Complex Iteration Applet for Polynomials* - is used to draw the basin of infinity for any polynomial.

7. *Mandelbrot Set Builder Applet* - is used to illustrate how the Mandelbrot set is constructed.

8. *Parameter Plane and Julia Set Applet* - is used for investigating both the parameter plane and dynamic plane pictures for the families of functions $z^2 + c, z^d + c, ce^z, c\sin(z), c\cos(z)$, and $z^d + c/z^m$.

1.2. Newton's Method

Solving equations, finding solutions to ordinary differential equations, finding eigenvalues of a matrix - all of these are very important mathematical procedures. However, each of these can be done *exactly*[1] only in very restrictive cases. When we come across a situation that is not one of these special cases, often the solution must be *approximated* via a numerical method, instead of computed *exactly*. And often, the numerical method is iterative in nature.

For example, consider the problem of finding a *root* of a complex valued function $f(z)$, i.e., a value α such that $f(\alpha) = 0$. If the function is the quadratic $f(z) = az^2 + bz + c$ where $a \neq 0$, then there are two roots given by the quadratic formula $\alpha_\pm = (-b \pm \sqrt{b^2 - 4ac})/2a$. If the function is a cubic or quartic polynomial, then there also exists formulas (or, more precisely, procedures) for *exactly* finding the roots. However, if $f(z)$ is a quintic polynomial $f(z) = az^5 + bz^4 + cz^3 + dz^2 + ez + h$, then there is not, in general, a procedure that will *exactly* find any of its roots.[2] The same is true for many so-called transcendental functions such as $h(z) = \cos z - z$ and $g(z) = e^z - 4z$.

---

[1]By computing a value *exactly* we usually mean being able to express the value in terms of standard mathematical operations and functions, e.g., $\sqrt{\cos(\pi/12)}$.

[2]Note that any polynomial $p(z)$ of degree n has, by the Fundamental Theorem of Algebra, n roots in the complex plane. However, this theorem does not help us to actually find them.

In such cases, one often must give up on finding *exact* roots and resort to *approximation* methods.

When considering real valued functions of a real variable, there are approximation methods for root finding based on the Intermediate Value Theorem, such as the bisection method. When considering a complex valued function $f(z)$, one can sometimes approximate a root of f using Rouche's Theorem (see [**1**], p. 294). Specifically, if f and g are analytic on and inside a simple closed curve C with $|f| > |g|$ on C, then f and $f + g$ have the same number of zeroes (counting multiplicities) inside C. Hence, if g can be chosen such that $f + g$ has a known root inside of C, then so must f. However, even if f is a nice function (e.g., a polynomial), this can often be a difficult method to implement (and still hope to find good estimates). So we look for a better method. One of the best methods to apply in either situation (real or complex) is Newton's method. Often it allows one to approximate roots with extreme accuracy and extreme speed, if we have access to a computer. In this section we examine Newton's method for both real valued and complex valued functions, with the goal of understanding when it will succeed and when it will fail.

1.2.1. Real Newton's Method. If one seeks to find a root α of a differentiable real valued function $f(x)$ defined for a real variable x, then one can apply Newton's method as follows. We start with an initial guess x_0 **close to** α and define $x_1 = x_0 - \frac{f(x_0)}{f'(x_0)}$, $x_2 = x_1 - \frac{f(x_1)}{f'(x_1)}$, $x_3 = x_2 - \frac{f(x_2)}{f'(x_2)}$, and, in general, $x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}$.

The geometric reasoning behind Newton's method, as illustrated in Figure 1.1, is as follows: Given an approximation x_0 to the root α, one considers a linear function $\tilde{f}(x)$ which approximates the function $f(x)$ near x_0. The best linear approximation in this case will be given by the first order approximation $\tilde{f}(x) = f(x_0) + f'(x_0)(x - x_0)$ whose graph is the tangent line L to the graph of $f(x)$ at the point x_0. The root of the approximating function $\tilde{f}(x)$, i.e., the x-intercept of L, is then the definition of x_1.

EXERCISE 1.1. Before going on, compute $\tilde{f}(x)$ for $f(x) = x^3 - 2x$ and $x_0 = 2$ as in Figure 1.1. Then use $\tilde{f}(x)$ to find the equation of the tangent line L and also check that x_1, given by the above formula, is the x-intercept of L. ***Try it out!***

EXERCISE 1.2. Verify that for general $f(x)$ the formula given for $\tilde{f}(x)$ has a root at x_1 as defined above. ***Try it out!***

In general (as illustrated in Figure 1.1), we *expect* that this root x_1 of $\tilde{f}(x)$ will be a better approximation to α, the sought after root of $f(x)$, than the initial guess x_0. This process is repeated now using x_1 as the initial guess to generate a new, hopefully improved, approximation x_2. We then iterate this process, i.e., apply this procedure over and over again, to generate successive approximations x_n, for $n = 0, 1, 2, \ldots$. For various reasons it will be useful for us to express this process in terms of iteration of the following function.

DEFINITION 1.3. The function $F(x) = x - \frac{f(x)}{f'(x)}$ is called the *Newton Map for f*.
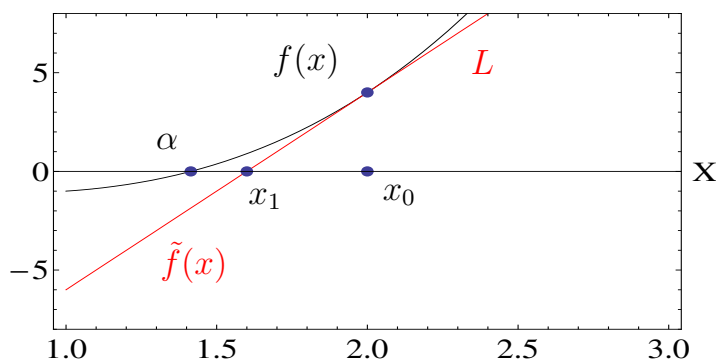
FIGURE 1.1. An illustration of the first step in Newton's method where $f(x) = x^3 - 2x$ and $x_0 = 2$. Newton's method can be succinctly described as "from x_0 move to the graph of $f(x)$, slide along the tangent until you hit the x-axis, and then repeat".

Thus, the *orbit*, i.e., sequence of iterates $F(x_0), F(F(x_0)), F(F(F(x_0))), ...$, is the same as the sequence x_n generated above, and it will be proven that x_n converges to the sought after root α whenever our original guess x_0 is "close enough" to α (see Proposition 1.21). Of course, this issue of what does it mean to be "close enough" is very important. We will come back to this later (see Remark 1.24), but we first explore this method with some examples.

EXAMPLE 1.4. Consider $f(x) = x^2 - 3x + 2 = (x-1)(x-2)$, which clearly has roots at 1 and at 2. Let's apply Newton's method to see how it works. We first compute the Newton Map for f which is $F(x) = x - \frac{x^2 - 3x + 2}{2x - 3} = \frac{x^2 - 2}{2x - 3}$. If we make an initial guess $x_0 = 0.5$, then using your calculator (do this now) you can compute $x_1 = F(x_0) = 0.875, x_2 = F(F(x_0)) = F(x_1) = 0.9875$, and $x_3 = F(F(F(x_0))) = F(x_2) = 0.99984756$. Since using the calculator is drudgery and computers are so efficient at such tasks, we have created a *Real Newton Method Applet* for you to use. Use this now to confirm the calculations above and then, by taking many iterates of the Newton map using this applet, convince yourself that with the *initial value* (a term we use interchangeably with the terms *seed value* and *starting point*) $x_0 = 0.5$, we have $x_n \to 1$. **Try it out!**

Now use the *Real Newton Method Applet* to determine the behavior using an initial guess $x_0 = 3$. **Try it out!**

What you learned above is that some initial guesses for x_0 *find* the root 1 (i.e., have the corresponding x_n limit to 1) while other initial guesses find the root 2. This begs the question: Given an initial guess x_0, how do we know which root it will find?

EXPLORATION 1.5. Make an intelligent prediction about which seed values for x_0 in Example 1.4 will find the root 1 and which will find the root 2. Are there any seed values for which Newton's method fails to find either root? Experiment with the *Real Newton Method Applet* to test your prediction. ***Try it out!***

EXPLORATION 1.6. Using the function $f(x) = x(x-1)(x+1)$ determine the Newton Map for f and then analyze Newton's method using different initial guesses for x_0 in the *Real Newton Method Applet*. Describe (as best as you can) the set of seed values for x_0 which find the root -1, and then do the same for for the roots 0, and 1. ***Try it out!***

1.2.2. Global Picture of Real Newton Method Dynamics. Identifying how the orbits of *all* seed values behave, as you attempted in Explorations 1.5 and 1.6, is what we mean by looking at the *global* dynamics of Newton's method. In some cases, this can be done without too much work, but in other cases, it turns out to be very complicated. In order to help with one labor intensive approach to this problem, use the **Graph basins of attraction** feature in the *Real Newton Method Applet* to display by using different colors which initial guesses will "find" which roots of $f(x)$ when Newton's method is applied.

EXERCISE 1.7. For each of Explorations 1.5 and 1.6, use the *Real Newton Method Applet* to get a one picture snapshot of what the dynamics of Newton's method are for *all* starting values. ***Try it out!***

What you see when using the *Real Newton Method Applet* with $f(x) = (x-1)(x-2)$ is, perhaps, what one might expect. The initial values which are closer to the root at 1 will find 1, and the initial values which are closer to the root at 2 will find 2. One should also check that Newton's method fails when the initial value $x_0 = 1.5$ is used. *Analytically* we see this in the formula because $f'(1.5) = 0$ leads to a zero in the denominator when one attempts to calculate x_1. *Geometrically* we see this by noting that the tangent line at $x_0 = 1.5$ becomes horizontal, never crosses the x-axis, and thus leaves x_1 undefined. We can also understand this *dynamically*, at least in a heuristic way. The point 1.5 separates those points which are pulled or attracted to 1 and those points which are attracted to 2, and so by an informal use of symmetry, it would *seem* that something must fail to work out at exactly $x_0 = 1.5$.

Though this reasoning is informal, it does seem to capture an important idea at play here. As a rule, we encourage the reader to often make use of and even create your own heuristic ideas to explain or describe mathematics. Sometimes it's hard to be formal with all of your mathematical ideas. But don't let that stop you from thinking of and sharing great mathematical thoughts, even if you can't make them precise or formal. Some of the best mathematics, if not all mathematics, starts off as raw unformed ideas with no foundation whatsoever in formalism. Later, one can (and should) try to be formal with their ideas.

The case when $f(x) = x(x-1)(x+1)$ is much different from the quadratic case. It is not simply that a seed for Newton's method will find the root it is nearest to. For example, $x_0 = 0.55$ will find the root at -1, even though it is closer to both the roots at 0 and 1. Indeed, in this case the set of initial guesses on the real line is divided into intricate regions of points that find the various roots. In fact, if you zoom in near the point $x = 0.4472135951871958$ you will see a cascade of ever shrinking and alternating colored intervals of blue and turquoise (see Figure 1.2). It turns out, in fact, that this pattern goes on forever (see Small Project 1.8). Use the **Zoom** feature of the *Real Newton Method Applet* to observe this.

FIGURE 1.2. On the left is a cascade of intervals in the global picture of Newton method dynamics for $f(x) = x(x-1)(x+1)$, and on the right is a magnification centered at $x = 0.4472135951871958$.

SMALL PROJECT 1.8. Prove the existence of the infinite cascade of ever shrinking and alternating colored intervals of blue and turquoise found in Figure 1.2. Hint: First use the applet to understand what each of these colored intervals means dynamically, and then try to give a proof for what you witness in the applet.

So we see that Newton's method gets really complicated to understand globally when we switch from a quadratic to a cubic, specifically from $f(x) = (x-1)(x-2)$ to $f(x) = x(x-1)(x+1)$. This begs the questions: Why? How? Is there a way to know when a system will necessarily be simple or complicated ahead of time?

Answers will come from taking the advice of Jacques Hadamard who once said, "The shortest path between two truths in the real domain passes through the complex domain." So let's look at Newton's method applied to complex valued functions of a complex variable. There are many wonderful theorems and structures at our disposal when we consider complex analytic maps instead of mere real valued differentiable maps. Let's take advantage of some of these.

Before we investigate Newton's method applied to complex valued functions, however, we first take some time to develop some important concepts that we'll need, not just to understand Newton's method, but also any kind of iterative dynamics.

1.2.3. Orbits, Examples, and Fixed Points. The general questions we consider in iterative dynamical systems concern describing and predicting what we shall call *orbits*. Let g be a function mapping its domain set domain(g) into itself,[3] which we take to be a subset of the Riemann sphere[4] $\overline{\mathbb{C}}$. We then denote the nth iterate of g by g^n. Thus $g^n(z) = (g \circ \cdots \circ g)(z)$ where the function g is applied n times, e.g., $g^3(z) = g(g(g(z)))$. Note that in this chapter $g^3(z)$ does NOT denote the value $g(z)$ raised to the third power, which would instead be denoted $[g(z)]^3$. We also define g^0 to be the identity map, i.e., $g^0(z) = z$. Furthermore, for any starting (seed) value $z_0 \in$ domain(g), the sequence of points $z_n = g^n(z_0)$, for $n = 1, 2, 3 \ldots$, is called the **orbit** of z_0 (under the map g).

When trying to predict the behavior of the evolution of a seed value, we inquire about such things as: Does the orbit converge, fall into a repeating pattern, show no signs of following any pattern at all? What happens for different starting points z_0? Do we get the same (or even similar) behavior if we choose starting points near z_0? We keep these questions in mind as we consider the following examples.

EXAMPLE 1.9. Say $f(x) = e^x$ for all $x \in \mathbb{R}$. Thus the iterates are $f^0(x) = x$, $f^1(x) = e^x, f^2(x) = e^{e^x}, f^3(x) = e^{e^{e^x}}$, and so on. Experiment with this on the *Real Function Iterator Applet* to convince yourself that $f^n(x) \to +\infty$ no matter what $x \in \mathbb{R}$ we start with. Additional Exercise 1.155 asks for a formal proof.

EXAMPLE 1.10. Let $f(x) = \sin x$ where $x \in \mathbb{R}$ is given in radians. Experiment with this on the *Real Function Iterator Applet* to convince yourself that $f^n(x) \to 0$ for *any* real number x. Additional Exercise 1.156 asks for a formal proof.

EXAMPLE 1.11. Let $f(x) = \cos x$ where $x \in \mathbb{R}$ is given in radians. Experiment with this on the *Real Function Iterator Applet* to convince yourself that $f^n(x) \to 0.739085....$ for *any* real number x. Additional Exercise 1.157 asks for a formal proof.

EXAMPLE 1.12. Let $f(x) = x^2 - 1$ and $x_0 = 0.9$. Use the *Real Function Iterator Applet* to convince yourself that the *tail end* of the orbit $x_n = f^n(0.9)$ *appears* to oscillate back and forth between 0 and -1.

EXAMPLE 1.13. Let $f(x) = 4x(1-x)$ and $x_0 = 0.2$. Use the *Real Function Iterator Applet* to see that the orbit $x_n = f^n(0.2)$ *appears* to have no patten to it at all, even after the first 25,000 orbit points are plotted. Zoom in on the orbit points to see how it *appears* that they are *dense* in the interval $[0, 1]$, that is, for every open interval (a, b) which meets $[0, 1]$, there is some orbit point $x_n \in (a, b)$.

---

[3]The **domain set** of a function g is the set of possible inputs and is denoted domain(g).

[4]The Riemann sphere $\overline{\mathbb{C}}$ is the natural setting for the functions in this chapter, and so the reader needs to be familiar with all the material in Appendix B starting on page 419.

Though many types of behavior can be exhibited in orbits, we wish to focus on one special type for the moment. In each of Examples 1.9, 1.10 and 1.11, we see that there is a point which "attracts" the orbits of many points in the domain of the function. This is a fundamental notion (and is at the heart of what we witnessed in the Newton's method examples) and so we provide the following definition.

DEFINITION 1.14 (Attracting Basin). Let $w \in \overline{\mathbb{C}}$. For any complex valued function g mapping its domain set into itself, we define the *attracting basin* of w (also called *basin of attraction*) under the function g, to be the set $A_g(w)$ of all seed values whose orbit limits to the point w, i.e., $A_g(w) = \{z \in \mathrm{domain}(g) : g^n(z) \to w\}$.

Note that the point w in the above definition does not necessarily have to lie domain(g) (as in Example 1.9). However, if $w \in \mathrm{domain}(g)$ and g is continuous (as in Examples 1.10 and 1.11), then the following result shows that w must necessarily be a ***fixed point*** of g, i.e., $g(w) = w$, whenever $A_g(w)$ is non-empty.

THEOREM 1.15. Let $f : \mathrm{domain}(f) \to \mathrm{domain}(f)$ be a continuous map where domain(f) $\subseteq \overline{\mathbb{C}}$. Suppose a and x_0 are both in domain(f) and $f^n(x_0) \to a$. Then $f(a) = a$.

PROOF. Since the sequence $f^n(x_0) \to a$ and f is continuous at a, we must have $f(a) = f(\lim_{n \to \infty} f^n(x_0)) = \lim_{n \to \infty} f(f^n(x_0)) = \lim_{n \to \infty} f^{n+1}(x_0) = a$. $\qquad\square$

Fixed points play a major role in dynamical systems and so we will be careful to pay special attention to them whenever they arise. In particular, we note from our previous examples that roots of f always *appear* to be fixed points of the corresponding Newton map $F(z)$, a fact we formally prove in Proposition 1.21. Additional Exercise 1.159 will shed some light on the extent to which the converse holds.

We call the fixed points in Examples 1.10 and 1.11 *attracting* fixed points because seed values near the the respective fixed points will iterate toward the respective fixed points. In order to be more precise we give a formal definition, but before we do, we remind the reader of a key relationship between the Euclidean metric on $\mathbb{C}$ and spherical metric σ on $\overline{\mathbb{C}}$, which is stated in Proposition B.7 on page 424. Namely, for points $z, w \in \mathbb{C}$, we have $|z| > |w|$ if and only if $\sigma(z, \infty) < \sigma(w, \infty)$.

DEFINITION 1.16 (Attracting Fixed Point). Let f be a map from its domain set $\Omega \subseteq \overline{\mathbb{C}}$ into itself (note that Ω could be a subset of $\mathbb{R}$).
(a) We call a (finite) fixed point $a \in \mathbb{C}$ an *attracting fixed point* of f if there exists a neighborhood U of a such that for any point $z \in \Omega \cap U \setminus \{a\}$, we have $|f(z) - a| < |z - a|$, i.e., the action of f is to move each point in $\Omega \cap U \setminus \{a\}$ *closer to a*.
(b) Suppose $f(\infty) = \infty$. We call ∞ an *attracting fixed point* of f if there exists a neighborhood $U \subseteq \overline{\mathbb{C}}$ of ∞ such that for any point $z \in \Omega \cap U \setminus \{\infty\}$, we have $|f(z)| > |z|$, i.e., the action of f is to move each point in $\Omega \cap U \setminus \{\infty\}$ *closer to ∞* (as measured by the spherical metric).

In Definition 1.16(a) we used the Euclidean metric to describe when the action of f moves points closer to a, but we could have equivalently used the spherical metric for this purpose by writing $\sigma(f(z), a) < \sigma(z, a)$. The reader should become comfortable with understanding when the particular metric used in a definition or result could equivalently be changed to another standard metric. We will often make use of the simplest metric at our disposal, relying on the reader to know when another metric could also be used. A good way to try to become comfortable with this concept is to frequently consider how sets appear when visualized in the flat plane $\mathbb{C}$ and in the sphere $\overline{\mathbb{C}}$. The reader will find an exposition of this material in Appendix B on page 419.

If a is an attracting fixed point of a continuous map f, then there necessarily exists some neighborhood U of a such that $U \subseteq A_f(a)$. The proof of this does not require that f be (real or complex) differentiable at a, but without such a differentiability condition the proof is more technical. Since we are only interested in specific differentiable functions in this chapter, we provide the proof only for the case that $|f'(a)| < 1$ (see Theorem 1.18 below).

REMARK 1.17. It is worth noting the curious fact that it may be the case that f has a fixed point a' such that $A_f(a')$ contains a whole neighborhood of a' without a' being attracting according to Definition 1.16 (see Example 1.50).

THEOREM 1.18. Suppose $\Omega \subseteq \mathbb{R}$ or $\Omega \subseteq \mathbb{C}$. Let $f : \Omega \to \Omega$ be such that $f(a) = a$ and $|f'(a)| < 1$. Then a is an attracting fixed point of f. Furthermore, there exists some $\epsilon > 0$ such that[5] $\triangle(a, \epsilon) \cap \Omega \subseteq A_f(a)$.

Note that the proof given below applies equally well to both cases $\Omega \subseteq \mathbb{R}$ or $\Omega \subseteq \mathbb{C}$, where f' denotes, respectively, the real or complex derivative.

PROOF. Since $|f'(a)| < 1$ we may select some β such that $|f'(a)| < \beta < 1$. Since, by definition, $f'(a) = \lim_{z \to a} \frac{f(z) - f(a)}{z - a}$, there exists $\epsilon > 0$ such that for any $z \in \Omega \setminus \{a\}$ for which $|z - a| < \epsilon$, we have

$$\left| \frac{f(z) - a}{z - a} \right| = \left| \frac{f(z) - f(a)}{z - a} \right| < \beta.$$

Hence for $z \in \Omega$ with $|z - a| < \epsilon$, we know that $|f(z) - a| \leq \beta|z - a| < \epsilon$. This says that for points z *near* a (within a distance of ϵ), the function f moves z closer to a by a factor of at least $\beta < 1$. Hence a is an attracting fixed point by Definition 1.16. If we then iterate the map f at z, generating the orbit of z, we know that each application of f takes the corresponding orbit point a step closer to a (by a factor of β). Hence we may use induction to show that $|f^n(z) - a| \leq \beta^n |z - a| \leq \beta^n \epsilon \to 0$ whenever $z \in \Omega$ with $|z - a| < \epsilon$. Thus $\triangle(a, \epsilon) \cap \Omega \subseteq A_f(a)$. $\qquad\square$

---

[5]Here, $\triangle(z_0, r) = \{z \in \mathbb{C} : |z - z_0| < r\}$ is the open Euclidean disk with center z_0 and radius r.

REMARK 1.19. Note that the smaller the value of $|f'(a)|$ (and hence the smaller the value of β may be chosen) in the above theorem, the faster the convergence is. In particular, if $f(a) = a$ and $f'(a) = 0$, then the β in the above proof can be taken to be extremely small leading to very fast convergence. Hence, in such a case the fixed point a is called **super attracting**.

EXAMPLE 1.20. Use the *Real Function Iterator Applet* to compare the rates of convergence given by the following maps. For $f(x) = \sin x$ consider the rate at which $f^n(\frac{1}{2})$ converges to 0. For $g(x) = x^2$ consider the rate at which $g^n(\frac{1}{2})$ converges to 0. For $h(x) = \frac{1}{2}x$ consider the rate at which $h^n(\frac{1}{2})$ converges to 0. Now compare the absolute value of the derivative at each function's fixed point and note the correspondence with Remark 1.19. **Try it out!**

By Theorem 1.18, in order to check if a fixed point is attracting, one can simply check the absolute value (or modulus) of the derivative at the fixed point. Note that in Example 1.11 one has $|f'(0.739085....)| < 1$ which proves that this fixed point is indeed attracting. When considering the Newton's method dynamics, it appears, in each of the cases we explored experimentally, that the roots of f are attracting fixed points for the Newton map F. In fact, this is exactly the case, and we will prove this in Proposition 1.21 by showing that $|F'| < 1$ at each root of f. But before we do, we give a word of caution.

We must be a bit careful with the use of Theorem 1.18 since it is not, in general, an if and only if statement. Consider the fixed point $a = 0$ for the map $f(x) = \sin x$ in Example 1.10. Here $f'(0) = 1$, but this function *as a real map*, defined only for all real numbers x, does have a genuine attracting fixed point at $a = 0$ (see Additional Exercise 1.156). However, if we consider the *complex map* $g(z) = \sin z$, defined for all complex numbers z, the fixed point $a = 0$ is no longer attracting. Indeed, $g^n(\pm i\epsilon) \to \infty$ for any $\epsilon > 0$. The reader is asked to prove this formally in Additional Exercise 1.158. However, the reader can see this illustrated by using the *Complex Function Iterator Applet*. **Try it out!**

1.2.4. Complex Newton's Method. We now return to our investigation of the dynamics of Newton's method when we allow our variables and output values to be complex valued. As we saw with many earlier examples, our experimentation with the applets suggest that the Newton map always has an attracting fixed point at each root of f. We can now carefully state and prove this fact in both the real and complex cases.

PROPOSITION 1.21 (Attracting Property of Newton's Method). Given any non-constant *real analytic*[6] or complex analytic function f with a root at $\alpha \in \mathbb{C}$, the point α is an attracting fixed point of the Newton map F and thus there exists $r > 0$ such

---

[6]A real valued function of a real variable is said to be *real analytic* if it possesses derivatives of all orders and agrees with its Taylor series in a neighborhood of every point in its domain set.

that all points within a distance r of α are necessarily in the set $A_F(\alpha)$. Put another way, for all initial values z_0 that are close enough to α (specifically, within a distance r) the successive approximations $F^n(z_0)$ converge to α, i.e., starting from such a z_0 and defining $z_{n+1} = z_n - \frac{f(z_n)}{f'(z_n)}$, we must have $z_n \to \alpha$.

PROOF. We consider the case that $f(z)$ is complex analytic, noting that the same proof applies when f is real analytic. In light of Theorem 1.18, it suffices to show that for $F(z) = z - \frac{f(z)}{f'(z)}$, we have $|F'(\alpha)| < 1$. We first note that $F'(z) = \frac{f''(z)f(z)}{(f'(z))^2}$.

If $f'(\alpha) \neq 0$, then clearly $F'(\alpha) = 0$ (since $f(\alpha) = 0$). However, if $f'(\alpha) = 0$, that is, α is a multiple root of f, then more care must be taken. Suppose this is the case and express $f(z) = (z - \alpha)^k h(z)$ where $h(\alpha) \neq 0$, and $k \in \mathbb{N}$ is the multiplicity of the root of f at α (see Lemma A.21 on page 415). Since $f'(\alpha) = 0$, the quotient $\frac{f(z)}{f'(z)}$ appearing in the definition of $F(z)$ is not formally defined at α. However, we can overcome this difficulty by showing that $\frac{f(z)}{f'(z)}$ has a removable singularity at α. Indeed, note that

$$\frac{f(z)}{f'(z)} = \frac{(z-\alpha)^k h(z)}{k(z-\alpha)^{k-1}h(z) + (z-\alpha)^k h'(z)} = \frac{(z-\alpha)h(z)}{kh(z) + (z-\alpha)h'(z)}$$

which equals 0 for $z = \alpha$. This lets us define $F(\alpha) = \alpha$ even in the case that $f'(\alpha) = 0$.

We leave it to the reader to use the ideas from above to show that $F'(z)$ also has a removable singularity at α which, in particular, allows us to define $F'(\alpha) = \frac{k-1}{k}$. Since $|F'(\alpha)| < 1$ we have completed the proof. $\qquad\square$

EXERCISE 1.22. Provide the details in the above proof that $F'(\alpha) = \frac{k-1}{k}$ when f has a root of order k. Also, note that α is a super attracting fixed point of the Newton map F if and only if $k = 1$, i.e., α is a simple root[7] of f. **Try it out!**

EXPLORATION 1.23 (Convergence rates for Newton's method). Let $f(z)$ and $g(z)$ be analytic, and consider the corresponding Newton Maps $F_f(z) = z - \frac{f(z)}{f'(z)}$ and $F_g(z) = z - \frac{g(z)}{g'(z)}$. Suppose f has a root at α of order k and g has a root at α of order m, where $k < m$. The rate of convergence of F_f^n near α is faster than the rate of convergence of F_g^n near β since $|F_f'(\alpha)| = \frac{k-1}{k} < \frac{m-1}{m} = |F_g'(\alpha)|$ (recall Remark 1.19 which relates the rate of convergence to the derivative at the attracting fixed point). Let's explore this in the real variable case with the functions $f(x) = x^k$ and $g(x) = x^m$. Use the *Real Newton Method Applet* to visualize how the degree of the root of f (respectively, g) influences the tangent lines used in Newton's method. In particular, note the effect the degree of the root has on the curvature of the graph near the root and how this provides a visual way to understand the relative rates of convergence of F_f^n and F_g^n near α. **Try it out!**

---

[7]It is interesting to note that the Newton map of $f(z)/f'(z)$ (as opposed to the Newton map of $f(z)$) always has super attracting fixed points at the roots of f regardless of the order of the root of f (see Additional Exercise 1.160).

REMARK 1.24. We see that the value r in Proposition 1.21 gives us a lower bound on how close an initial guess z_0 must be to an actual root α for Newton's method to be *guaranteed* to find α. Because of this, we call any such r a ***radius of convergence*** (for F and α) and the corresponding circle $C(\alpha, r) = \{z : |z - \alpha| = r\}$ is called a ***circle of convergence***. In practice, it is very useful to gauge r so that we can guarantee the success of Newton's method. There is no universal estimate for r that always works since it very much depends on the particular map f being used (see Exercise 1.161). However, for certain classes of maps f we can provide useable estimates for r (see Exercise 1.162 and Small Projects 1.163 and 1.164).

Let us now explore some specific examples using complex functions.

EXAMPLE 1.25. For constants $\alpha, \beta \in \mathbb{C}$, consider the map $f(z) = (z - \alpha)(z - \beta)$ which has roots at α and β. The Newton Map for f is then $F(z) = \frac{z^2 - \alpha\beta}{2z - (\alpha+\beta)}$.

EXPLORATION 1.26. Set $\alpha = 0$ and $\beta = 1 + i$ in Example 1.25 and consider the iterates of the Newton Map F with starting values $z_0 = 2$, $z_0 = -3 - 2i$, and $z_0 = i + 1$, which you can compute using the *Complex Newton Method Applet*. As expected we see that different starting values for z_0 find different roots of $f(z)$. Can you make a guess as to which seed values will find which root? Try to determine if there are there any seed values z_0 for which Newton's method fails to find either root? Experiment with the *Complex Newton Method Applet* to test your predictions. ***Try it out!***

EXPLORATION 1.27. Using the function $f(z) = z^3 - 1$, determine $F(z)$ the Newton Map for f and then analyze Newton's method using different initial guesses for z_0 in the *Complex Newton Method Applet*. Describe (as best as you can) which seed values z_0 will find which of the roots $1, e^{2\pi i/3}$, and $e^{-2\pi i/3}$. Thus we are asking you to describe (as best as you can) the attracting basins $A_F(1), A_F(e^{2\pi i/3})$, and $A_F(e^{-2\pi i/3})$. ***Try it out!***

1.2.5. Global Picture of Complex Newton Method Dynamics. As with the real valued maps, we wish to achieve an understanding of the global dynamics of Newton's method, i.e., an understanding of how the orbits behave for *all* seed values. Thus we employ the **Graph basins of attraction** feature of the *Complex Newton Method Applet*, which uses different colors to display which initial guesses will find which roots, and so gives us a one picture snapshot of the dynamics of Newton's method.

EXERCISE 1.28. Use the *Complex Newton Method Applet* to view the basins of attraction for the Newton maps in Explorations 1.26 and 1.27. ***Try it out!***

Looking carefully at the picture of the two attracting basins corresponding to Exploration 1.26, it appears that the boundary between these regions is a straight line; points on one side look closer to the root of f on that side than to the other root. That is, this boundary appears to be the perpendicular bisector (denoted by L in Figure 1.3

below) of the line segment from α to β. This notion suggested by the picture is actually true (pictures, however, can sometimes be misleading) and we can prove it using a very useful technique of *global conjugation.*

1.2.6. Global Conjugation. One learns in Linear Algebra that a change of basis can be used to, among other things, greatly facilitate certain calculations, procedures, and overall give you a better understanding of the field. In particular, the notion of similarity of matrices plays a key role (recall, matrices A and B are called *similar* when there is an invertible matrix P such that $A = PBP^{-1}$). A direct analog of this idea often used in dynamics is a type of *change of coordinates* provided by what we call *conjugation.*

DEFINITION 1.29. Let ϕ be a *Möbius*[8] map. We say that *rational*[9] maps f and g are *globally conjugate* (by the map ϕ) if $g = \phi \circ f \circ \phi^{-1}$.

Often the point of conjugating a map f to a map g is that g is easier to work with than f. And, as we will see, the information we usually want from f can be quickly obtained by studying the simpler map g. In particular, we note that the iterates are related as such $g^n = (\phi \circ f \circ \phi^{-1}) \circ (\phi \circ f \circ \phi^{-1}) \circ \cdots \circ (\phi \circ f \circ \phi^{-1}) \circ (\phi \circ f \circ \phi^{-1}) = \phi \circ f^n \circ \phi^{-1}$. Hence ϕ in this way transfers information between the iterates of f and the iterates of g. In particular, fixed points and their corresponding derivatives are transferred in the following way.

EXERCISE 1.30. Suppose maps f and g are *globally conjugate* by the Möbius map ϕ, i.e., $g = \phi \circ f \circ \phi^{-1}$. Prove that for any $a \in \overline{\mathbb{C}}$, $f(a) = a$ if and only if $g(\phi(a)) = \phi(a)$. Furthermore prove that if $a, \phi(a) \in \mathbb{C}$ and $f(a) = a$, then we also have $f'(a) = g'(\phi(a))$. **Try it out!**

Also, see Additional Exercises 1.165–1.168.

REMARK 1.31. There is also a very useful technique called *local* conjugation which can greatly simplify calculations in many situations. In fact, an important question is when we can locally conjugate a map of the form $f(z) = a_1 z + a_2 z^2 + \ldots$ having a fixed point at 0 to another map which is simply $z \mapsto a_1 z$. This is called *linearizing* the map f near 0. It can always be done when $0 < |a_1| \neq 1$, but only sometimes when $|a_1| = 1$. The interested reader can pursue such results and their proofs in the literature (e.g., [**1, 2, 24**]).

1.2.7. Analysis of the Newton map of a quadratic polynomial. We now use this powerful global conjugation technique to greatly simplify the analysis of the Newton map $F(z) = \frac{z^2 - \alpha\beta}{2z - (\alpha + \beta)}$ in Example 1.25. We first choose a Möbius map which sends α and β to 0 and ∞, respectively, and then analyze the much simpler map

---

[8]Recall that a *Möbius* map is a map of the form $z \mapsto \frac{az+b}{cz+d}$ where $ad - bc \neq 0$, and as such is a bijection of $\overline{\mathbb{C}}$ onto itself.

[9]Recall that a *rational* map is a quotient of two polynomials.

obtained by conjugation. In particular, the map $\phi(z) = \frac{z-\alpha}{z-\beta}$ conjugates F to the map $g(z) = \phi \circ F \circ \phi^{-1}(z) = z^2$, a calculation we leave to the reader. We also leave it to the reader to show that $g^n(z) = z^{2^n}$.

Our goal is to show that if $z \in \mathbb{C}$ is closer to α than to β, then $F^n(z)$ iterates to α, i.e., $|z - \alpha| < |z - \beta|$ implies $F^n(z) \to \alpha$. Let $|z - \alpha| < |z - \beta|$ and note that this implies $|\phi(z)| < 1$. Since $|g^n(\phi(z))| = |\phi(z)^{2^n}| = |\phi(z)|^{2^n} \to 0$, we have $g^n(\phi(z)) \to 0$. By the conjugation property above we have $F^n(z) = \phi^{-1}(g^n(\phi(z))) \to \phi^{-1}(0) = \alpha$. Thus we have shown that the points $z \in \mathbb{C}$ which are closer to α than to β are indeed in $A_F(\alpha)$.

EXERCISE 1.32. Show that the points $z \in \mathbb{C}$ which are closer to β than to α are in $A_F(\beta)$. ***Try it out!***

We illustrate the conjugation used above by the following diagram, called a *commutative diagram* because each of the maps $\phi \circ F$ and $g \circ \phi$ from the upper left to the bottom right are equal.
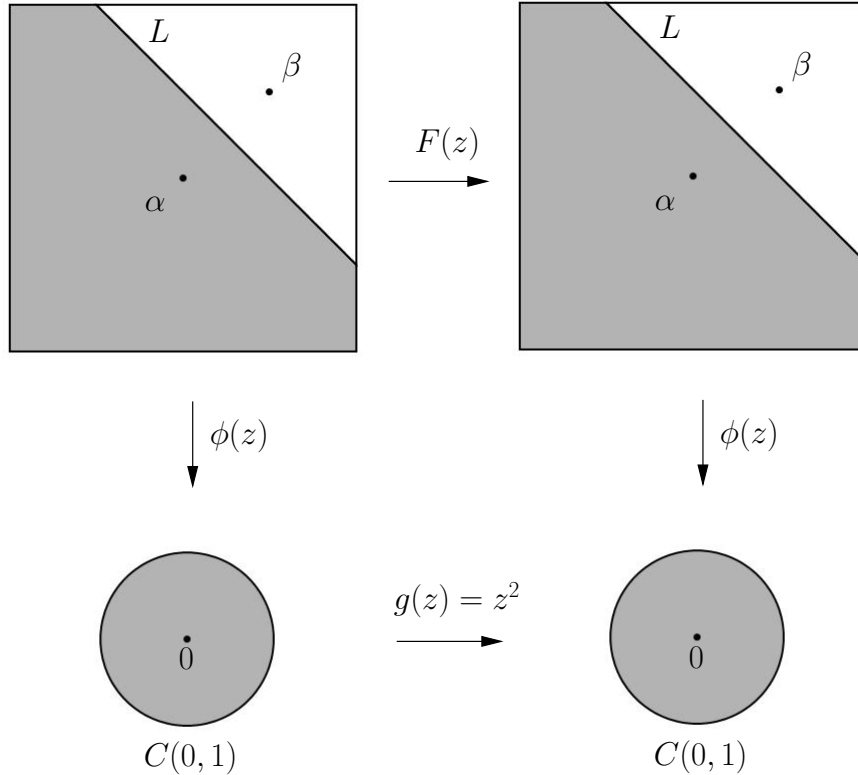

FIGURE 1.3. Commutative diagram for global conjugation of Newton map $F(z) = \frac{z^2 - \alpha\beta}{2z - (\alpha+\beta)}$.

Note how the points α and β are moved by ϕ to the points 0 and ∞, and that the line $L \cup \{\infty\}$ in the top pictures is transformed to the unit circle $C(0,1)$ in the bottom pictures. By using this conjugation, we are able to analyze the relatively simple dynamics of g to get corresponding information about the dynamics of F, in particular, $A_f(\alpha) = \phi^{-1}(A_g(0))$ and $A_f(\beta) = \phi^{-1}(A_g(\infty))$.

Let's return to the question of whether Newton's method can fail in this example. Are there initial values for which the Newton's method orbit never finds any root of f? In Example 1.5, we saw that Newton's method fails (or more properly, the *formula* for Newton's method fails) when an initial value x_0 is such that $f'(x_0) = 0$. This, what we term an *analytic* obstruction of having a zero in the denominator, however, is overcome when one allows ∞ to take its equal place with all the values in the Riemann Sphere $\overline{\mathbb{C}}$. In particular, even though in Example 1.25, we have $f'(\frac{\alpha+\beta}{2}) = 0$, the Newton Map gives $F(\frac{\alpha+\beta}{2}) = \infty$ (see Appendix Section B.4 on page 424 for a review of a discussion on functions defined at ∞). Also, since $F(\infty) = \infty$, we see that an initial value $z_0 = \frac{\alpha+\beta}{2}$ will never find either root α or β because this seed generates the following sequence of Newton "approximations": $\frac{\alpha+\beta}{2}, \infty, \infty, \infty, \ldots$. Note that the actual obstruction for the success of Newton's method starting with seed value z_0 is not that the function F cannot be appropriately defined at z_0, but that such a definition forces the map F to never iterate z_0 to a root of $f(z)$ (since $F^n(z_0) = \infty$ for all $n \geq 1$). The seed $\frac{\alpha+\beta}{2}$ is not the only complex number which fails to find a root of $f(z)$ though. We can show, again with the help of the conjugation above, that the boundary line L which divides $A_F(\alpha)$ from $A_F(\beta)$ consists of exactly those points in $\mathbb{C}$ for which Newton's method fails to find either root. As witnessed in the real variable examples, we can understand this *dynamically* since this line divides those points which are pulled or attracted to α and those points which are attracted to β. And so by an informal use of symmetry, it would stand to reason that something must fail to work out for points exactly on this line. Of course, we have not proven this fact carefully yet. We have only looked at compelling computer generated evidence, which, as is a common theme in this chapter, is always to be viewed with a healthy bit of skepticism.

A formal argument can be made however. For a point $z \in L$, we have $|\phi(z)| = \frac{|z-\alpha|}{|z-\beta|} = 1$ and thus by the conjugation property above we have $|g^n(\phi(z))| = |\phi(z)^{2^n}| = |\phi(z)|^{2^n} = 1$ for all $n \in \mathbb{N}$. Hence for all $n \in \mathbb{N}$, we see that $F^n(z) = \phi^{-1}(g^n(\phi(z))) \in \phi^{-1}(C(0,1)) = L \cup \{\infty\}$. In particular, $F^n(z)$ neither converges to α nor β.

EXERCISE 1.33. The above analysis applies to any *monic* quadratic polynomial $p(z)$ with distinct roots, but what if the leading coefficient is not 1? Also, what if the quadratic polynomial $p(z)$ has a double root instead of two distinct roots? Analyze what happens in these situations. ***Try it out!***

1.2.8. Analysis of the Newton map of a cubic polynomial. The behavior found in Exploration 1.27 with the cubic function $f(z) = z^3 - 1$ is far more complicated

than the quadratic case. It is not simply the case that a starting point under Newton's method will find the root to which it is nearest. If it were, then the picture of the global dynamics would look like the picture in Figure 1.4. However, the better picture to represent the actual dynamics is given in Figure 1.5, showing that the set of initial guesses in the complex plane is divided into very intricate regions of points that find the various roots.
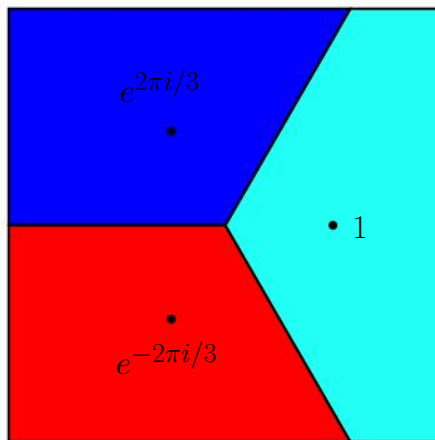
FIGURE 1.4. A reasonable (but false) guess for the picture description of the global dynamics of $F(z)$, the Newton Map for $f(z) = z^3 - 1$.

Let's experiment with the zoom feature of the *Complex Newton Method Applet* to investigate this. In particular, notice that when you zoom in on any point that is on the boundary of one colored region (attracting basin), you always find tiny "bulbs" of the other two colors (attracting basins) nearby. In fact, this happens no matter how much you zoom in! This shocking feature is why we call such sets *fractals*.[10] In this case, at a large scale there are extremely tiny bulbs (smaller than a single pixel) that are not revealed in the picture unless one zooms in far enough to see them.

So again we see that Newton's method became very complicated to understand globally when the original function $f(z)$ changed from a quadratic to a cubic map. Earlier we posed the questions: Why? How? Is there a way to know when a system will necessarily be simple or complicated ahead of time? With the use of our deep knowledge and fancy tools in complex analysis we can give some good reasons why the pictures, and hence the dynamics they represent, must be so complicated. We first remind ourselves of some key concepts.

DEFINITION 1.34. The *boundary of a set* E in $\overline{\mathbb{C}}$ is $\partial E = \overline{E} \cap \overline{\overline{\mathbb{C} \setminus E}}$, which is equal to the set of points $z \in \overline{\mathbb{C}}$ which have the property that every open disk (in the

[10]A *fractal* is a set which when you zoom in reveals new features not seen from the coarse larger scale picture.
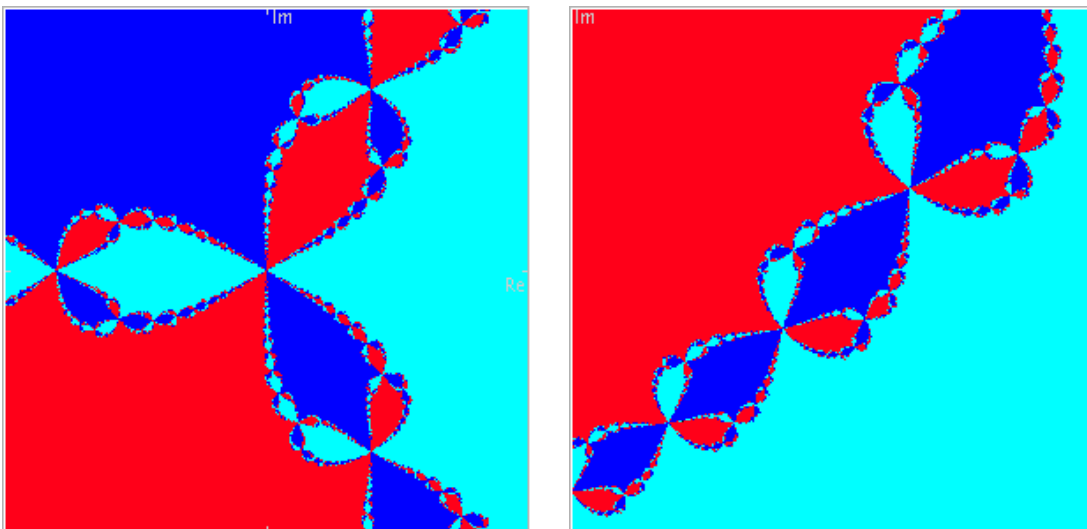
FIGURE 1.5. A more accurate picture of the global dynamics of $F(z)$, the Newton Map for $f(z) = z^3 - 1$ (with magnification on the right). Here the turquoise represents $A_F(1)$, the blue represents $A_F(e^{2\pi i/3})$ and the red represents $A_F(e^{-2\pi i/3})$.

spherical metric) $\triangle_\sigma(z, r)$ intersects both E and the complement of E no matter how small $r > 0$ is.

We can now describe the fractal features we already observed in Figure 1.5 by saying $\partial A_F(1) = \partial A_F(e^{2\pi i/3}) = \partial A_F(e^{-2\pi i/3})$. In fact, we see this same phenomenon in Example 1.4 and Explorations 1.6 and 1.26 as well when we consider the complex versions of all these maps. It turns out that, in general, all attracting basins of a Newton Map F must share the exact same set of boundary points. Specifically, we have the following Common Boundary Condition.

THEOREM 1.35 (Common Boundary Condition). Let $f(z)$ be an analytic function such that its Newton Map $F(z) = z - \frac{f(z)}{f'(z)}$ is a rational map. If w_1 and w_2 are roots of $f(z)$, then $\partial A_F(w_1) = \partial A_F(w_2)$.

PROOF. This result can be proven using Proposition 1.21 and the forthcoming Theorem 1.60. □

Let's examine how Theorem 1.35 forces the dynamics illustrated in Figure 1.5 to necessarily be complicated. According to Proposition 1.21 there exists some $r > 0$ such that $\triangle(1, r) \subseteq A_F(1), \triangle(e^{2\pi i/3}, r) \subseteq A_F(e^{2\pi i/3})$, and $\triangle(e^{-2\pi i/3}, r) \subseteq A_F(e^{-2\pi i/3})$. And so starting from the picture shown in Figure 1.6 we need to consider how to color in the rest of the points in $A_F(1), A_F(e^{2\pi i/3})$, and $A_F(e^{-2\pi i/3})$ with turquoise, blue, and red, respectively, *while also being certain to make sure that the boundary of each color exactly matches the boundary of the other two colors*. A moment's thought tells

19

us that this Common Boundary Condition forces the picture to necessarily be *very* complicated, and also rules out having the global dynamics behaving as suggested by Figure 1.4 (since, for example, a point on the negative real axis in that picture lies on the boundary of only two of the three color basins).

$$\triangle(e^{2\pi i/3}, r) \subset A_F(e^{2\pi i/3})$$

$$\triangle(1, r) \subset A_F(1)$$

$$*$$
$$0$$

$$\triangle(e^{-2\pi i/3}, r) \subset A_F(e^{-2\pi i/3})$$
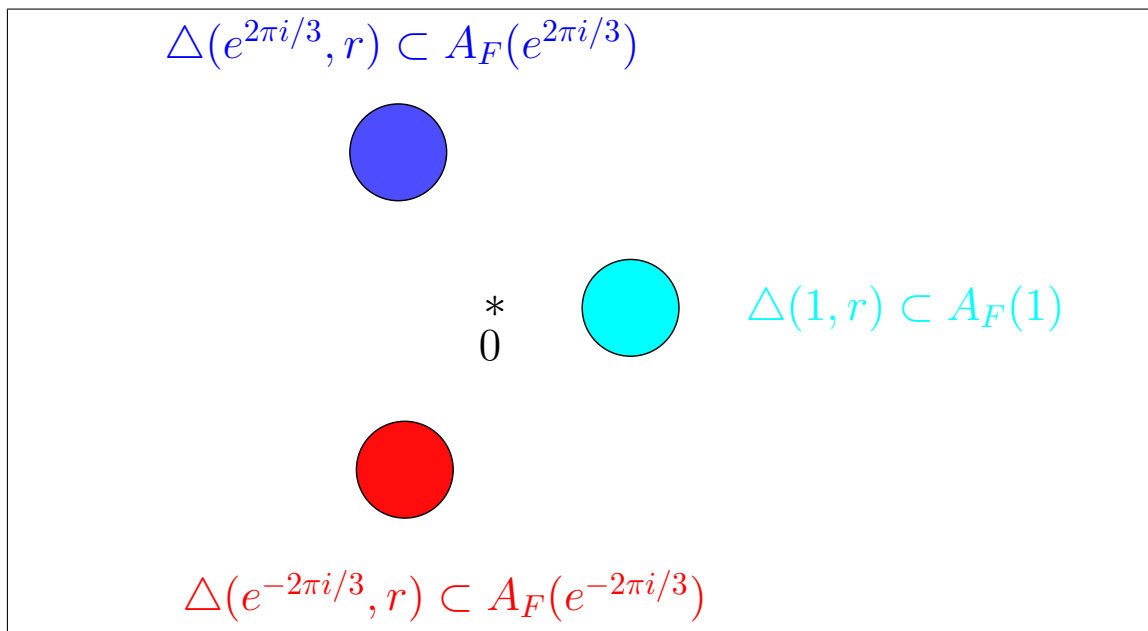
FIGURE 1.6. Consequence of the Attracting Property of Newton's Method (Proposition 1.21).

This Common Boundary Condition is also at the heart of what we will learn to call *chaos* in the dynamics of Newton's method. Consider any point z on the boundary of any basin and then draw a tiny disk B around it. According to the Common Boundary Condition, this tiny disk B must contain all three colors in it. So if we wish to determine the fate of the orbit of z with a computer, what we will find is that any tiny error (like roundoff error) from inputting the coordinates of z could lead to drastically different results. We see that by changing z even by the slightest amount, we can change the orbit of z tremendously as z could slip into any of the red, turquoise, or blue regions. So with such a point z, we see that the behavior of the orbit in the long run will be quite different even if we change z only a little. This is the essence of what we call chaos. Although a more formal and more thorough understanding of chaos will come later in the text, use this understanding of this notion to explain why each point on the line L that appeared in the analysis of the quadratic Newton method case has this behavior. Also, explain why each point in one of the attracting basins (which has a tiny disk of all the same color) is NOT such a point. ***Try it out!***

REMARK 1.36. We mention here a remarkable fact about sets which share a complicated boundary. The three sets $A_F(1), A_F(e^{2\pi i/3})$, and $A_F(e^{-2\pi i/3})$ above share the same boundary because each set is broken up into an infinite number of pieces which are then arranged in the complicated fractal pattern you see in Figure 1.5. However, it is true that three open sets (or even n open sets for any fixed $n \in \mathbb{N}$) can all share the same boundary set and also have the property that each set is connected.[11] Such sets are complicated indeed! The interested reader is encouraged to read about the Lakes of Wada in [**16**], p. 143.

Let us now consider the question of whether or not there are any starting (seed) values for which the Newton Map for $f(z) = z^3 - 1$ fails to find a root of $f(z)$. In all the previous examples there always exist such points; however, these were, in some sense, relatively few. You might guess that each point in $\partial A_F(1) = \partial A_F(e^{2\pi i/3}) = \partial A_F(e^{-2\pi i/3})$ is such a point...and you would be right. Heuristically, we can use our intuition to imagine that such points fail to find any root since any such point is, in some sense, being pulled or attracted by each of the three different roots with equal force. You are asked to prove this in Additional Exercise 1.169. Note however, that even assuming the result that each point in $\partial A_F(1) = \partial A_F(e^{2\pi i/3}) = \partial A_F(e^{-2\pi i/3})$ fails to iterate under $F(z)$ to any of the roots of $f(z)$, we still do not necessarily know the fate of ALL starting points. Specifically, we wonder if it is true that all points in $\overline{\mathbb{C}}$ lie in either one of the attracting basins or on the common boundary of these sets. We pose this and a related question as follows.

1. Is there a point not in $\partial A_F(1) = \partial A_F(e^{2\pi i/3}) = \partial A_F(e^{-2\pi i/3})$ which fails to find any root of $f(z)$?

2. Can there be a whole open disk of such points?

When considering the particular map $f(z) = z^3 - 1$ (or any of the maps f mentioned thus far in this chapter) and its related Newton Map $F(z)$, the answer to both of these questions is no (see Additional Exercise 1.180). However, in general, it is possible for the answer to the second question (and hence also the first) to be yes.

EXPLORATION 1.37. Consider the map $f(z) = z * (z-1) * (z - .909 - .416i)$ and its related Newton Map $F(z)$. Using the the *Complex Newton Method Applet* you can find regions of seed values colored black which fail to find any root of $f(z)$ under Newton's method. For example you will find such a region of points by zooming in on the point $0.64 + 0.14i$. Use the applet to select such points and then iterate $F(z)$ to explore the behavior. Experiment with many seed values chosen from the black regions to see what type of behavior you can find. Use the zoom feature to see whether or not it appears that the boundary of the black region matches the boundary of the attracting basins for the three roots of $f(z)$. ***Try it out!***

---

[11]Recall, an open set A is called *connected* when given any two points $z, w \in A$ there exists a *polygonal line* in A which connects z to w. See Appendix B.3 on page 422.

The reader can investigate Newton's method applied to other cubic polynomials in Additional Exercise 1.170.

1.2.9. Newton's method applied to any cubic polynomial. In this section we study Newton's method applied to an arbitrary cubic polynomial. We begin by introducing a (nearly) representative class for all cubic polynomials. Let $\mathcal{F}$ denote the collection of all polynomials $p_\rho(z) = z(z-1)(z-\rho)$ where ρ is in $D = \{\rho : \operatorname{Im}\rho \geq 0, |\rho| \leq 1, |\rho - 1| \leq 1\}$.
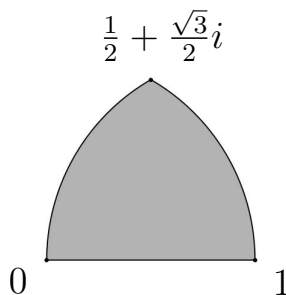
FIGURE 1.7. Region D of ρ values corresponding to the maps p_ρ in $\mathcal{F}$.

This class of polynomials $\mathcal{F}$ is then representative of all polynomials with three distinct roots in the following sense.

PROPOSITION 1.38. For each cubic polynomial $p(z)$ with three distinct roots, there exists $\rho \in D$ such that the Newton map F_p is globally conjugate by a linear map T to the Newton map F_{p_ρ} of $p_\rho(z)$, i.e., $T \circ F_p \circ T^{-1} = F_{p_\rho}$.

In order to prove the above proposition we require the following proposition, which we note applies to polynomials of any degree.

PROPOSITION 1.39. Let $p(z)$ be a polynomial and let $T(z) = az + b$ for $a \neq 0$ where $a, b \in \mathbb{C}$. Then for the polynomial $q(z) = p(T(z))$, we have

$$T \circ F_q \circ T^{-1} = F_p$$

where F_q and F_p are the Newton maps of q and p respectively.

We leave the proof of Proposition 1.39 to the reader. However, we say a few words about its meaning and usefulness. Note that the polynomial q (which is sometimes called the "rescaling" of p) will have the same degree as p. Furthermore, if p has roots at $r_1, \ldots, r_d$, then q will have roots at $T^{-1}(r_1), \ldots, T^{-1}(r_d)$. Hence the result of the proposition says that we can "move" the roots of p by choosing T^{-1} appropriately and generating a new polynomial q. Furthermore, studying the dynamics of F_q will be *essentially* the same as studying the dynamics of F_p since these are globally conjugate to each other.

22

EXAMPLE 1.40. Consider $p(z) = (z + i/2)(z - 1)(z + 1)$. We wish to illustrate Proposition 1.38 by finding $\rho \in D$ so that F_{p_ρ} is conjugate to F_p. Consider the triangle formed by the roots of $p(z)$ at $-i/2, 1$, and -1. We now construct a linear map T^{-1} to transform this triangle into a triangle with longest side being the interval $[0, 1]$ and third vertex (which will be our choice of ρ) in the upper half plane $\{\text{Im } z \geq 0\}$. These constrictions then force $T^{-1}(1) = 0$ and $T^{-1}(-1) = 1$, which in turn determine that $T^{-1}(z) = \frac{-1}{2}z + \frac{1}{2}$ and $\rho = T^{-1}(-i/2) = 1/2 + i/4$ (which is in D).

Let's verify that this choice of ρ works. Note that $q(z) = p(T(z))$ must have roots at 0, 1, and ρ. Since q and p_ρ share the exact same roots, they must agree up to a multiplicative constant, i.e., we must have $q(z) = cp_\rho(z)$ for some constant $c \neq 0$. This, however, implies that $F_q = F_{p_\rho}$ (why?). Hence, by Proposition 1.39 we have that $T \circ F_{p_\rho} \circ T^{-1} = T \circ F_q \circ T^{-1} = F_p$ as desired.

The reader should use the *Complex Newton Method Applet* to compare the pictures of the attracting basins for F_p and F_{p_ρ} noting the similarities one would expect from the fact that these maps are globally conjugate. In particular, see if you can see the effects of the specific map T when you compare the two pictures. ***Try it out!***

EXERCISE 1.41. Find $\rho \in D$ so that for $p(z) = (z - 4)(z + i)(z + 4)$ we have F_{p_ρ} is globally conjugate to F_p. ***Try it out!***

In the same manner as Example 1.40, one can prove Proposition 1.38, and so we omit the details. However, we leave it to the reader to investigate what can be said about the dynamics of F_p when $p(z)$ is a cubic polynomial with a double or triple root. In addition, we invite the reader to consider generalizations such as the following.

SMALL PROJECT 1.42. Consider how Proposition 1.39 is used to "move roots" by a linear transformation T^{-1} in the proof of Proposition 1.38. With this idea in mind, can you justify the statement that all cubic polynomials whose roots form an equilateral triangle have globally conjugate Newton maps (and hence essentially the same dynamics)? What about quartic polynomials whose roots form a square? Does this generalize to higher degree polynomials? Do the roots need to form regular n-gon? What can you say about generalizing the class $\mathcal{F}$ (and the set D) from above when considering polynomials of fixed higher degree?

Aside from the exceptional cases mentioned above, we can study the Newton's method dynamics of all cubic polynomials by studying just the maps p_ρ where $\rho \in D$. We have created the *Cubic Polynomial Complex Newton Method Applet* to help. This applet allows the user to generate the pictures of the attracting basins for Newton's method applied to any p_ρ (not just $\rho \in D$). Also, it allows the user to investigate the *parameter plane* of ρ values since each such value will be colored according to the corresponding dynamics of F_{p_ρ}.

A note about the coloring of the parameter plane of ρ values. You will see in Section 1.4 that the orbit of all points where $F'_{p_\rho}(z) = 0$ are critically important to understanding the dynamics of F_{p_ρ}. Since $F'_{p_\rho}(z) = \frac{p_\rho(z)p''_\rho(z)}{(p'_\rho(z))^2}$, we see that $F'_{p_\rho}(z) = 0$

only at the roots of $p_\rho(z)$ or when $p_\rho''(z) = 0$. Since the roots of $p_\rho(z)$ are attracting fixed points for F_{p_ρ}, their orbits are understood. However, since $p_\rho''(z) = 0$ for $z = (1 + \rho)/3$ we see that this, so-called, *free critical point* is important. Hence in the *Cubic Polynomial Complex Newton Method Applet*, we track, for each ρ, the *critical orbit* of $z_0 = (1 + \rho)/3$. When this critical orbit is attracted to one of the roots of p_ρ, we color the point ρ in the parameter plane the corresponding color of the attracting basin. To illustrate, click on the **Show Critical Orbit** checkbox to see this critical orbit appear as white dots in the right picture (dynamical plane). Thus, if this critical orbit converges to, say, the red root of p_ρ, then the corresponding parameter ρ (marked by a +) is colored red in the left picture (parameter plane). If this orbit is not attracted to any of the roots of p_ρ, we color the point ρ in the parameter plane black. Try clicking in the parameter plane to a select different colored ρ value, and then witness the convergence of the critical orbit to a different colored root of p_ρ.

EXPLORATION 1.43. Use the *Cubic Polynomial Complex Newton Method Applet* to investigate the dynamics of Newton's method applied to any p_ρ. Look for symmetries and experiment with the dynamical behavior you find. Make conjectures, and then see if you can prove them! ***Try it out!***

1.3. Iteration of an Analytic Function

In this section we focus our study on the dynamics of analytic maps which do not necessarily arise as Newton Maps of polynomials. One goal is to understand which dynamical features of Newton's method extend to such a wider class of maps. We pay particular attention to the iteration of polynomial maps of the form $z^2 + c$ where c is a complex parameter since these are the simplest maps which still produce a rich variety of dynamical behaviors.[12] We also note that with the help of the global conjugation technique (see Section 1.2.6), every quadratic map is globally conjugate, and in some sense dynamically equivalent, to exactly one map of the form $z^2 + c$ (see Additional Exercise 1.167). Hence, in this way the dynamics of maps $z^2 + c$ represent the dynamics of all quadratic maps.

One need not have worked through all of Section 1.2; however, it is critical that Sections 1.2.3 and 1.2.6 be read before proceeding with this section.

Though the quadratic maps of the form $z^2 + c$ are genuinely very simple and well understood *as functions*, we will see that their *dynamics* can be extremely complicated. We will explore this carefully, not only the dynamics of each map $z^2 + c$, but we will also study how the dynamics of such maps change as the parameter c changes, leading us into the study of what is called *bifurcation theory*.

In Section 1.6, we consider the dynamics of more exotic transcendental entire complex analytic maps such as ce^z, $c \sin z$, and $c \cos z$.

---

[12]The dynamics of the iteration of Möbius maps is simpler to study because such maps have such relatively simple dynamical behavior. See Section 1.B for a classification of the dynamics of Möbius maps.

1.3.1. Classification of Fixed Points for Analytic maps. Having noted the importance of fixed points in Section 1.2.3, where we defined attracting fixed points, we extend our discussion here to understand the two remaining types of fixed points, *repelling* and *indifferent*. We begin with an example.

EXAMPLE 1.44. Let $f(x) = x^2$ and $g(x) = \sqrt{x}$ each restricted so that $f, g : \mathbb{R}^+ \to \mathbb{R}^+$ where $\mathbb{R}^+ = \{x \geq 0\}$. We leave it to the reader to quickly check that 0 and 1 are each fixed points of both f and g. Further, we have $f^n(x) \to 0$ if $0 < x < 1$, and $f^n(x) \to \infty$ if $x > 1$. Also we have $g^n(x) \to 1$ for all $x \in \mathbb{R}^+ \setminus \{0\}$. Note that 0 is an attracting fixed point for f (why?). However, we see that for x values close to, but not equal to 1, we have that the orbit $f^n(x)$ moves away from 1. We then call 1 a *repelling* fixed point for f (which we formally define below). We also note that the function $g(x)$ has an attracting fixed point at 1 and a repelling fixed point at 0. We illustrate the dynamics of these two maps graphically in Figure 1.8.
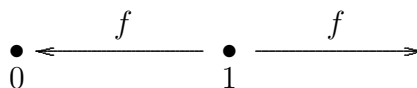
Dynamics of $f(x) = x^2$ has attracting fixed point 0 and repelling fixed point 1.

Dynamics of $g(x) = \sqrt{x}$ has repelling fixed point 0 and attracting fixed point 1.

FIGURE 1.8. Graphical description of the dynamics of f and $g = f^{-1}$.
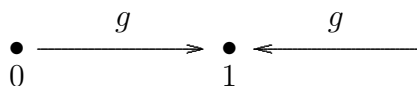
Note that $f = g^{-1}$ and $g = f^{-1}$. Giving some thought to this fact that f and g "undo" each other in this fashion, it stands to reason that an attracting fixed point for f must be a repelling fixed point for g, and vice versa.

Inspired by this example, we formally define a repelling fixed point as follows.

DEFINITION 1.45 (Repelling Fixed Point). Let f be a map with domain set $\Omega \subseteq \overline{\mathbb{C}}$ (note that Ω could be a subset of $\mathbb{R}$).

(a) We call a (finite) fixed point $a \in \mathbb{C}$ a *repelling fixed point* of f if there exists a neighborhood U of a such that for any point $z \in \Omega \cap U \setminus \{a\}$, we have $|f(z) - a| > |z - a|$, i.e., the action of f is to move each point in $\Omega \cap U \setminus \{a\}$ *farther from* a.

(b) Suppose $f(\infty) = \infty$. We call ∞ a *repelling fixed point* of f if there exists a neighborhood $U \subseteq \overline{\mathbb{C}}$ of ∞ such that for any point $z \in \Omega \cap U \setminus \{\infty\}$, we have $|f(z)| < |z|$, i.e., the action of f is to move each point in $\Omega \cap U \setminus \{\infty\}$ *farther from* ∞ (as measured by the spherical metric).

THEOREM 1.46. Suppose $\Omega \subseteq \mathbb{R}$ or $\Omega \subseteq \mathbb{C}$. Let $f : \Omega \to \Omega$ such that $f(a) = a$ and $|f'(a)| > 1$. Then a is a repelling fixed point of f. In fact, there exists $\epsilon > 0$ such that

for all $z \in \Omega \cap \triangle(a, \epsilon) \setminus \{a\}$ the orbit $f^n(z)$ eventually leaves $\triangle(a, \epsilon)$, i.e., there exists N such that $f^N(z) \notin \triangle(a, \epsilon)$.

REMARK 1.47. The proof is given by a quick modification of the proof of Theorem 1.18, and so we leave the details to the reader. Note, however, that the theorem does not preclude the case that the orbit of z, after leaving $\triangle(a, \epsilon)$, might reenter $\triangle(a, \epsilon)$ (see Additional Exercise 1.171).

REMARK 1.48. Again we note that Theorem 1.46 is not, in general, an if and only if result. In particular, the reader is encouraged to find a **real valued function** $f : \mathbb{R} \to \mathbb{R}$ such that 0 is a repelling fixed point which has $|f'(0)| = 1$. Hint: Consider maps of the form $x \mapsto x \pm x^n$.

Above we provided examples of *real valued* maps which show that the converses of Theorems 1.18 and 1.46 do not, in general, hold. However, the next theorem shows that such examples cannot be found for *complex analytic* maps.

THEOREM 1.49. Let $f(z)$ be an analytic map on an open set $\Omega \subseteq \mathbb{C}$ such that $f(a) = a$. Then we have the following,
  (i) a is an attracting fixed point if and only if $|f'(a)| < 1$, and
  (ii) a is a repelling fixed point if and only if $|f'(a)| > 1$.

Theorems 1.18 and 1.46 provide two of the four implications. The proofs of the remaining two, outlined in the Exercise 1.172, use special properties of complex analytic maps.

We note that Theorem 1.49 only applies to finite fixed points and so we ask if there is a corresponding result when ∞ is fixed. Let us examine the following examples in preparation of handling this issue.

EXAMPLE 1.50. Consider the map $h(z) = z+1$ on $\overline{\mathbb{C}}$, which fixes ∞ and shifts each point in the complex plane one unit to the right. Note that for z near ∞ with very large and *positive* real part, for example $z = 10^6 + 2i$, we have $|h(z)| > |z|$, which makes it appear that ∞ is attracting. However, for z near ∞ with very large and *negative* real part, for example $z = -10^6 + 5i$, we see that $|h(z)| < |z|$, which makes it appear that ∞ is repelling. We then are left to conclude that for h, the fixed point at ∞ is neither attracting nor repelling. Also note that the n-th iterate $h^n(z) = z + n$ and so for any point z in $\overline{\mathbb{C}}$ we have $h^n(z) \to \infty$ (always moving parallel to the x-axis). Thus $A_h(\infty)$ is equal to the entire Riemann Sphere $\overline{\mathbb{C}}$. This is somewhat surprising given that ∞ is not even an attracting fixed point. We graphically represent the dynamics both on $\mathbb{C}$ and $\overline{\mathbb{C}}$ in Figure 1.9.

EXAMPLE 1.51. Consider the map $g(z) = z/2$ on $\overline{\mathbb{C}}$. Each point in $\mathbb{C}$ is mapped to a point with one half the modulus, but with the same argument. Clearly then the origin is an attracting fixed point and $A_g(0) = \mathbb{C}$. We also note that $g(\infty) = \infty$ and if z is near ∞, then $g(z)$ moves away from ∞, i.e., $|g(z)| < |z|$. Thus ∞ is a repelling

26

FIGURE 1.9. Graphical representations of the dynamics of $h(z) = z+1$ on the plane $\mathbb{C}$ (left) and on the Riemann sphere $\overline{\mathbb{C}}$ (right).

fixed point. Note also that $g^n(z) = z/2^n$. We graphically represent the dynamics both on $\mathbb{C}$ and $\overline{\mathbb{C}}$ in Figure 1.10.

EXAMPLE 1.52. Consider the map $f(z) = z^2$ on $\overline{\mathbb{C}}$. Convince yourself that $f^n(z) = z^{2^n}$ for all $n = 1, 2, \dots$. Thus, we see that $|z| < 1$ implies both $|f(z)| = |z|^2 < |z|$ and $|f^n(z)| = |z^{2^n}| = |z|^{2^n} \to 0$. Also, $|z| > 1$ implies both $|f(z)| = |z|^2 > |z|$ and $|f^n(z)| = |z^{2^n}| = |z|^{2^n} \to +\infty$. Thus we conclude that both 0 and ∞ are attracting fixed points with attracting basins $A_f(0) = \triangle(0,1)$ and $A_f(\infty) = \overline{\mathbb{C}} \setminus \overline{\triangle(0,1)}$. We represent the dynamics graphically in Figure 1.11, being careful to note that the angle doubling property of the z^2 map is not represented in this picture. To be more precise, if $z_0 = re^{i\theta}$ in polar form we have that $f(z_0) = r^2 e^{i2\theta}$. Thus, for example, starting with seed $z_0 = 0.999e^{i\pi/100}$, the orbit z_n will converge towards 0, doubling the angle at each step, which we encourage the reader to visually see this using the *Complex Function Iterator Applet* (using the **Polar seed form** and the **Polar computation mode** of the applet).

The last three examples were chosen to illustrate fixed points at ∞, but we should note that they can be misleading in another regard. Namely, in each example we were to be able to calculate a formula for f^n. We caution the reader that this is very rare. We also make the point that having an actual formula for the iterates f^n was really unnecessary. We can analyze the dynamics without appealing directly to these formulas.
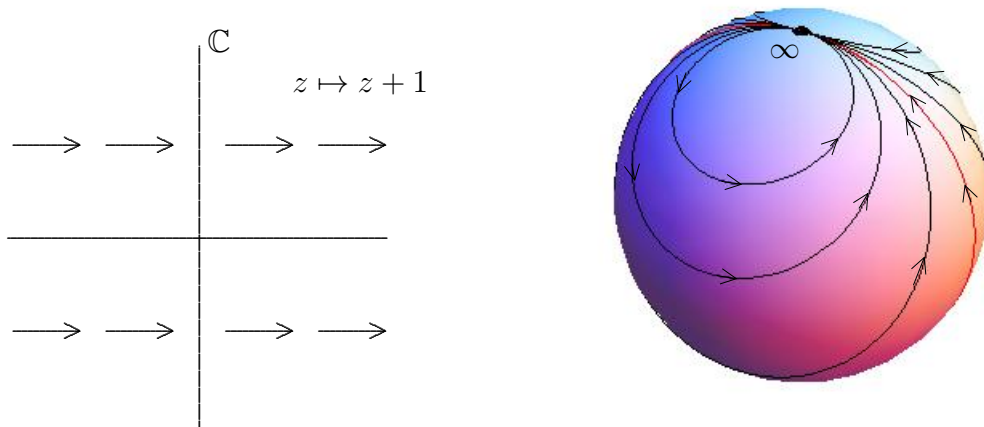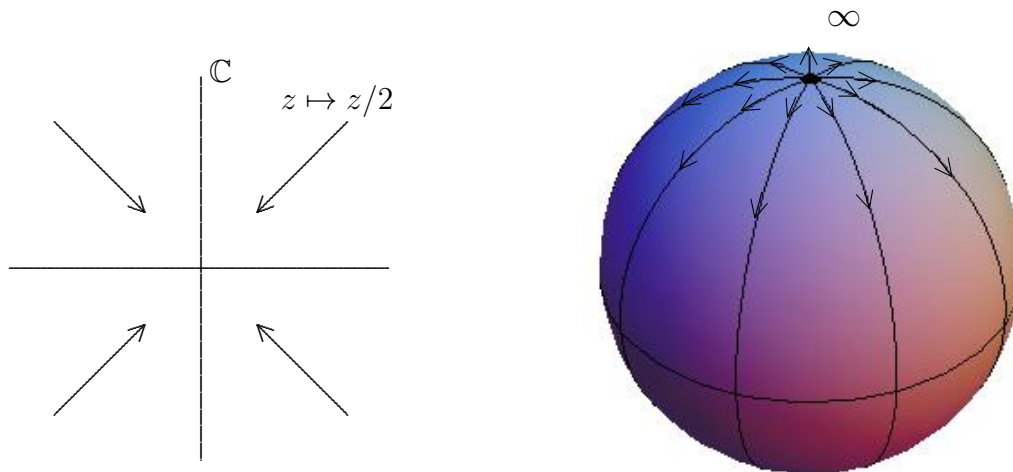
FIGURE 1.10. Graphical representations of the dynamics of $g(z) = z/2$ on the plane $\mathbb{C}$ (left) and on the Riemann sphere $\overline{\mathbb{C}}$ (right).
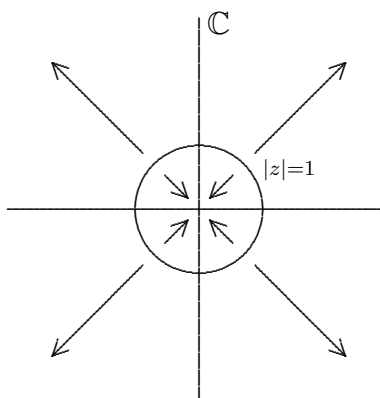
FIGURE 1.11. Graphical representation of the dynamics of $f(z) = z^2$.

In each of Examples 1.50, 1.51, and 1.52 we see that ∞ is a fixed point that we were able to classify by carefully examining the dynamics near ∞. However, we wonder (in light of Theorem 1.49) if we can also make use of the derivative to classify fixed points at ∞. It turns out that we can; however, we must be careful. Above we saw that $|h'(\infty)| = 1$ corresponded to a fixed point that was neither attracting nor repelling,

$|g'(\infty)| = 1/2$ corresponded to a repelling fixed point, and $f'(\infty) = \infty$ corresponded to an attracting fixed point. Clearly, the derivative evaluated at ∞ does not play the same role as it does in Theorem 1.18 for finite points. However, as we now see, it is not far off.

DEFINITION 1.53 (Multiplier of a fixed point). Let f be analytic at ∞ such that $f(\infty) = \infty$ (see Definition B.13 on page 425 for the definition of what is means for a map to be analytic at ∞). We define the *multiplier* λ at ∞ to be $1/f'(\infty) = \lim_{z \to \infty} 1/f'(z)$. If a is a finite fixed point in $\mathbb{C}$, then we define the *multiplier* λ at a to be $f'(a)$.

DEFINITION 1.54 (Classification of Fixed Points). With this definition and with Theorem 1.49 in mind, we now classify all fixed points of analytic maps, whether finite or not, based on their multiplier λ. Suppose $f : \Omega \to \Omega$ is analytic where Ω is an open subset of $\overline{\mathbb{C}}$ and a is a fixed point with multiplier λ. Then a is called
  a) *super attracting* if $\lambda = 0$
  b) *attracting* if $0 < |\lambda| < 1$
  c) *repelling* if $|\lambda| > 1$
  d) *indifferent* if $|\lambda| = 1$.
We write that a is (super)attracting, when it is known that either case (a) or (b) holds.

The reader can check that these classifications match what we found in Examples 1.50, 1.51, and 1.52. Further, we note the motivation for the definition of the multiplier when f has a fixed point at ∞. In this case, by globally conjugating f by $\phi(z) = 1/z$ we obtain the map $k(z) = 1/f(1/z)$ which has a fixed point at 0. We chose to define the multiplier of f at ∞ to be the multiplier of k at 0, which is shown in Lemma B.19 on page 426 to be $k'(0) = 1/f'(\infty)$. In Additional Exercise 1.173 you are asked to prove that this multiplier, thusly defined, does correctly correspond to the notions of attracting and repelling fixed points as given in Definitions 1.16 and 1.45.

REMARK 1.55. We note that indifferent fixed points can exhibit both a partial attracting nature and a partial repelling nature. For example, we saw in Section 1.2.3 that the indifferent fixed point at the origin for the *complex* map $g(z) = \sin z$ is "attracting" for real valued seeds, but "repelling" for purely imaginary seeds. Likewise, the indifferent fixed point at ∞ for the map $h(z) = z + 1$ "attracts" points with large and positive real part, but "repels" (at least initially) points with large and negative real part. However, sometimes an indifferent fixed point, like the origin under the map $f(z) = e^{\sqrt{2}\pi i}z$, acts in neither an attracting nor repelling manner. Indifferent fixed points of analytic maps can exhibit many different types of dynamical behavior and their study can be quite complicated. In fact, it is complicated enough that we will not say more about them here, but will only refer the interested reader to [**1, 2, 24**].

1.3.2. A Closer look at the Dynamics of $f(z) = z^2$. Let us return to considering the dynamics of the map $f(z) = z^2$. First we note that if $|z| = 1$, then $|f^n(z)| =$

$|z^{2^n}| = |z|^{2^n} = 1^{2^n} = 1$, i.e., if a point is on the unit circle $C(0,1) = \{z : |z| = 1\}$, then its entire orbit $\{f^n(z)\}_{n=1}^{\infty}$ is also on the unit circle. It might appear from this and from our previous work in Example 1.52 that all the mysteries concerning the dynamics of this map have been solved. But as we will see this is not at all close to the truth!

Consider the question, if we know the orbit of one seed value, will all nearby seed values have similar orbits? If z lies within the unit disk, then it and all the nearby points within the unit disk have the same behavior – namely each orbit converges to 0. Likewise, if z lies outside of the closed unit disk, then it and all the nearby points outside of the closed unit disk have the same behavior, namely each orbit converges to ∞. However, the story is quite different for seed values on the unit circle. In particular, for any seed on $C(0,1)$ we can find other seed values *arbitrarily* close which have drastically different orbits (namely which limit to either ∞ or 0).

Let's examine this behavior using the computer to iterate for us. For the map $f(z) = z^2$ use the *Complex Iteration Applet* to iterate the seed value $z_0 = 1 + i$ *as well as various seed values very close to z_0*. You can see that this seed value and all of the nearby seed values have the same behavior. Now repeat this by iterating z_0 for the values $-0.4 + 0.5i, 1, i$, and $0.6 + 0.8i$, being sure to test various nearby seed values that you choose yourself.

REMARK 1.56 (A word of caution about using the computer). We must always keep in mind that if we are trying to calculate the orbit under the map $f(z) = z^2$ of a point on $C(0,1)$, we may have some serious problems getting our computer to provide accurate results. Let's test this out now on the *Complex Iteration Applet*. For the map $f(z) = z^2$ try iterating seed value $z_0 = 0.6 + 0.8i$ (which lies on $C(0,1)$) by entering in the seed value in **Euclidean form** and using the **Euclidean computation mode**. Iterating 20 times, you will see what you would expect. However, after iterating about 55 times you will see the orbit move inside of the unit disk (where it will then iterate to 0). Why? Well, the problem is that if the computer truncates, rounds, or approximates any of the values in the orbit (as computers often do), these small errors will likely push the orbit outside or inside of the unit circle, causing the computer to mistakenly calculate the orbit as tending to either ∞ or 0. Thus, when using a computer it is important to know (a) if the computer will approximate values it uses, and (b) whether these approximations will be significant or not in the end result.

As an illustration of the subtle issues that can wreak havoc on your computations, we show the problems that can arise from the fact that $z * z$ and z^2 are not always equivalent. Of course we know that the expressions z^2 and $z * z$ are *mathematically* equivalent, but these expressions are not *computationally* equivalent. The former is evaluated as $e^{2\text{Log} z}$, where $\text{Log} z$ is the principal logarithm, and the latter is evaluated through usual complex multiplication. Thus each will incorporate different rounding errors at times. The end result of this very subtle difference is quite evident when iterating the seed $0.6 + 0.8i$ in **Euclidean computation mode** under each of these maps. ***Try it out!***

In line with the concerns one has when using the computer, specifically item (b) above, we see that it is important to clearly identify those seed values where approximations (or more precisely errors introduced by using approximations in place of exact values) would lead to significant errors in future calculations of orbit values. Such seed values are said to be in the **chaotic set**. The chaotic set in this context is usually called the **Julia set**, in honor of the mathematician Gaston Julia who in 1918, at the age of 25, published his 199 page masterpiece titled "Mémoire sur l'itération des fonctions rationnelles" describing the iteration of complex rational functions (see [**17**]).

REMARK 1.57. Note that our notion of Julia set is not tied only to errors or approximations that a computer might introduce. The issue is to know whether or not a tiny error in the seed value (no matter what or who created the error) *could* produce a significant error in some orbit value. For example, the seed $z_0 = i$ would be in the Julia set of $f(z) = z^2$ even though its orbit $i, -1, 1, 1, 1, \ldots$ is computed *exactly* by a computer. However, we still say it is a chaotic seed value since given *any* allowable error in the seed value (even an error as small as 10^{-631}), we can always find another seed z_0' *close to* $z_0 = i$ (i.e., within the tiny allowable error) such that z_0' has a drastically different orbit from the orbit of z_0. Thus, even if a computer wouldn't introduce an error of any kind in its orbit calculation for a specific seed z_0, the seed value might still be in the Julia set.

From the above discussion we see that the Julia set for $f(z) = z^2$ is $C(0, 1)$ because the dynamics there are *chaotic*, i.e., for any $z_0 \in C(0, 1)$ there is a point z_0' arbitrarily close by that has a drastically different orbit. As we have seen, the computer often fails to be accurate here since we would need the computer to store such a value (and each point in its orbit) with an *infinite* degree of accuracy. To put it another way, approximation of the starting value (or some future iterate value) ruins our confidence in the calculations of the long term behavior. This is called **sensitive dependence on initial conditions**, and this is the defining feature in what is called *chaos*.

We call the complement $\overline{\mathbb{C}} \setminus C(0, 1)$ the **stable set** for $f(z) = z^2$ because orbits are stable there, i.e., for any $z_0 \in \overline{\mathbb{C}} \setminus C(0, 1)$ its orbit z_n will behave like the orbit z_n' of any seed value z_0' chosen sufficiently close to z_0. In honor of Pierre Fatou the stable set is commonly called the **Fatou set** because of his role in developing the theory of complex function iteration in 1917 (see [**12, 13, 14**]). In honor of their pioneering work in the field, dynamics of complex analytic functions is often called the Fatou-Julia theory.

EXERCISE 1.58. Use your intuitive understanding of the meanings of the Fatou and Julia sets to determine each of these sets for the maps $h(z) = z + 1, g(z) = z/2$ and $k(z) = 3z$. **Try it out!**

EXERCISE 1.59. Try on your own to write down a precise mathematical definition of the Fatou set and the Julia set that will work for any rational function $g(z)$ defined on $\overline{\mathbb{C}}$. In particular, try to formulate what it means for orbits to be drastically or significantly different. **Try it out!**

Notation: For a rational function $g(z)$, we employ the notation $F(g)$ for the *Fatou set* of g and $J(g)$ for the *Julia set* of g. Formal definitions are given in chapter Appendix 1.A on page 96 . These should be read, but note that it is not necessary to know these formal definitions well to continue on in the text; an intuitive understanding of these ideas will allow you to progress through this text just fine.

We showed that $J(f) = C(0,1)$ for the map $f(z) = z^2$ by noting that for any seed $z_0 \in C(0,1)$ there is a point arbitrarily close to z_0 whose orbit tends to ∞ (and is thus drastically different from the orbit of z_0 which must remain on $C(0,1)$). However, it is also true that f is even chaotic on $C(0,1)$ if we restrict ourselves to only using seed values from $C(0,1)$. To see this, consider $z_0 = e^{i\theta}$ written in polar form. For any *nearby point* $z_0' = e^{i\alpha}$ we must, due to the angle doubling nature of f, have that the distance $|f(z_0) - f(z_0')| > |z_0 - z_0'|$. Indeed, denoting the (arclength) distance between z_0 and z_0' along the unit circle by $\beta = |\theta - \alpha|$, we see that the distance between $f(z_0)$ and $f(z_0')$ along the unit circle will then be 2β. Inductively, we see that the distance between $f^n(z_0)$ and $f^n(z_0')$ along the unit circle will be $2^n\beta$, as long as $2^n\beta < \pi$. Clearly then, no matter how close z_0 and z_0' start out, i.e., no matter how small β is, corresponding orbit points will eventually be "far" apart. More precisely, as soon as $2^n\beta > \pi/3$ we must have $|f^n(z_0) - f^n(z_0')| > 1$. Thus arbitrarily close seed values on $C(0,1)$ do not have corresponding orbit values that *forever* stay arbitrarily close. Use the *Complex Function Iterator Applet* (using both **Polar computation** mode and **Polar seed form**) to witness this sensitive dependence. In particular, simultaneously iterate (one step at a time) **Seed 1** $z_0 = e^{2.18i}$ and **Seed 2** $z_0' = e^{2.19i}$.

We close this section by reflecting on some commonalities between the Newton Map dynamics we have seen in Section 1.2.4 and the dynamics of $f(z) = z^2$. As we saw in each Newton's method example, so too for the $f(z) = z^2$ dynamics, any two basins of attraction of attracting fixed points share a common boundary. In the case of $f(z) = z^2$, the basins $A_f(0)$ and $A_f(\infty)$ share the common boundary which turns out the be exactly the Julia set $J(f) = C(0,1)$ where, informally speaking, the attractive "pull" of each attracting fixed point is balanced by the other. The reader who read all of Section 1.2 should now go back and consider each complex Newton's method example to convince himself or herself that in those cases too the Julia set of the Newton map is exactly the common boundary of any attracting basin. This is no coincidence and we state the precise result as follows.

THEOREM 1.60. Let $f(z)$ be a rational map. If w is an attracting fixed point of $f(z)$, then $\partial A_f(w) = J(f)$.

REMARK 1.61. Theorem 1.60 together with Proposition 1.21 implies Theorem 1.35.

REMARK 1.62. In Theorem 1.60, if we do not demand that the fixed point be attracting, then the conclusion might not follow. We saw this in Example 1.50 where $A_h(\infty) = \overline{\mathbb{C}}$ which has empty boundary, but $J(h) = \{\infty\}$ which we leave the reader to show.

AN INCOMPLETE SKETCH OF THE PROOF OF THEOREM 1.60. Showing the full details of this proof requires both a precise definition of Julia set and a little advanced complex analysis beyond the level of this text. Instead we provide just a sketch of some of the arguments. By Additional Exercise 1.169, $A_f(w)$ is an open set. Thus each point in $A_f(w)$ has a whole neighborhood of points whose orbits all act the same way, showing (albeit informally) $A_f(w) \subseteq F(f)$. Likewise, any point in $\partial A_f(w)$ contains points arbitrarily close which iterate to w and points arbitrarily close whose orbit points stay far away from w. Thus such a point must be in the Julia set, i.e., $\partial A_f(w) \subseteq J(f)$. What remains is to show that $\overline{\mathbb{C}} \setminus \overline{A_f(w)}$ is in the Fatou set. By showing $f(\overline{\mathbb{C}} \setminus \overline{A_f(w)}) \subseteq \overline{\mathbb{C}} \setminus \overline{A_f(w)}$, one may use Montel's Theorem (Theorem 1.208) to conclude $\overline{\mathbb{C}} \setminus \overline{A_f(w)} \subseteq F(f)$. We leave it to the interested reader to pursue the finer details of the proof in p. 58 of [**2**]. $\qquad \square$

1.3.3. Dynamics of maps of the form $f_c(z) = z^2 + c$**.** We have carefully investigated the dynamics of the map $f(z) = z^2$ (though there is still more to say since a finer analysis of the dynamics on $C(0,1)$ reveals some very interesting behavior - see Remark 1.212 and Additional Exercises 1.171 and 1.206). Now we investigate what effect there will be when we change the map by adding a constant.

One justification for studying this (although the only justification a mathematician usually requires is that the resulting problem be interesting) is in line with our theme of wanting to know what happens when the system we are studying is tweaked a little bit, or has some error causing us to believe that we cannot be 100% certain that our mathematical model is *exactly* correct (as opposed to being just a very good approximation). In the dynamics of $f(z) = z^2$ above, we identified which seed values z_0 were stable (i.e., in the Fatou set) or chaotic (i.e., in the Julia set) by studying the effects of allowing arbitrarily small errors in the *seed value*. Now we will investigate what type of stability may or may not be present when we allow for an error or *perturbation* in the *map* we are iterating.

Hence, we will consider the dynamics of maps of the form $f_c(z) = z^2 + c$ where c is a complex parameter. Thus, we will fix a c_0 value and study the dynamics of f_{c_0}. Then we will vary the c parameter a little and study whether or not the dynamics of the resulting maps f_c have similar behavior. Just as we saw the seed values z for a fixed map get naturally split into the stable Fatou set and the chaotic Julia set, so too will we see the set of all parameter values c be naturally split into *stable* parameters and what we will call *bifurcation* parameters.

We begin by noting a common aspect of the dynamics of f_c for all values of c. Each map f_c has a super attracting fixed point at ∞ (which the reader should formally verify by either checking the behavior of f_c near ∞ or by checking the multiplier of the fixed point at ∞). Thus, if any part of the orbit $z_n = f_c^n(z_0)$ of some seed value z_0 should be *large enough* (by this we mean that $|z_n|$ is large enough), then the orbit will tend to ∞. A careful calculation in Additional Exercise 1.189 shows that for the maps f_c, we have $|z_k| > \max\{2, |c|\}$ for some $k \in N$ if and only if $z_n \to \infty$.

We begin by considering a small perturbation of the map $f_0(z) = z^2$. If we let c be very small (by which we mean $|c|$ is small), then we might expect the dynamics of f_c to be very similar to the dynamics of $f_0(z) = z^2$. This turns out to be true in many respects.

EXPLORATION 1.63. Fix $c = 0.1$ and use the *Complex Function Iterator Applet* to study the dynamics of $f_{0.1}(z) = z^2 + 0.1$. Try many different seed values including $z_0 = 0, 1 + i, -2 - 0.5i, ...$ and record the different types of behavior you are able to find. ***Try it out!***

As in the f_0 case, it seems we have only two types of long term behavior. In particular, iterates of $f_{0.1}$ seem to either approach ∞ or they approach the attracting fixed point $p \approx 0.1127$. The intuition we've built up, however, tells us that this cannot be the whole story. If there are two attracting fixed points, then we expect there to be some "tension" between attracting basins, and that points on the boundaries of these basins will not be attracted to either fixed point. Recall, that the boundaries of the two attracting basins for the map $f_0(z) = z^2$ were both equal to the chaotic set $J(f_0) = C(0,1)$. Similarly, this same phenomenon occurs with the map $f_{0.1}$, except that the chaotic set $J(f_{0.1})$ is not a circle (though some advanced mathematics can show that it is a simple closed curve, see [**2**], p. 126). Let's now use the *Global Complex Iteration Applet for Polynomials* to see a picture of $J(f_{0.1})$. This applet will color each seed in the basin of attraction of ∞ for the map $f_{0.1}$ based on how many iterates it takes for the orbit to become strictly larger than $\max\{2, |0.1|\} = 2$, and it will color the remaining points black. By Theorem 1.60, the basin $A_{f_{0.1}}(\infty)$ has boundary set equal to $J(f_{0.1})$. Experiment with iterating seed values (two at a time) near $J(f_{0.1})$ to see the sensitive dependence on initial conditions. ***Try it out!***

The set of points colored in black in the applet has a special name which we present here in general terms for any polynomial.

DEFINITION 1.64. For a polynomial $g(z)$, we define the *filled-in Julia set* $K(g)$ to be the set of points which do not iterate to ∞, i.e.,

$$K(g) = \{z \in \mathbb{C} : \{g^n(z)\}_{n=0}^{\infty} \text{ is bounded in } \mathbb{C}\} = \overline{\mathbb{C}} \setminus A_g(\infty).$$

REMARK 1.65. For a polynomial $g(z)$ of degree greater than or equal to two (which must then have a super attracting fixed point at ∞), it is true that $\partial A_g(\infty) = \partial K(g)$, and so by Theorem 1.60 we have $J(g), \partial A_g(\infty)$ and $\partial K(g)$ are all identical sets. The reader is asked to prove this in Additional Exercise 1.174.

Returning to the dynamics of the map $f_{0.1}$, we ask if we could have predicted ahead of time that there would be a finite attracting fixed point? The answer is yes, and here is how. We find the two finite fixed points of $f_{0.1}$ by solving the equation $f_{0.1}(z) = z$ (do you see why?). Doing this we then show that the two finite fixed points $p = (1 - \sqrt{0.6})/2$ and $q = (1 + \sqrt{0.6})/2$ are attracting and repelling, respectively, since $|f'_{0.1}(p)| < 1$ and $|f'_{0.1}(q)| > 1$.

One way to relate the dynamics of f_0 and $f_{0.1}$ is to say that when the c parameter moves from $c = 0$ to $c = 0.1$, (a) the super attracting fixed point at 0 with multiplier $\lambda_0 = 0$ becomes an attracting fixed point at $p \approx 0.1127$ with multiplier $\lambda_p = f'_{0.1}(p)$, (b) the circle $J(f_0)$ becomes a slightly distorted circle $J(f_{0.1})$, and (c) the super attracting fixed point at ∞ persists, i.e., remains a super attracting fixed point for $f_{0.1}$. Thus the small change in the c parameter led to only a small change in the dynamics.

Let us explore other c values to decide which c values have similar dynamics to f_0 and which do not.

EXERCISE 1.66. Fix $c = 0.2 + 0.2i$ and use the *Global Complex Iteration Applet for Polynomials* to study the dynamics of $f_{0.2+0.2i}(z) = z^2 + 0.2 + 0.2i$ and to see the global picture of the attracting basins and the Julia set. Calculate by hand the *exact* value of the attracting fixed point of this map. You can test your calculation by using the applet to iterate a nearby seed. ***Try it out!***

Now let us see what happens if we move the parameter c somewhat far from 0.

EXERCISE 1.67. Fix $c = -1$ and use the *Global Complex Iteration Applet for Polynomials* to study the dynamics of $f_{-1}(z) = z^2 - 1$. Are the dynamics similar to that of f_0? ***Try it out!***

EXERCISE 1.68. By generalizing the calculations done in the $c = 0.1$ case, mathematically describe the set K_1 of all c values such that f_c has a finite attracting fixed point in $\mathbb{C}$. Start by solving an equation to find the fixed points, and then consider what conditions need to be met to make one of these fixed points attracting. Note that for all c values, f_c always has fixed points. Our goal is to determine for what c values will there be a fixed point which is *attracting*. You can see the picture of K_1, which is called a *cardioid* (heart shaped region) in Figure 1.13. ***Try it out!***

By examining the attracting basins and Julia sets for maps f_c where $c \in K_1$ (using the *Global Complex Iteration Applet for Polynomials*) you will see that each such function has dynamics similar to the dynamics of f_0. Thus we have witnessed our first example of **stability in the parameter**, i.e., for any c in this set of parameters, the dynamics does not fundamentally change when you move the parameter around a little bit. Another way of saying this is that K_1 is an *open* set of parameter values.

We know that for each $c \in K_1$ there is an attracting fixed point p_c of the map f_c. Let's call $\lambda(c)$ the multiplier at p_c and thus, using Theorem 1.49 which says that the multiplier of a (super) attracting fixed point must have modulus strictly less than 1, we may regard λ as a map from K_1 into $\triangle(0, 1)$. If we follow the calculations in the above Exercise 1.68 carefully, we see that this **multiplier map** $\lambda : K_1 \to \triangle(0, 1)$ is one-to-one, continuous, and onto (onto means that for every $\lambda_0 \in \triangle(0, 1)$ there is a parameter $c_0 \in K_1$ such that $\lambda(c_0) = \lambda_0$). In fact, this map can be *extended*[13] to be

---

[13]We say a map g defined on its domain set D can be *extended* to a larger set $\tilde{D} \supset D$, if there exists an extension map $\tilde{g}$ on $\tilde{D}$ such that $\tilde{g} = g$ on D. By a slight abuse of notation we use g to denote the extension, which simply means that we assume g itself is defined on the larger set $\tilde{D}$.

defined and continuous on all of $\overline{K_1}$. However, for $c \in \partial K_1$ we only have $\lambda(c) \in C(0,1)$ and the corresponding fixed point p_c is indifferent.

Let us explicitly compute this multiplier map and consider the meaning and significance of its inverse map. By a direct calculation, we see that $p_c = (1 - \sqrt{1 - 4c})/2$ and $q_c = (1 + \sqrt{1 - 4c})/2$ are the fixed points of f_c. Note that p_c will be attracting for an appropriate choice of c, but q_c will always be repelling (why?). Hence, the multiplier map is given by $\lambda = \lambda(c) = f'_c(p_c) = 2p_c = 1 - \sqrt{1 - 4c}$, when $c \in K_1$. The inverse of the multiplier map, as you can readily compute, is then $c = c(\lambda) = \frac{\lambda}{2} - \frac{\lambda^2}{4}$, which gives the c value for a fixed point with prescribed multiplier λ. Since c values in K_1 must correspond to (super)attracting fixed points, the set K_1 must be the image of the set of all $\lambda \in \triangle(0,1)$ under the map $c(\lambda)$. This produces the set K_1 illustrated in Figure 1.13 which you should take a moment to verify using the *ComplexTool Applet*.

EXERCISE 1.69. The Multiplier map and its Inverse.
(a) Find the c value such that f_c has an attracting fixed point p_c with multiplier $\lambda = 0.7e^{\pi i/4}$, and then use an applet to illustrate the attractiveness of p_c by iterating points near it.
(b) Find the c value such that f_c has an indifferent fixed point p_c with multiplier $\lambda = e^{\pi i/3}$, and then use an applet to illustrate the behavior of orbits of points near p_c.

Additional Exercise 1.175 will help the reader gain a better understanding of the role the multiplier plays in the dynamics near a fixed point. The reader may go on without doing this exercise now, but for a deeper understanding the reader should take the time to do it now.

1.3.4. Cycles for the map $f_c(z) = z^2 + c$. As we saw above, the orbit of 0 under the map $f_{-1}(z) = z^2 - 1$ is $-1, 0, -1, 0, -1, 0, \ldots$. We summarize this situation by saying that f_{-1} has a 2-cycle $\{0, -1\}$. Also, as we saw (for example, by iterating the seed $z_0 = 0.2 - 0.3i$), this 2-cycle seems to be *attracting* (actually, we discuss later that this cycle can even be properly labeled *super attracting*).

EXERCISE 1.70. Modify Definition 1.16 of an attracting fixed point to come up with your own definition of an attracting 2-cycle. ***Try it out!***

EXERCISE 1.71. Modify Definition 1.14 of an attracting basin of a point to come up with your own definition of an attracting basin for a 2-cycle. ***Try it out!***

Using an applet to produce the orbit of several seed values, we again seem to have only two types of long term behavior for the map f_{-1}. In particular, iterates of f_{-1} seem to either approach ∞ or they "approach" the 2-cycle $\{0, -1\}$. Our intuition, however, suggests that this cannot be the whole story. Tension between attracting "basins" probably leads to points which are neither attracted to ∞ nor to the 2-cycle $\{0, -1\}$. Use the *Global Complex Iteration Applet for Polynomials* to see these basins, and the corresponding Julia set. ***Try it out!***

EXAMPLE 1.72. Let $c = -0.9 - 0.1i$ and use one of the applets to study the dynamics of the map $f_{-0.9-0.1i}(z) = z^2 + (-0.9 - 0.1i)$ to verify that this map also has an attracting 2-cycle. ***Try it out!***

EXAMPLE 1.73. Let $c = -0.13 + 0.73i$ and use one of the applets to study the dynamics of the map $f_{-0.13+0.73i}(z) = z^2 + (-0.13 + 0.73i)$ to verify that this map has what we should naturally call an attracting 3-cycle. ***Try it out!***

Ultimately we would like to calculate (by hand, or at least only use the computer when we know we will get trustworthy results) which c parameters will lead to the map $f_c(z) = z^2 + c$ having an attracting 2-cycle, an attracting 3-cycle, an attracting 4-cycle, and so on. We will be able to make some progress on this question, but our methods, as you will see, will hit the familiar road block of trying to find the roots of a polynomial of high degree. We begin by first defining cycles of any length and then we see how to classify cycles as attracting/repelling/indifferent by using the derivative (or more precisely the multiplier).

1.3.5. p-Cycles and their Classification.

DEFINITION 1.74 (Cycles). A point $w \in \overline{\mathbb{C}}$ is called *periodic with period p* for the map f if $f^p(w) = w$ and $w, f(w), \ldots, f^{p-1}(w)$ are distinct points. In this case we call the set $\{w, f(w), \ldots, f^{p-1}(w)\}$ a *p-cycle* for the map f.

Thus we see that periodic points correspond exactly to fixed points of higher iterates f^p of the map f. For example, the periodic point of period 2 at $w = 0$ for the map f_{-1} is a fixed point of the second iterate f_{-1}^2. We can then use this fact to classify a *cycle* as attracting/repelling/indifferent based on the multiplier of the corresponding iterate.

DEFINITION 1.75 (Multiplier for Cycles). Suppose the set $\{w_0, \ldots, w_{p-1}\}$ forms a p-cycle for the map f. We define the *multiplier* λ of this cycle (also called the multiplier of each point $w_0, \ldots, w_{p-1}$ of period p) to be the multiplier of the map f^p at its fixed point w_0. Then the p-cycle $\{w_0, \ldots, w_{p-1}\}$ of the map f is called
  a) *super attracting* if $\lambda = 0$
  b) *attracting* if $0 < |\lambda| < 1$
  c) *repelling* if $|\lambda| > 1$
  d) *indifferent* if $|\lambda| = 1$.
As for fixed points, we say the cycle is (super)attracting when it is known that either case (a) or (b) holds.

EXAMPLE 1.76. The 2-cycle $\{0, -1\}$ for the map f_{-1} is super-attracting since $\lambda = (f_{-1}^2)'(0) = f_{-1}'(0) \cdot f_{-1}'(f_{-1}(0)) = f_{-1}'(0) \cdot f_{-1}'(-1) = 0$. We note also that instead of using 0 we could have use the other point in the 2-cycle to calculate $\lambda = (f_{-1}^2)'(-1) = f_{-1}'(-1) \cdot f_{-1}'(f_{-1}(-1)) = f_{-1}'(-1) \cdot f_{-1}'(0) = 0$.

Since the multiplier λ of a p-cycle is defined via the derivative of f^p, the p-fold composition of f with itself, is very important to understand the use of the chain rule, as seen in Example 1.76, in the classification of p-cycles. And so we now examine it more closely.

Chain Rule in $\mathbb{C}$: If f and g are analytic functions at **finite points** z_0 and z_1, respectively, and if $z_0 \overset{f}{\mapsto} z_1 \overset{g}{\mapsto} z_2$, then $(g \circ f)'(z_0) = g'(f(z_0))f'(z_0) = g'(z_1)f'(z_0)$. In other words, in order to compute the derivative of the composite function, we simply multiply the derivatives of each function (evaluated at the appropriate point) along the way.
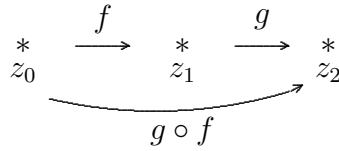
Now suppose that set of **finite** points $\{w_0, \ldots, w_{p-1}\}$ forms a p-cycle for the analytic map f as pictured in Figure 1.12.
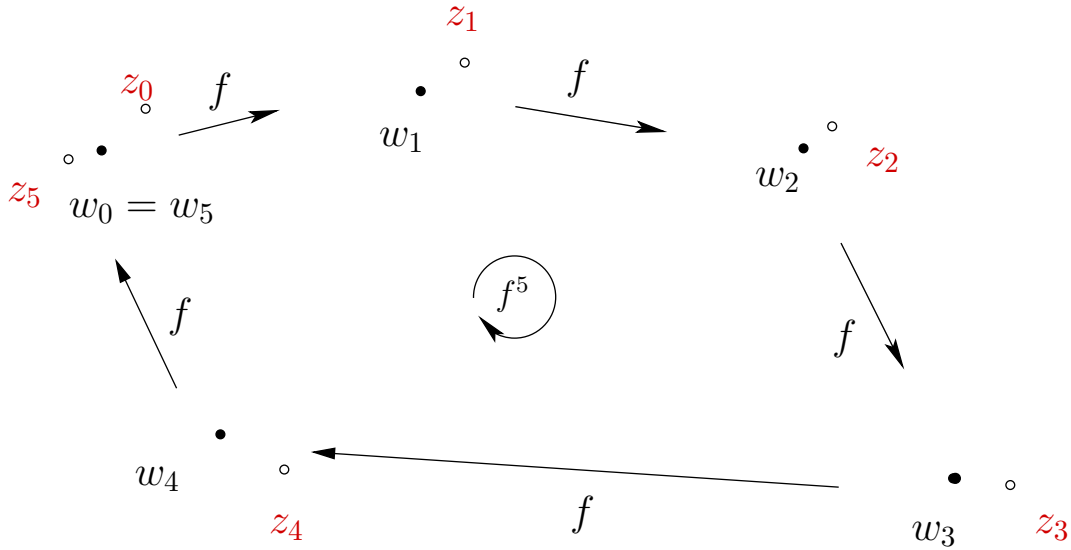
FIGURE 1.12. Illustration of a 5-cycle $\{w_0, w_1, \ldots, w_4\}$ along with the partial orbit of a point z_0 chosen near w_0.

We have $f^p(w_0) = w_0$, and so by the chain rule we compute

$$(1) \qquad \lambda = (f^p)'(w_0) = f'(w_0)f'(w_1)\ldots f'(w_{p-1}) = (f^p)'(w_j), \text{ for } j = 0, \ldots, p-1,$$

which shows, among other things, that the definition of the multiplier in Definition 1.75 is well defined since the derivative of f^p is the same at any point in the cycle.

We now wish to understand the relationship between the classification of the cycle as attracting/repelling/indifferent (determined by λ) and the dynamics of the map f near the cycle. We first note that simply by *continuity* of the map f, for any seed z_0 sufficiently close to w_0, the orbit points $z_1, \ldots, z_{p-1}, z_p$ will be close to the points $w_1, \ldots, w_{p-1}, w_0$, respectively. Supposing that $|\lambda| < 1$, the map f^p has an attracting fixed point at w_0. Thus, if we choose a seed z_0 sufficiently close to w_0, we must have that $f^p(z_0)$ is closer to w_0 than z_0 is (i.e., $|f^p(z_0) - w_0| < |z_0 - w_0|$) by Definition 1.16 applied to the fixed point w_0 of the map f^p. We also note that this argument works equally well for any of $w_1, \ldots, w_{p-1}$ as it does for w_0, and so one way to describe such an attracting cycle is to say that each time you apply the map f for a total of p times, points near any w_k will move around the cycle only to return closer to w_k. See Figure 1.12 where the orbit of z_0 exhibits this behavior. In a similar way we justify the classification of repelling cycles.

Note that the above calculations require only minor modifications when one of the points w_k in the cycle is ∞. In the spherical metric, where ∞ does not play a more special role than any other point in $\mathbb{C}$, one sees that the dynamic behavior (attraction/repulsion) of the cycle behaves in the same fashion as described above for cycles in the finite plane $\mathbb{C}$.

1.3.6. Attracting cycles for the maps $f_c(z) = z^2 + c$. Let us return to investigating the dynamics of the maps f_c. We have seen two examples of maps of the form $f_c(z) = z^2 + c$ with attracting 2-cycles. Let us now determine the set K_2 of all c values such that f_c has an attracting 2-cycle.[14] Any point in a 2-cycle must (a) be a fixed point of f_c^2, and (b) **not** be a fixed point of f_c. Thus we wish to solve the equation A: $(z^2 + c)^2 + c = z$ and exclude the solutions of equation B: $z^2 + c = z$. Since each solution to B, rewritten as $z^2 + c - z = 0$, is a solution to A, rewritten as $(z^2 + c)^2 + c - z = 0$, we have that $z^2 + c - z$ must divide $(z^2 + c)^2 + c - z$. After doing some long division, we can then rewrite A as $(z^2 + z + 1 + c)(z^2 + c - z) = 0$. Thus a 2-cycle $\{u, v\}$ must be such that both u and v solve the equation $(z^2 + z + 1 + c) = 0$, i.e., $(z - u)(z - v) = z^2 + z + 1 + c$, which after expanding and comparing coefficients implies $uv = 1 + c$. According to Equation (1) the multiplier of the 2-cycle $\{u, v\}$ is $\lambda = f_c'(u) f_c'(v) = 4uv = 4(1 + c)$. Hence this 2-cycle will be attracting exactly when $4|1 + c| = |\lambda| < 1$, i.e., when $|c - (-1)| < 1/4$. Hence $K_2 = \triangle(-1, 1/4)$, which is the disk pictured in Figure 1.13.

EXPLORATION 1.77. Test various c values in K_2 by using the *Global Complex Iteration Applet for Polynomials* to see that you do indeed get an attracting 2-cycle. Which seed values in the picture produced by the applet seem to "find" the attracting 2-cycle? *Try it out!*

In Additional Exercise 1.176 the reader is asked to investigate the relationship between the multiplier and the convergence towards the 2-cycle. And in Additional

---

[14]As 2-cycles generally exist, the issue here is to determine when such a cycle will be *attracting*.
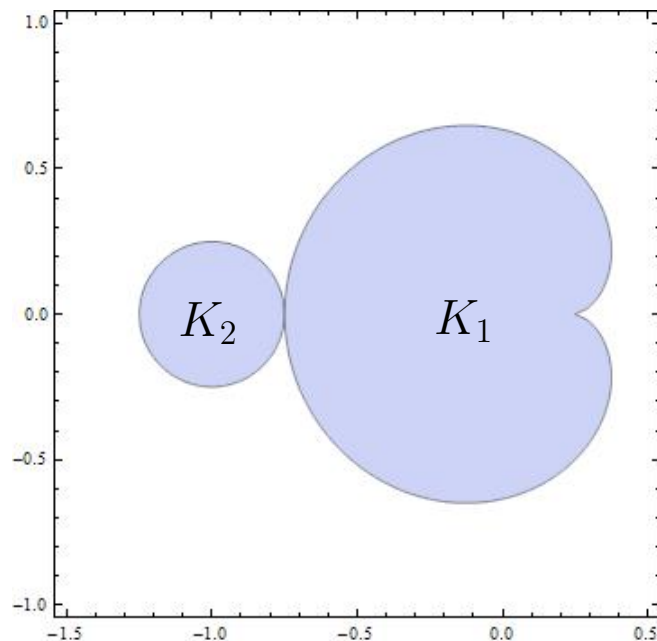
FIGURE 1.13. Parameter space of c values showing the cardioid K_1 with cusp at $c = 1/4$ and the disk $K_2 = \triangle(-1, 1/4)$. The boundaries of K_1 and K_2 meet at $c = -3/4$.

Exercise 1.177 the reader is asked to investigate another multiplier map (this time defined on K_2) and its inverse.

Let us now try to determine the set K_3 of all c parameters which lead to attracting 3-cycles for $f_c(z) = z^2 + c$. Any point in a 3-cycle must (a) be a fixed point of f_c^3, and (b) **not** be a fixed point of f_c, and so must solve (when substituted for z) the eighth degree polynomial $[(z^2 + c)^2 + c]^2 + c - z = 0$, but not be a root of $z^2 + c - z = 0$. After long division, as above, we are still left with a degree six polynomial to solve if we are to find the points of the 3-cycle. Thus we see that since we cannot, in general, solve a polynomial of degree five or greater, we will have considerable more difficulty determining the attracting 3-cycles for maps of the form $f_c(z) = z^2 + c$. However, in the above calculation we were able to locate the c values that correspond to attracting 2-cycles without ever having to explicitly solve for the points of the cycle (although we could have since it is just a matter of applying the quadratic formula). We wonder then if it is possible to use similar techniques, or devise new ones, to describe as much as we can of this set K_3.

LARGE PROJECT 1.78. Is it possible to determine the set K_3 precisely, as was done for K_1 and K_2? Even though we might not be able to explicitly solve the degree six polynomial mentioned above, perhaps one can use other root solving techniques (such as Newton's method) to approximate roots to a high enough degree to be useful. Such

roots could then be tested with the *Global Complex Iteration Applet for Polynomials* to see their role in the dynamics of f_c. Related questions of interest are: Can you show the fact (not proven in this chapter) that K_3 is an open set? It turns out that K_3 has more than one connected component (as opposed to K_1 and K_2 which were connected). Can you determine how many connected components K_3 has and what relationship these components have to each other? Can you find the c parameters which lead to *super attracting* 3-cycles? The roots of the aforementioned degree six polynomial must include three points which make up the attracting 3-cycle, but what do the other three points represent? Will any of these techniques yield useful results for the problem of finding the c parameters which yield attracting 4, 5, 6, ... cycles?

We present a definition and theorem (whose proof is beyond the scope of this text) which might be of some use in understanding the above project.

DEFINITION 1.79 (Hyperbolic Components). For each $n \in \mathbb{N}$, we define the set K_n to be the set of parameters c such that f_c has an attracting n-cycle. We call any connected component[15] W of some K_n a *hyperbolic component* of K_n.

Although K_1 and K_2 are connected sets, it is true that some of the K_n are disconnected. Each piece, that is, connected component of K_n, however, has the following nice property corresponding to its multiplier map.

THEOREM 1.80 (Multiplier Map Theorem ([**2**], p. 134)). Let W be a hyperbolic component of some K_n. Let $\lambda : W \to \triangle(0, 1)$ be the multiplier map which takes each parameter $c \in W$ to the multiplier $\lambda(c)$ of the associated attracting n-cycle. Then the map λ is one-to-one, analytic, and onto, i.e., it maps W conformally onto $\triangle(0, 1)$. Furthermore, the map λ extends[16] to be a one-to-one continuous map of $\overline{W}$ onto $\overline{\triangle(0, 1)}$.

DEFINITION 1.81. For a hyperbolic component W of K_n, we call the unique $c \in W$ the *center* of W if $\lambda(c) = 0$, i.e., it is the unique c in W for which f_c has a *super attracting* n-cycle.

EXAMPLE 1.82. We have already observed that the center of K_1 is $c = 0$ and the center of K_2 is $c = -1$.

One can **explicitly** show this theorem to be true for the parameter sets K_1 and K_2 by actually writing down and studying the multiplier map (see Additional Exercise 1.177). For the other sets K_n, however, it is not so easy. The proof of Theorem 1.80 uses key properties of the multiplier map without explicitly constructing it. One important application of this result is that it proves that each hyperbolic component W of K_n is open since it is the *conformal*[17] image of an open set, namely $W = \lambda^{-1}(\triangle(0, 1))$. Hence, each K_n is also open.

---

[15]See Appendix B.3 on page 422 for the definition of a *component*.
[16]See Footnote 13 on p. 35.
[17]Recall, that a map is called *conformal* when it is both one-to-one and analytic.

As we have seen, if we move the parameter c within the cardioid K_1, the dynamics of $f_c(z) = z^2 + c$ do not change much. Similarly, if we move the parameter c within the disk K_2, the dynamics of $f_c(z) = z^2 + c$ do not change much. For these reasons we call K_1 and K_2 *stable* regions of parameter space. It is also true that each K_n is a stable region of parameter space, which, given its definition, is really just another way of saying that K_n is an *open* set. This follows from above, but also seems reasonable without appealing to Theorem 1.80. Note that if you gently tweak a function with an attracting cycle (keeping in mind the strict inequality condition on the multiplier), then it seems reasonable the new function will still have an attracting cycle (i.e., will have a strict inequality condition on the multiplier) in roughly the same place (see Additional Exercise 1.178). This is exactly what we witness when we move the c parameter by small amounts within the cardioid K_1 or within the disk K_2. We **cannot**, however, expect this to happen with an indifferent fixed point (or cycle) since by tweaking the multiplier of an indifferent fixed point, the modulus could easily be strictly less than one (attracting) or strictly greater than one (repelling), instead of remaining exactly equal to one.

As opposed to the stable parameters found in the sets K_n, we call a parameter c **unstable** if there are parameters arbitrarily close for which the maps $f_c(z) = z^2 + c$ have fundamentally different dynamics. For example, $c = -3/4$ is an unstable parameter.

EXPLORATION 1.83. The reader should now pause before reading further to write down several reasons why $c = -3/4$ is an unstable parameter. Use the *Global Complex Iteration Applet for Polynomials* to explore the dynamics when c is close to $c = -3/4$, paying special attention to the dynamics for $c_1 = -0.75 + .05, c_2 = -0.75 - .05$ and $c_3 = -0.75 + .05i$. ***Try it out!***

Now that the reader has provided their own reasons, we go on to illustrate the unstable nature of the parameter $c = -3/4$ by describing three particular ways in which the dynamics changes at this c value. We note how the type of attracting cycle, the Julia set, and the orbit of the origin all undergo *fundamental* changes. We call $c = -3/4$ a *bifurcation point*, since it is the parameter on the boundary of two regions in parameter space where the corresponding dynamics undergoes a fundamental change.

1. ***Attracting Cycle:*** We first note that we don't even have to know the dynamics exactly at the point $c = -3/4$ to show it is unstable. It is enough to know that there are parameters arbitrarily close to and less than $c = -3/4$ (in K_2) which give rise to an attracting 2-cycle (but no attracting fixed point) and there are parameters arbitrarily close to and greater than $c = -3/4$ (in K_1) which give rise to an attracting fixed point (but no attracting 2-cycle). The reader can verify that as c decreases to and then past $-3/4$ the attracting fixed point becomes indifferent and then "splits" into an attracting 2-cycle.[18]

---

[18] A different way to look at this is to instead say that as c decreases to $-3/4$ the attracting fixed point merges with a repelling 2-cycle to form an indifferent fixed point exactly at $c = -3/4$. Then

2. ***Julia Set:*** We can also see the result of this fundamental change in dynamics by looking at the Julia sets $J(f_c)$ as we vary c. For example, if we start at $c = 0$ and then slowly decrease c, we see the Julia sets $J(f_c)$ change from a circle to a distorted circle. Decreasing c further towards $c = -3/4$, we see the distorted circle $J(f_c)$ begins to have infinitely many "bulbs" partially forming as the distorted circle $J(f_c)$ starts "pinching in". Exactly at $c = -3/4$ pinching in for each of the infinitely many distinct bulbs *simultaneously* becomes complete. Thus we have gone from having one bounded component of $F(f_c)$ for $c \in (-3/4, 1/4)$ consisting of one attracting basin of an attracting fixed point to the situation for $c \in (-5/4, -3/4)$ where there exist infinitely many bounded components of $F(f_c)$ consisting of the attracting basin of an attracting 2-cycle.

3. ***Orbit of the origin:*** Lastly, we notice another change in the dynamics as the parameter c moves from K_1 to K_2. The orbit of the origin changes from being attracted to an attracting fixed point to being attracted to an attracting 2-cycle. As this will be a key aspect to keep in mind, we point out that in each of the examples we have considered where the map $f_c(z) = z^2 + c$ had an attracting cycle, the origin was "absorbed" into the cycle in the sense that the tail end of the orbit (formally $\{f_c^n(0) : n > N\}$ for very large N) is nearly identical to the attracting cycle. Another way to say this is simply that the origin was ***attracted to the cycle***. It turns out that this is true in general (see Remark 1.94 below), and the key fact, as we shall see, is that $z = 0$ is a critical point, i.e., $f_c'(0) = 0$.

EXERCISE 1.84. The reader should verify that $c = 1/4$, the cusp of the cardioid K_1, is also an unstable parameter. Use the *Global Complex Iteration Applet for Polynomials* to investigate this with respect to items 1-3 above. ***Try it out!***

EXPLORATION 1.85. What do you think happens to the attracting fixed point and repelling 2-cycle when the parameter c decreases to $-3/4$, but then "makes a right turn" and starts heading towards $-0.75 + 0.05i$? Investigate this with respect to item 1 above and Footnote 18 using the *Global Complex Iteration Applet for Polynomials.*

1.4. Critical Points and Critical Orbits

In this section, we are able to address some of the following natural questions you may have been asking yourself as we looked at the dynamics of the maps $f_c(z) = z^2 + c$.

(1) Does every such map have an attracting cycle (other than ∞)?
(2) How many attracting cycles can f_c have?
(3) In all the examples we looked at, it seems that the orbit of the origin is always attracted to the attracting cycle. Is this true in general?

---

as c decreases further, the 2-cycle re-emerges as an attracting 2-cycle and the fixed point becomes repelling (see Exercise 1.179). Thus, the fixed point and the 2-cycle sort of exchange "polarity" in this transition.

In this section we see that the key to answering many of these questions involves the notion of *critical points*, and investigating how the orbit of such points largely determines many key dynamical features. We begin with a definition.

Recall that when the power series of an analytic map $f(z)$ at $z_0 \in \mathbb{C}$ has the form $f(z) = f(z_0) + a_k(z - z_0)^k + a_{k+1}(z - z_0)^{k+1} + \ldots$, where $a_k \neq 0$, we say that z_0 maps to $f(z_0)$ with **degree** $v_f(z_0) = k$ (we also call $v_f(z_0)$ the **multiplicity** or **valency**).

REMARK 1.86. The condition that z_0 and $f(z_0)$ are both in $\mathbb{C}$ will always be met in the examples in which we are concerned, and so we leave it to the interested reader to make the customary modifications to this definition when z_0 and/or $f(z_0)$ is infinity. When the term critical point is applied below to a point z_0, however, it will be understood that we do include the possibilities that z_0 and/or $f(z_0)$ is infinity.

DEFINITION 1.87. We call z_0 a *critical point* of f if $v_f(z_0) > 1$.

If z_0 and $f(z_0)$ are both in $\mathbb{C}$, then z_0 is a critical point exactly when $f'(z_0) = 0$ (just like we define in Calculus I). Also, as described in Appendix Section A.6.1 on page 414, since near z_0 the map f is locally a $v_f(z_0)$-to-one mapping, we see that z_0 is a critical point exactly when f is not locally one-to-one.

DEFINITION 1.88. Let f be a rational or entire map. If w is a (super) attracting fixed point of f, then we define the *immediate basin of attraction* $A_f^*(w)$ to be the connected component of $A_f(w)$ which contains w. We point out (but leave it to the interested reader to show) that $A_f^*(w)$ is the component of the Fatou set $F(f)$ which contains w.

THEOREM 1.89. Let w be a (super) attracting fixed point of a non Möbius rational map f. Then there exists a critical point $z_0 \in A_f^*(w)$ and hence $f^n(z_0) \to w$.

The proof of Theorem 1.89 is beyond the scope of this text, but the interested reader can find it as Theorem 7.5.1 in [1]. We do note, however, that we have witnessed this result in action many times now.

EXAMPLE 1.90. For c in K_1 (see Figure 1.13), we know that $f_c(z) = z^2 + c$ has a finite (super) attracting fixed point p_c. Since the origin is the only critical point (other than ∞, which is fixed) we must then have by Theorem 1.89 that $0 \in A_{f_c}^*(p_c)$ and $f_c^n(0) \to p_c$. This is exactly what we observed (without proof) using the applets in many examples above.

In order to present the corresponding result for attracting *cycles* we will need the following definition.

DEFINITION 1.91. Let f be a rational or entire map which is non Möbius. For a (super) attracting p-cycle $w_0, \ldots, w_{p-1}$ of f we define the immediate basin of this cycle to be the union of components $\cup_{j=0}^{p-1} F_j$ where F_j is the component of the Fatou set $F(f)$ containing w_j.

EXERCISE 1.92. The map $f_{-1}(z) = z^2 - 1$ has a super attracting 2-cycle $\{0, -1\}$ and so the immediate basin of this cycle consists of two components of $F(f_{-1})$, one of which contains the point 0 and the other contains the point -1. Use the *Global Complex Iteration Applet for Polynomials* to investigate the picture shown in Figure 1.14. What are the dynamic properties of the other black "bulbs", i.e., components of $F(f_{-1})$? Try to find a pattern to the bulbs so you can experimentally approximate points z_0 such that $f_{-1}^3(z_0) = 0$. How many such points are there? Do the same, but with changing 3 to 5. Can you generalize this? ***Try it out!***
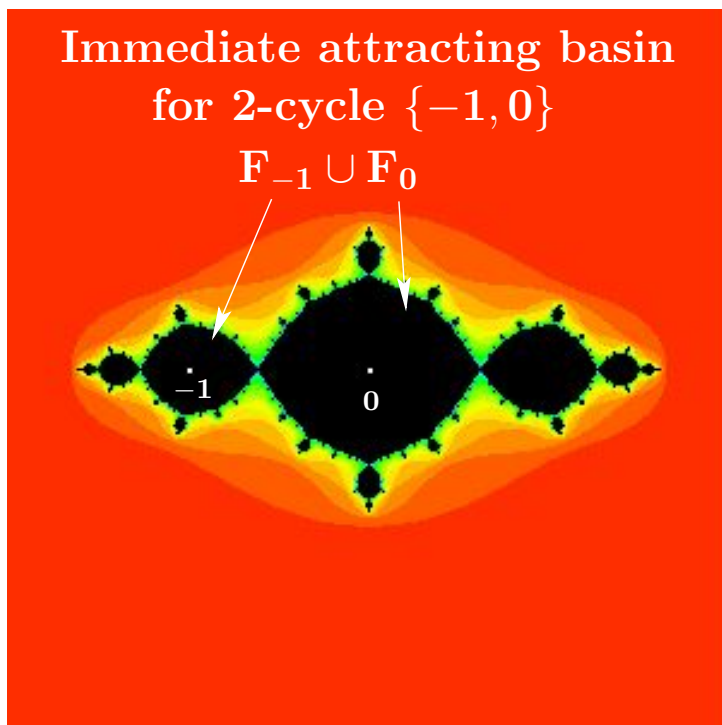
FIGURE 1.14. The immediate attracting basin of the 2-cycle $\{-1, 0\}$ for $f_{-1}(z) = z^2 - 1$ consists of the 2 components of the Fatou set containing the cycle points.

THEOREM 1.93. Let f be a rational map which is non Möbius. Then the immediate basin of each (super) attracting cycle contains a critical point of the map f.

The proof of this result takes advantage of the fixed point version of this theorem given in Theorem 1.89, the chain rule, and a couple other results which are slightly more than we want to take on at this point. The interested reader should consult [**1**] or [**2**]. However, it is pertinent to our discussion as it leads to the following remark.

REMARK 1.94. Since all maps of the form $f_c(z) = z^2 + c$ have only one finite critical point (namely $z = 0$), Theorem 1.93 implies that each such map can have at

most one **finite** (super) attracting cycle and any such cycle must "absorb" the orbit of the critical point at the origin. Thus we have answered Question 2 above and answered Question 3 in the case that f_c actually does have an attracting cycle.

DEFINITION 1.95. Both because 0 is a critical point of the map f_c and because its orbit is so important, we call $\{f_c^n(0)\}_{n=1}^\infty$ the *critical orbit* of the map f_c.

We have seen that the critical orbit $\{f_c^n(0)\}_{n=1}^\infty$ plays a special role in understanding the dynamics of the maps $f_c(z) = z^2 + c$. For some c values the critical orbit gets attracted to an attracting cycle (e.g., $c = 0, -1, 0.278 + 0.534i$) and for some other c values it gets attracted to the super attracting fixed point at ∞ (e.g., $c = 0.3, 4+i, -2-0.3i$). One may wonder then, is it true that for *every* c value the critical orbit becomes attracted to some (super) attracting cycle? Here again we can use our intuition to say that this is probably not true. If we consider the **parameter plane**, also called the c-plane, of all c parameters for the maps f_c, we can think of a new type of tension created by those c values whose critical orbits are attracted to ∞ and those c values whose critical orbits are not. It seems that maybe there are c values where the pull of the critical orbit towards ∞ and the pull of the critical orbit to stay bounded is balanced. This informal reasoning can be the basis for a good guess, but since it is far from a formal proof, we take the easier route and settle the question by looking at the following example.

EXAMPLE 1.96. Show formally with paper and pencil (or informally with an applet) that for $c = i, 1/4, -5/4, -2$, and $-3/4$, the critical orbit under $f_c(z) = z^2 + c$ is neither attracted to ∞ nor attracted a finite *attracting* cycle. Thus we have answered Question 1 from above. ***Try it out!***

In the case $c = i$, you noticed that the critical orbit $i \mapsto i - 1 \mapsto -i \mapsto i - 1 \mapsto -i \mapsto \dots$ *became* cyclic, but the cycle did not include 0. These types of orbits have an important role in dynamics and so we give them a special name, as well as the c values which lead to these types of critical orbits.

DEFINITION 1.97. We call a point z_0 *pre-periodic* (or *eventually periodic, but not periodic*) under the map f if it is not periodic, but some point on the orbit of z_0 is periodic.

DEFINITION 1.98. We call a parameter c a *Misiurewicz* point if 0 is pre-periodic under f_c.

Thus we see that $c = i$ and $c = -2$ are Misiurewicz points. However, $c = 0$ is not since $f_0(z) = z^2$ has a critical orbit $0 \mapsto 0 \mapsto \dots$ which is not pre-periodic. However, $f_0(z) = z^2$ does have non-critical pre-periodic points, and in Additional Exercise 1.181 you are asked to find them.

It turns out that if c is a Misiurewicz point, then $J(f_c) = K(f_c)$ (see [**2**], p. 133), in which case we call $J(f_c)$ a *dendrite*. See Figure 1.15 for an example.
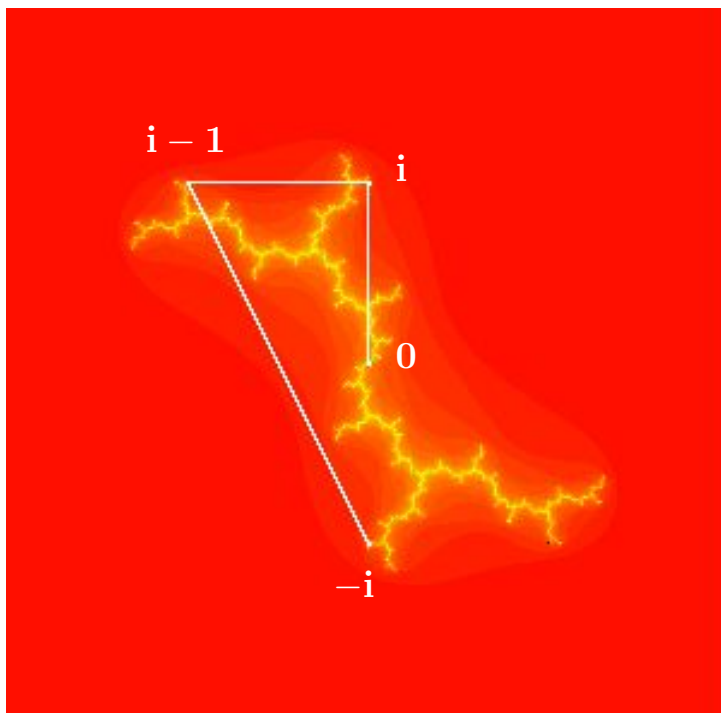
FIGURE 1.15. The dendrite $J(f_i)$ along with the critical orbit $i, i - 1, -i, i - 1, -i, \ldots$

REMARK 1.99. In order to correctly understand the concepts of pre-periodic points and Misiurewicz points we note the important distinction between orbits that are *pre-periodic* (which we informally described above as those which *became* cyclic) and those orbits that are *attracted* to a cycle. The difference is the same as the difference between a sequence approaching a value and a sequence eventually *being* a value. For example, the sequence $1, 1/2, 1/3, 1/4, \ldots$ approaches 0, but never becomes 0. However, the sequence $2, 1, 1/2, 0, 0, 0, \ldots$ eventually becomes 0 (and stays at 0). When using technology, such as the provided applets, it can be very difficult, if not impossible, to distinguish between these two concepts. For example, set $c = -0.9$ and look at the numerical values of the first 100 points of the orbit of $z_0 = 0$ under f_c. Focusing on the 75th and higher orbit values, we see that the data bounces back and forth between what appears to be the same two values. This might falsely lead you to conclude that $z_0 = 0$ is pre-periodic. The problem is that the *true* orbit in this case never actually *exactly* bounces back and forth between the *exact* same values. In fact, one can show (which we leave to the reader) that $z_1 < z_3 < z_5 < \ldots$ and $z_0 > z_2 > z_4 > \ldots$ with strict inequalities everywhere for all indices. However, the odd terms approach some value $-0.8872983346207418\ldots$ and the even terms approach $-0.1127016653792581\ldots$. Since the applet truncates the data for each z_n it *appears* that the odd sequence and even sequence do eventually become constant. The moral

47

of the story here is that when trying to discern whether a point is pre-periodic or not, technology might very well mislead you. In such matters, careful analysis and proof needs to be used.

Let us return to the question we addressed above, namely, does the critical orbit for every map $f_c(z) = z^2 + c$ get attracted to an attracting cycle? We saw that for $c = i, 1/4, -5/4, -2$, and $-3/4$ this is not the case. However, in each case the critical orbit either became cyclic or was attracted to a cycle (though not necessarily an attracting cycle). As mathematicians we then must wonder, is this always the case? Is it possible for some c value to be such that the critical orbit neither becomes cyclic nor is attracted to any cycle? The answer happens to be yes. We cannot explain the deep mathematics behind this answer in these limited pages, but we do note that asking good questions like this, whether or not we can answer them, is an important part of contributing to mathematics.

So it is not true that every c parameter has a critical orbit that either becomes cyclic or is attracted to a cycle. What *is* true, however, is that every c parameter has a critical orbit that is either attracted to ∞ or remains bounded. It is exactly this dichotomy which leads us to consider one of the most beautiful objects in all of mathematics, the Mandelbrot set.

DEFINITION 1.100. The *Mandelbrot set* is defined as $M = \{c \in \mathbb{C} : f_c^n(0) \nrightarrow \infty\}$.

We have already encountered some important aspects of M, namely, it contains both K_1 and K_2 pictured in Figure 1.13. We also know that it must contain K_n for every n (since for f_c to have an attracting n-cycle, the critical orbit must be attracted to this cycle by Theorem 1.93, and thus not be attracted to ∞). The calculation of the sets K_n is, however, a very arduous task and a complete description of all such sets has for years stumped mathematicians, and continues to stump us. Let us therefore use the computer to draw M for us and *experimentally* investigate the sets K_n and M (always keeping in mind that limitations of the sort discussed in Remarks 1.56 and 1.99 force us to moderate the confidence that we can place in such pictures). We begin by first using the following applet to construct a picture of M.

The *Mandelbrot Set Builder Applet* will color each selected point c in the parameter plane either red if $c \notin M$ or black if $c \in M$. Thus, if the the critical orbit (under the map f_c) limits to ∞, then the point c is colored red, otherwise it is colored black. Of course, we cannot compute the infinite number of points in the critical orbit, so the applet will compute only the number of iterates allowed in the **Maximum Iterations** input box. Setting this value to 100 will produce nice results (however, we encourage the reader to experiment with this value and investigate the effect it has on the picture). Also, the computer applet will color a selected c value red if and only if one of the calculated critical orbit points lands outside of $\overline{\triangle(0,2)}$. This is justified by the following lemma, which you are asked to prove in Additional Exercise 1.189.

LEMMA 1.101. If the critical orbit $f_c^n(0)$ ever *escapes* the closed disk of radius two centered at the origin, then the critical orbit must necessarily converge to ∞.

EXPLORATION 1.102. Take some time to experiment with the *Mandelbrot Set Builder Applet* to get a feel for the mathematics which defines the Mandelbrot set. **Try it out!**

One feature of the Mandelbrot set which stands out is that it is symmetric about the x-axis (see Additional Exercise 1.182). Another feature you may have observed is that M has no "holes" in it (see Additional Exercise 1.183). The *Mandelbrot Set Builder Applet* is a nice tool for visualizing M and can lead us as above to pursue some of its interesting features, but there are some important properties of M it cannot help with, such as whether M is a closed set or not. To do this we need more formal mathematics.

LEMMA 1.103. The Mandelbrot set M is a closed set.

PROOF. Suppose $c_k \to c^*$ where each $c_k \in M$. We will show that $c^* \in M$, thus proving that M is closed. By observing that for f_c the terms of the critical orbit are $0, c, c^2 + c, (c^2 + c)^2 + c, \ldots$, we see that the nth term can be written $Q_n(c)$ for some polynomial Q_n. Fix some $n \in \mathbb{N}$ and note that Q_n is continuous. Since $|Q_n(c_k)| \leq 2$ for each $k \in \mathbb{N}$ by Lemma 1.101, we must have $|Q_n(c^*)| \leq 2$ since $|Q_n(c_k)| \to |Q_n(c^*)|$ by the continuity of Q_n. Since this holds for every n, we have shown that the critical orbit of f_{c^*} is contained within $\overline{\triangle(0, 2)}$, and thus $c^* \in M$. $\qquad\square$

We close this section by presenting (without proof) two interesting facts about M. First, the set of Misiurewicz points, each which is clearly in M (why?), is dense in the boundary ∂M. This means that given any open set U which contains a point in ∂M, the set U must also contain a Misiurewicz point (see [**2**], p. 133). Second, ∂M is contained in the closure of the centers of the hyperbolic components defined in Definition 1.81 (see [**9**], p. 100). Thus, any open set which contains a point in ∂M, must also contain the center of some (small) hyperbolic component. Taking into account the fact that each center of a hyperbolic component is in the interior of M and each such center is separated away from any other such center, we see that ∂M necessarily must be quite complicated. We encourage the reader to take a moment to reflect and fully digest the previous statement. Start by explaining how the set $K_1 \cup K_2$ in Figure 1.13 fails to have this property, but that by attaching many tiny (what we will call) "bulbs" we can generate a set with this property.

REMARK 1.104. Although a lot is known about M, one very important question that has stumped mathematicians thus far, is whether or not M contains an open set which does not meet any K_n, that is, an open set of c values for which no f_c has an attracting cycle (other than ∞). The conjecture that asserts that this cannot happen is known as the *density of hyperbolicity conjecture* and remains the focus of much intense research.

1.5. Exploring the Mandelbrot Set M

The Mandelbrot set has been called one of the most beautiful objects in mathematics, but our red and black picture created by the *Mandelbrot Set Builder Applet* does not do it justice. The set M has many tiny "hairs" and "bulbs" that are hard to see with this *bichromatic* picture, and the complexity of the picture cries out for a way to zoom in on these intricate tiny hairs and bulbs.

Initially it was conjectured that the Mandelbrot set is disconnected. This was motivated by the low resolution pictures that did not show the fine details and thin filaments that connect all the parts of M. However, it turns out that M is connected (see [**1**], p. 239). We cannot prove this here, but a much more sophisticated applet will give us pictures which certainly hint that this might be the case.

From now on we use the *Parameter Plane and Julia Set Applet*, which still colors points in M black, but colors the parameters $c \notin M$ a different shade based on how many iterates it takes for the critical orbit to escape the disk $\overline{\triangle(0, 2)}$. This gives a much better feel for the immense detail of the Mandelbrot set M (see Figure 1.16). The applet also shows, for each selected c value, a picture of the corresponding Julia set.
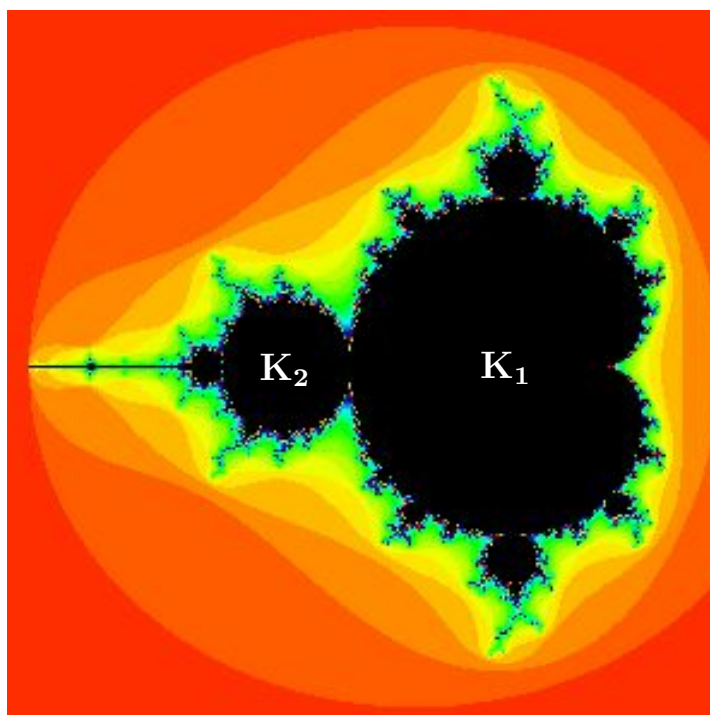

FIGURE 1.16. The Mandelbrot set.

EXPLORATION 1.105. The reader should now experiment and play with the *Parameter Plane and Julia Set Applet*, zooming in on various parts of M and the corresponding Julia sets. Look at not just the geometry of the pictures you see, but also at the dynamics you see. What do the bulbs and hairs mean dynamically? Is there a relationship between the geometry and the dynamics? Start looking for patterns and discovering new fine details in the hairs and bulbs. Describe what you find and make a list of observations, questions, and conjectures. You have the tools to explore an infinitely complex world. There are millions of new features of M and the related Julia sets to be found by using this applet. In fact, after exploring for a sufficient amount of time, zooming in on the fine details of the sets, you will likely find a picture that no other human has ever seen before! Take your time to explore this fascinating new world. ***Try it out!***

We have quite a menagerie of pictures to see and investigate, including Rabbits, Dragons, and Elephants. We even have Star Clusters, Galaxies, and baby Mandelbrot sets (see Figure 1.17).

In the next section we investigate such star cluster sets more carefully so that we can better understand exactly what we are seeing (or not seeing) in such pictures.

1.5.1. Cantor dust sets. As mentioned before we must always be careful when using technology to represent mathematics. What we see is not always an accurate representation of what we are trying to see. For example, using $c = 0.21 + 0.64i$ in the *Parameter Plane and Julia Set Applet*, it might appear that $J(f_c)$ is empty and that all points iterate to ∞ under f_c. In fact, if one displays the picture in Black and White via the **Dynamic plane black/white plot** checkbox on the **Settings tab**, you will see an all white screen, which taken at face value would mean $J(f_c)$ is empty. That, however, is very far from the truth. There are infinitely many (in fact, uncountably many) points which do not iterate to ∞ (see Additional Exercise 1.184). However $J(f_c)$, especially for having so many points, is rather small in the sense that it does not show up on the computer screen very well. To see such points better we view the color picture and then zoom in very far on the non-red parts (e.g., center each zoom in the middle of the largest "star cluster") to eventually see regions of black, which represent points which do not iterate to ∞. However, even these regions of black are not what they seem to be. Because we are only using a finite number of **Dynamic plane max iterations** some of these black points would ultimately iterate to ∞ if we would increase the number of iterations used. However, other of these black points will not iterate to ∞ and so should truly be colored black.

EXERCISE 1.106. Adjust the **Dynamic plane max iterations** value to make $J(f_c)$ (using $c = 0.21 + 0.64i$) harder to see and more accurate or easier to see and less accurate. Also, use the Black and White picture feature to see how it can sometimes be used to give you a much better picture of $J(f_c)$. Of course, we must always keep
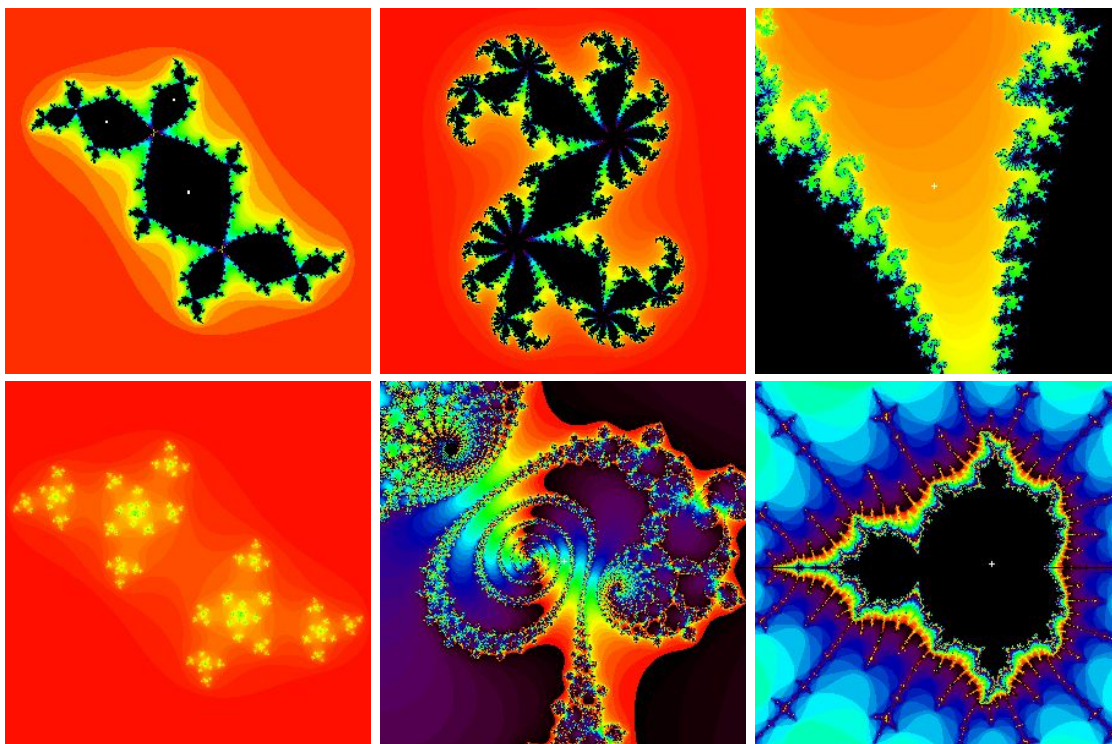
FIGURE 1.17. From top left to bottom right we see a Rabbit (Julia set when $c = -0.12 + 0.75i$), a Dragon (Julia set when $c = 0.36 + 0.1i$), Elephants (zoom in on the Mandelbrot set centered at $c = -0.77 + .173i$), a Star cluster (Julia set when $c = -0.4387 + 0.784i$), a Galaxy (zoom in on the Mandelbrot set centered at $c = -0.75623053 + 0.06418323i$), and a baby Mandelbrot set (zoom in on the Mandelbrot set centered at $c = -1.625$ and adjust the **Parameter plane max iterations** to 150 for a better resolution).

in mind that what we see on the computer screen is only an approximation to the real thing. ***Try it out!***

The star cluster set $J(f_{0.21+0.64i})$ and other such hard to see though infinite sets of points are what we call *Cantor dust* sets. It is hard to see such sets depicted with an applet because each point in the set is ***disconnected*** from any other point, that is, for any two points z_0 and w_0 in the set, there is a simple closed curve which never meets the Cantor dust set, but such that the curve winds around z_0 without winding around w_0. Such a set is called ***totally disconnected*** because the only connected subsets are single points. We explore one method to prove that certain Julia sets have this property in Section 1.5.2. However we first use our applet to illustrate many examples of this phenomenon.

EXPLORATION 1.107. Experiment with the *Parameter Plane and Julia Set Applet* to find many (apparent) examples of Cantor dust Julia sets $J(f_c)$. Can you make a conjecture about which c values correspond to such Julia sets? ***Try it out!***

1.5.2. Connectedness Locus. One of the many properties of M that you may have observed in your experimentation is that for each $c \in M$, the Julia set $J(f_c)$ is connected, and for each $c \notin M$, the Julia set $J(f_c)$ is disconnected.[19] Move the c value around the parameter plane in the *Parameter Plane and Julia Set Applet* to observe this. This is not a coincidence, and it is because of this that we call M the **connectedness locus** for the family of maps $\{f_c : c \in \mathbb{C}\}$, i.e., $M = \{c \in \mathbb{C} : J(f_c) \text{ is connected}\}$. This fact is proven by using the following theorem and noting that f_c has its sole finite critical point at the origin.

THEOREM 1.108 (See [**1**], p. 202). Let f be a polynomial of degree greater than or equal to two. Then every *finite* critical point of f has a bounded orbit if and only if both $J(f)$ and $K(f)$ are connected.

The proof of this result is beyond this text, but for the maps f_c we will be able to illustrate the reasoning behind the following related dichotomy.

THEOREM 1.109 (See [**2**], p. 67). For the map f_c, either
  (i) $\{f_c^n(0)\}_{n=1}^{\infty}$ is bounded in $\mathbb{C}$ and $J(f_c)$ is connected, or
  (ii) $f_c^n(0) \to \infty$ and $J(f_c)$ is totally disconnected.

We illustrate the method of proof, though a few technical details requiring advanced methods must be left for the interested reader to find in the literature. We first make an important definition.

DEFINITION 1.110. A *compact topological disk (ctd)* is a compact set[20] in $\mathbb{C}$ whose boundary is a simple closed smooth path.

The importance of this definition comes from the following lemma and from noting that any ctd is connected.

LEMMA 1.111. If E is a ctd and $c \notin \partial E$, then $f_c^{-1}(E)$ consists of
  (1) one ctd containing 0, if $c \in \text{Int}(E)$,
  (2) two ctd's, neither containing 0, if $c \notin E$.

We do not prove this lemma, but with the help of our applet we will illustrate it as it is used in the following.

---

[19]Your informal understanding of connectedness will suffice in this text; however, the following is a formal definition. A compact set E in $\mathbb{C}$ is **disconnected** if there exists a simple closed curve in $\mathbb{C} \setminus E$ which winds around some points of E, but not all points of E.
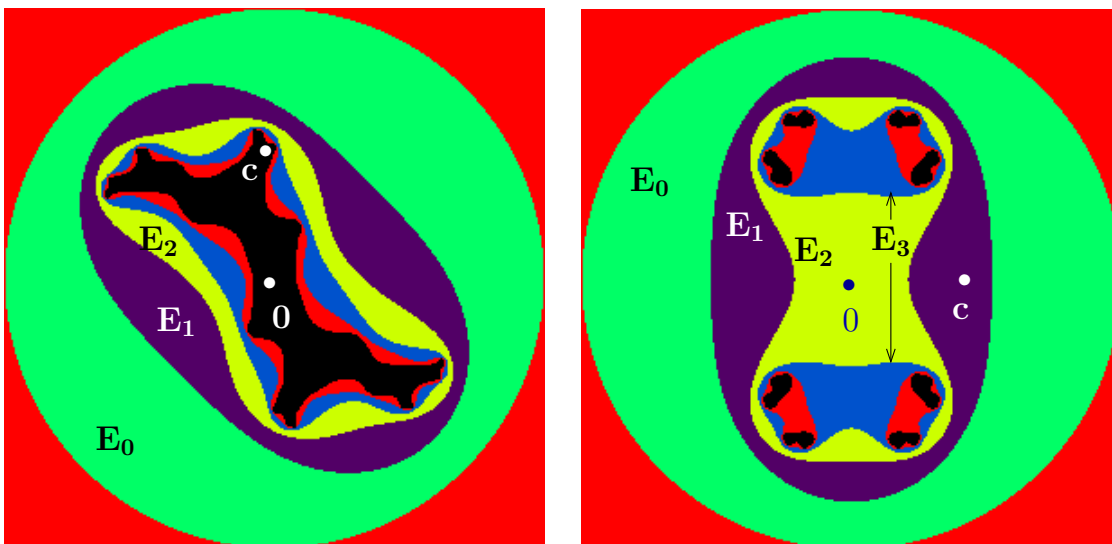[20]Recall that a compact set in $\mathbb{C}$ is a closed and bounded set.

FIGURE 1.18. The sets $E_0, \ldots, E_5$ drawn using the *Parameter Plane and Julia Set Applet* with Dynamic plane escape radius $r = 2.1$, Dynamic plane max iterations $= 5$, Color sample rate $= 7$. In both pictures $E_0 = \overline{\Delta(0, r)}$. Left picture shows $c = i$ with each E_n connected. Right picture shows $c = 0.9$ with each E_0, E_1 and E_2 connected, but subsequent E_n disconnected.

SKETCH OF PROOF OF THEOREM 1.109. Choose some $r > \max\{2, |c|\}$ with the property that $|f_c^n(0)| \neq r$ for all $n = 1, 2, \ldots$. Now define $E_0 = \overline{\Delta(0, r)}$ and define $E_n = \{z \in \mathbb{C} : f_c^n(z) \in E_0\}$ for each $n = 1, 2, \ldots$. Thus $E_n = f_c^{-1}(E_{n-1})$ for each $n = 1, 2, \ldots$ and $c \notin \partial E_n$ for each $n = 0, 1, 2, \ldots$. From Additional Exercise 1.189, we know that $|z| > r$ implies both $|f_c(z)| > |z| > r$ and $f_c^n(z) \to \infty$. Armed with this information the reader can now provide the details to show $E_0 \supset E_1 \supset E_2 \supset \ldots$ and $K(f_n) = \cap_{n=0}^{\infty} E_n$. We now investigate cases (i) and (ii) of the theorem by analyzing the sets E_n.

Suppose $\{f_c^n(0)\}_{n=1}^{\infty}$ is bounded in $\mathbb{C}$, i.e., $0 \in K(f_c)$. Since $c = f_c(0)$, we must also have $c \in K(f_c)$, and so $c \in E_n$ for all $n = 0, 1, 2, \ldots$. Since $c \notin \partial E_n$, we may then also conclude $0 \in \text{Int}(E_n)$ for all $n = 0, 1, 2, \ldots$. Applying Lemma 1.111 inductively we may then show that each E_n is connected as it consists of a single ctd (see Figure 1.18 (left)). Though we omit the details, it follows from a more advanced topological argument that $K(f_c) = \cap_{n=0}^{\infty} E_n$ must also be connected, which by another advanced topological argument implies $J(f_c) = \partial K(f_c)$ is connected.

Now suppose $f_c^n(0) \to \infty$, i.e., $0 \notin K(f_c)$. By the choice of r we know $c \in \text{Int}(E_0)$. However, since 0 does not lie in every E_n (else 0 would lie in $K(f_c)$), there is a positive integer m such that $c \in \text{Int}(E_{m-1})$, but $c \notin E_m$. Thus by Lemma 1.111 we see that E_m consists of one ctd, but $E_{m+1} = f_c^{-1}(E_m)$ consists of two ctd's (see Figure 1.18 (right)

where $m = 2$). Applying Lemma 1.111 to each ctd in E_{m+1}, we then see that E_{m+2} consists of four ctd's. Proceeding inductively we see that E_{m+k} must then consist of 2^k ctd's. Although we do not have the tools here to prove it, the size of the ctd's which make up E_{m+k} shrinks to zero as $k \to \infty$, and so we can conclude that any two distinct points z and w in $K(f_c)$ must lie in distinct ctd's of E_{m+k} when k is sufficiently large. Hence the boundary of such a ctd containing z, which lies in $A_{f_c}(\infty)$ (why?), must then be a curve which encloses z but not w. Hence we conclude that $K(f_c)$ is totally disconnected, and thus so is $J(f_c)$ since by utilizing Remark 1.65 one can show in this case that $J(f_c) = \partial K(f_c) = K(f_c)$. □

So, modulo a few details, we were able to understand how Cantor dust sets naturally arise when considering $J(f_c)$. The following is another topological property which you may have observed in your explorations; however, in this case we have all the tools to completely prove this property holds. This result says that $K(f)$ has no *holes*, where we define a **hole** as a bounded domain U in $\overline{\mathbb{C}} \setminus K(f)$ such that $\partial U \subseteq K(f)$.[21] We prove this as follows by appealing directly to the definition of $K(f)$ and applying the Maximum Modulus Theorem.

LEMMA 1.112. *Let f be a polynomial of degree greater than or equal to two. Then $K(f)$ has no holes.*

PROOF. Suppose the set U is a hole in $K(f)$. Let $R > 0$ be such that $|z| > R$ implies $|f(z)| > |z| > R$ (such an R must exist given that the map f is a polynomial of degree greater than or equal to two). From this, one can show $f^n(z) \to \infty$ whenever $|z| > R$. Thus we see that we must have $|f^n(z)| \leq R$ for all $n \in \mathbb{N}$ for any point z with a bounded orbit. This means, by definition of $K(f)$, we have $K(f) = \{z \in \mathbb{C} : |f^n(z)| \leq R \text{ for all } n \in \mathbb{N}\}$. Since $\partial U \subseteq K(f)$, we see that for any $n \in \mathbb{N}$, the polynomial f^n is bounded by R on ∂U. By a version of the Maximum Modulus Theorem given as Corollary A.18 on page 414, we then conclude that f^n is bounded by R on U as well. Since this works for all $n \in \mathbb{N}$, we see that $U \subseteq K(f)$ contrary to our assumption that U is a hole in $K(f)$ (which does not meet $K(f)$). This contradiction proves the result. □

1.5.3. Self-similarity and symmetry. One of the properties you may have noticed in each dynamical plane picture for the maps f_c is that each picture is symmetric about the origin. Specifically, each set $A_{f_c}(\infty)$, $J(f_c)$, and $K(f_c)$ has the property that it contains z if and only if it also contains $-z$ (see Additional Exercise 1.185).

Another property you may have noticed in each set $J(f_c)$ is that small parts of $J(f_c)$ look very much like other larger parts of $J(f_c)$. We call this property *self-similarity* and note that it actually is a property of all Julia sets.

---

[21]The advanced reader will recognize the absence of holes in $K(f)$ to be equivalent to the set $\overline{\mathbb{C}} \setminus K(f)$ being connected. However, to formally prove this equivalency we would need to delve more deeply into some more advanced topological results, and so we do not undertake such a proof here.

EXPLORATION 1.113. In the *Parameter Plane and Julia Set Applet*, set $c = 0.112 + 0.74667i$ and in the dynamic plane (z-plane) window set the x range to be 0.22026 to 0.28746 and set the y range to be 0.4648 to 0.532 (and then hit the **Update** button). Notice how similar the picture looks when we zoom in further by setting the x range to be 0.2277184 to 0.2475200 and setting the y range to be 0.50166592 to 0.52146752. Continue zooming in on any point of $J(f_c)$ to see how at all scales (i.e., depth of zoom) the picture, after rotating, looks very much like the first picture. (Keep in mind that you can click on the thumbnails at the bottom to see any previously created pictures.) Now repeat this with a few other c values to get a sense that this is a general property of Julia sets. ***Try it out!***

How about the Mandelbrot set M? Is it self-similar too? The answer is a definite...sort of. There are places in M which have small pieces that look like larger pieces, but it is not the case, as in the Star Cluster sets above, that the whole set M looks like a bunch of small copies of just one piece of itself. The Mandelbrot set is sometimes called *quasi*-self-similar for this reason.

EXPLORATION 1.114. In the *Parameter Plane and Julia Set Applet* set $c = 0.41491386 + 0.60134804i$ and begin zooming in on the Mandelbrot set, centering the zoom at this c value. What you will see are pictures that appear to look the same as you zoom in. In fact, if you zoom in on the tips of any of the "antennae" of the Mandelbrot set you will see a type of self-similarity. Now see what you find when you repeatedly zoom in on a point in the middle of such an antenna. ***Try it out!***

Now zoom in on the point $c = -1.2418406 - 0.32366967i$ to find what we call a *baby Mandelbrot set*. Set the **Parameter plane max iterations** to a higher value (such as 200 or 300) to see this picture better. You can also adjust the **Color sample rate** to adjust the color scheme to help create a nicer picture. What you have found is not an exact copy of the full Mandelbrot set, as you can tell by the long antennae coming out of it, but it certainly does have the unmistakeable look of M. It turns out, due to a deep result, that these baby Mandelbrot sets are actually ***dense*** in the boundary of M, that is, given any neighborhood of a point in ∂M, no matter how small, there exists a baby Mandelbrot set in that neighborhood. Of course, to see it, we usually have to zoom in pretty far, but it's there. This phenomenon, and more, is further discussed in Section 1.7.

EXPLORATION 1.115. Try zooming in on an arbitrary point in the ∂M, and then see if you can find a baby Mandelbrot set hiding in there. Keep in mind, however, that you might need to adjust the **Parameter plane max iterations** to give you a better view. Also, keep in mind that using a computer means there are limitations to how far we can zoom in and still get reasonable results. ***Try it out!***

1.5.4. Bulbs in M. Let's now try to systematically classify some of what we see in M, in terms of both geometry and dynamics. Recall that each point c in the parameter plane (also called the c-plane) corresponds to a function f_c whose Julia set

$J(f_c)$, Fatou set $F(f_c)$, and dynamics are viewed in the dynamic plane (also called the z-plane). With this in mind we will see that M, and all the parts which make it up, provide us with a sort of "dictionary" or index of the types of dynamics we may find for maps of the form $f_c(z) = z^2 + c$.

First we point out that we have already seen that the cardioid K_1 (see Figure 1.16) consists of c values corresponding to maps with attracting fixed points. Off this main cardioid K_1 there are infinitely many "bulbs" attached which we would like to understand both by their geometrical properties as sets in the c-plane and by the corresponding dynamic properties in the z-plane.

The most prominent bulb off K_1 is the disk $K_2 = \triangle(-1, 1/4)$ representing the parameters corresponding to maps with attracting 2-cycles. The next largest bulb off K_1 is near the top. Select $c = -0.12 + 0.77i$ from this bulb and consider the critical orbit in the dynamic plane to see that there exists an attracting 3-cycle. Use the *Parameter Plane and Julia Set Applet* to observe this by iterating the critical orbit one iterate at a time (by checking the **Show critical orbit** box, then checking the **Iterate orbits** box, and then hitting the **+** button that appears next to the **Iterate orbits** box). Also, note that for this c, the Julia set $J(f_c)$ in the dynamic plane is pinched in such a way that 3 bulbs meet at every pinch point. Let's call the pinch point which corresponds to the immediate basin of this attracting 3-cycle the ***main pinch point*** (which in this case is near $-0.282 + 0.492i$). Then by iterating the origin one step at a time you can see that each iterate makes roughly a 1/3 rotation around this main pinch point.[22] Remember that you can use the **Connect orbit points** checkbox to help you track the path of the orbit. Trying other c values from this same bulb in the c-plane we find that we always get this same type of behavior. For this reason we call this bulb in M the 1/3 bulb.

EXERCISE 1.116. Now use the *Parameter Plane and Julia Set Applet* to investigate the dynamics for $c = -0.513 + 0.5693i$. Pause for a moment to decide what fraction p/q would best describe the bulb containing this point. ***Try it out!***

In the dynamic plane we see an attracting 5-cycle, which cycles around a main pinch point at which 5 bulbs meet. Further, we see by iterating the origin one step at a time, the iterates make a 2/5 rotation about the main pinch point in each step. For this reason we call the bulb of M (in the c-plane) which contains $c = -0.513 + 0.5693i$ the 2/5 bulb. Again, you should experiment to see this same type of behavior for other c values within this 2/5 bulb. ***Try it out!***

DEFINITION 1.117. Let B be a hyperbolic component of K_q, whose boundary meets the boundary of the main cardioid K_1. We call B the p/q *bulb*, and denote it $B_{p/q}$, if for each $c \in B$ the corresponding map f_c has an attracting q cycle and each step in

---

[22]When referencing rotations, we use the standard convention that a positive rotation is counterclockwise. Hence, a 1/3 rotation means a 120 degree rotation (one-third of a full rotation) in the counterclockwise direction.

the critical orbit in the dynamic plane makes roughly a p/q rotation about the main pinch point.

From what follows it will become clear that for each rational number $0 \le p/q \le 1$, there is *only* one p/q bulb and so we are justified in calling $B_{p/q}$ ***the*** p/q bulb, as opposed to ***a*** p/q bulb.

EXERCISE 1.118. Experiment and prove a relationship between the p/q bulb and its conjugate bulb, that is, the bulb reflected over the x-axis (see Additional Exercise 1.182). In particular, find p' and q' if the conjugate of the p/q bulb is the p'/q' bulb. ***Try it out!***

REMARK 1.119. Theorem 1.80 shows that the multiplier map $\lambda_{p/q}$ maps $B_{p/q}$ conformally onto $\triangle(0,1)$. Furthermore, the map $\lambda_{p/q}$ extends to be a one-to-one continuous map of $\overline{B_{p/q}}$ onto $\overline{\triangle(0,1)}$.

It turns out that $\overline{B_{p/q}}$ meets the boundary of the main cardioid ∂K_1 in just a single point, which we denote $c_{p/q}$, and call the ***root of the*** p/q ***bulb***. For example, $c_{1/2} = -3/4$ is the root of the $1/2$ bulb $B_{1/2} = K_2$. In Additional Exercise 1.186 you are asked to investigate the multiplier maps evaluated at the root (note that since the root $c_{p/q}$ of the p/q bulb lies on the boundary of two hyperbolic components, namely K_1 and $B_{p/q}$, we see that there are two multiplier maps (namely, $\lambda : \overline{K_1} \to \overline{\triangle(0,1)}$ and $\lambda_{p/q} : \overline{B_{p/q}} \to \overline{\triangle(0,1)}$) that are defined at each $c_{p/q}$). These root points happen play a special role and so further understanding of them is now warranted. In particular, it is important to understand the following dynamic property.

EXERCISE 1.120. The root $c_{p/q}$ of the p/q bulb is the c parameter for which f_c has an indifferent fixed point with multiplier exactly equal to $e^{2\pi i p/q}$, i.e., $c_{p/q} = c(e^{2\pi i p/q})$ where $c(\lambda)$ is the inverse of the multiplier map $\lambda : \overline{K_1} \to \overline{\triangle(0,1)}$. Since $c_{p/q}$ is a bifurcation parameter, you should describe three dynamical changes (as done in Exercise 1.84) which occur as the parameter c moves from the main cardioid K_1 into $B_{p/q}$ passing through the root $c_{p/q}$. ***Try it out!***

EXPLORATION 1.121. Explore and label other bulbs in the same way as done in Exercise 1.116. We recommend that you print a large image of M and label each bulb as you go. ***Try it out!***

While doing Exploration 1.121 you may have stumbled upon an interesting pattern that shows how one can quickly compute p/q for the largest bulb (measured by area) between two bulbs. We explain here the pattern we see, but for the detailed proofs we direct the interested reader to [**6**]. Looking at M we see $B_{1/2}$ and $B_{1/3}$ each attached to the main cardioid, with infinitely many p/q bulbs in between decorating the boundary of the main cardioid. The largest such bulb, as we have seen, is $B_{2/5}$. We can correctly "do the math" quickly in this situation by what is known as Farey addition to compute

$$\frac{1}{2} \oplus \frac{1}{3} = \frac{2}{5}.$$

In Farey addition, one "adds" fractions in the *unusual* way of simply adding the numerators and adding the denominators (much easier than that arduous task which involves finding a common denominator!). We call the two addends the **Farey parents** (e.g., $\frac{1}{2}$ and $\frac{1}{3}$) and the resulting fraction the **Farey child** (e.g., $\frac{2}{5}$). Correspondingly, we will call the bulbs $B_{1/2}$ and $B_{1/3}$ the Farey parents of the Farey child bulb $B_{2/5}$.

EXAMPLE 1.122. The Farey child of $B_{2/5}$ and $B_{1/3}$, that is, the largest bulb between these two, is $B_{3/8}$ since

$$\frac{2}{5} \oplus \frac{1}{3} = \frac{3}{8}.$$

A quick check with the *Parameter Plane and Julia Set Applet* will confirm this result. Remember to use the **Connect orbit points** checkbox to help you track the path of the orbit. **Try it out!**

REMARK 1.123. We must be careful with our use of Farey addition of bulbs. The main rule we must be sure to adhere to is: **Two bulbs can only be Farey parents if all the bulbs between them are smaller than they are.** For example, one can check that the largest bulb between the 4/11 bulb (the one containing $c = -0.292 + 0.633i$) and the 2/5 bulb is the 3/8 bulb. Here the Farey addition clearly does not work. However, since the supposed child 3/8 bulb is clearly larger than the supposed parent 4/11 bulb (which the reader should check on the applet), we know that Farey addition is not applicable in this instance.

There is another issue to be dealt with if this Farey addition is to truly help us compute the bulb fraction for all the p/q bulbs. Between the $B_{1/3}$ and the cusp of the cardioid K_1 there is a largest bulb, but how can one use Farey addition to determine it? The key is to treat the cusp itself like a Farey parent which is larger than all other bulbs. We leave it to the reader in Additional Exercise 1.188 to experimentally determine what Farey fraction should be used to represent the cusp.

1.5.5. Sub-bulbs of M. Just as the Main Cardioid K_1 of M has many p/q bulbs attached to it, so does each p/q bulb have many "sub-bulbs" attached. Let's investigate these small sub-bulbs. Use the *Parameter Plane and Julia Set Applet* to view the change in dynamics as we let c vary from $c_1 = -0.16097811 + 0.80545706i$ within the 1/3 bulb to $c_2 = -0.17462417 + 0.8296561i$ in an attached smaller sub-bulb. We see that the attracting 3-cycle became an attracting 15-cycle. Note also how the picture of the Julia set with the attracting 15-cycle has features common to both Julia sets with an attracting 3-cycles and Julia sets with an attracting 5-cycle. We see that 3 bulbs of $K(f_{c_1})$ meet at each pinch point. In changing c to c_2, we see that these pinch points from $K(f_{c_1})$ persisted, but also, within each bulb of $K(f_{c_1})$, new pinching occurs, such that at each of these newly formed pinch points 5 different bulbs of $K(f_{c_2})$ meet. See Figure 1.19 and also take some time to use the zooming features in the *Parameter Plane and Julia Set Applet* to see this pattern repeat at all scales (remembering to adjust the **Dynamic plane max iterations** value as needed).
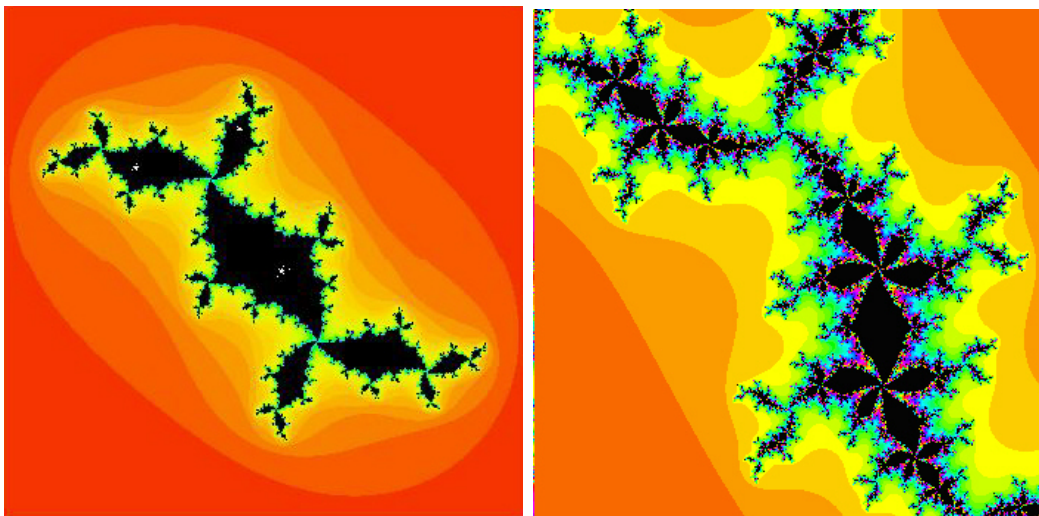
FIGURE 1.19. Left shows $J(f_{c_1})$ and right shows $J(f_{c_2})$, slightly magnified to make it easier to see.

SMALL PROJECT 1.124. Investigate the behavior witnessed above by entering into a variety of sub-bulbs attached to a variety of p/q bulbs. Can you discern a pattern? If shown an example of a Julia set such as the one in Figure 1.20 below, can you identify which sub-bulb the c value came from?

1.5.6. Limbs and Antennae in M. It turns out that the Mandelbrot set M is a connected set. Your intuitive notion of what connected means will suffice in this chapter. However, we can also more formally describe the notion in this case by saying that any simple closed curve in the complement of M must either wind around no point of M or must wind around all of M. The fact that M is connected is not easy to prove and so we shall not attempt to do so here, but we will avail ourselves of this fact in order to describe some other important aspects of M. In particular, we make use of the fact that we can naturally "disconnect" M into two connected pieces by removing any root $c_{p/q}$ of a p/q bulb (a fact we also will not prove here).

DEFINITION 1.125. The set $M \setminus \{c_{p/q}\}$ consists of two connected sets, one which contains the cardioid K_1 and the other containing $B_{p/q}$, which we call the p/q *limb*.

On each p/q limb there is the main p/q bulb with infinitely many sub-bulbs attached. Each of these sub-bulbs in turn has an infinite number of tinier sub-sub-bulbs, and so on. However, the p/q limb contains more than just these bulbs. It also contains what we informally call ***antennae*** made out of thin filaments that, sort of, reach out from the p/q bulb. Looking straight above the 1/3 bulb there is a junction where 3 equally spaced filaments, which we informally call ***spokes***, meet (see Figure 1.21). The ***main*** or ***principal spoke*** is the one which attaches to the 1/3 bulb, and the
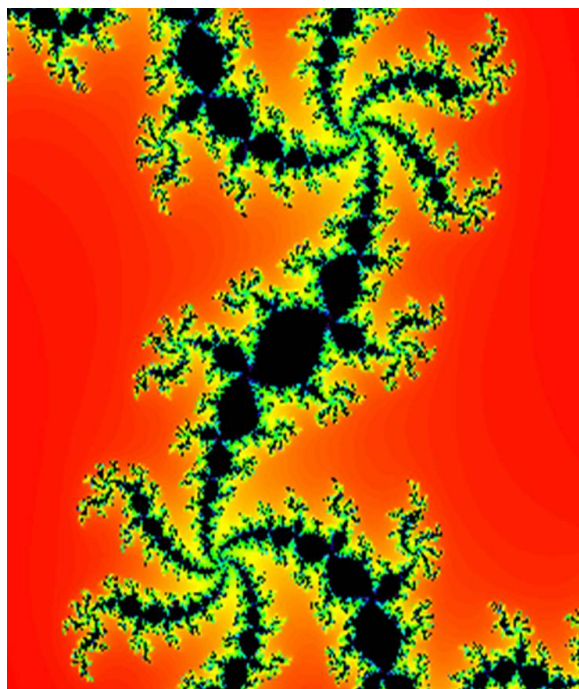
60

FIGURE 1.20. How can you tell where to find the corresponding c parameter for this picture of $J(f_c)$?

shortest spoke is a $1/3$ rotation about the junction point from the main spoke (and if you were wondering, yes, the $1/3$ rotation here is, in fact, related to the $1/3$ designation of the bulb). It turns out that this relationship between the short spoke and the p/q value occurs quite frequently, but not necessarily always (e.g., examine $B_{1/5}$ which contains the point $c = 0.39 + 0.33i$ using the *Parameter Plane and Julia Set Applet*). This generalization has been made into a formal theorem, but to do so, the notion of "shortest" had to be changed slightly (see [**6**]). We also note that if you zoom in on other junctions found on the antennae on the $1/3$ limb, you will very often (but not always) see 3 spokes meeting at the junction. This type of (more than) coincidence cries out for further exploration indeed. First we make a definition.

DEFINITION 1.126. When q spokes meet at a junction point, we will call this a junction point of *order q*.

EXPLORATION 1.127. Investigate other p/q limbs and their principal spokes to get an idea of how often the spoke that is a p/q rotation from the main spoke is the shortest. Is there a similar pattern for the longest spoke? ***Try it out!***

EXPLORATION 1.128. Investigate other p/q limbs and the junction points within them to get an idea of how often the junctions there are of order q. What is the order
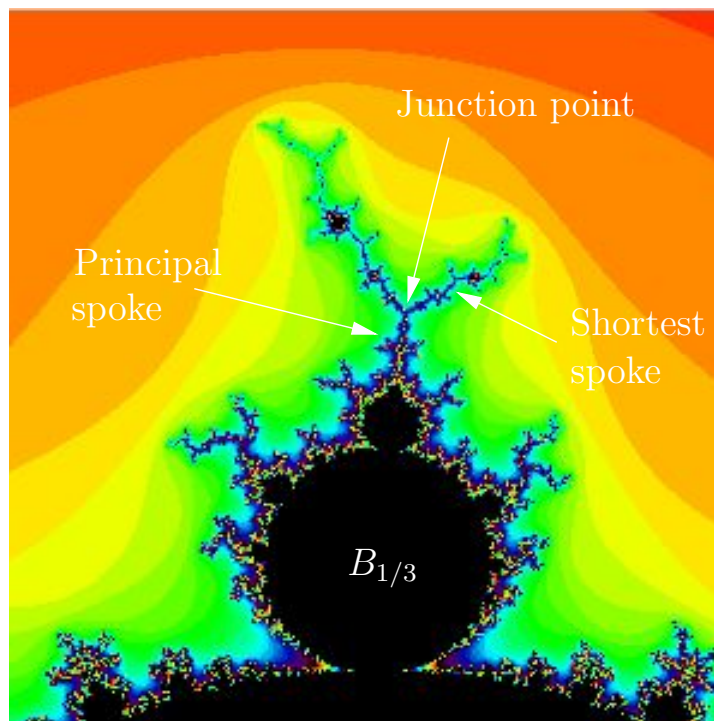
FIGURE 1.21. Illustration of main junction point of $1/3$ limb of M.

of the other junctions? Is there a pattern to the types of junction orders one can find on a given p/q limb? **_Try it out!_**

In the course of exploring the p/q limbs you may have stumbled upon many baby Mandelbrot sets like the one in Figure 1.22. In order to see these baby Mandelbrot sets clearly you will usually want to adjust the **Parameter plane max iterations** value to 300, 400, 500, or even 1,000 or more depending on the size of baby Mandelbrot set you want to see. Also, adjusting the **Color sample rate** can help. We mentioned before that these baby Mandelbrot sets are everywhere (to be more precise, they are dense in ∂M); however, even though they are all unmistakable "copies" of the original M set, they are not all the same. They each have different decorations (antennae sprouting from them). If we pay close attention to these antennae and how many meet at the various junction points, can we get an idea how the given baby Mandelbrot set is related to the p/q limb it lives in? For example, in Figure 1.22 we see that the junction points near the "tips" of the antennae are all of order 3. Furthermore, we see that along these filaments, closer to the baby Mandelbrot set, we have junction points of order 5. Can this information be a clue to help you find where this baby Mandelbrot set lives? Imagine playing a game where your friend shows you a picture of a baby Mandelbrot set. Can you win the game by telling your friend where it lives? Note: These are very hard questions, but maybe by investigating these you will be able

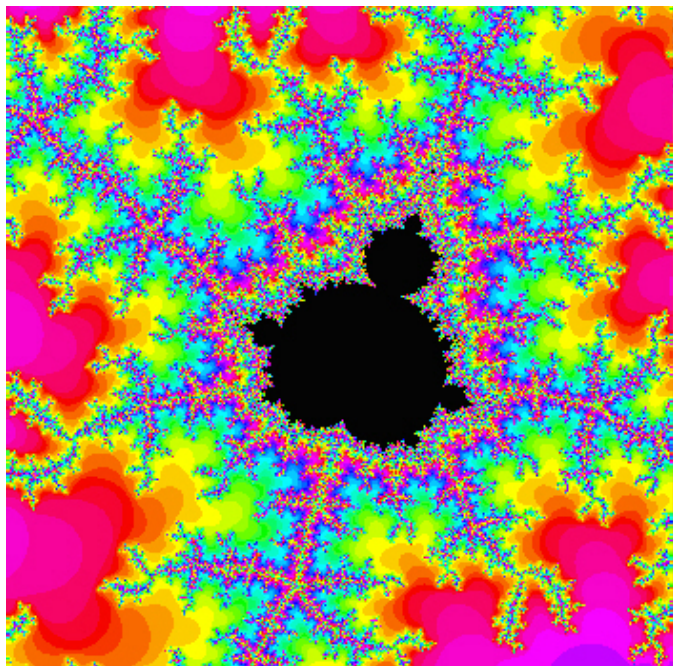to come up with some partial answers, connections, or maybe pose some interesting questions of your own.

FIGURE 1.22. Which p/q limb contains this baby Mandelbrot set? Off of which sub-bulb of $B_{p/q}$ does it live?

In your investigation of the antennae, bulbs, and limbs of M you may have also noticed that some of the parameter plane pictures (specifically enlarged areas of the tips of the antennae of M) look very much like some of the dynamic plane pictures (specifically enlarged areas of the tips of certain Julia sets). In fact, these pictures can look so much alike that in can be confusing which is which (see Figure 1.23). This general phenomenon was proven in [20] where it was shown that zooming in the parameter plane near a Misiurewicz point c (see Definition 1.98) will show a portion of the Mandelbrot set which is a rotation of an enlargement of the Julia set of f_c in the dynamic plane near c. This fact is quite curious given that the parameter plane and the dynamic plane are, on the face of things, really different animals.[23] Why do parts of the Mandelbrot set seem identical to parts of certain Julia sets? The proof is too complicated for these pages, but we will be sure to witness and admire examples of it. Use the *Parameter Plane and Julia Set Applet* to zoom in very far on the point $c = 0.4711819 + 0.3541484i$ in **both the parameter plane and the dynamic plane**

---

[23]This reminds the author of a comic strip which said, in paraphrased form, "Most mathematical discoveries are not accompanied by a shout of 'Eureka!', but rather a quietly spoken 'Huh, now that's curious.'"

(note that in the dynamic plane, c is the the first point in the critical orbit). Then compare the portion of M near c to the portion of $J(f_c)$ near c. Up to a rotation it seems that these images are indeed identical.
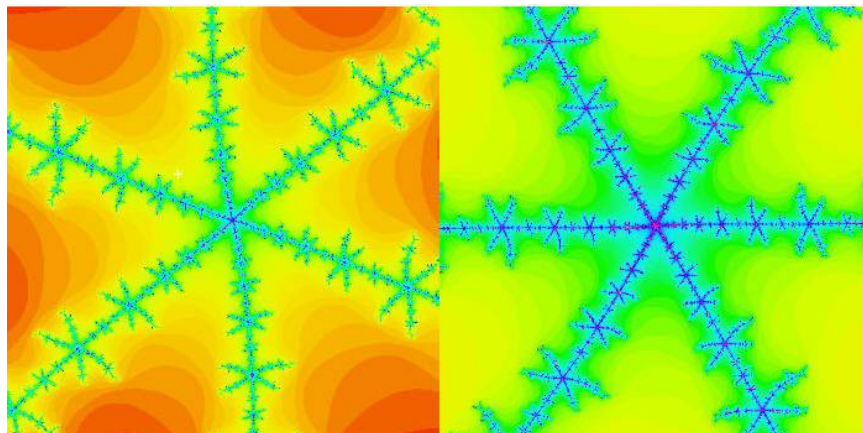

FIGURE 1.23. Which is part of the Mandelbrot set and which is part of a Julia set?

1.5.7. Fixed points for $f_c(z) = z^2 + c$. In this section we make a few remarks regarding the dynamics near the fixed points of $f_c(z) = z^2 + c$.

EXAMPLE 1.129. Using the *Parameter Plane and Julia Set Applet* set $c = -0.513 + 0.5693i$ in the parameter plane and then choose seed $z_0 = -0.4034 + 0.2896i$ (very near the main pinch point) in the dynamical plane. Zoom in close to the fixed point (with the **Plot fixed points** box, you can color these purple to more easily identify them) and then iterate this seed value one step at a time and observe the behavior. Try iterating several other seeds near the pinch point to see the same behavior. From this behavior we see that the main pinch point is a repelling fixed point whose multiplier λ has argument very close to $(2/5) * 2\pi$.

EXERCISE 1.130. Using paper and pencil (along with your calculator/computer) solve for the fixed points of f_c where $c = -0.513 + 0.5693i$ as in Example 1.129. Compute their multipliers and verify that the "pinch point" multiplier λ does indeed have argument very close to $(2/5) * 2\pi$. Also, use the *Complex Function Iterator Applet* or *Global Complex Iteration Applet for Polynomials* to iterate the map $z \mapsto \lambda z$ for various seed values near the repelling fixed point at the origin. Compare the dynamics of $z \mapsto \lambda z$ near the origin and $z \mapsto f_c(z)$ near the "pinch point". ***Try it out!***

EXPLORATION 1.131. Try investigating other c values in the 2/5 limb to see if it is always the case that the repelling fixed point has a multiplier with argument very close to $(2/5) * 2\pi$. You can do this experimentally by checking the behavior of orbits which start near the repelling fixed point (use the **Plot fixed points** feature in the

Parameter Plane and Julia Set Applet to locate the fixed points). Relate the argument of the multiplier to the dynamics near the repelling fixed point, specifically, note how this argument corresponds to the spiralling we see in $K(f_c)$ near the repelling fixed point. How would this behavior change in other p/q limbs? ***Try it out!***

Above we witnessed the role of the fixed point of f_c which also serves as a main "pinch point" in the set $K(f_c)$. What can you observe about the other fixed point? In what way does it appear to be different from the main "pinch point"? Does it have a special dynamic role? Can you prove anything or make a conjecture about this fixed point?

1.5.8. Concluding remarks about the family of maps $f_c(z) = z^2 + c$. Though we have witnessed and discussed many aspects of the Mandelbrot set there are many more things to know. We encourage the reader to delve deeper into the topics we discussed here and certainly pursue your own line of questions. There are many fascinating things left to discover in this infinite playground of mathematics. ***Go enjoy it!***

We end this section with a remark about one of the goals we stated in the introduction to this chapter. This goal is to introduce, investigate, and understand what we can about chaotic dynamical systems. Systems involving the weather, the stock market, or the motion of all the heavenly bodies in the Milky Way galaxy are easy to believe to be chaotic. There are so many uncontrollable variables in these systems, and changing one, even just a little, will impact all the other variables. However, what we see with this family of maps $f_c(z) = z^2 + c$ is also chaos. But with the family $f_c(z) = z^2 + c$, chaos ensued without a hundred unknown variables or any random processes. This simple system of one complex variable and one complex parameter turns out to produce chaotic behavior of an unimaginably wide and rich sort. This gives us small taste of what is truly a global phenomenon – chaos is all around us, even in what seems like the simplest of systems. For a wonderful and non-technical layperson's look at some history of chaos, chaos theory, and the chaoticians who investigate it, we recommend the book *CHAOS: MAKING A NEW SCIENCE* by James Gleick [**15**].

1.5.9. Other Uni-critical families of polynomials. In this section we investigate the dynamics of polynomials of the form $P_c(z) = z^d + c$ where $d = 2, 3, 4, \ldots$ (noting that we have already extensively studied the $d = 2$ case). Since these maps also have only one finite critical point (at the origin), we can analyze the parameter space of these maps in much the same way as we did for the maps $f_c(z) = z^2 + c$. In particular, we define the *d-th degree Mandelbrot set* to be $M_d = \{c \in \mathbb{C} : P_c^n(0) \nrightarrow \infty\}$, where it is understood that polynomial P is of fixed degree d. As we were told for the Mandelbrot set, the condition that the critical orbit $\{P_c^n(0)\}$ tends to ∞ is equivalent to the condition that some point in this orbit escape $\overline{\triangle(0, 2)}$. Steps to prove this, and more, are given in Additional Exercise 1.189.

The reader is encouraged to explore the dynamics of each such family of maps using the *Parameter Plane and Julia Set Applet*. Where the function $z^2 + c$ appears, use the drop down menu to select $z^d + c$, and then enter in any integer $d > 2$. Immediately, you will notice that the parameter plane and the dynamic plane exhibit certain symmetries (see Additional Exercise 1.190). Furthermore, you will recognize M_d as the connectedness locus for this new family as well (see Additional Exercise 1.191).

LARGE PROJECT 1.132. Investigate these families of maps using the *Parameter Plane and Julia Set Applet* in the same way as as we did for the family $f_c(z) = z^2 + c$. Try to find patterns and relationships in the bulbs and antennae. Develop your own questions, make conjectures, and describe M_d.

1.6. Transcendental Dynamics

In this section we investigate the dynamics of three different families of maps $E_c(z) = ce^z$, $S_c(z) = c \sin z$, and $C_c(z) = c \cos z$, where again $c \in \mathbb{C} \setminus \{0\}$ is a parameter. We first consider the dynamic properties of these maps for fixed c, then we study the parameter plane, that is, we investigate what changes occur in the dynamics when the parameter c is varied. These functions all have one very striking difference from the maps $P_c(z) = z^d + c$ studied above. They are *transcendental entire maps*,[24] and as such they not only fail to have an attracting fixed point at ∞, they fail to even be defined at ∞. In fact, these maps have an *essential singularity* at ∞ which, among other things, means that they cannot be defined at ∞ in any continuous way (see Example B.12 on page 425). However, it turns out that ∞, or more specifically the basin of ∞, still plays a central role. In fact, it turns out that instead of the basin of ∞ being in the Fatou set like we have for polynomials, we have quite the opposite.

PROPOSITION 1.133. For each of the maps $E_c(z) = ce^z$, $S_c(z) = c \sin z$, and $C_c(z) = c \cos z$, where $c \in \mathbb{C} \setminus \{0\}$, the Julia set is equal to the closure of the attracting basin of ∞, e.g., $J(E_c) = \overline{A_{E_c}(\infty)}$.

The reader should pause for a moment to take in the striking difference asserted by this proposition between polynomial dynamics and the dynamics of the given transcendental maps. Though the proof of this is beyond the scope of this text, we will make use of this result (see [**5, 8**] as a general references on the dynamics of E_c). In particular, we use this fact to program the computer to attempt to illustrate the Julia sets of these maps. However, in order to do so, we must first understand how, or rather in what direction, a point z_0 can iterate to ∞ under each of these maps. We first investigate the map $E_c(z) = ce^z$.

REMARK 1.134. For those readers who are familiar with *Picard's Theorem* (see [**1**], p. 242), we give some idea why the Julia set of certain transcendental entire functions

---

[24]A *transcendental entire* map is defined to be a map which is analytic on all of $\mathbb{C}$ (entire), but which is not a polynomial.

is related to the closure of the points which iterate to ∞. Picard's Theorem says that given any entire map g with an essential singularity at ∞ and given *any* neighborhood U of ∞ (no matter how small), g will map $U \setminus \{\infty\}$ *infinitely often* onto the entire complex plane $\mathbb{C}$ minus at most one point. This means that with the exception of at most one point $w \in \mathbb{C}$, there are infinitely many points $z \in \mathbb{C}$ such that $g(z) = w$.[25] Hence, there must be a very high degree of sensitive dependence as some points very near ∞ must have very different orbits even in the first application of the map (let alone after repeated iteration of the map).

1.6.1. Exponential dynamics. In this section we investigate the dynamics of the complex exponential maps $E_c(z) = ce^z$. Recall that $E_c(z) = ce^z = ce^x e^{iy}$ where $x = \operatorname{Re} z$ and $y = \operatorname{Im} z$, and so $|E_c(z)| = |c|e^x$. Hence we see that $E_c(z)$ is very large, and thus close to ∞, when $x = \operatorname{Re} z$ is large. In particular, if $\operatorname{Re} z > 50$, then $|E_c(z)| > |c|e^{50}$ is *extremely* large indeed (for any c that is not so small that a computer would recognize it as 0). We use this to justify the algorithm implemented in the *Parameter Plane and Julia Set Applet*, for drawing $J(E_c)$, which you can access by selecting ce^z from the function drop down menu. In particular, a point in the dynamic plane is colored based on how many iterates it takes to "escape", by which we mean have its real part become greater than 50. Thus the colored points represent $A_{E_c}(\infty)$, which by Proposition 1.133, must visually look the same as the Julia set.[26] Points colored black do not escape, at least not after iterating the number of times set in the **Dynamic plane max iterations** box, and so these points represent the Fatou set. Since $E_c^n(z_0) \to \infty$ if and only if $\operatorname{Re} E_c^n(z_0) \to +\infty$ (see Additional Exercise 1.192), we say that points that iterate to ∞, do so *in the direction of the positive real axis*.

EXAMPLE 1.135. We show that for $c = 0.2$ the Julia set $J(E_{0.2})$ is what we will call a *Cantor Bouquet*. We note here that our proof will not make use of the applet, nor of its algorithm, but will employ only the notion of sensitive dependence on initial conditions. However, we will find it useful to get a visual by first viewing the Julia set as drawn by the *Parameter Plane and Julia Set Applet* (see Figure 1.24).

Recalling that points in black do not escape and other points iterate to ∞ (in the direction of the positive real axis), we see that the colored set depicts $J(E_{0.2})$. It appears from the picture that $J(E_{0.2})$ contains large open sets, but it turns out that this is simply an artifact of our algorithm's inability to iterate infinitely many times. Were we able to set the **Dynamic plane max iterations** equal to infinity we would not find any open sets in $J(E_{0.2})$. We explain why in a moment, but first use the

---

[25]As an example, consider the map e^z which has such an exceptional value at $w = 0$, but also is easily seen to have infinitely many preimages of any point in $\mathbb{C} \setminus \{0\}$.

[26]According to the algorithm just described a point z may be identified to be in the Julia set when in truth it is not. However, one can show that if such were to occur, there would have to be a point *very* close, depending on the parameter settings of the applet, to z which does truly lie in the Julia set (see Additional Exercise 1.193). Hence for the purpose of creating a visual representation of the Julia set, this technical issue does not pose a serious concern.

zooming feature to see that what appears to be tiny "fingers" in $J(E_{0.2})$ are actually made of tinier fingers, which are themselves made of even tinier fingers and so on. ***Try it out!*** So we see self similarity again. We now explain why the picture of $J(E_{0.2})$ shows these fingers inside of fingers.
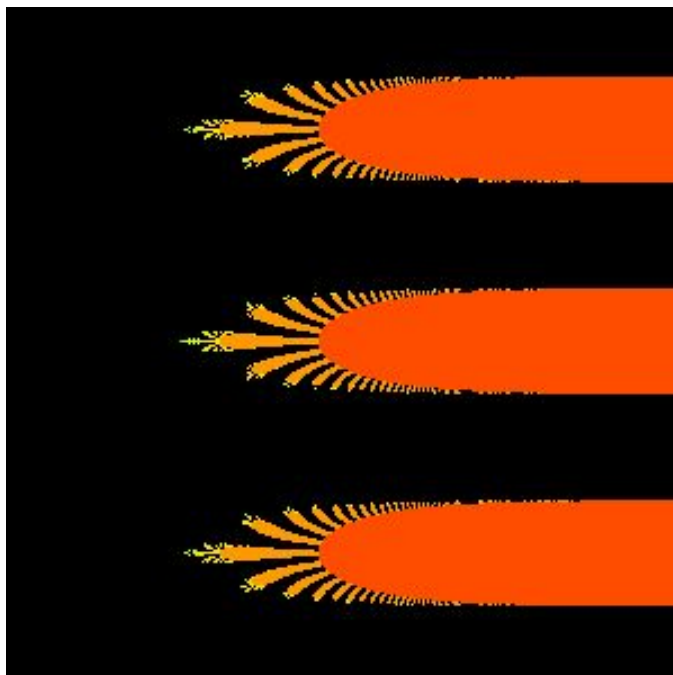

FIGURE 1.24. A portion of the Julia set of $E_{0.2}$ shown for $0 \le x \le 10$ and $-10 \le y \le 10$.

We begin by showing that $E_{0.2}$ has a real attracting fixed point. Using the Intermediate Value Theorem we note that $E_{0.2}$ has a fixed point p for some real value $0 < p < 1$ since the $E_{0.2}(x) - x$ is positive for $x = 0$ and negative for $x = 1$. Also, we see that p is attracting since $|E'_{0.2}(p)| < 1$.

We now proceed to show that the half plane $H = \{\operatorname{Re} z < 1\}$ is contained in the attracting basin $A_{E_{0.2}}(p)$. Set $\eta = |E'_{0.2}(1)| = 0.2e$ and note that $|E'_{0.2}(z)| < \eta < 1$ for all $z \in H$. Thus, for all $z \in H$, we have $|E_{0.2}(z) - p| = |E_{0.2}(z) - E_{0.2}(p)| = |\int_p^z E'_{0.2}(s)\,ds| \le \eta|z - p|$, where the straight line path is used in the integral. Hence the action of $E_{0.2}$ is to move points in H closer to p by a factor of at least η.[27] Formally using induction, along with the fact that $E_{0.2}(H) \subseteq H$ (verify), the reader can show that for all $z \in H$ we have $E_{0.2}^n(z) \to p$, i.e., $H \subseteq A_{E_{0.2}}(p)$.

---

[27]Using the *Parameter Plane and Julia Set Applet* with $E_{0.2}(z) = 0.2e^z$, the reader can see this contraction in action. In the Dynamic Plane window with $-2 \le \operatorname{Re} z \le 2$ and $-2 \le \operatorname{Im} z \le 2$ (which will be all black), pick any two points and iterate the map one step at a time. You can clearly see that after each step the new points are much closer together.

The technique just employed which shows how a bound on the derivative can be used to show "contraction" of a map is a general, and very useful, one. We interrupt our proof of the Cantor bouquet result in order to state this as a lemma for future use (whose proof is left to the reader in Additional Exercise 1.194).

LEMMA 1.136 (Contraction Lemma). Let f be analytic and map $convex^{28}$ domain D into itself. Suppose that $|f'(z)| < \eta < 1$ for all $z \in D$. Then f contracts distances by a factor of at least η on D, i.e., $|f(z) - f(w)| < \eta|z - w|$ for all $z, w \in D$. Furthermore, $|f^n(z) - f^n(w)| < \eta^n|z - w|$ for all $z, w \in D$. Hence, if $a \in D$ is a fixed point of f, then $D \subseteq A_f(a)$.

Returning to the proof that $J(E_{0.2})$ is a Cantor bouquet we now consider $E_{0.2}^{-1}(H) = \{z \in \mathbb{C} : E_{0.2}(z) \in H\}$. The reader should use the transformation properties of the exponential map to justify the picture in Figure 1.25 showing that $\mathbb{C} \setminus E_{0.2}^{-1}(H)$ consists of infinitely many components, which we call **fingers**. In particular, we label these fingers by C_k for each $k \in \mathbb{Z}$, noting that each C_k is just a translate by $2\pi i k$ of C_0. We also note that $H \subseteq E_{0.2}^{-1}(H)$.

We note that each C_k is mapped $conformally^{29}$ onto the half plane $\{\operatorname{Re} z \geq 1\}$, and thus each C_k must contain a preimage under $E_{0.2}$ of each C_j for all $j \in \mathbb{Z}$. These preimages, shown in red in Figure 1.26, are each sub-fingers of C_k with what we informally call **gaps** (in black) in between. We note that these gaps, just as in Figure 1.25, extend all the way to ∞ (in the direction of the positive x-axis).

We continue to take inverse images of H to see that, for any $n \in \mathbb{N}$, the set $E_{0.2}^{-n}(H)$ has a complement $\mathbb{C} \setminus E_{0.2}^{-n}(H)$ consisting of fingers inside of fingers inside of fingers, and so on. More precisely, we make the following definition.

DEFINITION 1.137. We call each component of $\mathbb{C} \setminus E_{0.2}^{-n}(H)$ a *stage n finger*.

For example, Figure 1.25 depicts stage 1 fingers in red and the bottom picture in Figure 1.26 depicts stage 2 fingers in red. The gaps (in black) are then portions of $E_{0.2}^{-2}(H)$ which separate the fingers from each other. We encourage the reader to investigate these stage n fingers more closely using the *Parameter Plane and Julia Set Applet* by setting both the **Dynamic plane max iterations** and the **Dynamic plane min iterations** to n, while setting the **Escape criterion** to $\operatorname{Re} z > 1$. Note that each stage n finger is contained in a stage $n-1$ finger (for $n \geq 2$). Also note that the boundary of each stage n finger (for $n \geq 2$) is mapped by $E_{0.2}$ onto the boundary of some stage $n-1$ finger and thus is mapped by $E_{0.2}^n$ onto the vertical line $\{\operatorname{Re} z = 1\}$.

It will be important to understand how "thick" these fingers can be and so we make the following definition.

---

[28]A set D is called **convex** if for any points $z, w \in D$, the line segment connecting z and w is a subset of D.

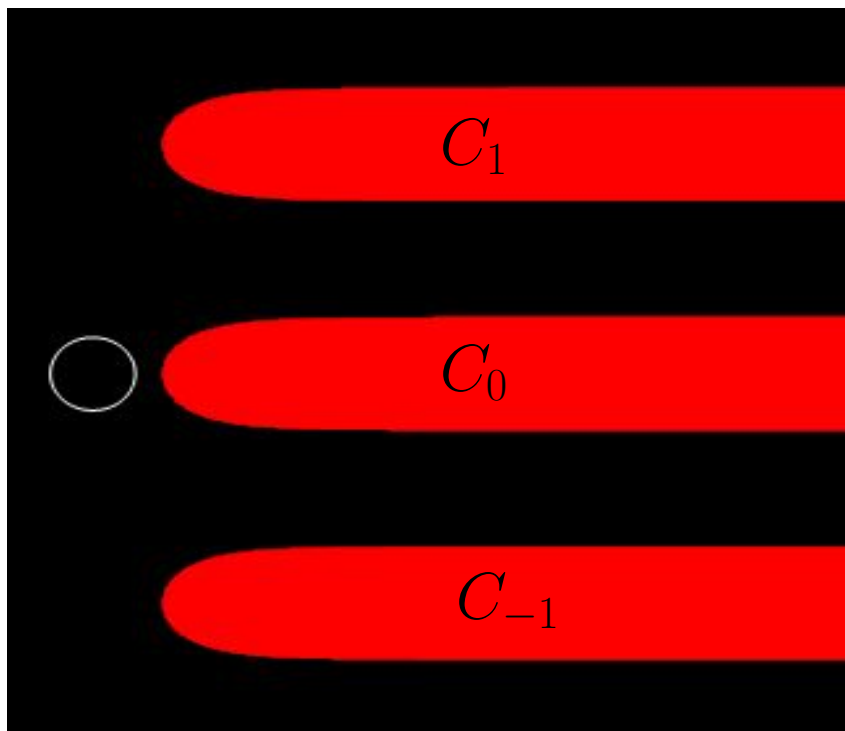[29]Recall, that a conformal map is one-to-one and analytic.

FIGURE 1.25. In this window of the complex plane, the set $E_{0.2}^{-1}(H)$ is colored in black with components of its complement in red for $-2 \leq x \leq 18$ and $-10 \leq y \leq 10$. For reference, the unit circle is also shown in white.

DEFINITION 1.138. Let F be a stage n finger. We define the *thickness* of F to be the special value of t such that F cannot contain an open Euclidean disk with diameter strictly larger than t, but for any value $s < t$, the set F does contain an open Euclidean disk with diameter s.[30]

EXERCISE 1.139. Use the mapping properties of the exponential map and the definition of stage 1 finger to prove that the thickness of each stage 1 finger C_k is π. ***Try it out!***

Note that the set $F_n = \mathbb{C} \setminus E_{0.2}^{-n}(H)$ is the union of all stage n fingers and because each stage n finger is contained in a stage $n-1$ finger we see that $F_1 \supset F_2 \supset \ldots$, and in particular, $F_1 \cap F_2 \cap \cdots \cap F_n = F_n$. Thus in some loose sense we can regard $\cap_{n=1}^{\infty} F_n$ as the union of all of the stage "infinity" fingers. It is these stage "infinity" fingers together with the point at ∞ that comprise the Julia set, a proposition we state as follows.

---

[30]Those familiar with the concept of supremum may recognize that $t = \sup\{2r > 0 : \Delta(z,r) \subseteq F \text{ for some } z \in F\}$, the *supremum* of the diameters all open disks contained in F.
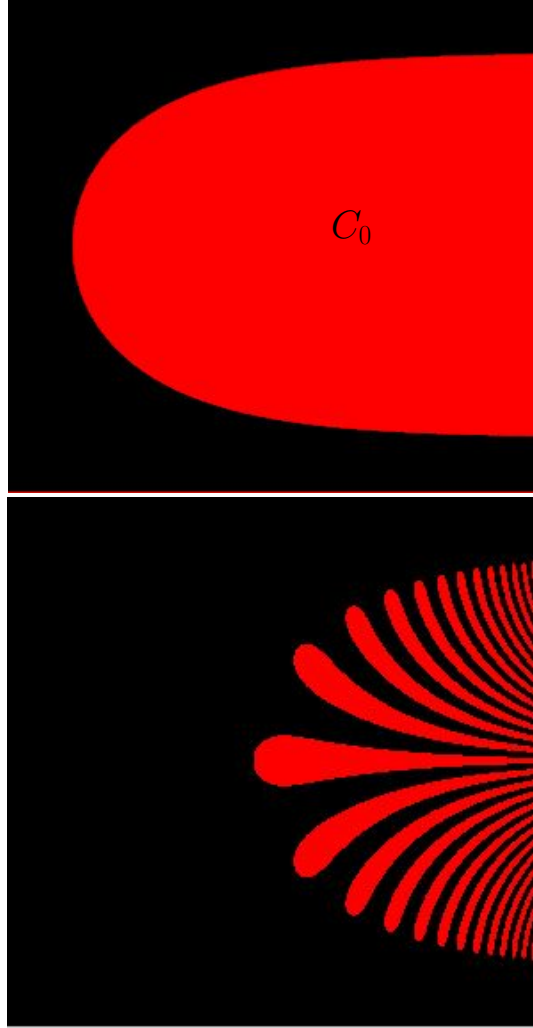
FIGURE 1.26. On top you can see the tip of the finger C_0 and on bottom you can see a portion of $E_{0.2}^{-2}(H) = E_{0.2}^{-1}(E_{0.2}^{-1}(H))$ for $1 \leq x \leq 6$ and $-2 \leq y \leq 2$. The red sub-fingers each sit well inside of C_0.

PROPOSITION 1.140. The Julia set $J(E_{0.2}) = \cap_{n=1}^{\infty} F_n \cup \{\infty\} = \overline{\mathbb{C}} \setminus A_{E_{0.2}}(p)$.

PROOF. We set $J = \cap_{n=1}^{\infty} F_n \cup \{\infty\}$ and outline the proof as follows. First, note the following facts, whose proof we leave to the exercises: (a) any two fingers are always separated by an infinitely long gap (of black) points in $A_{E_{0.2}}(p)$ (see Additional Exercise 1.195), and (b) the thickness of the stage n fingers shrinks to zero as n goes to infinity (see Additional Exercise 1.196). Facts (a) and (b) can then show that the set J cannot contain any open set (see Additional Exercise 1.197). Hence for any point $z \in J$ we see that there are are points arbitrarily close which lie in $\cup_{n=1}^{\infty} E_{0.2}^{-n}(H)$, and thus iterate to p. Since the orbit of every $z \in J$ lies forever in $\{\text{Re}\, z \geq 1\}$ by the

definition of J (and thus is quite different from an orbit which converges to p), we get that z is sensitive to initial conditions and so $z \in J(E_{0.2})$. Likewise, for a point not in J, it and all the nearby points which also lie in the open set $A_{E_{0.2}}(p)$ have the same dynamic behavior (in that they all iterate towards p), placing such a point in the Fatou set $F(E_{0.2})$. Hence we have shown $J = J(E_{0.2})$ as desired. □

EXPLORATION 1.141. Use the *Parameter Plane and Julia Set Applet* to track the orbit of nearby points in $J(E_{0.2})$ to witness this sensitivity to initial conditions. It is a bit difficult to pick points that stay in the viewing window after more than a few steps, but by looking at the data in the **Orbits 1 and 2** tab at the bottom, you can see the orbits separate from each other. You can also zoom in on the picture so that you can choose your initial seed values very close to each other, and then zoom out to see their orbits. ***Try it out!***

1.6.2. Hairs and endpoints. Thinking again about this construction of $J(E_{0.2})$ we now describe this "Cantor Bouquet" by describing the "hairs" (previously called the stage "infinity" fingers) which make up this set. This section is not crucial to moving forward with the rest of the text, and we will not provide all of the details here to prove everything claimed, but we won't let that get in the way of discussing some very fascinating aspects of what we have stumbled upon in the above example. This section is more about "points of interest" that may inspire the motivated reader to pursue such ideas further.

Pick a stage 1 finger C_{k_1}, and then from the infinite number of sub-fingers within C_{k_1} pick one sub-finger $C_{k_1 k_2}$. Then from the infinite number of sub-sub-fingers within that, pick one sub-sub-finger $C_{k_1 k_2 k_3}$. Continuing on in this fashion so that each $C_{k_1 k_2 \ldots k_n}$ is a stage n finger which lives in $C_{k_1 k_2 \ldots k_{n-1}}$. We see that $\gamma = \cap_{n=1}^{\infty} C_{k_1 k_2 \ldots k_n}$ must, by (b) above, be infinitely thin (i.e., contain no open disk of any positive radius), but which also stretches to ∞ in the positive x-direction. We call this intersection γ a ***hair***, and note that there are infinitely many (actually uncountably many) of these hairs in $J(E_{0.2})$ corresponding to all the different ways one can choose $k_1 k_2 \ldots$. It turns out that each hair γ is actually a curve, that is, an image of a continuous map $h_\gamma : [0, \infty) \to \mathbb{C}$ such that $h_\gamma(t) \to \infty$ as $t \to \infty$. We call the point $h_\gamma(0)$ the ***endpoint*** of the hair γ. There are a few important details to be shown to prove all this. In particular, an important detail we have yet to address at all is: how do we know that there is anything left in the set $\gamma = \cap_{n=1}^{\infty} C_{k_1 k_2 \ldots k_n}$, or even in the larger set in $\cap_{n=1}^{\infty} F_n$? Perhaps, the left "tips" of the stage n fingers move farther and farther to the right in a way that the set $\cap_{n=1}^{\infty} F_n$ is actually empty. We leave it to the interested reader to pursue these matters in [**5**]. However, we do have another way to see that we have infinitely many hairs (clearly shown to have finite endpoints) in $J(E_{0.2})$ and the interested reader can pursue this in Additional Exercise 1.198.

As one can see in the construction above, each hair is separated from any other hair by an infinite (though possibly very thin) gap which stretches out to ∞ in the direction of the positive x-axis. Thus in the plane $\mathbb{C}$ we would say these hairs are all

separated from each other. However, by including ∞, where all these hairs "meet", we produce a set $J(E_{0.2}) \cup \{\infty\}$ which is a connected subset of the sphere $\overline{\mathbb{C}}$.

We conclude this section with remarks about two truly fascinating properties of the Cantor Bouquet $J(E_{0.2})$. In order to discuss this properly we will need the following more general definition of what it means for a (not necessarily open or closed) subset of $\overline{\mathbb{C}}$ to be connected.

DEFINITION 1.142. A set C is *connected* in $\overline{\mathbb{C}}$ if one cannot find open subsets U and V of $\overline{\mathbb{C}}$ such that (i) $U \cap C \neq \emptyset$, (ii) $V \cap C \neq \emptyset$, (iii) $C \subseteq U \cup V$ and (iv) $U \cap V \cap C = \emptyset$. Another, perhaps more intuitive definition is as follows: A set $C \subseteq \overline{\mathbb{C}}$ is connected in $\overline{\mathbb{C}}$ if we cannot we break up (*disconnect*) C into a disjoint union of non-empty sets A and B (thus $C = A \cup B$) such that no sequence from A converges to a point in B, and vice versa.

We can now present (without proof) two startling properties of the Cantor Bouquet $J(E_{0.2})$.

Property 1. Let $\mathcal{E}$ denote the set of all endpoints of all the hairs in $J(E_{0.2})$. Then the set $\mathcal{E}^* = \mathcal{E} \cup \{\infty\}$ is *connected* in $\overline{\mathbb{C}}$. Informally, this says that these endpoints *together with* ∞ are somehow bunched up together so tightly in such a way that it is impossible to disconnect this set in the manner described in Definition 1.142. However, and here comes the strange part, it turns out that removing ∞ from $\mathcal{E}^*$ creates a *totally disconnected* set $\mathcal{E}$ (i.e., the only connected subsets of $\mathcal{E}$ are sets which contain just a single point). In fact, we can show this second part quite easily by noting that between any two hairs there is a *gap* in $F(E_{0.2})$ between them, thus showing that the endpoints of these hairs cannot both lie in the same connected subset of $\mathcal{E}$. This gap, however, extends to ∞ and so cannot be used to create a disconnection of $\mathcal{E}^*$. However, it still remains to be shown that no other method of disconnecting $\mathcal{E}^*$ can work either. Since this is too complicated for the present text we leave the interested reader to view the details in [**21**].

Property 2. The Hausdorff dimension is a concept that allows one to assign a number to each set in $\overline{\mathbb{C}}$ which relates to the "size" of the set. The number is defined (see [**11**] for further details) in such a way that it shares many of the properties of our usual notion of dimension and so we refer to it as a "dimension" even though it need not be an integer. In particular, the Hausdorff dimension of a set measures, in some sense, "how much space the set fills up". With this, the Hausdorff dimension of any *smooth* curve, in particular, any hair in $J(E_{0.2})$ (which are shown to be smooth in [**4**]), has dimension 1, whether the endpoint is included or not. Let $\mathcal{H}$ denote the union of all **open** hairs, that is, each hair with its endpoint removed. Thus $\mathcal{H} = J(E_{0.2}) \setminus \mathcal{E}^*$. These hairs, as you will recall, are each separated by gaps that prevent them from accumulating too much and "filling up" a lot of space. So, at least heuristically, it seems plausible that $\mathcal{H}$ has Hausdorff dimension equal to 1, which it turns out it does.

What is astounding is that the set of endpoints $\mathcal{E}$ has dimension equal to 2. Thus the endpoints seem to bunch up in a strange fashion so that the set of all of them "fills up" two dimensions worth of space. Since for each infinitely long hair there is only one endpoint, it seems impossible that the set of endpoints could be larger than the set of open hairs (as measured by Hausdorff dimension). But this is exactly what happens. The interested reader should see [**18, 19**] for the details.

1.6.3. Critical orbits, Exploding Julia sets, Parameter space for E_c. As we have seen, the role of critical points dictates much about the overall dynamics of a rational map, especially for the family of polynomials $f_c(z) = z^2 + c$. The exponential maps $E_c(z) = ce^z$, however, have no critical points (verify), but each such map does have 0 as what we call an *asymptotic value* and this will play a critical role very much like a critical point for a rational map does.

DEFINITION 1.143. An *asymptotic value* for an entire function f is a value $a \in \overline{\mathbb{C}}$ such that there exists a curve $\gamma : [0, +\infty) \to \mathbb{C}$ tending to ∞ (i.e., $\gamma(t) \to \infty$ as $t \to +\infty$) such that $f(\gamma(t)) \to a$ as $t \to +\infty$.

We note that for the case of E_c we may choose the curve to be the negative real axis, along which the values $E_c(z) \to 0$ since $\operatorname{Re} z \to -\infty$. The following theorem, whose proof can be found in [**5**], is a result which parallels Theorems 1.89 and 1.93 given for rational maps.

THEOREM 1.144. Let f be a transcendental entire map. Then the immediate basin of each (super) attracting cycle of f contains a critical point of f or an asymptotic value of f.

From this result we see that the orbit $\{E_c^n(0)\}_{n=1}^\infty$ plays a very special role and so we call it the **critical orbit**.

Fact 1. Any attracting cycle of E_c must attract the *critical orbit* $\{E_c^n(0)\}_{n=1}^\infty$.

Fact 1 follows from Theorem 1.144. This next fact, however, has a proof beyond the scope of this text, but we shall nonetheless make frequent use of it.

Fact 2. If the *critical orbit* $E_c^n(0) \to \infty$, then $J(E_c) = \overline{\mathbb{C}}$.

Note that Fact 2 is strikingly different from anything we have yet seen. In fact, due to the super attracting fixed point at ∞ for any polynomial we know that we could never have such a Fatou set be empty.

Fact 2 also leads us to discover the fascinating *exploding* Julia sets we can see when we vary the parameter c in the map E_c. As we saw in Example 1.135, $J(E_{0.2})$ is a Cantor Bouquet of hairs and $F(E_{0.2})$ consists of exactly the attracting basin of a finite attracting fixed point $p \approx 0.26$. However, if we increase c from $c = 0.2$ to $c = 0.5$ we see that $E_{0.5}^n(0) \to +\infty$ and so, by Fact 2, we must have $J(E_{0.5}) = \overline{\mathbb{C}}$. Hence some real value c^* between $c = 0.2$ and $c = 0.5$ must be a bifurcation point where the dynamics

drastically change. What we see is that as c grows from being smaller than c^* to larger than c^* the Julia set goes from being confined to $\{\operatorname{Re} z \geq 1\}$ to *exploding* to be all of $\overline{\mathbb{C}}$ (see Additional Exercise 1.199).

So we see that we get very different dynamic behavior for E_c depending on whether or not the critical orbit escapes to ∞. We use this dichotomy to color the *parameter plane* based on how many iterates it takes for the critical orbit to escape (that is, have real part become greater than 50), leaving the c parameter black if the critical orbit does not escape. You can view and investigate this parameter space using the *Parameter Plane and Julia Set Applet*. Using this applet, set the window in the Parameter plane to show $[-3.2, 3.2] \times [-3, 3]$ to find a cardioid shape region, with many bulbs attached. You will likely want to set the **Parameter plane max iterations** value to 40, 60, 80, or 100 to see this cardioid in more detail. However, due to the computing power needed to compute the complex exponential iterates, each picture may take a little while for the applet to complete.

EXPLORATION 1.145. Take some time to experiment with this applet to see if you can guess the dynamical significance of this cardioid and some of the attached bulbs. ***Try it out!***

Let's investigate this exponential parameter plane much like we did for the family $f_c(z) = z^2 + c$ by calculating the set L_1 of c values for which E_c has an attracting fixed point p. In order for p to be an attracting fixed point we require $E_c(p) = p$ and $|E_c'(p)| < 1$. The first condition yields $ce^p = p$ which used with the second gives $|p| = |ce^p| < 1$. Notice in this case p is both the fixed point and the multiplier. Thus we are looking for c such that $c = pe^{-p}$ for $|p| < 1$. By letting p move around the unit circle, the corresponding c values trace out the cardioid-like set in the parameter plane pictured in Figure 1.27. Use the *ComplexTool* applet to draw this.

This cardioid-like L_1 has many similarities with the cardioid K_1 found in the Mandelbrot set. For example, one can find p/q bulbs attached to its boundary. (Note that the letter p when used to refer to an attracting fixed point of E_c is a complex number, whereas the p used in the numerator of p/q to denote a particular bulb is always a positive integer (as is q). Context will always make it clear which is which and so no confusion should arise.) These p/q bulbs might not be located where you first expect. For example, the $1/3$ bulb in the Mandelbrot set M was found at the top of the cardioid K_1 whereas the top of L_1 has the $2/5$ bulb attached.

EXPLORATION 1.146. Use the *Parameter Plane and Julia Set Applet*[31] to investigate the "new" locations of the p/q bulbs. One way to determine the location of these p/q bulbs is to experiment long enough to find a pattern (can you use Farey addition

---

[31]Be sure the settings for the applet are returned to values which will produce good pictures. In particular, you may want to check the values for the Escape criterion and max/min iterations which you adjusted earlier (or simply reload the applet).
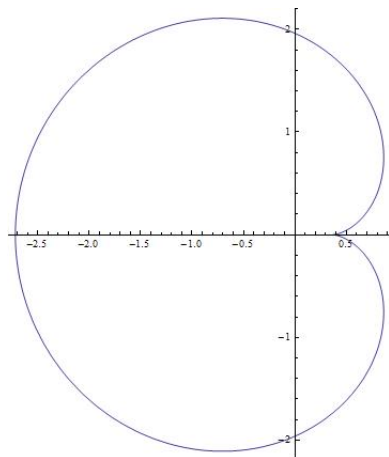
FIGURE 1.27. This cardioid-like shape is the boundary of L_1, the set of c parameters which give rise to an attracting fixed point for the map $E_c(z) = ce^z$.

here too?). Another way is to use the multiplier map, or more precisely, the inverse of the multiplier map (see Additional Exercise 1.200). ***Try it out!***

EXPLORATION 1.147. Where is the 1/2 bulb located? Does it have a sub-bulb which corresponds to attracting 4-cycles? ***Try it out!***

EXERCISE 1.148. Whether or not you think you can solve it, try to come up with an interesting mathematical question about what you see in the dynamics of exponentials. Try to draw parallels, or show where parallels do not hold, with the family $f_c(z) = z^2 + c$. ***Try it out!***

1.6.4. The Trigonometric maps. Recall that $\sin z = \frac{-i}{2}(e^{iz} - e^{-iz})$ and $\cos z = \frac{1}{2}(e^{iz} + e^{-iz})$, and set $S_c(z) = c \sin z$ and $C_c(z) = c \cos z$. We leave it to the reader to use these facts to show that orbits under the maps $S_c(z)$ and $C_c(z)$ "escape" to ∞ in the direction of the positive or negative imaginary axis (see Additional Exercise 1.201). Thus, using Proposition 1.133 to note $J(S_c) = \overline{A_{S_c}(\infty)}$ and $J(C_c) = \overline{A_{C_c}(\infty)}$, the algorithm in the *Parameter Plane and Julia Set Applet* for creating the Julia sets for the exponential maps $E_c(z) = ce^z$ can easily be modified. In particular, an orbit under either S_c or C_c is deemed to limit to ∞ if any point in the orbit has imaginary part larger than 50 in absolute value. It is known that neither S_c nor C_c has a finite asymptotic value (see Additional Exercise 1.202), however, both have infinitely many critical points. At first glance this may appear to provide us with considerable difficulty, since a map having more than one critical orbit would pose a real challenge to coloring and analyzing the parameter space as easily as we have done for the maps $z^d + c$ and ce^z. However, as we ask you to verify in Additional Exercise 1.203, each map S_c and

76

C_c has just one critical orbit. In particular, $\{S_c^n(c)\}_{n=1}^{\infty}$ is the critical orbit for S_c and $\{C_c^n(c)\}_{n=1}^{\infty}$ is the critical orbit for C_c.

As in the exponential case we have key facts that will greatly aid in our understanding of the dynamics of these Trigonometric functions.

Fact 1. Any attracting cycle of S_c must attract its *critical orbit* $\{S_c^n(c)\}_{n=1}^{\infty}$.

Fact 2. Any attracting cycle of C_c must attract its *critical orbit* $\{C_c^n(c)\}_{n=1}^{\infty}$.

Fact 3. If the *critical orbit* $S_c^n(c) \to \infty$, then $J(S_c) = \overline{\mathbb{C}}$.

Fact 4. If the *critical orbit* $C_c^n(c) \to \infty$, then $J(C_c) = \overline{\mathbb{C}}$.

Because the orbit $\{S_c^n(c)\}_{n=1}^{\infty}$ is critical to the dynamics of the map S_c, we can then draw the parameter plane much like we did for the family of maps E_c. In particular, the *Parameter Plane and Julia Set Applet* will color each c parameter based on how many iterates it takes for the critical orbit to escape, in this case in the direction of the positive or negative imaginary axis. As before, the c parameter is left black if the critical orbit does not escape. As always, we must heed the usual warnings that the computer is illustrating only an approximation (hopefully a good one) to the actual dynamics we are trying to study. The applet for C_c works similarly. Use this applet now to experiment and investigate the dynamics you find with the functions S_c and C_c. If you look closely in the parameter planes, you will find a familiar friend. ***Try it out!***

1.7. The Mandelbrot Set is Universal

You have probably noticed that the parameter planes for the maps $f_c(z) = z^2 + c$, $C_c(z) = c\cos z$, and $S_c(z) = c\sin z$ drawn by the *Parameter Plane and Julia Set Applet* all show show black regions that contain sets which resemble the Mandelbrot set M. See Figure 1.28.

The appearance of copies of the Mandelbrot set in all of these diverse settings is quite unexpected. Why should the Mandelbrot set, a picture of the fundamental dichotomy in the study of the parameter space of quadratic maps $f_c(z) = z^2 + c$, have anything to do with the parameter space of the transcendental maps $C_c(z) = c\cos z$ and $S_c(z) = c\sin z$? It turns out that the Mandelbrot set, or at least various "copies" of it appear in so many parameter planes that it is truly a fundamental mathematical object (much like the numbers π and e) which arises in many unexpected places. Just like we saw that baby Mandelbrot sets are dense in the full Mandelbrot set M, so too are they dense in many parameter spaces. The main reason behind all of this, which we can only state loosely in this text, is that very often iterates of maps have dynamic behavior very much like the dynamic behavior of iterates of a quadratic polynomial, when these maps are considered on relatively small domains. Thus, as a parameter is changed, we see the same type of changing behavior in the parameter plane of this family of maps as we do in in the original family $f_c(z) = z^2 + c$ corresponding the
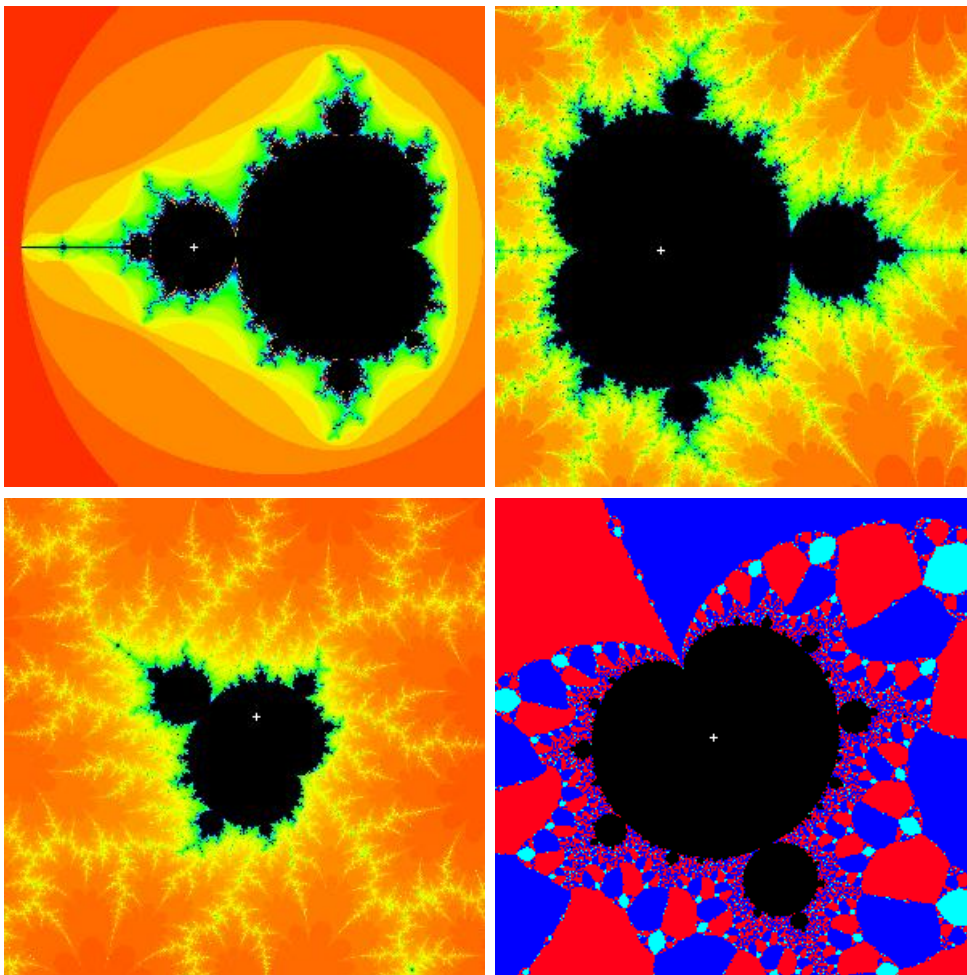
FIGURE 1.28. The original Mandelbrot set M (upper left), a "copy" of M containing $c = 3.2$ in the parameter space for $c \cos z$ (upper right), a "copy" of M containing $c = 2.17 + 1.3i$ in the parameter space for $c \sin z$(lower left), and a "copy" of M containing $\rho = 0.906 + 0.423i$ in the parameter space for Newton's method applied to the map $p_\rho(z) = z(z-1)(z-\rho)$.

original Mandelbrot set. The interested reader can look to [**22**] where it is shown that small Mandelbrot sets are dense in the bifurcation locus for what are called *holomorphic families* of rational maps.

1.8. Concluding remarks and new directions

We have investigated the chaos that arises through a number of different complex dynamical systems. We saw how Newton's method can (and very often must) produce

very chaotic behavior, represented by beautiful fractal images. In order to investigate such chaos in a more general setting, we turned our eyes to iteration of analytic maps that were not necessarily generated as the Newton map of some polynomial. There we found that even in the seemingly simple class $f_c(z) = z^2 + c$ there was a richness and complexity we could not have imagined. Many of these features were also present in the dynamics of the entire maps $E_c(z) = ce^z, S_c(z) = c\sin z$, and $C_c(z) = c\cos z$, though some new phenomena also appeared. Throughout, an important part of our investigations was to track, if we could, how the behavior of these systems changed when the maps were changed. We did this with the above families by varying the parameter c. In some sense, however, these were really mild perturbations of the maps since many of the salient features of the maps (such as the super attracting fixed point at ∞ for $f_c(z) = z^2 + c$ or the essential singularity at ∞ for the transcendental maps) persisted no matter how the c values were changed.

1.8.1. Perturbation with a pole. There are, however, many ways to perturb a map that warrant our attention and can pique our interest in what turns out to also produce very beautiful mathematics. For example, the maps $F_c(z) = z^d + c/z^m$ with fixed $d, m \in \mathbb{N}$ are, for $c \in \mathbb{C} \setminus \{0\}$, perturbations of the dynamically well understood map $z \mapsto z^d$. Unlike the $f_c(z) = z^2 + c$ perturbations of $z \mapsto z^2$, however, these maps F_c add a whole new dimension to the analysis due to the pole appearing at the origin. In particular, for c very small, the map F_c behaves very much like $z \mapsto z^d$, but only as long as z is sufficiently far from the origin. Near the origin the presence of the pole changes the dynamics considerably from $z \mapsto z^d$. It turns out that such systems lead to much fascinating mathematics that can sometimes be represented by pictures such as Figure 1.29 showing features not observed in any of the other systems studied in this chapter.

LARGE PROJECT 1.149. Investigate the dynamics of the maps of the form $F_c(z) = z^d + c/z^m$. You can use the *Parameter Plane and Julia Set Applet* to do this, as well as your own analytic paper and pencil techniques. Note that we have not described how the parameter plane is being drawn. It is up to you to determine this as well as figure out what you can on your own. Out of all the maps mentioned in this chapter, these are the newest and least studied. There is a lot of new territory waiting to be discovered here.

1.8.2. Random Dynamics. In each of the examples of perturbed maps throughout this chapter, even though the perturbation could be mild by changing a simple parameter or more severe by adding a pole as in F_c, there was one fundamental assumption always made regarding the perturbed system – once the map was perturbed, this map was fixed and iterated again and again to create the dynamics. A new way to perturb a system, however, is to allow the map to change at every step in the orbit.

For example, suppose we start with two maps f and g. The usual iteration dynamics says we consider the orbit generated by either $f^n(z_0)$ or by $g^n(z_0)$. However, it is natural
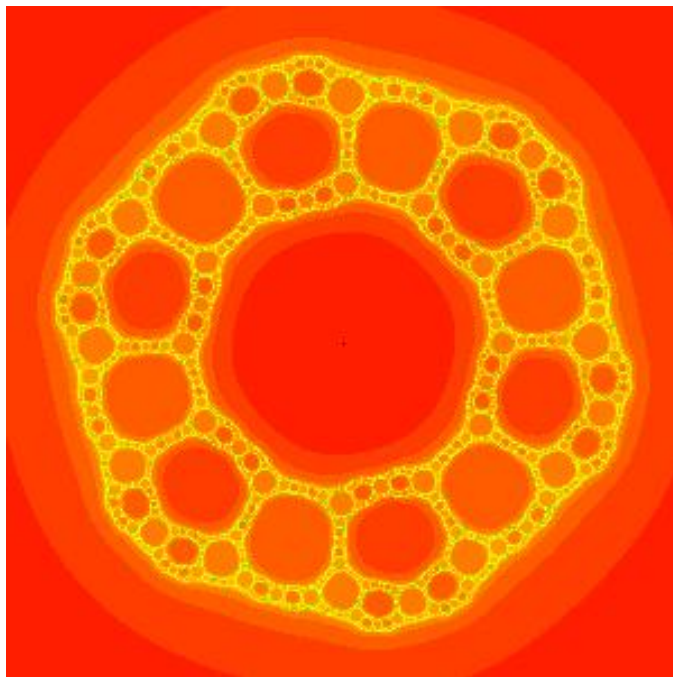
FIGURE 1.29. A Sierpinski curve Julia set for the function $F(z) = z^3 + 0.13i/z^3$. The regions of various shades of orange are in $A_F(\infty)$ and the remaining points in yellow represent $J(F)$.

to ask what happens if at each stage of the orbit *either* map f *or* g can be applied? In some cases the dynamics can be much more uncivilized (and more fun). Such dynamics investigates the behavior of orbits $h_{i_1}(z_0), h_{i_2}(h_{i_1}(z_0)), h_{i_3}(h_{i_2}(h_{i_1}(z_0))), \ldots$ where the maps h_i are chosen to be either f or g. If one chooses each map h_i at random at each stage of the orbit, then one enters the research area of so-called *random dynamics*. Such systems are directly connected to the study of what are known as "iterated functions systems" and their attractors sets (see [**7**]), such as the van Koch curve and Sierpinski triangle.

Instead of investigating such attractor sets, we look in another direction. We investigate what we call a *random basin of attraction* as follows. Fix a point $z_0 \in \mathbb{C}$ and randomly select the map h_1 to be either f or g (each with probability $1/2$). Then set $z_1 = h_1(z_0)$. Now randomly select the map h_2 to be either f or g and set $z_2 = h_2(z_1) = h_2(h_1(z_0))$. Continue in this fashion to produce what we call a ***random orbit*** $z_0, z_1, z_2, \ldots$. Now we consider whether or not such a random orbit can have a particular limit, and what the probability is of having that limit.

EXAMPLE 1.150 (Devil's Staircase). Let $f(x) = 3x$ and $g(x) = 3x - 2$ be defined on the real line $\mathbb{R}$. Now consider for each $x_0 \in \mathbb{R}$ generating a random orbit $x_0, x_1, x_2, \ldots$ as described above. Let $P(x_0)$ denote the probability that this random orbit converges

to $+\infty$. Note that here we do distinguish between convergence of this random orbit to $+\infty$ and $-\infty$. It is not hard to believe that if $x_0 > 1$, then we must have $P(x_0) = 1$ since no matter what choices of f and g we make at each step of the orbit, we will always have the orbit points growing larger (and positive). Similarly, one can believe that if $x_0 < 0$, then we must have $P(x_0) = 0$ since both maps f and g force the orbit points to grow larger (and negative). What happens for $0 \le x_0 \le 1$, however, is much more interesting, and we show the graph of P, called the Devil's Staircase, in Figure 1.30.
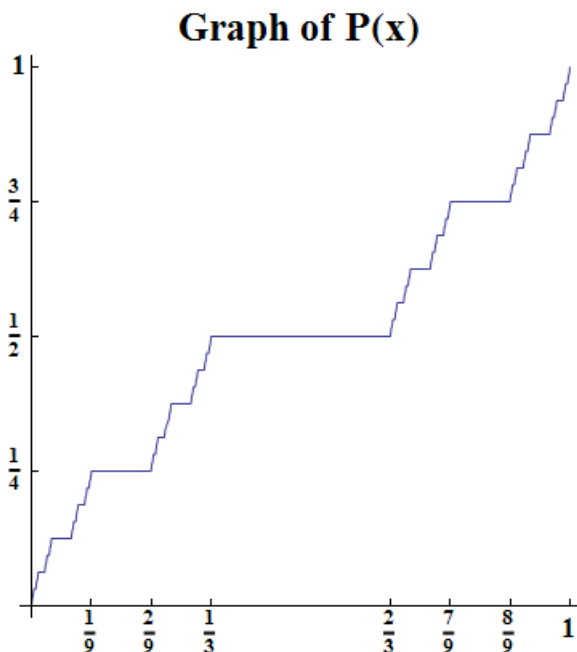
Graph of P(x)

FIGURE 1.30. The "Devil's Staircase" is the graph of the probability $P(x)$ that a random orbit generated by the maps $f(x) = 3x$ and $g(x) = 3x - 2$ tends to $+\infty$.

An interesting property of this graph is that between any two steps (which we define to be the horizontal sections of the graph, i.e., intervals where $P'(x) = 0$), we have infinitely many steps in between. Also, the sum of the lengths of all the steps between 0 and 1 is exactly 1 (for this reason we say that this graph is *almost always* flat, i.e., almost always has $P'(x) = 0$). However, this function is both continuous and increases (not strictly) from $y = 0$ to $y = 1$ as x goes from 0 to 1.

EXERCISE 1.151. Prove the following aspects of the function $P(x)$, the Devil's Staircase function, in Example 1.150.

(a) Show that $P(x) = 0$ for $x < 0$ and $P(x) = 1$ for $x > 1$.
(b) Show that $P(0) = 0$ and $P(1) = 1$.

(c) Show that $P(x)$ is increasing (not strictly), i.e., $x < y$ implies $P(x) \leq P(y)$.

(d) Show that on the open interval $(1/3, 2/3)$ both $P(x) = 1/2$ and $P'(x) = 0$.

(e) Show that on the open interval $(1/9, 2/9)$ both $P(x) = 1/4$ and $P'(x) = 0$.

(f) Find where $P(x) = 3/4$.

(g) In parts (d)-(f), you have located the largest step (corresponding to $(1/3, 2/3)$), and the two next largest steps. Describe the four next largest steps, and then describe the pattern relating all steps.

(h) Sum the lengths of all the steps to show this sum is 1.

(i) Show that $P(x)$ is continuous.

EXAMPLE 1.152 (Devil's Colosseum). The Devil's Colosseum is a higher dimensional analog of the Devil's Staircase, and is also understood through random dynamics. Instead of functions defined on just the real line, we consider the complex valued maps $g_1(z) = z^2/4$ and $g_2(z) = z^2 - 1$, and then using the second iterates of these maps set $f(z) = g_1^2(z)$ and $g(z) = g_2^2(z)$.

We now investigate the *random basin* of attraction for ∞ as follows. Fix a point $z_0 \in \mathbb{C}$. Consider the random orbit $z_0, z_1, z_2, \dots$ generated as described above, and let $P(z_0)$ be the probability that this random orbit z_n tends to ∞. Note that for z_0 large ($|z_0| > 10$ will do), we have $P(z_0) = 1$ since no matter what choices of f or g we make at each step, the orbit z_n will go to ∞. Though it is not as simple, one can also show $P(z_0) = 0$ for z_0 near zero. In between, however we get some very interesting behavior. In fact, this function $P(z)$ has a graph given in Figure 1.31 called the Devil's Colosseum, and it has many properties similar to those of the Devil's Staircase. In particular, if you started from the bottom and tried to climb out, you would have to walk a path very much like the Devil's Staircase.
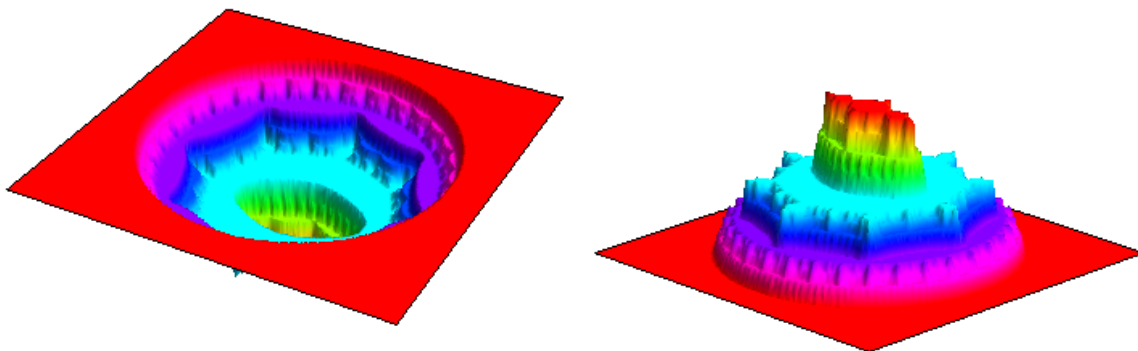

FIGURE 1.31. The left picture shows the Devil's Colosseum, the graph of the probability $P(z)$ that a random orbit as defined above tends to ∞. The picture on the right is the inverted Devil's Colosseum, sometimes called a fractal wedding cake. The first pictures of these objects were created and carefully studied by Hiroki Sumi (see [**25**]).

We can describe the dynamics which underlie the Devil's Colosseum in terms of stability of orbits, and in doing so will introduce the following extensions of the notions of the Fatou and Julia sets. Let $F(\langle f, g \rangle)$ denote the Fatou set of the random system described above, which we define to be those points $z_0 \in \mathbb{C}$ such that *every* randomly generated sequence of maps h_{i_n} produces a stable orbit (in the sense that all nearby seed values will have similar orbits when the sequence of maps h_{i_n} is used). We then define the Julia set of this random system to be $J(\langle f, g \rangle) = \overline{\mathbb{C}} \setminus F(\langle f, g \rangle)$. We can thus describe the oddity of the Devil's Colosseum like so. The function P is continuous on all of $\mathbb{C}$, yet it is constant on each component of the Fatou set, even though the Julia set contains no open set (see [**25**]). It is these components of the Fatou set which make up the various horizontal *levels* (or steps) you see in Figure 1.31. The Julia set is then the set on which $P(z)$ varies (since $P(z)$ does not vary, i.e., it is flat, on $F(\langle f, g \rangle)$). We see this Julia set pictured in Figure 1.32.
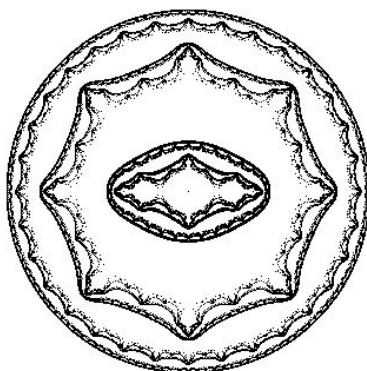

FIGURE 1.32. The Julia set $J(\langle f, g \rangle)$ in $\mathbb{C}$ is the the set on which $P(z)$ varies. This pictured was drawn via Julia 2.0 (see [**23**]).

REMARK 1.153. We note that the random dynamics displayed in the Devil's Colosseum produced something that cannot happen with the usual iteration dynamics. For example, given a polynomial $f(z)$, consider defining a function $P(z_0)$ to be the probability of an orbit with seed z_0 tending to ∞. Of course, since there is only one map to choose, there is no randomness involved at all, and so we see that $P(z_0)$ must be either 0 or 1. In particular, we have $P(z) = 1$ for $z \in A_f(\infty)$ and $P(z) = 0$ for $z \in K(f)$. The function P is then discontinuous at every point $z \in \partial A_f(\infty) = \partial K(f) = J(f)$. Thus we see that the random dynamics exhibited by the Devil's Colosseum, where $P(z)$ is continuous on all of $\mathbb{C}$, is a real change from iteration dynamics.

EXERCISE 1.154. The reader is encouraged to investigate such random dynamical systems visually by generating their own pictures similar to the Devil's Colosseum. Sample code is provided in Appendix 1.C.

There is a very large world of chaotic phenomena to explore. Whether you stay only within the confines of the topics mentioned in this book or venture into new areas, there remains an infinite amount of mathematics to discover. We encourage you to go and explore and make your own contributions to mathematics.

1.9. Additional Exercises

Orbits, Examples, and Fixed Points

EXERCISE 1.155. Prove $f^n(x) \to +\infty$ for all $x \in \mathbb{R}$ when $f(x) = e^x$. In fact, the advanced reader can actually show that the convergence is *uniform*,[32] i.e., given any $M > 0$, there exists an integer $N > 0$ such that $n > N$ implies $f^n(x) > M$ for all $x \in \mathbb{R}$.

EXERCISE 1.156. For $f(x) = \sin x$ where $x \in \mathbb{R}$ is given in radians prove that $f^n(x) \to 0$ for all $x \in \mathbb{R}$. Hint: Apply the Mean Value Theorem to show that $|\sin x| < |x|$ for all $x \in \mathbb{R} \setminus \{0\}$. The advanced reader can actually show that the convergence is *uniform*,[32] i.e., given any $\epsilon > 0$, there exists an integer $N > 0$ such that $n > N$ implies $|f^n(x) - 0| < \epsilon$ for all $x \in \mathbb{R}$.

EXERCISE 1.157. For $f(x) = \cos x$ where $x \in \mathbb{R}$ is given in radians prove that $f^n(x) \to 0.739085....$ for all $x \in \mathbb{R}$ where $x^* = 0.739085....$ is the number such that $\cos x^* = x^*$. Hint: Apply the Mean Value Theorem to show that $|\cos x - x^*| < |x - x^*|$ for all $x \in \mathbb{R} \setminus \{x^*\}$. The advanced reader can actually show that the convergence is *uniform*,[32] i.e., given any $\epsilon > 0$, there exists an integer $N > 0$ such that $n > N$ implies $|f^n(x) - x^*| < \epsilon$ for all $x \in \mathbb{R}$.

EXERCISE 1.158. For $g(z) = \sin z$, show that $g^n(\pm i\epsilon) \to \infty$ for any $\epsilon > 0$. Hint: First show that $g(iy)$ is purely imaginary and $|g(iy)| > |y|$ for each real $y \neq 0$.

Complex Newton's method

EXERCISE 1.159. What Newton's method can "find". Let F be the Newton map of a polynomial f of degree two or more.
(a) Suppose $F^n(z_0) \to a$ for some $z_0 \in \mathbb{C}$ and some *finite* point $a \in \mathbb{C}$. Prove $f(a) = 0$.
(b) However, show that we always have $F(\infty) = \infty$ and so we see that it is necessary that we consider only finite points in part (a).

EXERCISE 1.160. Prove that that the Newton map of $f(z)/f'(z)$ (as opposed to the Newton map of $f(z)$) always has super attracting fixed points at the roots of f regardless of the order of the root of f. Hint: Given a root α of f of order k, determine the order of the root of $f(z)/f'(z)$.

EXERCISE 1.161. Construct examples of analytic maps f to justify the statement that there is no universal $r^* > 0$ such that $\triangle(\alpha, r^*) \subseteq A_F(\alpha)$ for *all* analytic maps f with a root at α. Here, as usual, F denotes the Newton map for f.

EXERCISE 1.162. Radius of Convergence for polynomials.
Let $p(z) = (z - \alpha)^k (z - \alpha_1) \ldots (z - \alpha_s)$ be a polynomial of degree $n = k + s$, where the α_j need not be distinct. Follow the steps below to show that $r = \frac{d(2k-1)}{2n-1}$ is a radius

---

[32]Uniform convergence is not critical the to development of the text and so may be safely disregarded. However, advanced students should try to understand the concept in the given context.

of convergence for the Newton map F at α, where $d = \min_{j=1,\dots,s} |\alpha - \alpha_j|$ (the distance from α to the nearest other root of p).

(a) For the Newton map $F(z) = z - \frac{p(z)}{p'(z)}$, we wish to show that $|z - \alpha| < r$ implies $|F(z) - \alpha| < |z - \alpha|$. Note that this last inequality says that the action of F is to move points in the disk $\Delta(\alpha, r)$ closer to α. Show that such an inequality then implies that $F^n(z) \to \alpha$ as $n \to \infty$ when $|z - \alpha| < r$.

(b) Show that $|F(z) - \alpha| < |z - \alpha|$ is equivalent to $\frac{|F(z) - \alpha|}{|z - \alpha|} = \left| 1 - \frac{p(z)}{(z - \alpha)p'(z)} \right| < 1$.

(c) Show that $\left| 1 - \frac{1}{w} \right| = \left| \frac{w-1}{w} \right| = \left| \frac{w-1}{w-0} \right| < 1$ holds if and only if $\mathrm{Re}\, w > 1/2$.

(d) Justify the following. The term $w = \frac{(z-\alpha)p'(z)}{p(z)} = (z - \alpha)\left[\frac{k}{z-\alpha} + \sum_{j=1}^{s} \frac{1}{z-\alpha_j} \right] = k + \sum_{j=1}^{s} \frac{z-\alpha}{z-\alpha_j}$. Thus $\mathrm{Re}\, w > \frac{1}{2}$ exactly when $\mathrm{Re}\, \sum_{j=1}^{s} \frac{z-\alpha}{z-\alpha_j} > \frac{1}{2} - k$, which holds, in particular, when $\sum_{j=1}^{s} \left| \frac{z-\alpha}{z-\alpha_j} \right| < k - \frac{1}{2}$.

(e) Show that if $|z - \alpha| < r$, we then have $|z - \alpha_j| > d - r$, where $d = \min_j |\alpha - \alpha_j|$. Hence $\sum_{j=1}^{s} \left| \frac{z-\alpha}{z-\alpha_j} \right| < s\frac{r}{d-r} = (n - k)\frac{r}{d-r}$.

(f) Combine all the above to reach the desired conclusion. Also, note that for $z \in \Delta(\alpha, r)$ we have $|F(z) - \alpha| < |z - \alpha|$, which is, in general, a much stronger statement than merely saying that $\Delta(\alpha, r) \subseteq A_f(\alpha)$. For this reason, some may call the value $r = \frac{d(2k-1)}{2n-1}$ a radius of **contraction** for Newton's method at α.

SMALL PROJECT 1.163. Fix an integer $n \geq 2$ and let $f(z) = z^n - 1$. Express the value for r found in Exercise 1.162 in terms of only n. See if you can improve upon this value for the radius of *contraction* by carefully investigating the above proof as applied to this select class of maps $f(z) = z^n - 1$. You might (or might not) wish to to calculate an $r^* > 0$ such that $|F'(z)| < 1$ on $\triangle(\alpha, r^*)$ for each root α of f, and then apply the Contraction Lemma 1.136.

SMALL PROJECT 1.164. This is an open ended project to investigate whether the value for r found in Exercise 1.162 can be improved if we know something about the geometry of the roots of p. See if you can improve upon this value for the radius of *contraction* by carefully investigating the proof in Exercise 1.162 for specific polynomials f of degree 3 with distinct roots $\alpha_p, \beta_p, \gamma_p$. If the roots, all lie on a line, can you squeeze more from that proof? What if the roots form an equilateral triangle? Are there some geometric configurations of the roots that allow for better results than what is given in Exercise 1.162? Can you generalize this to higher degree polynomials? What if you allow for multiple roots of f? If a paper and pencil result is too hard, provide a conjecture based on examples you considered with the *Complex Newton Method Applet*. Whether or not you can answer them, come up with some related questions that one might consider. **Remember, asking questions is important, even if you can't answer them.**

Comment: You might (or might not) wish to to calculate an $r^* > 0$ such that $|F'(z)| < 1$ on $\triangle(\alpha, r^*)$ for each root α of f, and then apply the Contraction Lemma 1.136.

Global Conjugation

EXERCISE 1.165. Prove that if $ab \neq 0$, then the maps $z \mapsto az$ and $z \mapsto bz$ are globally conjugate if and only if either $a = b$ or $a = 1/b$.

EXERCISE 1.166. Prove that if $ab \neq 0$, then the maps $z \mapsto z + a$ and $z \mapsto z + b$ are globally conjugate.

EXERCISE 1.167. Prove that any quadratic map $f(z) = az^2 + bz + d$ is globally conjugate to one and only one map of the form $z \mapsto z^2 + c$.

EXERCISE 1.168. Prove that a rational map R is globally conjugate to a polynomial if and only if there exists $w \in \overline{\mathbb{C}}$ with $R^{-1}(\{w\}) = \{w\}$.

Analysis of the Newton map of a cubic polynomial

EXERCISE 1.169. Prove that the attracting basin of an *attracting* fixed point of a rational or entire map is an open set in $\overline{\mathbb{C}}$, which, in particular, does not contain any of its boundary points. Hint: Use Theorem 1.18 together with the fact that if f is a continuous map, then $f^{-1}(U)$ is open whenever U is open.

EXERCISE 1.170. Use the *Complex Newton Method Applet* to investigate the behavior of Newton's method applied to $f_1(z) = z * (z - 1) * (z - 0.908 - 0.423i)$ and $f_2(z) = z * (z - 1) * (z - 0.913 - 0.424i)$. Describe the behavior you see, especially for the black seed values where Newton's method fails.

Iteration of an Analytic Function

EXERCISE 1.171. Dense orbit in $J(f)$.
(a) Show that for the map $f(z) = z^2$ there is a point $z_0 \in C(0, 1)$ whose orbit is dense on $C(0, 1)$, that is, $\overline{\{f^n(z_0) : n \in \mathbb{N}\}} = C(0, 1)$. Hint: Note that on $C(0, 1)$, we have $f(e^{i2\pi\theta}) = e^{i4\pi\theta}$, that is, the angle $\theta \in [0, 1)$ is doubled (mod 1). Consider $\theta_0 \in [0, 1)$ given in binary form as $\theta_0 = 0.0\,1\,00\,01\,10\,11\,000\,001\,010\,011\,100\,101\,110\,111\,0000\,\ldots$ where spaces have been included in this binary expansion to help illustrate the pattern.
(b) Use (a) to justify the second claim in Remark 1.47.
(c) Notice, however, that due to computer limitations, *any* point you choose on $C(0, 1)$ will have a *computed* orbit that eventually becomes periodic (although it might take a very large number of orbit points to see this). Justify this statement.

EXERCISE 1.172. In this exercise you are asked to complete the proof of Theorem 1.49.
(a) Use Schwarz's Lemma (see Theorem A.20 on page 414) to show that an attracting fixed point $a \in \mathbb{C}$ of an analytic map f must have $|f'(a)| < 1$. Hint: Consider a

small disk $\triangle(a, r)$ where f is attracting and then investigate the map $g : \triangle(0, 1) \to \triangle(0, 1)$ defined by $g(z) = \frac{f(rz+a)-a}{r}$.

(b) Now suppose $a \in \mathbb{C}$ is a repelling fixed point of an analytic map f. Use the fact that f is locally one-to-one with a locally defined inverse to show $|f'(a)| > 1$. Hint: Since $|f'(a)| \geq 1$ (else a would be attracting), we must have that f is locally one-to-one (see Section A.6.1.1 on page 416 regarding the Open Mapping Theorem and the Inverse Function Theorem A.23). Letting h be f^{-1} defined on some small disk $\triangle(a, r)$, show that a is an attracting fixed point for h (since it was repelling for f) and then apply part (a) to h. Then use this to argue for your desired conclusion.

EXERCISE 1.173. Let f be a map that is analytic at ∞ such that $f(\infty) = \infty$. Consider the multiplier $\lambda = 1/f'(\infty)$. Prove that if $|\lambda| < 1$, then ∞ is an attracting fixed point according to Definition 1.16. Also, prove that if $|\lambda| > 1$, then ∞ is a repelling fixed point according to Definition 1.45. Hint: Noting that the map $k(z) = 1/f(1/z)$ has a fixed point at the origin also with multiplier λ (see Lemma B.19 on page 426), study how the map $\phi(z) = 1/z$ transfers information about k at the origin to information about f at ∞.

EXERCISE 1.174. Let $g(z)$ be a polynomial of degree greater than or equal to two (which must then have a super attracting fixed point at ∞). Prove that $\partial A_g(\infty) = \partial K(g)$, and then note that by Theorem 1.60 we have $J(g) = \partial A_g(\infty) = \partial K(g)$.

EXERCISE 1.175. Convergence toward an attracting fixed point depends on the multiplier.

(a) For $c = -0.467 + 0.513i$ determine the attracting fixed point p_c of f_c.
(b) Write the multiplier λ for p_c in polar form.
(c) Use the *Global Complex Iteration Applet for Polynomials* to iterate a seed z_0 near p_c one step at a time (use the zoom feature to get a better look) to notice the way in which the orbit approaches p_c. The way this orbit approaches p_c is directly related to the multiplier λ. Describe this connection in your own words.
(d) Use the *Complex Function Iterator Applet* to iterate the map $z \mapsto \lambda z$ for various seed values near the attracting fixed point at the origin. Describe the convergence.
(e) Compare (c) and (d), explaining in as much mathematical detail as you can what you find. Hint: Consider the Taylor series of f_c expanded around the point p_c.
(f) Choose various c in K_1 and then study the convergence of orbits to the attracting fixed point using the applet. Use this picture to approximate the argument of the multiplier. Note: some c values will be easy to work with here, while others are not as easy.

EXERCISE 1.176. Convergence toward the attracting 2-cycle.

(a) For $c = -0.92568 + 0.22512i$ determine the attracting 2-cycle $\{u_c, v_c\}$ of f_c.
(b) Write the multiplier λ for $\{u_c, v_c\}$ in polar form.
(c) Use the *Global Complex Iteration Applet for Polynomials* to iterate a seed z_0 near one of the cycle points u_c one step at a time (use the zoom feature to get a better

look) to notice the way in which every other orbit point approaches u_c. The way this orbit approaches u_c is directly related to the multiplier λ. Describe this connection in your own words.

(d) Use the *Complex Function Iterator Applet* to iterate the map $z \mapsto \lambda z$ for various seed values near the attracting fixed point at the origin. Describe the convergence.

(e) Compare (c) and (d), explaining in as much mathematical detail as you can what you find. Hint: Consider the Taylor series of the iterate f_c^2 expanded around the point u_c.

(f) Choose any c in K_2 and then study the convergence of orbits to the attracting 2-cycle using the applet. Use this picture to approximate the argument of the multiplier. Note: some c values will be easy to work with here, while others are not as easy.

EXERCISE 1.177. Multiplier map for attracting 2-cycles of f_c. Determine the exact form of the multiplier map on K_2. This function $\lambda : K_2 \to \triangle(0,1)$ maps each $c \in K_2$ to the multiplier of the attracting 2-cycle of the map f_c. Also show that this map can be extended to a map $\lambda : \overline{K_2} \to \overline{\triangle(0,1)}$ which is continuous, one-to-one, and onto.

EXERCISE 1.178. Without appealing to Theorem 1.80, prove that if f_{c_0} has an attracting n-cycle, then for all c close to c_0, the map f_c also has an attracting n-cycle. Conclude that K_n is an open set. Hint: One may use Rouché's Theorem (see [1], p. 294) to prove a general result that if the coefficients of the polynomial P are sufficiently close to the corresponding coefficients of the polynomial Q of the same degree, then the roots of P and the roots of Q are close. Then apply this result to the polynomials of $P_1(z) = f_{c_0}^n(z) - z$ and $Q_1(z) = f_c^n(z) - z$.

EXERCISE 1.179. Mathematically justify Footnote 18 on page 42.

Critical Points and Critical Orbits

EXERCISE 1.180. Let $f(z) = z^3 - 1$ and let F denote the corresponding Newton map. Show that F cannot have an attracting cycle of any order other than the fixed points at the roots of f. Hint: Check the behavior of each critical point of F and consider Theorem 1.89.

Note: More can be said about this situation. It is true that Newton's method fails to find a root of f only for seed values from $\partial A_F(1) = \partial A_F(e^{2\pi i/3}) = \partial A_F(e^{-2\pi i/3})$. In fact, we have $\overline{\mathbb{C}} = A_F(1) \cup A_F(e^{2\pi i/3}) \cup A_F(e^{-2\pi i/3}) \cup \partial A_F(1)$. The method of proof is along the same lines as above, however, it uses a more powerful set of results. These results imply that if Newton's method fails for all seed values in an open set of points, then there must be a critical point of F whose orbit does not converge to any of the roots of f. The key ingredients to this proof are Sullivan's No Wandering Domains Theorem, the Classification of forward invariant components of Fatou sets, and the role of critical points in parabolic domains, Siegel disks, and Herman rings. All of these results can be researched by the interested reader in [1].

EXERCISE 1.181. Show that the periodic points for the map $f(z) = z^2$ are exactly those points of the form $e^{i2\pi p/q}$ where $p, q \in \mathbb{N}$ with p/q in lowest terms and q odd. Show that the pre-periodic points have the same form, but for q even. Hint: Show that if $p/q = \frac{n}{2^k-1}$ for some integers $k \geq 1, n \geq 1$, then $e^{i2\pi p/q}$ has a period of k or some divisor of k. Then use the fact that every rational number p/q with q odd can be expressed in this form (which can be proved by using the fact that the element 2 in the multiplicative group consisting of those elements of $\mathbb{Z}_q$ which are relatively prime to q has a finite multiplicative order which we call k).

EXERCISE 1.182. Symmetry in M. One feature of the Mandelbrot set which stands out is that it is symmetric about the x-axis. Prove this by showing that $c \in M$ if and only if $\bar{c} \in M$. Hint: Compare the critical orbits $\{f_c^n(0)\}$ and $\{f_{\bar{c}}^n(0)\}$.

EXERCISE 1.183. Prove that the Mandelbrot set M has no "holes" (i.e., its complement $\overline{\mathbb{C}} \setminus M$ is connected) by appealing directly to its definition. Hint: Suppose that U is a bounded domain of $\overline{\mathbb{C}} \setminus M$ such that $\partial U \subseteq M$. Now apply the Corollary A.18 on page 414 (a version of the Maximum Modulus Theorem) to the maps Q_n on $\overline{U}$ used in the proof of Lemma 1.103.

Exploring the Mandelbrot Set M

EXERCISE 1.184. Let P be a polynomial of degree $n \geq 2$. Show that the Filled-in-Julia set $K(P)$ has infinitely many points as follows.

(a) Show that $P(z)$ has a fixed point z_0 by applying the Fundamental Theorem of Algebra.
(b) Show that if z_0 is such that $P^{-1}(\{z_0\}) = \{z_0\}$, then $P(z) = z_0 + a(z - z_0)^n$ for some $a \neq 0$ from which it can be shown that $K(P) = \overline{\Delta(z_0, r)}$ where $r = |a|^{1/(1-n)}$.
(c) Show that if $P^{-1}(\{z_0\}) \neq \{z_0\}$, then there exists an infinite sequence of distinct points z_{-n} such that $\ldots \mapsto z_{-n} \mapsto z_{-n+1} \mapsto \ldots \mapsto z_{-2} \mapsto z_{-1} \mapsto z_0$. Hence, $\{z_{-n}\} \subseteq K(P)$.

We note that it is true that $J(P)$ must contain uncountably many points, but to show this requires considerably more effort (see [**1**], p. 95).

EXERCISE 1.185. Prove that for all $c \in \mathbb{C}$, we have $z \in A_{f_c}(\infty)$ if and only if $-z \in A_{f_c}(\infty)$. Do the same for the sets $J(f_c)$ and $K(f_c)$.

EXERCISE 1.186. Since the *root* $c_{p/q}$ of the p/q bulb in the Mandelbrot set lies on the boundary of two hyperbolic components (namely K_1 and $B_{p/q}$), we see that Theorem 1.80 implies that there are two multiplier maps that are defined at $c_{p/q}$. In particular, the map $\lambda : K_1 \to \triangle(0,1)$ can be extended to be defined at $c_{p/q}$, but also the map $\lambda_{p/q} : B_{p/q} \to \triangle(0,1)$ can be extended to be defined at $c_{p/q}$. Does $\lambda(c_{p/q}) = \lambda_{p/q}(c_{p/q})$? What does is mean if these values are the same? different?

EXERCISE 1.187. Find a c value in $B_{5/32}$. Formal proof is not required.

EXERCISE 1.188. Farey addition.

(a) Experiment with the *Parameter Plane and Julia Set Applet* to determine what Farey fraction should be used to represent the cusp of the main cardioid K_1 in the Farey procedure for finding child bulbs. Keep in mind that we are to treat the cusp like a Farey parent which is larger than all other bulbs.

(b) It is true (though not so easy to prove) that starting with Farey parents $B_{1/2}$ and the cusp one can compute through Farey addition all of the p/q bulbs attached to the upper half of the main cardioid K_1. Illustrate this by producing a Farey "family" tree which contains the lineage of $B_{5/32}$ all the way up to the ancestors $B_{1/2}$ and the cusp of the main cardioid K_1.

(c) Use Exercise 1.118 and part (b) to show that *all* of the p/q bulbs attached to the main cardioid K_1 can now be identified through Farey addition.

Other Uni-critical families of polynomials

EXERCISE 1.189. For the maps $P_c(z) = z^d + c$ with $d = 2, 3, 4, \ldots$ show the following:

(a) If $|z| \geq |c|$ and $|z| > 2$, then there exists $\epsilon > 0$ such that $|P_c(z)| > |z|(1 + \epsilon)$.

(b) Use induction to show that if $|z| \geq |c|$ and $|z| > 2$, then $P_c^n(z) \to \infty$.

(c) Apply (b) to prove $P_c^n(0) \to \infty$ (as $n \to \infty$) if and only if $|P_c^k(0)| > 2$ for some $k \in \mathbb{N}$. In particular, $M_d \subseteq \overline{\triangle(0, 2)}$.

(d) Prove that the filled in Julia set of P_c is contained in $\overline{\triangle(0, 2)}$ when $|c| \leq 2$. Also show by example that this statement does not necessarily hold when $|c| > 2$, i.e., it is not the case that for all c we have $P_c^n(z) \to \infty$ whenever $|z| > 2$.

EXERCISE 1.190. Symmetries in the family $P_c(z) = z^d + c$.

(a) Use the *Parameter Plane and Julia Set Applet* to explore the parameter plane for the family $P_c(z) = z^d + c$, and then identify and explain the symmetry you see there.

(b) Use the *Parameter Plane and Julia Set Applet* to explore the Julia sets of maps in the family $P_c(z) = z^d + c$, and then identify and explain the symmetry you see there.

EXERCISE 1.191. For fixed integer $d > 2$, prove M_d is the connectedness locus for the family of maps $P_c(z) = z^d + c$, i.e., show that $M_d = \{c \in \mathbb{C} : J(P_c) \text{ is connected}\}$.

Transcendental Dynamics

EXERCISE 1.192. Carefully prove that $E_c^n(z_0) \to \infty$ if and only if $\mathrm{Re}\, E_c^n(z_0) \to +\infty$, where $E_c(z) = ce^z$ and $c \in \mathbb{C} \setminus \{0\}$.

LARGE PROJECT 1.193. Accuracy of the Exponential Julia set applet pictures. This large project is open ended and should only be attempted after all of Section 1.6.1 has been covered.

As mentioned in Footnote 26 on p. 67, the algorithm used to illustrate the Exponential Julia set $J(E_{0.2})$ can color some seed values z_0 non-black (which it does when at least one of the orbit points $E_{0.2}(z_0), \ldots, E_{0.2}^{20}(z_0)$ lands in the set $\{\operatorname{Re} z > 50\}$), but which is actually not in $J(E_{0.2})$. For example, consider any point z_0 which maps to $E_{0.2}(z_0) = 51 + \pi i$. Such a z_0 will not be left black. However, we see that $E_{0.2}^2(z_0) = -0.2e^{51} \in H = \{z : \operatorname{Re} z < 1\} \subseteq A_{E_{0.2}}(p)$, which implies $z_0 \in A_{E_{0.2}}(p) = F(E_{0.2})$. The goal of this project is to estimate how close such points like z_0 are to actual points in $J(E_{0.2})$. This is in some sense a measure of the error in the algorithm.

One method to approach this problem is to consider an open rectangle R centered at z_0 and consider its **expanded** image $E_{0.2}(R)$. If this expanded image meets some point which truly does have an orbit which tends to ∞, then there must be a corresponding point in R which lies in $J(E_{0.2})$. So one must consider how much such an R gets expanded, and how big this expansion needs to be before we are guaranteed that expanded image $E_{0.2}(R)$ contains some point with an orbit which tends to ∞. Hint: Consider the open interval $(q, +\infty)$ and its $2\pi i k$-translates where q is the repelling fixed point for $E_{0.2}$.

Also, one should consider what similar estimates one can get if w_0 is a point such that $\operatorname{Re} E_{0.2}(w_0) < 50$, but $E_{0.2}^2(w_0) = 51 + \pi i$. Thus, such a point w_0 takes two steps before the algorithm gives it a non-black color, as opposed to z_0 above which used only one step. As above, w_0 is also not in $J(E_{0.2})$, but there are points nearby which are. Can you estimate how nearby these points are?

Some related questions are: Is there a relationship between the size of the viewing window and the accuracy of our pictures? Can you quantify this either in general or in specific cases? How does this relationship depend on the value 50 that was chosen to determine our escape criterion or on the value 20 that determined how many orbit points might be checked? If we set the **Dynamic plane min iterations** to 2 (which means that a seed value is iterated twice before the escape condition is tested, i.e., a seed z_0 is given a non-black color only if at least one of the points $z_2, \ldots, z_{20}$ has real part greater than 50), do we get a more accurate or less accurate picture?

This is a very technical project. The reader may wish to see [**10**] where some of these questions are addressed.

EXERCISE 1.194. Contracting and Expanding maps.
(a) Prove the Contraction Lemma stated on page 69. Hint: Consider the argument made just preceding the statement of the lemma.
(b) Show that if f is a one-to-one analytic function such that $|f'(z)| > \gamma > 1$ for all z in some domain D, and $f(D)$ is convex, then f is "expanding" on D, i.e., $|f(z) - f(w)| \geq \gamma |z - w|$ for all $z, w \in D$. Hint: Consider f^{-1}, which by the Inverse Function Theorem A.23 on page 416 is analytic.

EXERCISE 1.195. Prove that any two fingers (of any stage) constructed in Example 1.135 are always separated by an infinitely long gap (of black) points in $A_{E_{0.2}}(p)$. Hint: Note that the horizontal lines $K_n = \{z \in \mathbb{C} : \operatorname{Im} z = (2n + 1)\pi\}, n \in \mathbb{N}$ each lie

in $A_{E_{0.2}}(p)$ and separate the stage 1 fingers. With this show that the inverse image of these lines under $E_{0.2}$ separate the stage 2 fingers, and then proceed inductively.

EXERCISE 1.196. Follow the steps below to prove that the thickness of each stage n finger is no greater than π/γ^{n-1} where $\gamma = \ln 5 > 1$.

(a) Prove that any stage n finger is contained in the right half plane $\{\operatorname{Re} z \geq \gamma\}$ by showing $\operatorname{Re} E_{0.2}(z) \leq |E_{0.2}(z)| < 1$ whenever $\operatorname{Re} z < \gamma$.

(b) Suppose $E_{0.2}(\triangle(z_0, r)) \subseteq \{\operatorname{Re} z > \gamma\}$ for some $r < \pi$. Show that $E_{0.2}(\triangle(z_0, r)) \supseteq \triangle(E_{0.2}(z_0), \gamma r)$ by filling in the details of the following argument. Define a branch L of the inverse of $E_{0.2}$ on the convex set $\{\operatorname{Re} z > \gamma\}$ such that $L(E_{0.2}(z_0)) = z_0$ (given by $L(z) = \operatorname{Log} z + \ln 5 + 2\pi i n$ for some $n \in \mathbb{Z}$). For $|z - z_0| = r$, we then have $r = |z - z_0| = |L(E_{0.2}(z)) - L(E_{0.2}(z_0))| \leq (1/\gamma)|E_{0.2}(z) - E_{0.2}(z_0)|$ since $|L'(z)| = |1/z| < 1/\gamma$ when $\operatorname{Re} z > \gamma$. Noting that $E_{0.2}$ is univalent on $\triangle(z_0, r)$, it follows that $E_{0.2}$ maps the circle $|z - z_0| = r$ to a simple closed curve C which encloses $\triangle(E_{0.2}(z_0), \gamma r)$.[33] Applying the Maximum Modulus Theorem A.17 on page 414 to $L(z) - z_0$ on $E_{0.2}(\triangle(z_0, r))$ we see that $L(\triangle(E_{0.2}(z_0), \gamma r)) \subseteq \triangle(z_0, r)$, and the result then follows.

(c) Use parts (a) and (b) to prove that if $\triangle(z_0, r)$ is contained in a stage n finger (for $n \geq 2$), then $\triangle(E_{0.2}(z_0), \gamma r)$ is contained in some stage $(n-1)$ finger.

(d) Use Exercise 1.139 and induction to complete the proof.

EXERCISE 1.197. Prove that the set J in Proposition 1.140 cannot contain any open set.

EXERCISE 1.198. Follow the steps below to construct infinitely many (though not all) of the "hairs" in $J(E_{0.2})$. Recall that each hair is the image of some continuous map $h : [0, \infty) \to \mathbb{C}$ such that $h(t) \to \infty$ as $t \to \infty$.

(a) Prove that there exists a repelling fixed point $q \in \mathbb{R}$ such that for any $x > q$ we have $E_{0.2}(x) > x$ and therefore $E_{0.2}^n(x) \to +\infty$. Use this to conclude that the interval $h_0 = [q, +\infty)$ is a hair in $J(E_{0.2})$ (which we call a straight hair since it extends to ∞ in a straight line).

(b) Use the $2\pi i$ periodicity of $E_{0.2}$ to show that for each $k \in \mathbb{Z}$ the set h_k defined to be the $k2\pi i$ translate of h_0 is also a hair in $J(E_{0.2})$. We call each h_k a stage 1 hair (which could be called the "main" hair in the corresponding stage 1 finger).

(c) Argue that inside of the stage 1 finger C_0, there exist infinitely many hairs $h_{0,k}$ in $J(E_{0.2})$ such that $E_{0.2}(h_{0,k}) = h_k$. Note that no $h_{0,k}$ is a straight hair (except for $h_{0,0}$ which equals $h_0 = [q, +\infty)$).

(d) Use the $2\pi i$ periodicity of $E_{0.2}$ again to show that for each $j, k \in \mathbb{Z}$ the set $h_{j,k}$ defined to be the $j2\pi i$ translate of $h_{0,k}$, is also a hair in $J(E_{0.2})$. We call each $h_{j,k}$ a stage 2 hair.

---

[33]At this point one could complete the proof by applying the Argument Principle A.19 on page 414 to see that the number of zeros of $E_{0.2}(z) - E_{0.2}(a)$ is equal to one for all values a in the interior of C.

(e) Repeat the above arguments to argue that there must exist infinitely many *stages* of hairs in $J(E_{0.2})$.

(f) We remark that if we let H_n denote all the stage n hairs generated as above, then $\cup_{n=1}^{\infty} H_n \cup \{\infty\}$ is not quite all of $J(E_{0.2})$. However, with a bit of more careful analysis one can show that $J(E_{0.2}) = \overline{\cup_{n=1}^{\infty} H_n \cup \{\infty\}}$ by showing that each point in the hairs described in Section 1.6.2 is a limit of points from the collection of stage n hairs generated as above.

EXERCISE 1.199. Determine the exact real bifurcation value c^* between $c = 0.2$ and $c = 0.5$ for the family of maps E_c. Also, show that for $c < c^*$ we have $\{\operatorname{Re} z < 1\} \subseteq F(E_c)$, but for $c > c^*$ we have $F(E_c) = \emptyset$.

EXERCISE 1.200. For each $c \in L_1$, the map $E_c(z) = ce^z$ has an attracting fixed point with a corresponding multiplier $\lambda(c)$ (see Figure 1.27). Thus we have a multiplier map $\lambda : L_1 \to \triangle(0, 1)$.

(a) Find the inverse of the multiplier map.

(b) Use the inverse of the multiplier map to find the c value where the 7/13 bulb attaches to the cardioid L_1. Hint: What kind of fixed point will there be at this c value.

EXERCISE 1.201. Show that orbits under the maps $S_c(z) = c \sin z$, and $C_c(z) = c \cos z$ "escape" to ∞ in the direction of the positive or negative imaginary axis by showing the following.

(a) Show that $S_c^n(z) \to \infty$ if and only if $|\operatorname{Im} S_c^n(z)| \to +\infty$.

(b) Show that $C_c^n(z) \to \infty$ if and only if $|\operatorname{Im} C_c^n(z)| \to +\infty$.

EXERCISE 1.202. Show that neither $S_c(z) = c \sin z$ nor $C_c(z) = c \cos z$ has a finite asymptotic value by following the steps below.

(a) Examine $|\sin(x + iy)|$ to show that if S_c were to have a finite asymptotic value, then the curve γ (along which S_c has a finite asymptotic value) would have to be vertically bounded, that is, live in some horizontal strip $\{|\operatorname{Im} z| \leq M\}$.

(b) Show that $\sin z$ maps the imaginary axis into itself.

(c) Show that $\sin z$ maps the vertical line $\{\pi/2 + iy : y \in \mathbb{R}\}$ into $[1, +\infty)$.

(d) Use the fact that $\sin z$ is 2π periodic along with parts (a), (b) and (c) to show that $\sin z$ cannot have a finite asymptotic value.

(e) Show that S_c cannot have a finite asymptotic value since $\sin z$ does not.

(f) Show that C_c cannot have a finite asymptotic value since $C_c(z) = S_c(z + \pi/2)$.

EXERCISE 1.203. Trig functions with one critical orbit.

(a) Explain in what sense $\{S_c^n(c)\}$ is the only critical orbit of the map S_c.

(b) Explain in what sense $\{C_c^n(c)\}$ is the only critical orbit of the map C_c.

Definitions and Properties of the Julia and Fatou sets

EXERCISE 1.204. Let a_n and b_n for $n \geq 0$ be sequences of positive real numbers such that $1 = a_0 > b_0 > a_1 > b_1 > \ldots$ and $a_n \to 0$ (which also implies $b_n \to 0$).

(a) Construct a strictly increasing function $g : [0, +\infty) \to [0, +\infty)$ such that g is differentiable on $(0, +\infty)$ and we have $g(a_n) = a_n$ and $g(b_n) = b_n$ with $g'(a_n) = 0$ and $g'(b_n) = 2$. Furthermore, construct g so that g does not fix any points other than the a_n and b_n and 0. Hence each a_n is a super attracting fixed point of g and each b_n is a repelling fixed point of g. Sketch a graph of such a g to convince yourself that this can be done. It is not necessary that you produce a formula for this map g.

(b) Define a function $f : \mathbb{C} \to \mathbb{C}$ given by $f(re^{i\theta}) = g(r)e^{i\theta}$. Consider the dynamics of points both on and near the circles $|z| = a_n$ and $|z| = b_n$ to see that each "repelling" circle $|z| = b_n$, which is fixed by f, must lie in $J(f)$.

(c) Show that $0 \in F(f)$. Hint: Use the fact that $f(\Delta(0, a_n)) \subseteq \Delta(0, a_n)$ for each $n \geq 0$.

(d) Show that (b) and (c), together with the fact that $b_n \to 0$, implies that 0 is not in the interior of $F(f)$.

EXERCISE 1.205. Use the definitions of Fatou and Julia set to prove the statements in Proposition 1.211 parts (a)-(d). Hint: For the complete invariance statements, you can use the fact that, when it is non-constant, f is an *open* map[34], i.e., if U is an open set in the domain of f, then the image $f(U)$ is also an open set in $\overline{\mathbb{C}}$. For (d), use Montel's theorem 1.208 and part (c).

EXERCISE 1.206. Show that the set of repelling cycles for the map $f(z) = z^2$ is dense in $J(f) = C(0, 1)$ by explicitly solving for the set of p-periodic points for each $p \in \mathbb{N}$.

---

[34]See the Open Mapping Theorem A.22 on page 416.

1.A. Appendix - Definitions and Properties of the Julia and Fatou sets

Here we present the formal definitions of the Fatou set and Julia set of a rational or entire map. We also a provide the statement of Montel's theorem, a major tool in complex dynamics, along with a proposition stating some of the properties of the Fatou and Julia sets.

Recall that $\sigma(z, w)$ denotes the spherical distance between the points $z, w \in \overline{\mathbb{C}}$ (see Appendix Section B.2 on page 420).

DEFINITION 1.207. Let f be a rational or entire map. The Fatou set is the set $F(f) = \{z \in \overline{\mathbb{C}} : \text{ for every } \epsilon > 0 \text{ there exists } \delta > 0 \text{ such that } \sigma(z, w) < \delta \text{ implies } \sigma(f^n(z), f^n(w)) < \epsilon \text{ for all } n \in \mathbb{N}\}$.

Thus, if $z \in F(f)$ and ϵ is small, then for w to have an orbit ϵ-*similar* to the orbit of z (by which we mean that corresponding orbit points are never more than ϵ apart) we just need to choose w close enough to z, i.e., within a distance δ. Another way to interpret this definition is to say that $F(f)$ is the set of points z such that for any $\epsilon > 0$ there exists $\delta > 0$ such that $f^n(\triangle_\sigma(z, \delta)) \subseteq \triangle_\sigma(f^n(z), \epsilon))$ for *all* $n \in \mathbb{N}$, i.e., a "tiny" neighborhood of z (of size δ measured with the spherical metric) will have an orbit that "stays tiny" (of size no greater than ϵ measured with the spherical metric) along the *entire* orbit (see Figure 1.33). The reader should reflect on how this fails to happen with the map $f(z) = z^2$ for any neighborhood of a point on the unit circle $C(0, 1)$ and also compare this definition to your own definition given earlier.

A major tool used in dynamics is *Montel's theorem* (see [**2**], p. 9, where it is stated in terms of the notion of *normal families*), which we state without proof in a setting most easily applied to the dynamics of interest to us here.

THEOREM 1.208 (Montel's theorem). Let $U \subseteq \overline{\mathbb{C}}$ be an open set in the domain set of a rational or entire map f. If the family of maps $\{f^n : n \in \mathbb{N}\}$ *omits* any three given points $z, w, v \in \overline{\mathbb{C}}$, i.e., $f^n(U) \cap \{z, w, v\} = \emptyset$ for all $n \in \mathbb{N}$, then $U \subseteq F(f)$. In particular, if $f(U) \subseteq U$ and $\overline{\mathbb{C}} \setminus U$ contains three or more points, then $U \subseteq F(f)$.

Montel's theorem can the make it very quick to show that $f(z) = z^2$ has $F(f) = \overline{\mathbb{C}} \setminus C(0, 1) = \Delta(0, 1) \cup (\overline{\mathbb{C}} \setminus \overline{\Delta(0, 1)})$. Since $f(\Delta(0, 1)) \subseteq \Delta(0, 1)$, Montel's theorem gives $\Delta(0, 1) \subseteq F(f)$. Similarly, one can show $\overline{\mathbb{C}} \setminus \overline{\Delta(0, 1)} \subseteq F(f)$. Also, from our work in Section 1.3.2 or by considering the above definition of Fatou set, we know that $C(0, 1)$ does not contain any points in the Fatou set $F(f)$. Hence, we conclude $F(f) = \overline{\mathbb{C}} \setminus C(0, 1)$.

REMARK 1.209. Though its proof is too advanced for this text, it is true that if f is a rational or entire map and $z_0 \in F(f)$, then there exists $r > 0$ such that $\triangle_\sigma(z_0, r) \subseteq F(f)$, i.e., $F(f)$ is an open set. However, if f is not rational or entire, then the set $F(f)$ need not be open. To see this, consider the map $f(z)$ which is defined to be zero when $z = x + iy$ with both $x, y \in \mathbb{Q}$ and $f(z) = z$ otherwise. In this case
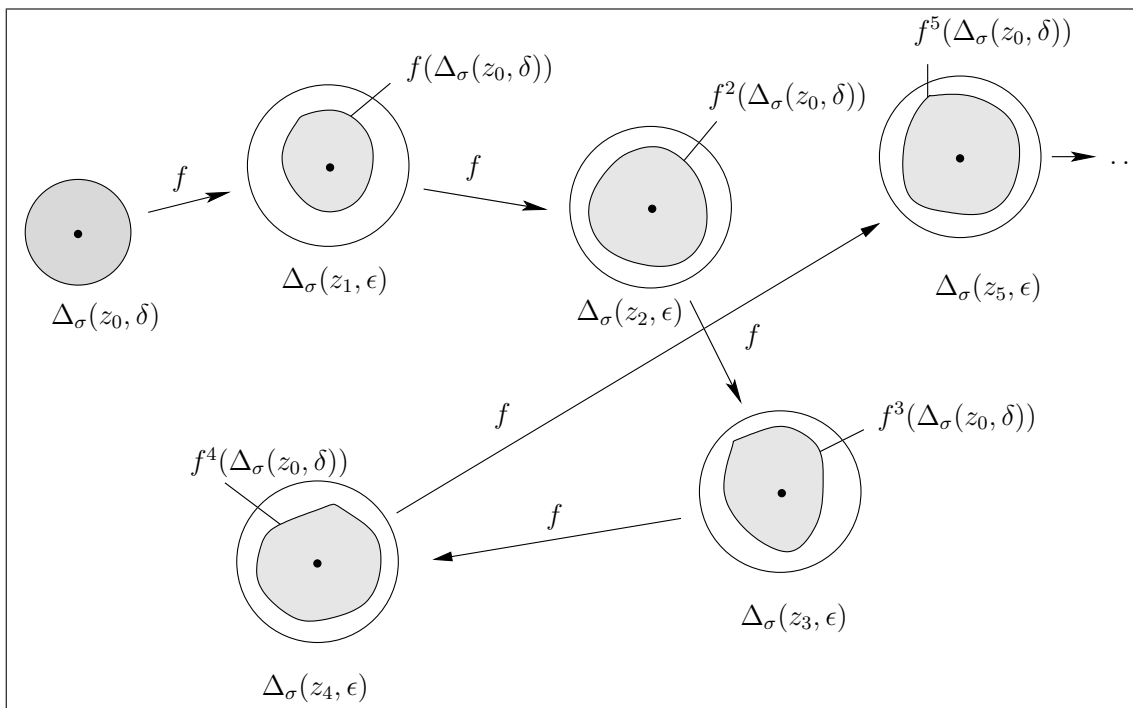
FIGURE 1.33. This picture illustrates how, for a point $z_0 \in F(f)$, a tiny neighborhood $\Delta_\sigma(z_0, \delta)$ has a forward image that stays tiny (within an ϵ neighborhood of each orbit point z_n) for the entire orbit. We illustrated this up to the fifth iterate, but the reader should take note of the last arrow with the dots indicating that this happens for *all* iterates f^n.

$f^n = f$ for all $n \in \mathbb{N}$ and $F(f) = \{0\}$. See also Additional Exercise 1.204 for such an example where f is continuous (but not analytic).

DEFINITION 1.210. The Julia set of a rational or entire map f is defined to be

$$J(f) = \overline{\mathbb{C}} \setminus F(f).$$

Hence, for a point z to be in $J(f)$ there must be points w which are arbitrarily close to z, but which fail to have orbits similar to the orbit of z. More precisely, for a point z to be in $J(f)$ there must be some $\epsilon > 0$ such that for every $\delta > 0$, there exists a point w within a distance δ of z such that the orbit of w is NOT ϵ-*similar* to the orbit of z.

Directly from the definitions and discussion above one can show the following properties (a)-(d) (see Additional Exercise 1.205). Properties (e) and (f), however, requires some tools from complex analysis which are a bit beyond the level of this text (see [**1**], p. 148, and [**24**], p. 38).

PROPOSITION 1.211 (Properties of the Julia and Fatou sets). Let f be a rational or entire map. Then we have the following:

(a) $F(f)$ is an open set in $\overline{\mathbb{C}}$ and thus $J(f)$ is a closed set in $\overline{\mathbb{C}}$.
(b) $F(f)$ is completely invariant, i.e., if $f(F(f)) \subseteq F(f)$ and $f^{-1}(F(f)) \subseteq F(f)$.
(c) $J(f)$ is completely invariant, i.e., if $f(J(f)) \subseteq J(f)$ and $f^{-1}(J(f)) \subseteq J(f)$.
(d) $J(f)$ contains an open set if and only if $J(f) = \overline{\mathbb{C}}$.
(e) The set of repelling periodic cycles of f is dense in $J(f)$. That is, each repelling periodic cycle of f is in $J(f)$ and for every open set U which intersects $J(f)$, there is a repelling periodic point z which lies in U.
(f) Let A be the set of z in $J(f)$ such that the orbit of z is *dense* in $J(f)$ (i.e., for every open set U which intersects $J(f)$, there is an orbit point z_n which lies in U). Then A is *dense* in $J(f)$ (i.e., for every open set V which intersects $J(f)$, there is a point $z \in A$ which lies in V).

REMARK 1.212. Additional Exercises 1.206 and 1.171 illustrate properties (e) and (f) explicitly for the map $f(z) = z^2$. Reflecting for a moment, we see, both in the $f(z) = z^2$ example and in general, the Julia set contains a dense set of points which in some sense are the ultimate in regular behavior. These are the periodic points - nothing could be more regular and predictable than to have the orbit follow the same finite set of points over and over again. However, the Julia set also contains points with a dense orbit in $J(f)$ which means that it never settles into any type of regularity. Thus we see that inside the sensitive dependence which defines chaos (and thus defines the Julia set $J(f)$) there lies a strange interwoven mix of regularity and unpredictability.

1.B. Appendix - Global Conjugation and Möbius map dynamics

In this appendix we show that global conjugation (see Section 1.2.6) can be used to quickly classify and understand the dynamics of any Möbius map, i.e., a map of the form $f(z) = \frac{az+b}{cz+d}$ where $ad - bc \neq 0$. The classification depends on the number and type of fixed points of f and so we begin with the following proposition.

PROPOSITION 1.213. Non identity Möbius maps can only have exactly one or exactly two fixed points in $\overline{\mathbb{C}}$.

PROOF. Let $f(z) = \frac{az+b}{cz+d}$ be a non identity Möbius map. First suppose that $f(\infty) \neq \infty$, i.e, $c \neq 0$. Solving for fixed points of f, i.e., solving the equation $f(z) = z$, yields the equation $cz^2 + (d - a)z - b = 0$, which has two distinct roots or one double root in $\mathbb{C}$. If $c = 0$, then f is a linear map which must have one fixed point at ∞ and possibly a second fixed point in $\mathbb{C}$. $\square$

We now describe the dynamics of a non identity Möbius map f based on how many fixed points it has.

Case 1: Suppose $f(z)$ fixes only ∞. Then one can quickly show that f has the form $f(z) = z + \beta$ for some $\beta \in \mathbb{C} \setminus \{0\}$. Thus $f^n(z) = z + n\beta$, and hence $f^n(z) \to \infty$

as $n \to \infty$ for all $z \in \overline{\mathbb{C}}$ much in the same way as Example 1.50. We also see that $J(f) = \{\infty\}$ in this case.

Case 2: Suppose $f(z)$ **fixes only** $w \in \mathbb{C}$. Let $\psi(z) = \frac{1}{z-w}$ and define $g(z) = \psi f \psi^{-1}(z)$, noting that g is also Möbius. Furthermore, the map $g(z)$ has only one fixed point at ∞ (see Exercise 1.30). So from Case 1 we see that $g(z) = z + \alpha$ for some $\alpha \in \mathbb{C} \setminus \{0\}$ and $g^n(z) \to \infty$ for all $z \in \overline{\mathbb{C}}$. Hence $f^n(z) = \psi^{-1} g^n \psi(z) \to \psi^{-1}(\infty) = w$ for all $z \in \overline{\mathbb{C}}$. So, if f is Möbius with unique fixed point w, then $f^n(z) \to w$ for all $z \in \overline{\mathbb{C}}$ and $J(f) = \{w\}$.

Case 3: Suppose that f **fixes** 0 **and** ∞.

Then $f(z) = kz$ for some $k \in \mathbb{C} \setminus \{1\}$ (verify) and thus $f^n(z) = k^n z$. We now have the following categories based on $|k|$.

(1) If $|k| < 1$, then $f^n(z) \to 0$ for all $z \in \overline{\mathbb{C}} \setminus \{\infty\}$ and $J(f) = \{\infty\}$.

(2) If $|k| > 1$, then $f^n(z) \to \infty$ for all $z \in \overline{\mathbb{C}} \setminus \{0\}$ and $J(f) = \{0\}$.

(3) If $|k| = 1$, then f is simply a rotation $z \mapsto e^{i\theta} z$ for some $\theta \in (0, 2\pi)$ whose dynamics are easy to understand. In particular, $F(f) = \overline{\mathbb{C}}$.

Case 4: Suppose that f **fixes** w_1 **and** w_2 **where** $w_1 \neq w_2$. Defining $\psi(z) = \frac{z-w_1}{z-w_2}$ and setting $g(z) = \psi f \psi^{-1}(z)$, we see that g falls into Case 3. As in Case 2, we can now understand the dynamics of f as a change of coordinates of the simple dynamics of g by noting that $f^n(z) = \psi^{-1} g^n \psi(z)$. In particular, we have one of the following:

(1) $f^n(z) \to w_1$ for all $z \in \overline{\mathbb{C}} \setminus \{w_2\}$ and $J(f) = \{w_2\}$.

(2) $f^n(z) \to w_2$ for all $z \in \overline{\mathbb{C}} \setminus \{w_1\}$ and $J(f) = \{w_1\}$.

(3) f is conjugate to a rotation $z \mapsto e^{i\theta} z$ for some $\theta \in (0, 2\pi)$, and thus $F(f) = \overline{\mathbb{C}}$.

REMARK 1.214. In Cases 2 and 4, we see that the Möbius map $\psi(z)$ was used to move the fixed points of f to more convenient locations so that the simple dynamics of Cases 1 and 3 could be related to the dynamics of f. This is one of the great advantages to using global conjugation; it allows us to reposition special points in more convenient places before we do our analysis. Not only can this technique be used to simplify the analysis of Möbius map dynamics, but we can also use it with higher degree maps as well. See Section 1.2.7 where a conjugation of the Newton map of a quadratic function greatly simplifies the analysis. Also see Additional Exercises 1.165–1.168.

1.C. Appendix - Code for drawing random dynamics pictures

Because of its complexity, the following computationally heavy algorithm is not well suited for an applet. However, we provide the code for Matlab, which one could use to quickly generate pictures like the Devil's Colosseum in Figure 1.31. These four files main.m, f.m, g.m, and proced2.m need to be created separately, but stored in the same folder. When main.m is compiled using Matlab, a picture will be generated.

```
% ********* This begins the file main.m *********************
% This file requires the files (functions) g.m, f.m, and proced2.m to be
% in the same folder when this file is compiled.  Upon compilation
% this file will generate a graph of the probability that a random orbit
% generated by the maps f and g will escape to infinity.

h = 0.1; % Determines step size in mesh of points to be plotted
Maxd = 10; % Determines how many random steps can be taken

cntx = 0; % Counter for x coordinate
clear x1;
for x = -5:h:5
 cntx = cntx + 1;
 cnty = 0; % Counter for y coordinate
 clear y1;
  for y = -5:h:5
    cnty = cnty + 1;
    n = 0;
    x1(cntx) = x;
    y1(cnty) = y;
    [z(cntx,cnty), n] = proced2(x, y, 1, n, Maxd);
  end
end

% % Uncomment the following lines to export/save data sets containing
% % the x(i) coordinates, y(j) coordinates, and z(i,j) values which
% % when plotted make up the graph.  These data sets can then be imported
% % into another application (such as Maple or Mathematica) and plotted and
% % otherwise manipulated there.
%
% save DevilDataX3.dat x1 -ASCII;
% save DevilDataY3.dat y1 -ASCII;
% save DevilDataZ3.dat z -ASCII;

figure
surfc(x1, y1, z)
lighting phong
shading interp %flat %interp

clear;
% ********* This ends the file main.m *********************
```

100

```
% ********* This begins the file f.m ********************
function z = f(x,y)
z1 = [x*x-y*y-1,2*x*y];
z = [z1(1)*z1(1)-z1(2)*z1(2)-1,2*z1(1)*z1(2)];
% ********* This ends the file f.m ********************

% ********* This begins the file g.m ********************
function z = g(x,y)
z1 = [(x*x-y*y)/4,x*y/2];
z = [(z1(1)*z1(1)-z1(2)*z1(2))/4,z1(1)*z1(2)/2];
% ********* This ends the file g.m ********************

% ********* This begins the file proced2.m ********************
% This recursively defined procedure/function will compute the value
% s which for proced2(x, y, 1, n, Maxd) (here I used q=1) represents the
% probability that the point (x,y) will have a random orbit
% escape (have modulus > K) in the first M steps of the orbit.

function [s, n] = proced2(x, y, q, n0, M)
n = n0 + 1;
p = 0.5; % Probability that map f is chosen at each step
K = 10;  % Escape radius for random orbit
s = 0.0;
z=f(x,y); % file f.m must be in folder next to this file
w=g(x,y); % file g.m must be in folder next to this file

if n < M  % M is the max number of recursion steps allowed
    if z(1)*z(1)+z(2)*z(2) > K*K
        s = s + p*q;
    else
        [d, n] = proced2(z(1), z(2), p*q, n, M);
        s = s + d;
    end

    if w(1)*w(1)+w(2)*w(2) > K*K
        s = s + (1-p)*q;
    else
        [d, n] = proced2(w(1), w(2), (1-p)*q, n, M);
        s = s + d;
    end
end
% ********* This ends the file proced2.m ********************
```

Bibliography

[1] Alan F. Beardon. *Iterations of Rational Functions*. Springer-Verlag, New York, 1991.

[2] Lennart Carleson and Theodore W. Gamelin. *Complex Dynamics*. Springer-Verlag, New York, 1993.

[3] Ruel V. Churchill and James Ward Brown. *Complex variables and applications*. McGraw-Hill Book Co., New York, eighth edition, 2009.

[4] M. Viana da Silva. The differentiability of the hairs of e$xp(Z)$. *Proc. Amer. Math. Soc.*, 103(4):1179–1184, 1988.

[5] R. L. Devaney. Complex exponential dynamics. To appear in Volume 3 of the Handbook of Dynamical Systems.

[6] R. L. Devaney and M. Moreno Rocha. Geometry of the antennas in the Mandelbrot set. *Fractals*, 10(1):39–46, 2002.

[7] Robert L. Devaney. *A first course in chaotic dynamical systems*. Addison-Wesley Studies in Non-linearity. Addison-Wesley Publishing Company Advanced Book Program, Reading, MA, 1992. Theory and experiment, With a separately available computer disk.

[8] Robert L. Devaney. A survey of exponential dynamics. In *Proceedings of the Sixth International Conference on Difference Equations*, pages 105–122, Boca Raton, FL, 2004. CRC.

[9] Robert L. Devaney and Linda Keen, editors. *Chaos and fractals*, volume 39 of *Proceedings of Symposia in Applied Mathematics*. American Mathematical Society, Providence, RI, 1989. The mathematics behind the computer graphics, Lecture notes prepared for the American Mathematical Society Short Course held in Providence, Rhode Island, August 6–7, 1988, AMS Short Course Lecture Notes.

[10] Marilyn B. Durkin. The accuracy of computer algorithms in dynamical systems. *Internat. J. Bifur. Chaos Appl. Sci. Engrg.*, 1(3):625–639, 1991.

[11] K. Falconer. *Fractal geometry - Mathematical Foundations and Applications*. John Wiley & Sons, Chichester, 1990.

[12] P. Fatou. Sur les équations fonctionnelles. *Bull. Soc. Math. France*, 47:161–271, 1919.

[13] P. Fatou. Sur les équations fonctionnelles. *Bull. Soc. Math. France*, 48:208–314, 1920.

[14] P. Fatou. Sur les équations fonctionnelles. *Bull. Soc. Math. France*, 48:33–94, 1920.

[15] James Gleick. *Chaos*. Penguin Books, New York, 1987. Making a new science.

[16] John G. Hocking and Gail S. Young. *Topology*. Dover Publications Inc., New York, second edition, 1988.

[17] Gaston Julia. Mémoire sur l'itération des fonctions rationnelles. *Journal de Math. Pure et Appl.*, pages 47–245, 1918.

[18] Bogusława Karpińska. Area and Hausdorff dimension of the set of accessible points of the Julia sets of λe^z and $\lambda \sin z$. *Fund. Math.*, 159(3):269–287, 1999.

[19] Boguslawa Karpińska. Hausdorff dimension of the hairs without endpoints for $\lambda \exp z$. *C. R. Acad. Sci. Paris Sér. I Math.*, 328(11):1039–1044, 1999.

[20] Tan Lei. *Ressemblance eintre l'ensemble de Mandelbrot set l'ensemble de Julia au voisinage d'un point de Misiurewicz*, pages 139–152. in Étude dynamique des polynômes complexes. Partie II by A. Douady and J. H. Hubbard.

[21] John C. Mayer. An explosion point for the set of endpoints of the Julia set of $\lambda \exp(z)$. *Ergodic Theory Dynam. Systems*, 10(1):177–183, 1990.

[22] Curtis T. McMullen. The Mandelbrot set is universal. In *The Mandelbrot set, theme and variations*, volume 274 of *London Math. Soc. Lecture Note Ser.*, pages 1–17. Cambridge Univ. Press, Cambridge, 2000.

[23] Rich Stankewitz, W. Conatser, T. Butz, B. Dean, Y. Li, and K. Hart. *JULIA 2.0 Fractal Drawing Program*. http://www.bsu.edu/web/rstankewitz/JuliaHelp2.0/Julia.html.

[24] Norbert Steinmetz. *Rational iteration*, volume 16 of *de Gruyter Studies in Mathematics*. Walter de Gruyter & Co., Berlin, 1993. Complex analytic dynamical systems.

[25] Hiroki Sumi. Random complex dynamics and semigroups of holomorphic maps. Preprint 2008, http://arxiv.org/abs/0812.4483.

CHAPTER 2

Soap films, Differential Geometry, and Minimal Surfaces

Michael Dorff

2.1. Introduction

Minimal surfaces are beautiful geometric objects with interesting properties that can be studied with the help of computers. Some standard examples of minimal surfaces in $\mathbb{R}^3$ are the plane, Enneper's surface, the catenoid, the helicoid, and Scherk's doubly periodic surface (see Figure 2.1; note that the images shown are just part of these surfaces and that each surface actually continues on forever).
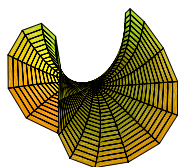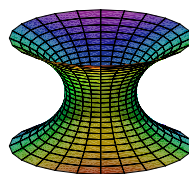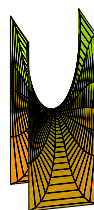

Enneper surface   catenoid
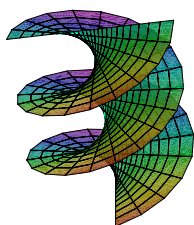
helicoid   Scherk's doubly periodic surface

Figure 2.1. Examples of some minimal surfaces

Minimal surfaces are related to soap films that result when a wire frame is dipped in soap solution. To get a sense of this connection, consider the following problem.

Steiner Problem: Four houses are located so that they form the vertices of a square that has sides of length one mile. These neighbors want to connect their houses with a road of least length. What should the shape of the road be?

FIGURE 2.2. What is the shortest path connecting these 4 vertices?

Some possible solutions include the following:

Length = 3   Length = 3   Length ≈ 3.41   Length ≈ 2.83
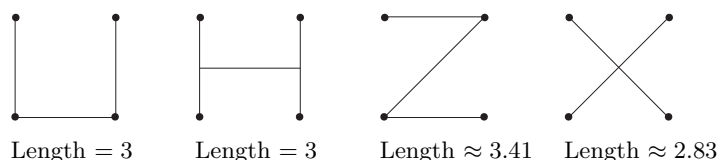
However, none of these is the solution. The correct solution has a length of $1 + \sqrt{3} \approx 2.7$ miles (see Figure 2.3). For more information about Steiner problems see [5] or [13].
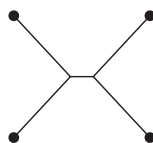
FIGURE 2.3. The shortest path connecting these 4 vertices.

How can we generalize this problem? One way is to have n-vertices. So the problem becomes, given n cities find a connected system of straight line segments of shortest length such that any pair of cities is connected by a path of line segments.

106

| A n-vertices Steiner problem | The solution |
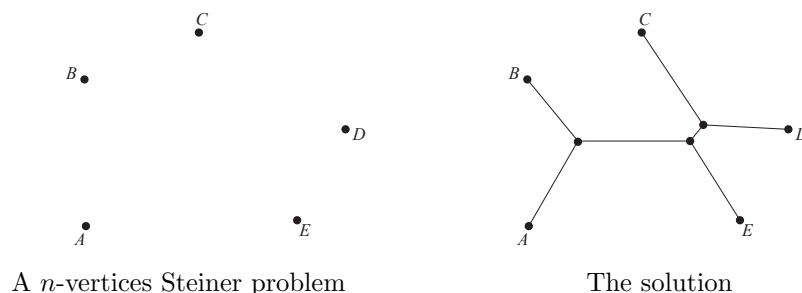
Another way to generalize this idea is to move up a dimension. What is the analogue of the Steiner problem in one dimension higher? The Steiner Problem minimizes distance (1-dimensional object) in a plane (2-dimensional object). Soap film minimizes area (2-dimensional object) in space (3-dimensional object).

Is there a connection between these two- and three-dimensional optimization problems? Consider the soap film created by dipping a cube frame into soap solution shown in Figure 2.4. The soap film creates the minimum possible surface area for a surface with a cube as its boundary. If we projected this surface onto the plane, the resulting shape would look like the solution to the Steiner problem with four vertices shown in Figure 2.3.
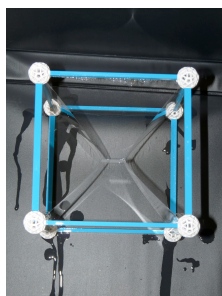
FIGURE 2.4. Soap film formed by a cube.

Water molecules exert a force on each other. Near the surface of the water there is a greater force pulling the molecules toward the center of the water. This force creates surface tension which tends to minimize the surface area of the shape. Soap solution has a lower surface tension than water and this permits the formation of soap films which also tend to minimize geometric properties such as length and area. For more information along this line see [**22**].

Minimal surfaces can be created by dipping wire frames into soap solution.

EXAMPLE 2.1. By dipping into soap solution a wire frame of a "slinky" (or helix) with a straw connecting the ends of the "slinky", we can create part of the minimal surface known as the helicoid.
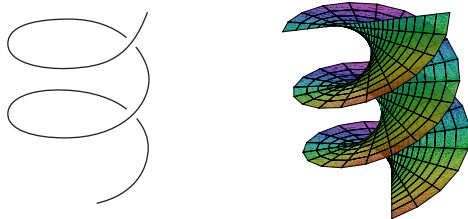
107

FIGURE 2.5. The wire frame of a slinky can be used to create part of the helicoid.

EXAMPLE 2.2. By dipping a 3-dimensional version of the wire frame shown below (a box frame missing two parallel edges on the top and two parallel edges on the bottom) into soap solution we can create part of the minimal surface known as Scherk's doubly periodic surface.
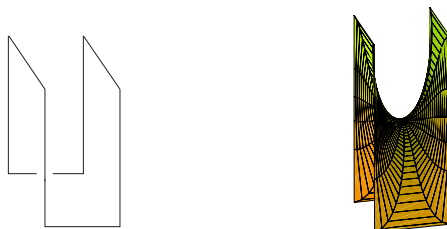

FIGURE 2.6. The wire frame of a box missing 4 edges can be used to create part of Scherk's doubly periodic surface.

EXPLORATION 2.3. Each of the minimal surfaces shown in Figure 2.1 can be formed by dipping a wire frame in soap solution. Determine the shape of the wire frame that creates: (a) Enneper's surface; and (b) the catenoid.
Try it out!

REMARK 2.4. To get a soap film of the part of Enneper's surface shown in Figure 2.1, we can dip a wire frame that matches the seams along a baseball. In fact, dipping such a wire frame in soap solution produces *two* minimal surfaces. The first is one half of the sphere of a baseball and the other is the complementary half of the "baseball." What is interesting is that if you start with one, you can deform it into the other by slightly and carefully blowing air into the soap film. There is a third, mysterious and unseen minimal surface one passes through while doing this, and this minimal surface is

unstable. In other words, it cannot actually exist or remain in existence–disturbances cause it to pop or "wiggle" into another surface.

One area of minimal surface theory that has seen a lot of interest and results recently is the study of complete embedded minimal surfaces. Basically, these are minimal surfaces that are boundaryless (complete) and have no self-intersections (embedded). The plane, the catenoid, the helicoid, and Scherk's doubly periodic surface are examples of complete embedded minimal surfaces. However, the Enneper surface is not embedded, because it has self-intersections as its domain increases (see Exploration 2.11).

To begin to understand minimal surfaces, we need some tools from differential geometry, and these are discussed in Section 2. Section 3 uses material from the previous section to define a minimal surface and discuss some examples and properties of minimal surfaces. Section 4 brings in complex analysis to study minimal surfaces and introduces the Weierstrass representation formula to efficiently describe and study properties of minimal surfaces. These three sections are fundamental and should be read first. In Sections 5 - 7, we begin to explore ideas that lead to beginning research problems for students. Sections 5 and 6 are independent of one another. In Section 5 we present the Weierstrass representation in the form of the Gauss map and height differential, which is the basis for much of the current research about minimal surfaces in $\mathbb{R}^3$. Section 6 connects ideas about minimal surfaces with planar harmonic mappings in geometric function theory (i.e., the study of complex analysis from a geometric viewpoint). Section 7 is a new area of investigation that combines the ideas of the previous two sections and has several problems that can be explored by beginning students. In this chapter, there are four applets used and they can be accessed online at http://www.jimrolf.com/explorationsInComplexVariables/chapter2.html:

- *DiffGeomTool* is used to visualize and explore basic differential geometry concepts in $\mathbb{R}^3$ such as the graph of a parametrization of a surface, curves on a surface, tangent planes on a surface, and unit normals on a surface.
- *MinSurfTool* is used to visualize and explore minimal surfaces in $\mathbb{R}^3$ by using various forms of the Weierstrass representation.
- *ComplexTool* is used to plot the image of domains in $\mathbb{C}$ under complex-valued functions.
- *LinComboTool* is used to plot and explore the convex combination of complex-valued harmonic polygonal maps.

Each section of this chapter contains examples, exercises, and explorations that involve using the applets. You should do all of the exercises and explorations, many of which present surfaces and concepts that will be used later in the chapter (there are additional exercises at the end of the chapter). In addition, there are short projects and long projects that are suitable as research problems for undergraduates to explore. The goal of this chapter is not to give a comprehensive or step-by-step approach to this topic, but rather to get the reader engaged with the general notions, questions, and techniques of the area – but even more so, to encourage the reader to actively pose as

well as pursue their own questions. To better understand the nature and purpose of this text, the reader should be sure to read the Introduction at the beginning of this book before proceeding.

2.2. Differential Geometry

Our goal is to develop the mathematics necessary to investigate minimal surfaces in $\mathbb{R}^3$. Such minimal surfaces minimize area locally and can be thought of as saddle surfaces. At each point, the bending upward in one direction is matched with the bending downward in the orthogonal direction. Such bending is known mathematically as curvature. So, to initially understand and investigate minimal surfaces, we must first understand curvature, a quantity that is measured using the tools of differential geometry. Differential geometry is a field of mathematics in which the ideas and techniques of calculus are applied to geometric shapes.

We will begin our discussion of differential geometry by looking at a surface in $\mathbb{R}^3$. Every point on a surface $M \subset \mathbb{R}^3$ can be designated by a point, $(x, y, z) \in \mathbb{R}^3$, but it can also be represented by two parameters. Let D be an open set in $\mathbb{R}^2$. Then the surface M can be represented by a function $\mathbf{x} : D \to \mathbb{R}^3$, where $\mathbf{x}(u, v) = (x_1(u, v), x_2(u, v), x_3(u, v))$ (that is, M is the image of $\mathbf{x}(D)$). We will require that $\mathbf{x}$ be differentiable. That is, each coordinate function $x_k(u, v)$ has continuous partial derivatives of all orders in D. Such a function or mapping is called a *parametrization*.

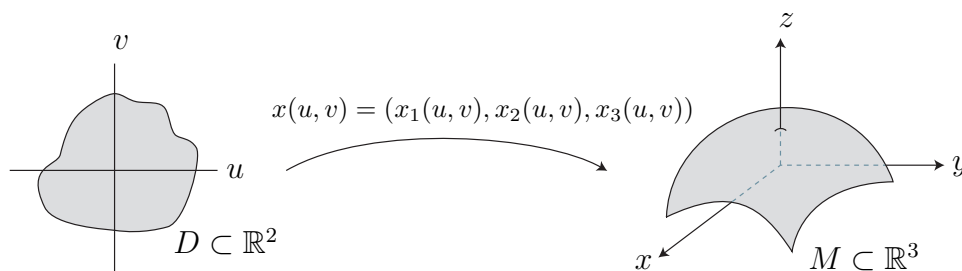$$x(u, v) = (x_1(u, v), x_2(u, v), x_3(u, v))$$

FIGURE 2.7. The parameterization of a surface

Let's consider two examples.

EXAMPLE 2.5. The Enneper surface is a minimal surface formed by bending a disk into a saddle surface. It can be parametrized by

$$\mathbf{x}(u, v) = \left(u - \frac{1}{3}u^3 + uv^2, v - \frac{1}{3}v^3 + u^2 v, u^2 - v^2 \right),$$

where u, v are in a disk of radius r. We can use the applet, *DiffGeomTool*, to graph this parametrization of the Enneper surface (see Figure 2.8). Open *DiffGeomTool* and

enter the coordinate functions of the parametrization as

$$X(u, v) = u - 1/3 * u \wedge 3 + u * v \wedge 2$$
$$Y(u, v) = v - 1/3 * v \wedge 3 + u \wedge 2 * v$$
$$Z(u, v) = u \wedge 2 - v \wedge 2$$

into the appropriate boxes. In the gray part on the right side of the applet, click on `Circular grid` with radius min: 0.0, radius max: 1.0, theta min: 0.0, and theta max: $2*$ pi. This is because we want our u, v values to be the unit disk. Then click the **Graph** button. To rotate the graph, place the cursor arrow on the image of the surface, and then click on and hold the left button on the mouse as you move the cursor. To increase the size of the image of the surface click on the left button on the mouse; to decrease the size, click on the right mouse button.
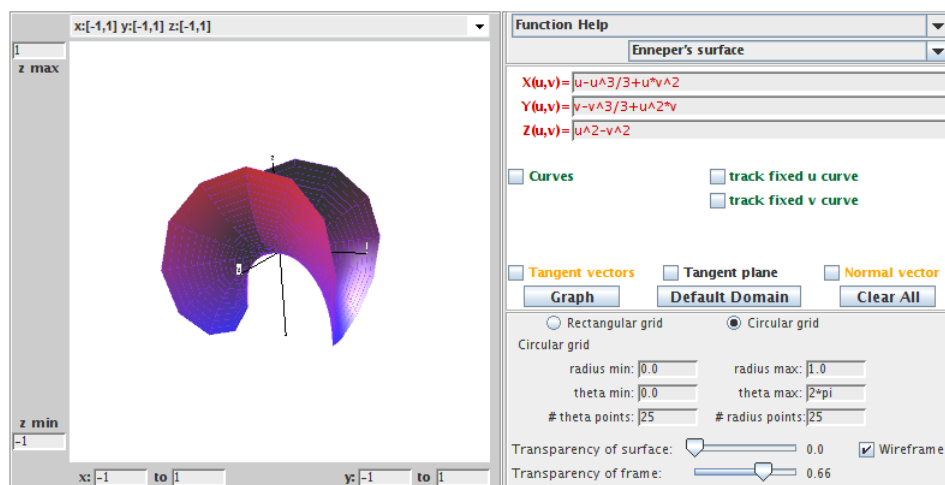

FIGURE 2.8. The Enneper surface.

EXAMPLE 2.6. If a heavy flexible cable is suspended between two points at the same height, then it takes the shape of a curve that can be described mathematically by the function $\alpha(t) = a \cosh(t/a)$. Such a curve is called a catenary from the Latin word that means "chain". In calculus, we discuss rotating a curve about a line to get a surface of revolution. We can apply this idea to the catenary curve to get a surface known as the catenoid. In particular, a *catenoid* is a surface that can be generated by rotating a catenary on its side about the x_3-axis (see Figure 2.9). A catenoid is also a minimal surface as we will verify in Section 2.3. How do we parametrize this catenoid? If we let $x_1 = a \cosh v$ ($-\infty < v < \infty$) and $x_3 = av$, then $r(v) = (a \cosh v, av)$ is a parametrization of the catenary curve on its side in the $x_1 x_3$-plane. Rotating a line about an axis is a circular motion, and a circle can be parametrized by $(\cos u, \sin u)$. So, we can parametrize this rotation of the catenary curve about the x_3-axis by multiplying
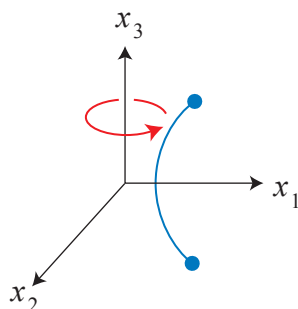
FIGURE 2.9. Creating a catenoid by rotating a catenary.

$a \cosh v$ by $\cos u$ for the x_1-coordinate function, and multiplying $a \cosh v$ by $\sin u$ for the x_2-coordinate function. Hence, we get the following parametrization for this catenoid surface:

$$\mathbf{x}(u, v) = \Big(a \cosh v \cos u, a \cosh v \sin u, av \Big).$$

Using *DiffGeomTool*, we can graph this parametrization of a catenoid with $a = 1$, clicking on `Rectangular grid`, and setting the boxes to $0 <= u <= 2$ pi and -2pi$/3 <= v <= 2$pi$/3$ (see Figure 2.10). Note that $\cosh v$, $\cos u$, and $\sin u$ should be entered as $\cosh(v)$, $\cos(u)$, and $\sin(u)$, respectively.
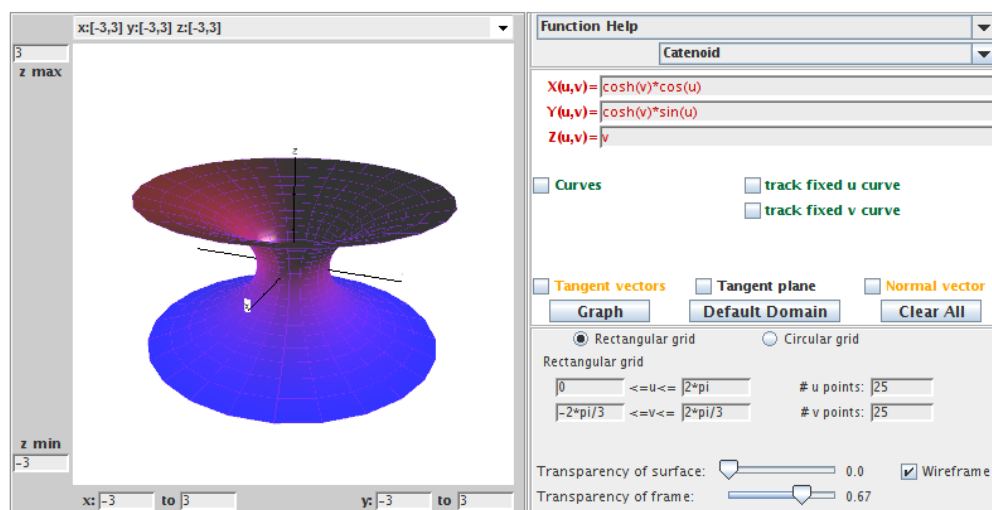

FIGURE 2.10. The catenoid.

Check out what happens if you change the u, v values. For example, try:

(a) pi $<=$ u $<=$ 2pi, -2pi/3 $<=$ v $<=$ 2pi/3;
(b) 0 $<=$ u $<=$ 2pi, 0 $<=$ v $<=$ 2pi/3;

(c) $0 <= u <= 2\text{pi}$, $-\text{pi}/4 <= v <= \text{pi}/4$;
(d) $0 <= u <= 2\text{pi}$, $-\text{pi} <= v <= \text{pi}$.

EXERCISE 2.7. A torus is a surface (but not a minimal surface) that can be formed by rotating a circle in the $x_1 x_3$-plane about the x_3-axis. Let this be a circle of radius b and whose center is a distance of a from the origin.
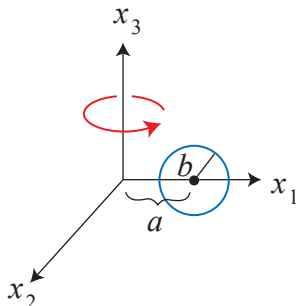

FIGURE 2.11. Creating a torus by rotating a circle.

Then the parametrization of this torus is

$$\mathbf{x}(u, v) = \Big((a + b\cos v)\cos u, (a + b\cos v)\sin u, b\sin v \Big),$$

where a, b are fixed, $0 < u < 2\pi$, and $0 < v < 2\pi$.

(a) Show how to derive this parametrization for a torus.
(b) Use *DiffGeomTool* to sketch the graph of this torus when $a = 3$ and $b = 2$; use `Rectangular grid` with $0 < u < 2\pi$, and $0 < v < 2\pi$.

Try it out!

In pre-calculus, we talk about a function of one variable $y = F(x)$ as one that satisfies the vertical line test. The graph of $F(x)$ is a 1-dimensional object, living in $\mathbb{R}^2$, created by plotting points of the form $(u, F(u))$ in the plane. Thus, it is parametrized by the map from $\mathbb{R}$ to $\mathbb{R}^2$ defined by $u \to (u, F(u))$. Analogously, we speak of a function of two variables $z = f(x, y)$, where the points (x, y) lie in a two-dimensional domain and f satisfies the vertical line test (here a line is vertical when it is parallel to the z-axis). The graph of $f(x, y)$ is a two-dimensional surface living in $\mathbb{R}^3$ with a height of $z = f(x, y)$ at a point (x, y) in its domain. An example of such a graph is the minimal surface known as Scherk's doubly periodic surface. It can be parametrized by

$$\mathbf{x}(u, v) = \left(u, v, \ln\left(\frac{\cos u}{\cos v} \right) \right).$$

EXERCISE 2.8.

(a) In this parametrization of Scherk's doubly periodic surface, what are the restrictions on the u and v values in the domain?

(b) Use *DiffGeomTool* and your answer from part (a) to sketch a graph of Scherk's doubly periodic surface with -0.48 pi <= u,v <= 0.48 pi.

(c) Scherk's doubly periodic surface is a particular example of the graph of a function. Now, let $f(x,y)$ be any function. Find a parametrization of the graph of f in general.

Try it out!

EXERCISE 2.9. Let r be a differentiable curve whose derivative does not vanish (i.e., $r'(v) \neq 0$ for all values v in the domain) and let r lie in some plane in $\mathbb{R}^3$. A *surface of revolution* is a surface that forms by rotating r about an axis in that plane such that the curve does not intersect the axis. The catenoid and torus are examples of this. For this exercise, let $r(v) = (f(v), 0, g(v))$ be such a curve in the $x_1 x_3$-plane.

(a) Find a parametrization for the surface of revolution generated by rotating this curve about the x_3-axis.

(b) Check that your answer to part (a) matches the parametrizations of the catenoid and the torus given above.

Try it out!

EXPLORATION 2.10. Consider the torus $T_{a,b}$ whose parametrization is given in Exercise 2.7. Use *DiffGeomTool* to plot $T_{3,2}$ again. Describe what happens to the shape of the torus $T_{a,b}$ as as a gets smaller and b gets larger. Explain this in terms of how we derived the parametrization of the torus. [HINT: in *DiffGeomTool*, plot each of the following tori:

$$T_{2.7,2} \qquad T_{2.4,2} \qquad T_{2,2} \qquad T_{3,2.4} \qquad T_{3,2.7} \qquad T_{3,3}$$

What happens when $a < b$?]

Try it out!

EXPLORATION 2.11. As mentioned earlier, the Enneper surface is not embedded; that is, it has self-intersections. Use *DiffGeomTool* and the parametrization given in Example 2.5 to graph the Enneper surface with the domain being a disk of radius 1.

(a) What happens to the Enneper surface as the radius r of the disk increases?

(b) Estimate the largest value of r for which the Enneper surface has no self-intersections.

(c) Assuming that the intersection occurs on the x_3-axis, prove your result from part (b).

Try it out!

So far we have discussed how a function (i.e., a parametrization) models a surface. Our goal is to determine the bending or curvature of curves on a surface. To do this, we next will need to use the parametrization of a surface to discuss the concepts of a tangent plane and a normal vector at a point on the surface. Suppose $\mathbf{x}(u,v)$ is a parametrization of a surface $M \subset \mathbb{R}^3$. If we fix $v = v_0$ and let u vary, then $\mathbf{x}(u, v_0)$

depends on one parameter and is known as a *u-parameter curve.* Likewise, we can fix $u = u_0$ and let v vary to get a *v-parameter curve* $\mathbf{x}(u_0, v)$ (see Figure 2.12). Tangent vectors for the u-parameter and v-parameter curves are computed by differentiating the component functions of $\mathbf{x}$ with respect to u and v, respectively. That is, $\mathbf{x}_u$ and $\mathbf{x}_v$ are the tangent vectors defined by

$$\mathbf{x}_u = \left(\frac{\partial x_1}{\partial u}, \frac{\partial x_2}{\partial u}, \frac{\partial x_3}{\partial u} \right), \qquad \mathbf{x}_v = \left(\frac{\partial x_1}{\partial v}, \frac{\partial x_2}{\partial v}, \frac{\partial x_3}{\partial v} \right).$$

Note that for each point $p = x(u_0, v_0)$ on the surface, we get two corresponding vectors by plugging u_0 and v_0 in $\mathbf{x}_u$ and $\mathbf{x}_v$.
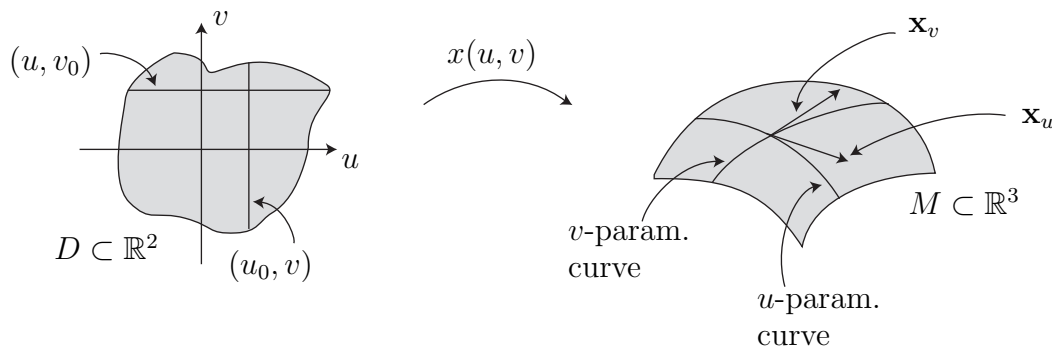

FIGURE 2.12. The u-parameter and v-parameter curves

Whenever we have a parametrization of a surface, we will require that $\mathbf{x}_u$ and $\mathbf{x}_v$ be linearly independent (i.e., not constant multiples of each other). Because of this, the span of $\mathbf{x}_u$ and $\mathbf{x}_v$ (i.e., the set of all vectors that can be written as a linear combination of $\mathbf{x}_u$, $\mathbf{x}_v$) forms a plane called the *tangent plane.* A tangent plane is an example of a vector space which are objects studied in linear algebra.

DEFINITION 2.12. The *tangent plane* of a surface M at a point p is
$$T_p M = \{\mathbf{v} | \mathbf{v} \text{ is tangent to } M \text{ at } p\}.$$

DEFINITION 2.13. The *unit normal* to a surface M at a point $p = \mathbf{x}(a, b)$ is
$$\mathbf{n}(a, b) = \left. \frac{\mathbf{x}_u \times \mathbf{x}_v}{\left| \mathbf{x}_u \times \mathbf{x}_v \right|} \right|_{(a,b)}.$$

Not every surface has a well-defined choice of a unit normal $\mathbf{n}$. Such surfaces are called *non-orientable.* An example of this is the Möbius strip. An orientable strip can be modeled by taking a strip of paper and taping the ends together. In this model, the strip has an outside and an inside–you <u>cannot</u> get from one side to the other without crossing over the edge. On the other hand, a Möbius strip is modeled by taking the
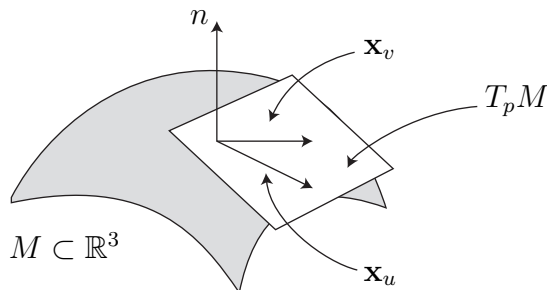
115

FIGURE 2.13. A tangent plane, T_pM, and unit normal vector, $\mathbf{n}$

same strip of paper, but twisting one end before taping the ends. This forms a strip in which you <u>can</u> get from one side to the other without crossing over the edge. This is the intuitive idea of a surface being non-orientable. The relationship between a Möbius strip and its unit normal $\mathbf{n}$ is further explored in Exercise 2.137 in the **Additional Exercises** at the end of this chapter. Note that the unit normal, $\mathbf{n}$, is orthogonal to the tangent plane at p (see Figure 2.13). Also, if the surface M is oriented, then geometrically there are two unit normals at each point $p \in M$ – an outward pointing normal and an inward pointing normal. However, the definition of $\mathbf{n}$ automatically chooses one of these normals.

EXAMPLE 2.14. Consider a torus parametrized by

$$\mathbf{x}(u, v) = \Big((3 + 2\cos v)\cos u, (3 + 2\cos v)\sin u, 2\sin v \Big),$$

where $0 < u, v < 2\pi$. For $v_0 = \frac{\pi}{3}$, the u-parameter curve is

$$\mathbf{x}\left(u, \frac{\pi}{3} \right) = (4\cos u, 4\sin u, \sqrt{3}).$$

For $u_0 = \frac{\pi}{2}$, the v-parameter curve is

$$\mathbf{x}\left(\frac{\pi}{2}, v \right) = (0, 3 + 2\cos v, 2\sin v).$$

Notice

$$\mathbf{x}_u(u, v) = (-(3 + 2\cos v)\sin u, (3 + 2\cos v)\cos u, 0)$$
$$\mathbf{x}_v(u, v) = (-2\sin v\cos u, -2\sin v\sin u, 2\cos v).$$

Now the u-parameter curve, $\mathbf{x}\left(u, \frac{\pi}{3}\right)$, and the v-parameter curve, $\mathbf{x}\left(\frac{\pi}{2}, v\right)$, intersect on the torus at $p = \mathbf{x}\left(\frac{\pi}{2}, \frac{\pi}{3}\right)$. Then the tangent vectors to the u- and v-parameter curves

at the point p are

$$\mathbf{x}_u\left(\frac{\pi}{2}, \frac{\pi}{3}\right) = (-4, 0, 0)$$

$$\mathbf{x}_v\left(\frac{\pi}{2}, \frac{\pi}{3}\right) = (0, -\sqrt{3}, 1).$$

These two vectors span the tangent plane, T_pM, at p. We compute that

$$\mathbf{x}_u\left(\frac{\pi}{2}, \frac{\pi}{3}\right) \times \mathbf{x}_v\left(\frac{\pi}{2}, \frac{\pi}{3}\right) = (-4, 0, 0) \times (0, -\sqrt{3}, 1) = (0, 1, 4\sqrt{3}).$$

Hence,

$$\mathbf{n}\left(\frac{\pi}{2}, \frac{\pi}{3}\right) = \left(0, \frac{1}{7}, \frac{4\sqrt{3}}{7}\right).$$

We can use *DiffGeomTool* to display this u-parameter curve, v-parameter curve, $\mathbf{x}_u$, $\mathbf{x}_v$, and $\mathbf{n}$. Enter the parametrization in this example for the torus. Then click `Curves`. A `Point location` box along with a `fixed u` and a `fixed v` boxes will appear. In the `Point location` box, enter pi/2 into the first box (i.e., the fixed u value) and pi/3 into the second box (i.e., the fixed v value). Make sure that the `fixed u` box is clicked but not the `fixed v` box. Then click the **Graph** button. The v-parameter curve will appear. If you now click the `track fixed u curve` box, a slider will appear. Moving the slider with the cursor will move the point along the v-parameter curve on the torus. Now, click on the `fixed u` box and then click the `Graph` button again. The u-parameter curve will appear. By clicking on the `track fixed v curve` box and moving the slider, the point along the v-parameter curve will move. Next, click on each of the following boxes separately followed by the **Graph** button: `Tangent vectors` box, `Tangent plane` box, and `Normal vector` box. This will cause these geometric objects to appear. You should convince yourself that the images of the vectors at $(u, v) = (\frac{\pi}{2}, \frac{\pi}{3})$ match the computed values done earlier in Example 2.14.

EXERCISE 2.15. For a surface of revolution (see Exercise 2.9) parametrized by

$$\mathbf{x}(u, v) = \big(f(v) \cos u, f(v) \sin u, g(v)\big)$$

the u-parameter curves are called *parallels* and are the curves formed by horizontal slices, while the v-parameter curves are called *meridians* and are the curves formed by vertical slices. Describe the parallels and meridians for the catenoid in Example 2.6 and the torus in Exercise 2.7.
  Try it out!

EXERCISE 2.16. Recall the parametrization of a catenoid

$$\mathbf{x}(u, v) = \Big(\cosh v \cos u, \cosh v \sin u, v\Big),$$

with $0 < u < 2\pi$ and $-\frac{2\pi}{3} < v < \frac{2\pi}{3}$.
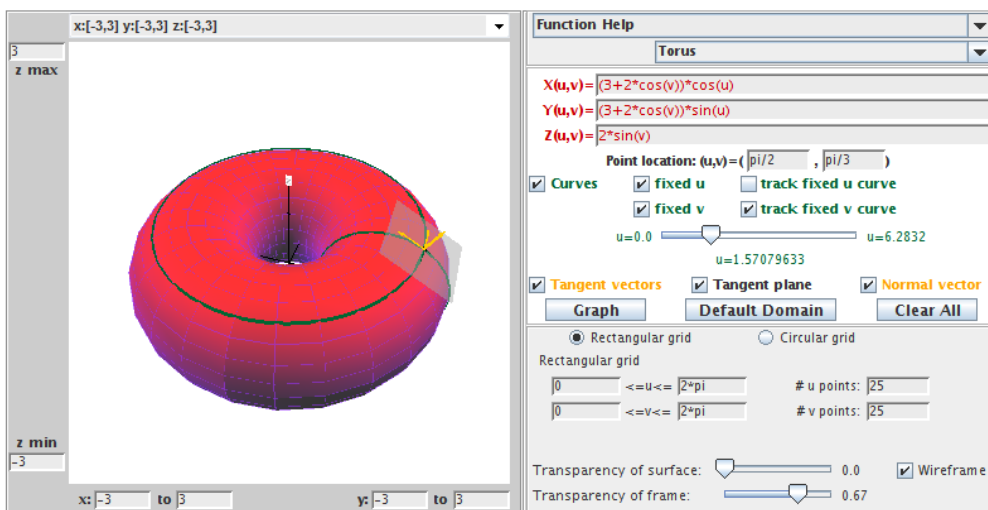
FIGURE 2.14. The torus with specific u, v-parameter curves, the tangent vectors, the tangent plane, and the normal vector.

(a) Use *DiffGeomTool* to sketch the u-parameter curve, $\mathbf{x}(u, 0)$, and the v-parameter curve, $\mathbf{x}(0, v)$ on the catenoid. Also, sketch the vectors $\mathbf{x}_u(0, 0)$, $\mathbf{x}_v(0, 0)$, and $\mathbf{n}(0, 0)$.

(b) Compute the vectors $\mathbf{x}_u(0, 0)$, $\mathbf{x}_v(0, 0)$, and $\mathbf{n}(0, 0)$.

Try it out!

EXERCISE 2.17. Recall the parametrization of Scherk's doubly periodic surface

$$\mathbf{x}(u, v) = \left(u, v, \ln\left(\frac{\cos u}{\cos v} \right) \right),$$

with $-0.48\pi < u, v < 0.48\pi$.

(a) Use *DiffGeomTool* to sketch the u-parameter curve, $\mathbf{x}\left(u, \frac{\pi}{4}\right)$, and the v-parameter curve, $\mathbf{x}\left(\frac{\pi}{4}, v\right)$ on Scherk's doubly periodic surface (make sure you use $-0.48\pi < u, v < 0.48\pi$). Also, sketch the vectors $\mathbf{x}_u\left(\frac{\pi}{4}, \frac{\pi}{4}\right)$, $\mathbf{x}_v\left(\frac{\pi}{4}, \frac{\pi}{4}\right)$, and $\mathbf{n}\left(\frac{\pi}{4}, \frac{\pi}{4}\right)$. Note that you can slide these vectors by clicking the `track fixed u curve` box. This collection of vectors, $(\mathbf{x}_u, \mathbf{x}_v, \mathbf{n})$, is known as a *moving frame* or *Frenet frame* of a curve. The way these vectors vary in $\mathbb{R}^3$ as the frame moves along the curve describes how the curve twists and turns in $\mathbb{R}^3$. For more details, see [**18**] or [**21**].

(b) Compute the vectors $\mathbf{x}_u\left(\frac{\pi}{4}, \frac{\pi}{4}\right)$, $\mathbf{x}_v\left(\frac{\pi}{4}, \frac{\pi}{4}\right)$, and $\mathbf{n}\left(\frac{\pi}{4}, \frac{\pi}{4}\right)$.

Try it out!

We will use the normal vector $\mathbf{n}$ to define the curvature of a curve on a surface. As we mentioned in the introduction, a minimal surface has very particular behavior

118

with respect to curvature: at each point on the surface, the bending upward in one direction is matched with the bending downward in the orthogonal direction. Hence, we will use this definition of curvature in our definition of a minimal surface.

Notice that any plane containing the normal **n** will intersect the surface M in a curve, α. For each curve α, we can compute its curvature, which measures how fast the curve pulls away from the tangent line at p. So let's now consider some ideas about the curvature of a curve. Any curve in $\mathbb{R}^3$ can be parametrized by a function of one variable, say $\alpha(t)$, where $\alpha : [a, b] \to \mathbb{R}^3$. However, this parametrization is not unique.

EXERCISE 2.18. Find two different parametrizations of the unit circle in the x_1x_2-plane.

Try it out!

This lack of uniqueness can cause difficulties in exploring the concept of curvature. To eliminate these difficulties we will standardize our parametrization by requiring it to be a unit speed curve.

DEFINITION 2.19. A curve α is a *unit speed curve* if $\left|\alpha'(t)\right| = 1$.

If our parametrization of our regular curve $\alpha(t)$ is not of unit speed, we can always reparametrize it by arclength to form a unit speed curve $\alpha(s)$ (you probably saw this in calculus 3 or vector calculus, and you may want to review how to reparametrize a curve by arclength). Because of this, we will assume that the curves we will be discussing are unit speed curves $\alpha(s)$. This assumption means that we are only interested in the geometric shape of a regular curve since reparametrizing does not change its shape.

Given a curve α, we want to discuss its curvature (or bending). We quantify the amount that the curve bends at each point p by measuring how rapidly the curve pulls away from the tangent line at p. In other words, we measure the amount of bending by measuring the rate of change of the angle θ that neighboring tangents make with the tangent at p. Thus, we are interested in the rate of change of the tangent vector (i.e., the value of the second derivative).

DEFINITION 2.20. The *curvature* of the unit speed curve α at s is $\left|\alpha''(s)\right|$.

EXAMPLE 2.21. Consider a torus parametrized by

$$\mathbf{x}(u, v) = \Big((3 + 2\cos v)\cos u, (3 + 2\cos v)\sin u, 2\sin v \Big),$$

where $0 < u, v < 2\pi$. Let's compute the curvature for the u-parameter curves and the v-parameter curves. All the v-parameter curves (or meridians) are the curves formed by vertical slices of the torus, and hence are circles of radius $b = 2$. To compute the curvature of these v-parameter curves, we start with the parametrization of these curves

$$\mathbf{x}(v) = \mathbf{x}(u_0, v) = \Big((3 + 2\cos v)\cos u_0, (3 + 2\cos v)\sin u_0, 2\sin v \Big),$$

where u_0 is a fixed value. We next need to reparametrize $\mathbf{x}(v)$ so that it is a unit speed curve. Differentiating $\mathbf{x}(v)$ with respect to v we get

$$\mathbf{x}'(v) = \left(-2\sin v \cos u_0, -2\sin v \sin u_0, 2\cos v \right).$$

Thus, $|\mathbf{x}'(v)| = 2$. To make $\mathbf{x}'(v)$ into a unit speed curve, we replace v with $\frac{s}{2}$. So, our reparametrized curve is

$$\mathbf{x}(s) = \left(\left(3 + 2\cos\left(\frac{s}{2}\right)\right)\cos u_0, \left(3 + 2\cos\left(\frac{s}{2}\right)\right)\sin u_0, 2\sin\left(\frac{s}{2}\right) \right).$$

Then we compute

$$\mathbf{x}'(s) = \left(-\sin\left(\frac{s}{2}\right)\cos u_0, -\sin\left(\frac{s}{2}\right)\sin u_0, \cos\left(\frac{s}{2}\right) \right)$$

$$\mathbf{x}''(s) = \left(-\frac{1}{2}\cos\left(\frac{s}{2}\right)\cos u_0, -\frac{1}{2}\cos\left(\frac{s}{2}\right)\sin u_0, -\frac{1}{2}\sin\left(\frac{s}{2}\right) \right).$$

Hence, the curvature of the v-parameter curves is

$$\left| \mathbf{x}''(s) \right| = \frac{1}{2}.$$

The u-parameter curves (or parallels) are the curves formed by horizontal slices of the torus, and so are circles of radius $3 + 2\cos v_0$, where $v_0 \in (0, 2\pi)$ is fixed; note that these radii vary between 1 and 5. These curves are parametrized by

$$\mathbf{x}(u) = \mathbf{x}(u, v_0) = \left((3 + 2\cos v_0)\cos u, (3 + 2\cos v_0)\sin u, 2\sin v_0 \right),$$

which are reparametrized to the unit speed curve

$$\mathbf{x}(s) = \left((3 + 2\cos v_0)\cos\left(\frac{s}{3 + 2\cos v_0}\right), (3 + 2\cos v_0)\sin\left(\frac{s}{3 + 2\cos v_0}\right), 2\sin v_0 \right).$$

Finally, computing the curvature of these u-parameter curves yields

$$\left| \mathbf{x}''(s) \right| = \frac{1}{3 + 2\cos v_0}.$$

So, the curvature of the these curves varies between $\frac{1}{5}$ and 1.

EXERCISE 2.22. Verify that replacing v with $\frac{s}{2}$ in $\mathbf{x}(u, v)$ from the previous example gives a unit speed curve.
Try it out!

EXERCISE 2.23. Compute the curvatures of the meridians and parallels of the catenoid parametrized by

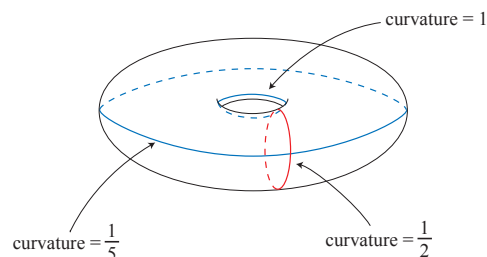$$\mathbf{x}(u, v) = \left(a\cosh v \cos u, a\cosh v \sin u, av \right).$$

Try it out!

FIGURE 2.15. The curvature of the meridians and parallels on a torus.

EXERCISE 2.24. The curve parametrized by $\alpha(t) = (a\cos t, a\sin t, bt)$ is known as a *helix* which is a spiral that rises with a pitch of $2\pi b$ on the cylinder $x^2 + y^2 = a^2$.

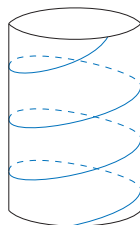

FIGURE 2.16. A helix in a cylinder.

We can create a surface by connecting a line from the axis $(0, 0, bt)$ through the helix $(a\cos t, a\sin t, bt)$ at each height bt. This ruled surface is a minimal surface known as a *helicoid*. All minimal surfaces including the helicoid can be parametrized in several ways. For our purposes, we will use the following parametrization of the helicoid:

$$\mathbf{x}(u, v) = \Big(a\sinh v \cos u, a\sinh v \sin u, au \Big).$$

(a) Compute the curvatures of the u-parameter curves and v-parameter curves of this helicoid (note: making the v-parameter curve into a unit speed curve is not easy, so in doing this computation you may need to be creative).
(b) Use *DiffGeomTool* to graph this helicoid with $a = 1$ (see Figure 2.17).

Try it out!

EXERCISE 2.25. From the results of Example 2.21 you may have conjectured that the curvature of a circle of radius r is $\dfrac{1}{r}$. This conjecture is correct. Prove this.

Try it out!

Now let's return to surfaces. Suppose we have a curve $\sigma(s)$ on a surface M. We can determine the unit tangent vector, $\mathbf{w}$ of σ at $p \in M$ and the unit normal, $\mathbf{n}$ of M at $p \in M$. Note that $\mathbf{w} \times \mathbf{n}$ forms a plane $\mathcal{P}$ that intersects M creating a curve $\alpha(s)$.
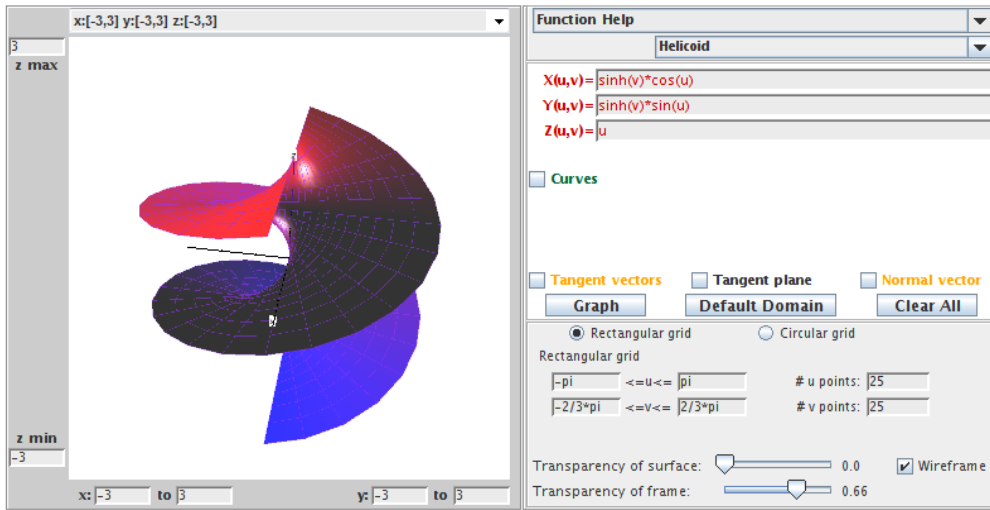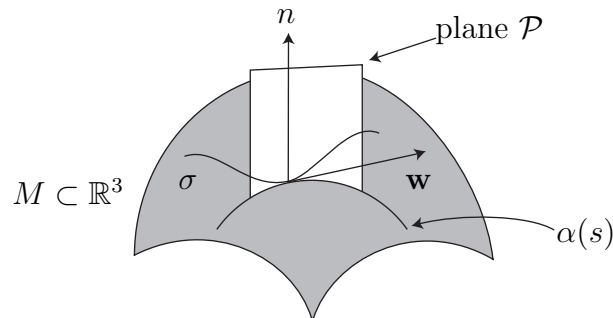
FIGURE 2.17. Helicoid.

FIGURE 2.18. The normal curvature

DEFINITION 2.26. The *normal curvature* in the **w** direction is

$$k(\mathbf{w}) = \alpha'' \cdot \mathbf{n}.$$

Recall $\alpha'' \cdot \mathbf{n} = |\alpha''||\mathbf{n}| \cos \theta$, where θ is the angle between $\mathbf{n}$ and α''. Hence $\alpha'' \cdot \mathbf{n}$ is the projection of α'' onto the unit normal (hence, the name *normal* curvature). Intuitively, the normal curvature measures how much the surface bends towards $\mathbf{n}$ as you travel in the direction of the tangent vector $\mathbf{w}$ starting at point p. As we rotate the plane about the normal $\mathbf{n}$, we will get a set of curves on the surface each of which has a value for its curvature. Let k_1 and k_2 be the maximum and minimum curvature values at p, respectively. The directions in which the normal curvature attains its absolute maximum and absolute minimum values are known as the *principal directions*.

122

DEFINITION 2.27. The *mean curvature* (i.e., average curvature) of a surface M at p is

$$H = \frac{k_1 + k_2}{2}.$$

It turns out that k_1 and k_2 come from two orthogonal tangent vectors. The mean curvature depends upon the point $p \in M$. However, it can be shown that H does not change if we choose any two orthogonal vectors and use their curvature values to compute H at p. Also, we will use the convention that if the principal curve with normal curvature k_j is bending toward the unit normal $\mathbf{n}$, then $k_j > 0$ and if it is bending away from $\mathbf{n}$, then $k_j < 0$.

EXAMPLE 2.28. At any point on a sphere of radius a, all the curves α are circles of radius a and hence have the same curvature value which can be computed to be $1/a$. Since these curves are bending away from $\mathbf{n}$, $k_1 = -1/a = k_2$. So the mean curvature is $-1/a$.

EXERCISE 2.29. Determine the mean curvature at all points on the cylinder parametrized by $\mathbf{x}(u, v) = (a \cos u, a \sin u, bv)$.
   Try it out!

EXERCISE 2.30. Determine if there are points on the torus $\mathbf{x}(u, v) = \big((a+b\cos v)\cos u, (a+b\cos v)\sin u, b\sin v\big)$ where $H > 0$, $H = 0$, and $H < 0$.
   Try it out!

In the next section of this chapter we will define a minimal surface in terms of mean curvature. However, using our current definition it is tedious to compute the mean curvature of a surface because we have to find the principle curvatures individually for every point on the surface. Instead, we would like to have an explicit expression to compute mean curvature at every point p on the surface simply by plugging p into the formula (or, more precisely, plugging in each point (u, v) in the domain of the parametrization of the surface). Fortunately, there is a more useful formula for mean curvature using the coefficients of the first and second fundamental forms for a surface. Recall that α is a unit speed curve. Hence

$$\begin{aligned}
1 =& |\alpha'|^2 = \alpha' \cdot \alpha' \\
=& (\mathbf{x}_u \, du + \mathbf{x}_v \, dv) \cdot (\mathbf{x}_u \, du + \mathbf{x}_v \, dv) \\
=& \mathbf{x}_u \cdot \mathbf{x}_u \, du^2 + 2\mathbf{x}_u \cdot \mathbf{x}_v \, dudv + \mathbf{x}_v \cdot \mathbf{x}_v \, dv^2 \\
=& E \, du^2 + 2F \, dudv + G \, dv^2.
\end{aligned}$$

(2)

The terms $E = \mathbf{x}_u \cdot \mathbf{x}_u$, $F = \mathbf{x}_u \cdot \mathbf{x}_v$, and $G = \mathbf{x}_v \cdot \mathbf{x}_v$ are known as the *coefficients of the first fundamental form*. These describe how lengths on a surface are distorted as compared to their usual measurements in $\mathbb{R}^3$.

Next, recall $k(\mathbf{w}) = \alpha'' \cdot \mathbf{n}$. Note that $\alpha' \cdot \mathbf{n} = 0$, and so $(\alpha' \cdot \mathbf{n})' = 0$, which implies $\alpha'' \cdot \mathbf{n} + \alpha' \cdot \mathbf{n}' = 0$, and thus $\alpha'' \cdot \mathbf{n} = -\alpha' \cdot \mathbf{n}'$. Similarly, $-\mathbf{x}_u \cdot \mathbf{n}_u = \mathbf{x}_{uu} \cdot \mathbf{n}$. So

$$
\begin{aligned}
k(w) &= -\,\alpha' \cdot \mathbf{n}' \\
&= -\,(\mathbf{x}_u\, du + \mathbf{x}_v\, dv) \cdot (\mathbf{n}_u\, du + \mathbf{n}_v\, dv) \\
&= -\,\mathbf{x}_u \cdot \mathbf{n}_u\, du^2 - (\mathbf{x}_u \cdot \mathbf{n}_v + \mathbf{x}_v \cdot \mathbf{n}_u)\, dudv - \mathbf{x}_v \cdot \mathbf{n}_v\, dv^2 \\
&= \mathbf{x}_{uu} \cdot \mathbf{n}\, du^2 + 2\mathbf{x}_{uv} \cdot \mathbf{n}\, dudv + \mathbf{x}_{vv} \cdot \mathbf{n}\, dv^2 \\
&= e\, du^2 + 2f\, dudv + g\, dv^2.
\end{aligned}
$$

The terms $e = \mathbf{x}_{uu} \cdot \mathbf{n}$, $f = \mathbf{x}_{uv} \cdot \mathbf{n}$, and $g = \mathbf{x}_{vv} \cdot \mathbf{n}$ are called *the coefficients of the second fundamental form.* These describe how much the surface bends away from the tangent plane.

EXAMPLE 2.31. Recall that a catenoid can be parametrized by

$$
\mathbf{x}(u, v) = \Big(a \cosh v \cos u, a \cosh v \sin u, av\Big).
$$

Using this parametrization, we compute that

$$
\begin{aligned}
\mathbf{x_u} &= (-a \cosh v \sin u, a \cosh v \cos u, 0) \\
\mathbf{x_v} &= (a \sinh v \cos u, a \sinh v \sin u, a).
\end{aligned}
$$

So, the coefficients of the first fundamental form are:

$$
\begin{aligned}
E &= \mathbf{x}_u \cdot \mathbf{x}_u = a^2 \cosh^2 v; \\
F &= \mathbf{x}_u \cdot \mathbf{x}_v = 0; \\
G &= \mathbf{x}_v \cdot \mathbf{x}_v = a^2 \cosh^2 v.
\end{aligned}
$$

What do these values for E, F, and G tell us? Let $(u_0, v_0) \in D$ be a point in the domain and let's take a small square with a vertex at this point. Because $\mathbf{x}_u \cdot \mathbf{x}_v = F = 0$, we know that the orthogonal lines from the u-parameter curve and the v-parameter curve will remain orthogonal on the catenoid. That is, small squares will be mapped to small rectangles. Next, because $E = G$, adjacent sides of the image rectangle will have the same length. So, in fact, small squares in the domain D will be mapped to small squares on the catenoid. Now suppose for simplicity sake that $a = 1$. Then $E = G = \cosh^2 v$. When $v = 0$, $E = G = 1$ and as v gets farther away from 0, E and G get larger. This means that a small square containing the u-parameter curve $v = 0$ will get mapped to a small square of the same size on the catenoid. But as v gets farther away from 0, the size of the side lengths of the image square will increase by a factor of $\cosh^2 v$. This can be seen in Figure 2.19 where we have used the `Transparency of surface` and the `Transparency of the frame` sliders to make the u- and v-parameter curves more distinct (note that the u-parameter curve with $v = 0$ gets mapped to a parallel on the neck of the catenoid, and the u-parameter curve with $v = \frac{2\pi}{3}$ gets mapped to the edge of the catenoid, as displayed in the figure).
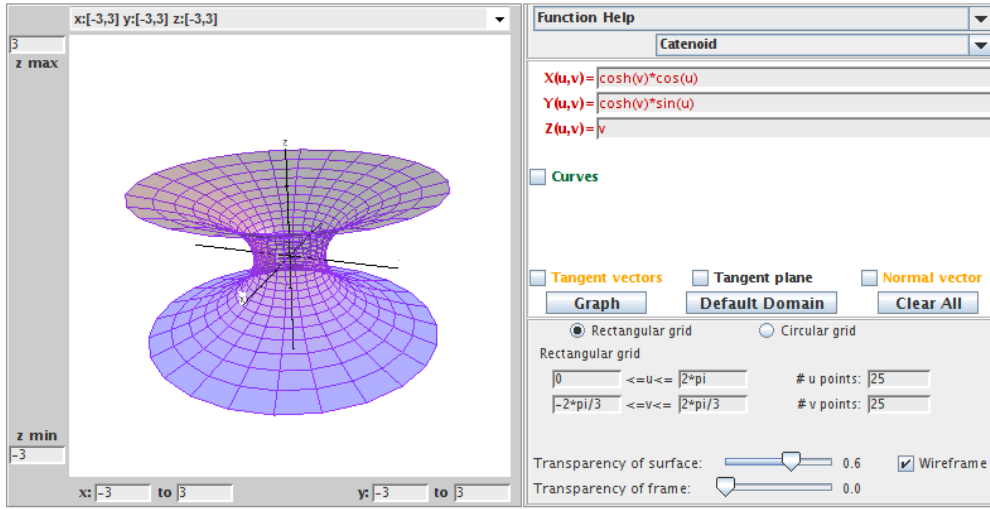
FIGURE 2.19. The catenoid.

In order to compute the coefficients of the second fundamental form, we need to compute $\mathbf{n} = \dfrac{\mathbf{x}_u \times \mathbf{x}_v}{|\mathbf{x}_u \times \mathbf{x}_v|}$. Now,

$$\mathbf{x}_u \times \mathbf{x}_v = (a^2 \cosh v \cos u, a^2 \cosh v \sin u, -a^2 \cosh v \sinh v),$$

and so

$$|\mathbf{x}_u \times \mathbf{x}_v| = a^2 \cosh^2 v.$$

Hence

$$\mathbf{n} = \left(\frac{\cos u}{\cosh v}, \frac{\sin u}{\cosh v}, -\frac{\sinh v}{\cosh v} \right).$$

Also, we can compute that

$$\mathbf{x}_{uu} = (-a \cosh v \cos u, -a \cosh v \sin u, 0)$$
$$\mathbf{x}_{uv} = (-a \sinh v \sin u, -a \sinh v \cos u, 0)$$
$$\mathbf{x}_{vv} = (a \cosh v \cos u, a \cosh v \sin u, 0)$$

Therefore, the coefficients of the second fundamental form are:

$$e = \mathbf{n} \cdot \mathbf{x}_{uu} = -a;$$
$$f = \mathbf{n} \cdot \mathbf{x}_{uv} = 0;$$
$$g = \mathbf{n} \cdot \mathbf{x}_{vv} = a.$$

What do these values for e, f, and g tell us? Again, let $(u_0, v_0) \in D$ be a point in the domain, and let $p \in M$ be the image of (u_0, v_0) on the surface. Then at p, the vectors $\mathbf{x}_u$ and $\mathbf{x}_v$ create the tangent plane T_pM and the unit normal $\mathbf{n}$. For this catenoid, the u-parameter curve is bending away from $\mathbf{n}$ while the v-parameter curve is bending

125

toward $\mathbf{n}$. However, both curves are bending the same amount away from the tangent plane. This is represented by the fact that $e = -g$ in this example.

EXERCISE 2.32. A torus has the parametrization

$$\mathbf{x}(u, v) = \Big((a + b \cos v) \cos u, (a + b \cos v) \sin u, b \sin v \Big).$$

(a) Compute the coefficients of the first and the second fundamental forms.
(b) Open *DiffGeomTool* and enter this parametrization for the torus with $a = 2$ and $b = 1$. Use the `Rectangular grid` with $0 <= $ u $<= 2*$ pi and $0 <=$ v $<= 2*$ pi. And set the # u points: to 20 and the # v points: to 20. Describe how the results from part (a) match with the image of the torus in *DiffGeomTool*. Use the `Transparency of surface` and the `Transparency of the frame` sliders and click off and on the `Wireframe` button to make the wireframe clearer.

Try it out!

EXERCISE 2.33. Compute the coefficients of the first and the second fundamental forms for Scherk's doubly periodic surface parametrized by

$$\mathbf{x}(u, v) = \left(u, v, \ln \left(\frac{\cos u}{\cos v} \right) \right).$$

Try it out!

Now we want to express the mean curvature H in terms of these coefficients of the first and second fundamental forms. In particular, we will show that

$$H = \frac{Eg + Ge - 2Ff}{2(EG - F^2)}.$$

PROOF. There is an elegant way to derive this formula. This approach requires some concepts that are interesting but beyond the scope of what we will need. So instead we will use a straightforward calculation that is in Oprea ([**22**], pp. 40-42). Although this calculation does not give much insight into the formula, it does provide a straightforward proof of this important formula. For a discussion involving the more advanced approach, see [**3**] or [**18**].

Let $\mathbf{w}_1, \mathbf{w}_2$ be any two perpendicular unit vectors and k_1, k_2 be their normal curvatures using the curves $\alpha_1(s) = (u_1(s), v_1(s))$ and $\alpha_2(s) = (u_2(s), v_2(s))$. Also, let $p_1 = du_1 + idu_2$ and $p_2 = dv_1 + idv_2$. Then using the second fundamental form to compute k_1 and k_2, we have

$$2H = k_1 + k_2 = e(du_1^2 + du_2^2) + 2f(du_1 dv_1 + du_2 dv_2) + g(dv_1^2 + dv_2^2)$$
$$= e(p_1 \, \overline{p_1}) + f(p_1 \, \overline{p_2} + \overline{p_1} p_2) + g(p_2 \, \overline{p_2}).$$

We want to further simplify this so that p_1 and p_2 do not appear in the expression. Recall eq (2):

$$1 = E \, du^2 + 2F \, dudv + G \, dv^2.$$

126

Also, note that $Edu_1du_2 + F(du_1dv_2 + du_2dv_1) + Gdv_1dv_2 = 0$, because $\mathbf{w}_1, \mathbf{w}_2$ are perpendicular and so $\mathbf{w}_1 \cdot \mathbf{w}_2\alpha_1'(s) \cdot \alpha_2'(s) = 0$.

Now consider

$$
\begin{aligned}
Ep_1^2 + 2Fp_1p_2 + Gp_2^2 &= E\left[du_1^2 - du_2^2 + i2du_1du_2\right] + 2F\left[du_1dv_1 - du_2dv_2 + i(du_1dv_2 + du_2dv_1)\right] \\
&\quad + G\left[dv_1^2 - dv_2^2 + i2dv_1dv_2\right] \\
&= 2i\left[Edu_1du_2 + F(du_1dv_2 + du_2dv_1) + Gdv_1dv_2\right] \\
&\quad + \left[Edu_1^2 + 2Fdu_1dv_1 + Gdv_1^2\right] - \left[Edu_2^2 + 2Fdu_1dv_2 + Gdv_2^2\right] \\
&= 0 + 1 - 1 \\
&= 0.
\end{aligned}
$$

Thus,

$$
p_1 = \frac{-2Fp_2 \pm \sqrt{4F^2p_2^2 - 4EGP_2^2}}{2E} = \left(-\frac{F}{E} \pm i\frac{\sqrt{EG - F^2}}{E}\right)p_2
$$

$$
\overline{p_1} = \left(-\frac{F}{E} \mp i\frac{\sqrt{EG - F^2}}{E}\right)\overline{p_2}.
$$

And so

(3)
$$
p_1\,\overline{p_1} = \left(\frac{F^2}{E^2} + EG - F^2E\right)p_2\,\overline{p_2} = \frac{G}{E}p_2\,\overline{p_2}, \text{ and}
$$

(4)
$$
p_1\,\overline{p_2} + \overline{p_1}p_2 = -\frac{2F}{E}p_2\,\overline{p_2}.
$$

Now we have

$$
2H = k_1 + k_2 = e(p_1\,\overline{p_1}) + f(p_1\,\overline{p_2} + \overline{p_1}p_2) + g(p_2\,\overline{p_2})
$$
$$
= \left[e\frac{G}{E} + f\left(\frac{-2F}{E}\right) + g\right]p_2\,\overline{p_2}.
$$

We just need to get rid of $p_2\,\overline{p_2}$. Again using eq (2), we have

$$
\begin{aligned}
Ep_1\,\overline{p_1} &+ F(p_1\,\overline{p_2} + \overline{p_1}p_2) + Gp_2\,\overline{p_2} \\
&= E(du_1^2 + du_2^2) + 2F(du_1dv_1 + du_2dv_2) + G(dv_1^2 + dv_2^2) \\
&= 1 + 1 = 2.
\end{aligned}
$$

Using eqs (3) and (4), we derive

$$
2 = E\left(\frac{G}{E}p_2\,\overline{p_2}\right) + 2F\left(\frac{-2F}{E}p_2\,\overline{p_2}\right) + Gp_2\,\overline{p_2}
$$

$$
\Rightarrow 2 = \left[2G - \frac{2F^2}{E}\right]p_2\,\overline{p_2}
$$

$$
\Rightarrow p_2\,\overline{p_2} = \frac{E}{EG - F^2}
$$

127

Therefore,

$$H = \frac{1}{2}\left[e\frac{G}{E} + f\left(\frac{-2F}{E}\right) + g \right] p_2 \, \overline{p_2} = \frac{Eg + eG - 2Ff}{2(EG - F^2)}.$$

$\square$

2.3. Minimal Surfaces

Now that we have a foundation of some essential ideas from differential geometry, we can begin to explore minimal surfaces. Earlier we mentioned that minimal surfaces can be thought of as saddle surfaces. That is, at each point the bending upward in one direction is matched with the bending downward in the orthogonal direction. This picture can be described mathematically with the following definition.

DEFINITION 2.34. A *minimal surface* is a surface M with mean curvature $H = 0$ at all points $p \in M$.

Make sure that you understand how this definition fits with the picture of a surface bending upward in one direction while also bending downward in the orthogonal direction. At this point we can use the results from the previous section. First, we can use the formula

$$(5) \qquad H = \frac{Eg + Ge - 2Ff}{2(EG - F^2)}.$$

to show that a surface with a specific parametrization is minimal.

EXAMPLE 2.35. We will use eq (5) to show that the catenoid is a minimal surface. Recall that a catenoid can be parametrized by

$$\mathbf{x}(u, v) = \Big(a \cosh v \cos u, a \cosh v \sin u, av \Big).$$

From Example 2.31

$$E = \mathbf{x}_u \cdot \mathbf{x}_u = a^2 \cosh^2 v,$$
$$F = \mathbf{x}_u \cdot \mathbf{x}_v = 0,$$
$$G = \mathbf{x}_v \cdot \mathbf{x}_v = a^2 \cosh^2 v,$$

and

$$e = \mathbf{n} \cdot \mathbf{x}_{uu} = -a,$$
$$f = \mathbf{n} \cdot \mathbf{x}_{uv} = 0,$$
$$g = \mathbf{n} \cdot \mathbf{x}_{vv} = a.$$

Hence

$$H = \frac{eG - 2fF + Eg}{2(EG - F^2)} = 0.$$

And so the catenoid is a minimal surface.

128

EXERCISE 2.36. Using the parametrization for the helicoid

$$\mathbf{x}(u, v) = (a \sinh v \cos u, a \sinh v \sin u, au),$$

prove that the helicoid is a minimal surface. Using *DiffGeomTool* we display the graph of this helicoid when $a = 1$.
Try it out!

EXERCISE 2.37. Using the parametrization for the torus

$$\mathbf{x}(u, v) = \big((a + b \cos v) \cos u, (a + b \cos v) \sin u, b \sin v\big),$$

prove that it is not a minimal surface.
Try it out!

EXERCISE 2.38. Suppose a surface M is the graph of a function $f(x, y)$ of two variables (see the paragraph before Exercise 2.8). Then M can be parametrized by

$$\mathbf{x}(x, y) = \big(x, y, f(x, y)\big),$$

where its domain is formed by the projection of M onto the xy-plane.
  (a) Compute the coefficients of the first and second fundamental forms for M.
  (b) A *minimal graph* is a minimal surface that is a graph of a function. Prove

(6) M is a minimal graph if and only if $f_{xx}\left(1 + f_y^2\right) - 2f_x f_y f_{xy} + f_{yy}\left(1 + f_x^2\right) = 0.$

Try it out!

In the paragraph before Exercise 2.8, we stated that Scherk's doubly periodic surface is a minimal graph. We will now use eq (6) to prove that. Applying this equation is usually not easy, because solving explicitly for f can be complicated. However, one case in which we can do this is when f can be separated into two functions, each of which is dependent upon only one variable. In particular, suppose $f(x, y) = g(x) + h(y)$. Then the minimal surface equation becomes:

$$g''(x)[1 + (h'(y))^2] + h''(y)[1 + (g'(x))^2] = 0.$$

This is a separable differential equation and hence can be solved. To do so, separate all the terms with the x variables from those with the y variables by putting them on opposite sides. This yields:

(7) $$-\frac{1 + (g'(x))^2}{g''(x)} = \frac{1 + (h'(y))^2}{h''(y)}.$$

What does this mean? If we fix y, the right side remains constant even if we change x in the left side. The same is true if we fix x and vary y. The only way such a situation can occur is if both sides are constant. So we have:

$$-\frac{1 + (g'(x))^2}{g''(x)} = k \qquad \Longrightarrow \qquad 1 + (g'(x))^2 = -kg''(x).$$

To solve this, let $\phi(x) = g'(x)$. Then $\frac{d\phi}{dx} = g''(x)$ and so

$$\int dx = -k \int \frac{d\phi}{1 + \phi^2}$$
$$\implies \quad x = -k \arctan \phi + C$$
$$\implies \quad \phi = -\tan\left(\frac{x + C}{k}\right).$$

For convenience, let $C = 0$ and $k = 1$. Since $\phi = g'$, we can integrate again to get:

$$g(x) = \ln[\cos x].$$

Completing the same calculations for the y-side of eq (7) yields:

$$h(y) = -\ln[\cos y].$$

Hence

$$f(x, y) = g(x) + h(y) = \ln\left[\frac{\cos x}{\cos y}\right]$$

which is an equation for Scherk's doubly periodic surface. Using *DiffGeomTool* we display the graph of Scherk's doubly periodic surface (see Figure 2.20).
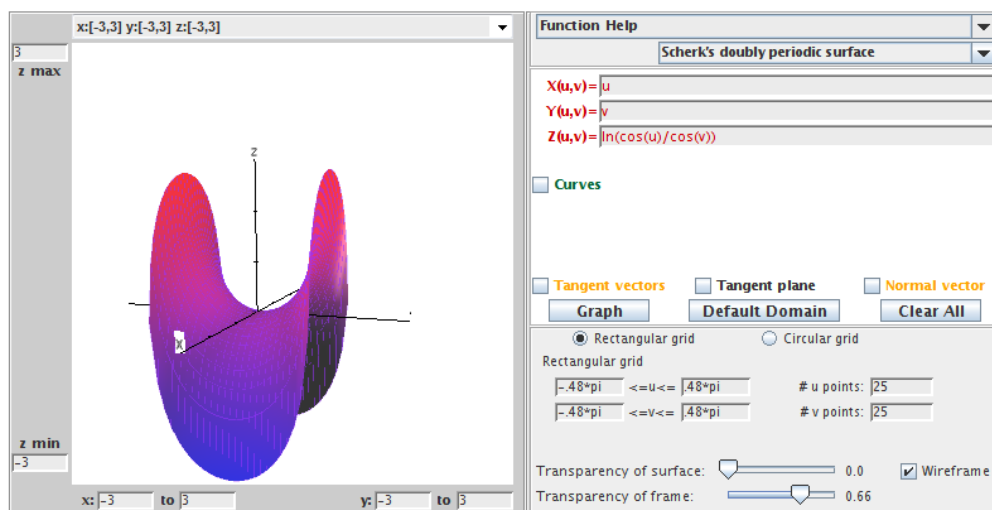

FIGURE 2.20. Scherk's doubly periodic surface.

Notice that $-\frac{\pi}{2} < x, y < \frac{\pi}{2}$ and so this surface is defined over a square with side lengths π centered at the origin. By a theorem known as the Schwarz Reflection Principle, we can fit pieces of Scherk's doubly periodic surface together horizontally and vertically to get a checkerboard domain (See Figure 2.21). Because one piece of this surface can be repeated or tiled in two directions, it is called a doubly periodic surface. It turns

out that there are many examples of periodic minimal surfaces, and this is currently an active area of research.
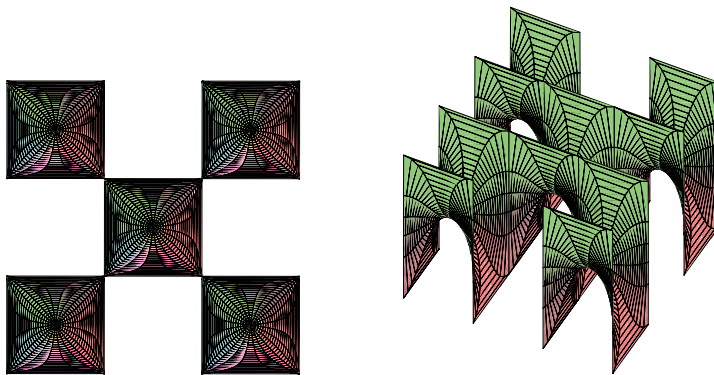

FIGURE 2.21. A tiling of Scherk's doubly periodic surface.

Let's summarize the minimal surfaces we have encountered so far, as well as a few other well-known examples.

(1) The plane:
$$\mathbf{x}(u,v) = (u, v, 0).$$
(2) The Enneper surface:
$$\mathbf{x}(u,v) = \left(u - \tfrac{1}{3}u^3 + uv^2, v - \tfrac{1}{3}v^3 + u^2v, u^2 - v^2\right).$$
(3) The catenoid:
$$\mathbf{x}(u,v) = (a \cosh v \cos u, a \cosh v \sin u, av).$$
(4) The helicoid:
$$\mathbf{x}(u,v) = (a \sinh v \cos u, a \sinh v \sin u, au).$$
(5) Scherk's doubly periodic surface:
$$\mathbf{x}(u,v) = \left(u, v, \ln\left(\frac{\cos u}{\cos v}\right)\right).$$
(6) Scherk's singly periodic surface:
$$\mathbf{x}(u,v) = (\operatorname{arcsinh}(u), \operatorname{arcsinh}(v), \arcsin(uv)).$$
(7) Henneberg surface:
$$\mathbf{x}(u,v) = (-1 + \cosh(2u)\cos(2v), -\sinh(u)\sin(v) - \tfrac{1}{3}\sinh(3u)\sin(3v),$$
$$-\sinh(u)\cos(v) + \tfrac{1}{3}\sinh(3u)\cos(3v)).$$
(8) Catalan surface:
$$\mathbf{x}(u,v) = (1 - \cos(u)\cosh(v), 4\sin(\tfrac{u}{2})\sinh(\tfrac{v}{2}), u - \sin(u)\cosh(v)).$$

In addition to the Enneper surface, the Henneberg surface and the Catalan surface are not embedded.

There is a fairly extensive list of minimal surfaces, and we have no way of listing all of them. So, instead, we often focus on trying to classify minimal surfaces. This

means, we try to find results that include all possibilities for minimal surfaces with specific properties. The simplest example of this is Theorem 2.39.

THEOREM 2.39. Any nonplanar minimal surface in $\mathbb{R}^3$ that is also a surface of revolution is contained in a catenoid.

As we have seen, determining if a surface is minimal basically involves solving second order differential equations. We can simplify these equations if we use a specific type of parametrization of a surface known as an isothermal parametrization.

DEFINITION 2.40. A parametrization $\mathbf{x}$ is *isothermal* if $E = \mathbf{x}_u \cdot \mathbf{x}_u = \mathbf{x}_v \cdot \mathbf{x}_v = G$ and $F = \mathbf{x}_u \cdot \mathbf{x}_v = 0$.

What does this mean? Recall that E, F, and G describe how lengths on a surface are distorted as compared to their usual measurements in $\mathbb{R}^3$. So if $F = \mathbf{x}_u \cdot \mathbf{x}_v = 0$ then the vectors $\mathbf{x}_u$ and $\mathbf{x}_v$ are orthogonal and if $E = G$, then the amount of distortion is the same in these two orthogonal directions. Thus, we can think of an isothermal parametrization as mapping a small square in the domain to a small square on the surface. Sometimes an isothermal parametrization is called a conformal parametrization, because the angle between a pair of curves in the domain is equal to the angle between the corresponding pair of curves on the surface.
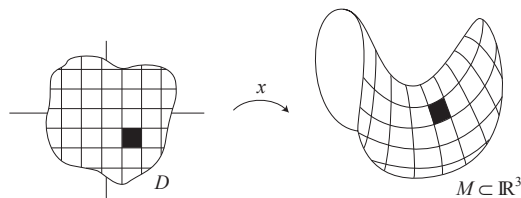

FIGURE 2.22. An isothermal parametrization maps small squares to small squares.

EXAMPLE 2.41. The parametrization

$$\mathbf{x}(u, v) = (a \cosh v \cos u, a \cosh v \sin u, av)$$

for the catenoid is isothermal, because in Example 2.35 we derived that $E = a^2 \cosh^2 v = G$ and $F = 0$. We can get a geometric sense that this parametrization is isothermal by using *DiffGeomTool* to graph this parametrization of the catenoid. Open *DiffGeom-Tool* and enter this parametrization with $a = 1$ and click the **Graph** button. Use the `Transparency of surface` and the `Transparency of the frame` sliders and click off and on the `Wireframe` button to make the curves more distinct. Notice that the grid of squares in the domain are pretty much mapped to a grid of squares as predicted above (see Figure 2.23).
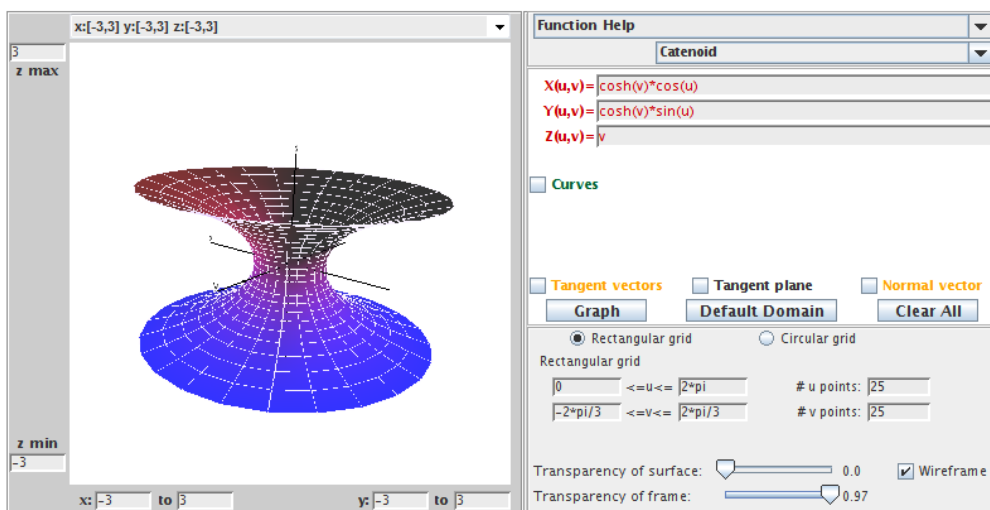
FIGURE 2.23. This parametrization of the catenoid is isothermal.

EXAMPLE 2.42. The parametrization

$$\mathbf{x}(u,v) = \big((a + b\cos v)\cos u, (a + b\cos v)\sin u, b\sin v\big)$$

for the torus is not isothermal. This is because in Exercise 2.32, you derived that

$$
\begin{aligned}
E &= (a + b\cos v)^2, \\
F &= 0, \\
G &= b^2.
\end{aligned}
$$

Because $F = 0$, the vectors $\mathbf{x}_u$ and $\mathbf{x}_v$ are orthogonal on the torus. But $E \geq G$ with equality only when $v = \pi + 2\pi k$, $(k \in \mathbb{Z})$ and either $a = 0$ or $a = 2b$. Thus the image of squares in the domain will be nonsquare rectangles whenever $v \neq \pi + 2\pi k$. Again, open *DiffGeomTool* and enter this parametrization for the torus with $a = 2$ and $b = 1$. Set the `Rectangular grid` values to $0 <= u <= 2*$ pi and $0 <= v <= 2*$ pi. Again, use the `Transparency of surface:` and the `Transparency of the frame:` sliders and click off and on the `Wireframe` button to make the curves more distinct. Notice that the grid of squares in the domain are mapped to a grid of mostly nonsquare rectangles as mentioned above. The ratio, $\frac{length}{height}$, of the sides of the rectangles is largest for the part of the torus farthest away from the origin. This occurs when $v = 0$ (or $v = 2\pi$) resulting in $E = 4$ while $G = 1$. On the other hand, the rectangles are squares for the part of the torus closest to the origin. This occurs when $v = \pi$ resulting in $E = 1$ while $G = 1$. This helps us see why this parametrization of the torus is not isothermal (see Figure 2.24).
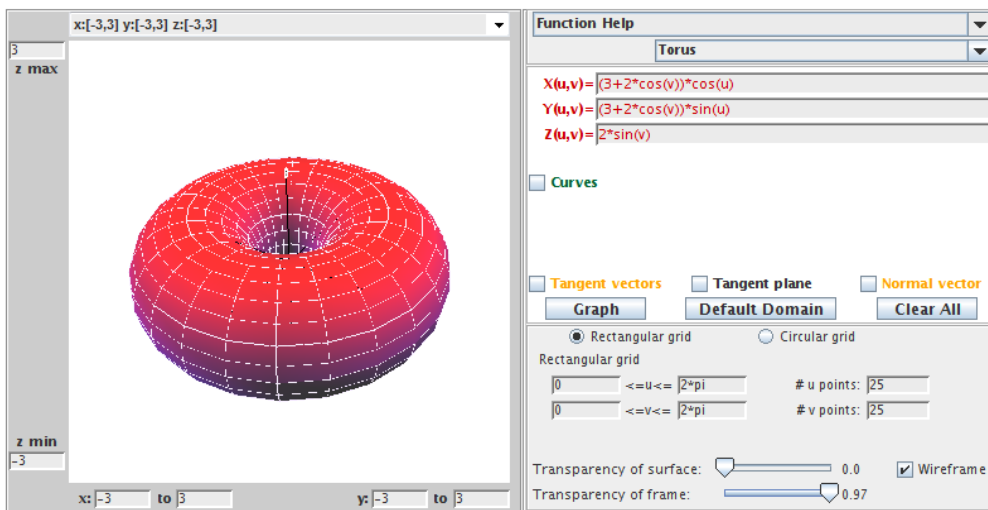
133

FIGURE 2.24. This parametrization of the torus is not isothermal.

EXERCISE 2.43. Using Definition 2.40 determine which of the following parametrizations of minimal surfaces is isothermal:

(a) The Enneper surface parametrized by
$$\mathbf{x}(u,v) = \left(u - \tfrac{1}{3}u^3 + uv^2, v - \tfrac{1}{3}v^3 + u^2v, u^2 - v^2 \right);$$

(b) Scherk's doubly periodic surface parametrized by
$$\mathbf{x}(u,v) = \left(u, v, \ln \left(\frac{\cos u}{\cos v} \right) \right);$$

(c) The helicoid parametized by
$$\mathbf{x}(u,v) = (a \sinh v \cos u, a \sinh v \sin u, au).$$

Try it out!

EXPLORATION 2.44. Use *DiffGeomTool* to check the reasonableness of your answers in Exercise 2.43 by graphing each parametrization in that exercise as was done in Examples 2.41 and 2.42. Set the # u points: to 20 and the # v points: to 20, and use the following values for the U/V domain boxes:

(a) Enneper surface:
$$-\tfrac{\pi}{3} <= \text{u,v} <= \tfrac{\pi}{3};$$

(b) Scherk's doubly periodic surface:
$$-\tfrac{\pi}{2} + 0.1 <= \text{u,v} <= \tfrac{\pi}{2} - 0.1;$$

(c) Helicoid:
$$-\pi <= \text{u,v} <= \pi.$$

Try it out!

From Exercise 2.43 and Exploration 2.44 you have seen there are parametrizations of minimal surfaces that are not isothermal. However, requiring minimal surfaces

to have an isothermal parametrization is not a restriction because of the following theorem.

THEOREM 2.45. Every minimal surface in $\mathbb{R}^3$ has an isothermal parametrization.

REMARK 2.46. See [**23**] for a proof of Theorem 2.45. In fact, every differentiable surface has an isothermal parametrization. This is a very interesting result. Unfortunately, a proof of this is beyond the scope of this text, but if you are interested, a proof is given in [**2**], pp 15-35.

Recall that in Example 2.35, we derived that the isothermal parametrization for the catenoid

$$\mathbf{x}(u, v) = \Big(a \cosh v \cos u, a \cosh v \sin u, av \Big)$$

has

$$e = -g.$$

In general, we have the following result.

THEOREM 2.47. Let M be a surface with isothermal parametrization. Then M is minimal if and only if $e = -g$.

EXERCISE 2.48. Prove Theorem 2.47.
Try it out!

EXPLORATION 2.49. Recall that $e = -g$ for the coefficients of the 2nd fundamental form represents that the u-parameter curve and the v-parameter curve are bending the same amount away from the normal $\mathbf{n}$ but in different directions. Use *DiffGeomTool* in connection with Theorem 2.47 to geometrically verify which of the following surfaces are minimal:

(a) Enneper surface:
$$\mathbf{x}(u, v) = \Big(u - \tfrac{1}{3}u^3 + uv^2, v - \tfrac{1}{3}v^3 + u^2 v, u^2 - v^2 \Big);$$
(b) Cylinder:
$$\mathbf{x}(u, v) = \big(\cos u, \sin u, v \big);$$
(c) Helicoid:
$$\mathbf{x}(u, v) = (a \sinh v \cos u, a \sinh v \sin u, av).$$
Try it out!

Now, here is an interesting and important result that brings in an idea from complex analysis. Recall from complex analysis that if $f(z) = x(u, v) + iy(u, v)$ is an analytic function, then the Cauchy-Riemann equations hold for f. That is,

$$x_u = y_v, \qquad x_v = -y_u.$$

In such a case, y is called the *harmonic conjugate* of x. Also, if f is analytic, then

(8)
$$f'(z) = x_u + iy_u.$$

This concept allows us to relate a minimal surface to another minimal surface, known as its conjugate minimal surface.

DEFINITION 2.50. Let $\mathbf{x}$ and $\mathbf{y}$ be isothermal parametrizations of minimal surfaces such that their component functions are pairwise harmonic conjugates. That is,

$$\text{(9)} \qquad\qquad \mathbf{x}_u = \mathbf{y}_v \qquad \text{and} \qquad \mathbf{x}_v = -\mathbf{y}_u.$$

In such a case, $\mathbf{x}$ and $\mathbf{y}$ are called *conjugate minimal surfaces*.

EXAMPLE 2.51. Let's find the conjugate surface of the catenoid parametrized by

$$\mathbf{x}(u,v) = (a \cosh v \cos u, a \cosh v \sin u, av).$$

Let $\mathbf{y}(u,v)$ be the parametrization of this conjugate surface. By the first part of eq (9), we know

$$\mathbf{y}_v = \mathbf{x}_u = (-a \cosh v \sin u, a \cosh v \cos u, 0).$$

Integrating this with respect to v yields

$$\mathbf{y} = (-a \sinh v \sin u + F_1(u), a \sinh v \cos u + F_2(u), F_3(u)),$$

where each $F_k(u)$ is a function independent of v. Similarly, by the second part of eq (9), we derive

$$\mathbf{y} = (-a \sinh v \sin u + G_1(v), a \sinh v \cos u + G_2(v), -au + G_3(v)).$$

Equating these two expressions for $\mathbf{y}$ we get that

$$\mathbf{y} = (-a \sinh v \sin u + K_1, a \sinh v \cos u + K_2, -au + K_3).$$

Using the substitution $u = \widetilde{u} - \frac{\pi}{2}$, $v = \widetilde{v}$, and letting $K_1 = 0$, $K_2 = 0$, and $K_3 = -a\frac{\pi}{2}$, does not affect the geometry of this minimal surface, and yields the parametrization of a helicoid

$$\mathbf{y}(\widetilde{u}, \widetilde{v}) = (a \sinh \widetilde{v} \cos \widetilde{u}, a \sinh \widetilde{v} \sin \widetilde{u}, -a\widetilde{u})$$

given in Exercise 2.24 (Note that the negative sign in the third component function just has the effect of reflecting the surface through the $x_1 x_2$-plane). Hence, the conjugate surface of this catenoid is a helicoid.

This idea of conjugate minimal surfaces gets really interesting. It turns out that any two conjugate minimal surfaces can be joined through a one-parameter family of minimal surfaces by the equation

$$\mathbf{z} = (\cos t)\mathbf{x} + (\sin t)\mathbf{y},$$

where $t \in \mathbb{R}$. Note that when $t = 0$ we have the minimal surface parametrized by $\mathbf{x}$, and when $t = \frac{\pi}{2}$ we have the minimal surface parametrized by $\mathbf{y}$. So for $0 \leq t \leq \frac{\pi}{2}$, we have a continuous parameter of minimal surfaces known as *associated surfaces*. In other words, we can continuously "morph" one minimal surface into another minimal surface so that all the in-between surfaces are also minimal.

In Example 2.51, we saw that the helicoid and the catenoid are conjugate surfaces. Images of them and certain associated surfaces are shown in Figure 2.25.
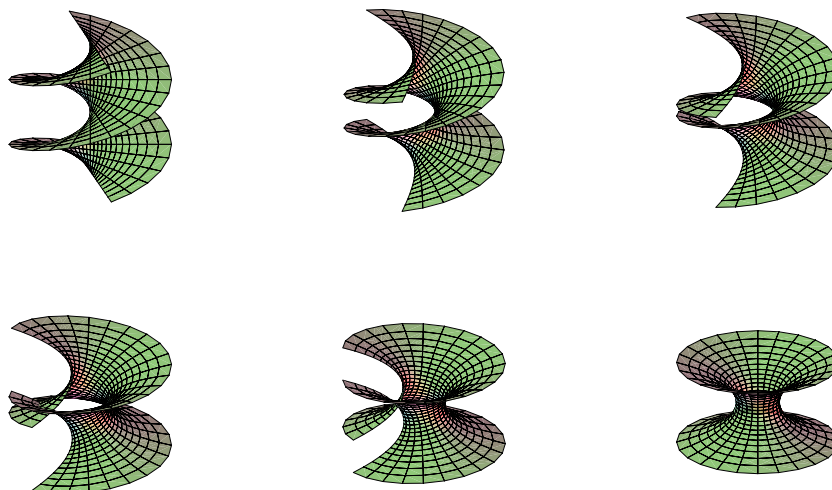
FIGURE 2.25. The helicoid, some associated surfaces, and the catenoid.

This is neat, but it is just the beginning. The rest of this section will explore properties of conjugate surfaces.

EXERCISE 2.52. Find the conjugate minimal surface for the Enneper surface

$$\mathbf{x}(u,v) = \left(u - \frac{1}{3}u^3 + uv^2, v - \frac{1}{3}v^3 + u^2v, u^2 - v^2 \right).$$

Try it out!

If we try to determine the conjugate minimal surface for Scherk's doubly periodic surface with the parametrization

$$\mathbf{x}(u,v) = \left(u, v, \ln\left(\frac{\cos u}{\cos v} \right) \right),$$

this method will not work, because this parametrization is not isothermal. However, later we will see that Scherk's doubly periodic surface does have a conjugate surface. It is Scherk's singly periodic surface (see Figure 2.26).

EXPLORATION 2.53. You can see the associated surfaces that occur between Scherk's doubly periodic surface and Scherk's singly periodic surface by using another applet for this chapter. This applet is called *MinSurfTool*, and it can be used to visualize and explore minimal surfaces in $\mathbb{R}^3$. Open the *MinSurfTool* applet. On the right-hand side near the top, there is a set of tabs for different features of this applet. For this exploration, we want to use the **W.E.(p,q)** feature, so make sure that the **W.E.(p,q)** tab is on top (if it is on top, it should be a different color than the other tabs). In the
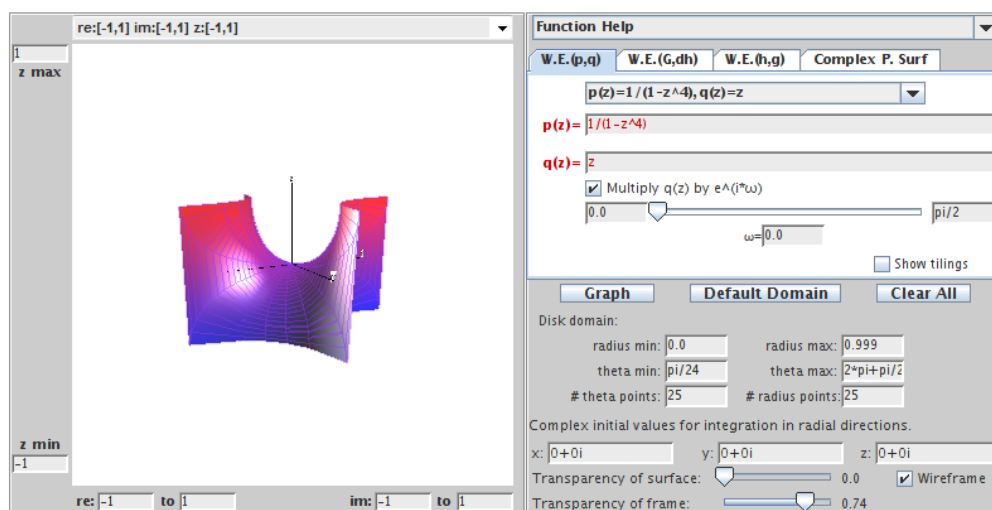
FIGURE 2.26. Scherk's singly periodic surface.

Pre-set functions window, choose $p(z) = 1/(1 - z^4)$, $q(z) = z$. Click on the **Graph** button, and one piece of Scherk's singly periodic surface will appear. Then move the slider arrow, that is above the **Graph** button, to the right to see how Scherk's singly periodic is transformed by way of the associated surfaces into Scherk's doubly periodic surface.

 Try it out!

Recall that individual pieces of Scherk's doubly periodic surface can be put together in the $x_1 x_2$-plane in a checkerboard fashion. So, these pieces repeat (or are periodic) in two directions, x_1 and x_2. Surfaces that repeat in one direction are called singly periodic. For example, individual pieces of Scherk's singly periodic surface can fit together creating a tower in the x_3 direction. You can visualize adding two pieces together by taking one piece of Scherk's singly periodic surface and adding it to another piece that has been reflected across the $x_1 x_2$-plane and shifted up in the x_3 direction. By continuing to do this, you can create a tower of several pieces (see Figure 2.27). Note that the helicoid is a singly periodic surface too.

Earlier in Example 2.31 we saw that the coefficients of the first fundamental form for the given parametrization of a catenoid are $E = a^2 \cosh^2 v$, $F = 0$, and $G = a^2 \cosh^2 v$. In Exercise 2.140 from the **Additional Exercises** at the end of the chapter it can be shown that for the given parametrization of Enneper's surface these coefficients are $E = (1 + u^2 + v^2)^2$, $F = 0$, and $G = (1 + u^2 + v^2)^2$. Clearly, the E's and G's do not match up. However, for any two conjugate minimal surfaces and their associated minimal surfaces, the coefficients of the first fundamental form are always the same. The following exercise will help you prove this surprising result.
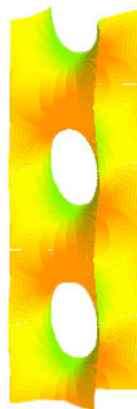
EXERCISE 2.54.

FIGURE 2.27. Scherk's singly periodic surface.

(a) Prove that given two conjugate minimal surfaces, $\mathbf{x}$ and $\mathbf{y}$, all surfaces of the one-parameter family

$$\mathbf{z} = (\cos t)\mathbf{x} + (\sin t)\mathbf{y}$$

have the same fundamental form: $E = \mathbf{x}_u \cdot \mathbf{x}_u = \mathbf{y}_u \cdot \mathbf{y}_u$, $F = 0$, $G = \mathbf{x}_v \cdot \mathbf{x}_v = \mathbf{y}_v \cdot \mathbf{y}_v$.

(b) Prove that all the surfaces in the one-parameter family $\mathbf{z}$ from part (a) are minimal for all $t \in \mathbb{R}$.

Try it out!

Finally, recall that the normal vector $\mathbf{n}$ at a point on a surface points orthogonally away from the surface. Since different minimal surfaces have different shapes, there is no reason to suspect that the normal vectors on one surface will be related to the normal vectors on another surface. However, for conjugate minimal surfaces and their associated minimal surfaces there is a strong connection. It turns out that for any point in the domain, the corresponding surface normal points in the same direction on all these minimal surfaces. The next theorem establishes this idea.

THEOREM 2.55. Let $\mathbf{x}, \mathbf{y} : D \to \mathbb{R}^3$ be isothermal parametrizations of conjugate minimal surfaces. Then for each $(u_0, v_0) \in D$, the corresponding surface unit normal is the same for all the associated surfaces.

PROOF. Let $(u_0, v_0) \in D$. Let $\mathbf{n}^{\mathbf{x}}$ and $\mathbf{n}^{\mathbf{y}}$ represent the surface normal for $\mathbf{x}$ and for $\mathbf{y}$, respectively. Then by the definition of conjugate surfaces, $\mathbf{x}$ and $\mathbf{y}$ have the same unit normal, because

$$\mathbf{n}^{\mathbf{x}} = \frac{\mathbf{x}_u \times \mathbf{x}_v}{\left|\mathbf{x}_u \times \mathbf{x}_v\right|} = \frac{\mathbf{y}_v \times -\mathbf{y}_u}{\left|\mathbf{y}_v \times -\mathbf{y}_u\right|} = \frac{\mathbf{y}_u \times \mathbf{y}_v}{\left|\mathbf{y}_u \times \mathbf{y}_v\right|} = \mathbf{n}^{\mathbf{y}}.$$

To show that this is true for the associated surfaces , let $\mathbf{z} = (\cos t)\mathbf{x} + (\sin t)\mathbf{y}$ be the parametrization of the associated surfaces. Then

$$
\begin{aligned}
\mathbf{z}_u \times \mathbf{z}_v &= \big(\cos t\mathbf{x}_u + \sin t\mathbf{y}_u \big) \times \big(\cos t\mathbf{x}_v + \sin t\mathbf{y}_v \big) \\
&= \cos^2 t\big(\mathbf{x}_u \times \mathbf{x}_v\big) + \cos t \sin t\big(\mathbf{x}_u \times \mathbf{y}_v\big) + \cos t \sin t\big(\mathbf{y}_u \times \mathbf{x}_v\big) + \sin^2 t\big(\mathbf{y}_u \times \mathbf{y}_v\big) \\
&= \cos^2 t\big(\mathbf{x}_u \times \mathbf{x}_v\big) + \cos t \sin t\big(\mathbf{x}_u \times \mathbf{x}_u\big) + \cos t \sin t\big(- \mathbf{x}_v \times \mathbf{x}_v\big) + \sin^2 t\big(- \mathbf{x}_v \times \mathbf{x}_u\big) \\
&= \cos^2 t\big(\mathbf{x}_u \times \mathbf{x}_v\big) + \cos t \sin t\big(\mathbf{0}\big) + \cos t \sin t\big(\mathbf{0}\big) + \sin^2 t\big(\mathbf{x}_u \times \mathbf{x}_v\big) \\
&= \mathbf{x}_u \times \mathbf{x}_v.
\end{aligned}
$$

$\square$

The following example and exploration helps us visualize this idea.

EXAMPLE 2.56. Using *DiffGeomTool* we can graph the catenoid and its conjugate surface, the helicoid, whose parametrizations are given in Example 2.51. If we plot the normal $\mathbf{n}$ at the point $\left(\frac{\pi}{3}, -\frac{\pi}{4}\right)$ on these conjugate surfaces, we see that both normals point in the same direction as guaranteed by Theorem 2.55 (see Figure 2.28 and Figure 2.29).
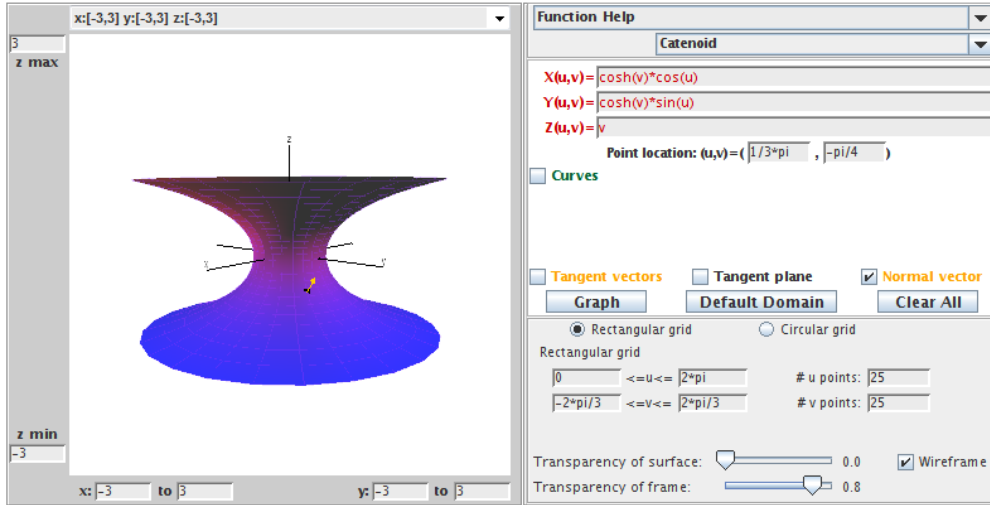

FIGURE 2.28. The catenoid with $\mathbf{n}$ at $\left(\frac{\pi}{3}, -\frac{\pi}{4}\right)$.

EXPLORATION 2.57. Open two separate windows of *DiffGeomTool*. In one plot the catenoid parametrized by

$$\mathbf{x}(u, v) = (\cosh v \cos u, \cosh v \sin u, v),$$

where $0 \le u \le 2\pi$ and $-\frac{2\pi}{3} \le v \le \frac{2\pi}{3}$. In the other plot its conjugate surface, the helicoid, with parametrization

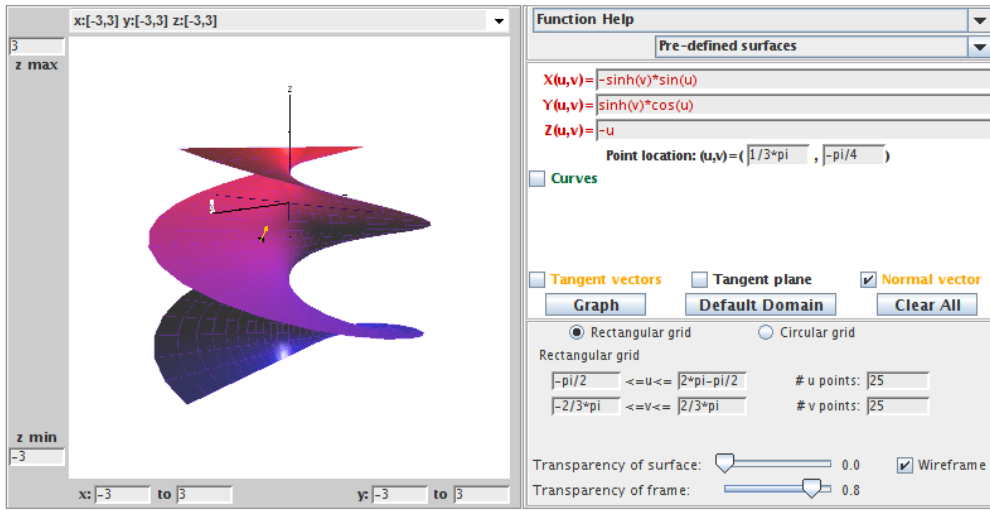$$\mathbf{y}(u, v) = (-\sinh v \sin u, \sinh v \cos u, -u),$$

FIGURE 2.29. The conjugate helcoid with $\mathbf{n}$ at $\left(\frac{\pi}{3}, -\frac{\pi}{4}\right)$.

where $-\frac{\pi}{2} \le u \le 2\pi - \frac{\pi}{2}$ and $-\frac{2\pi}{3} \le v \le \frac{2\pi}{3}$ (Note that the u values for the helicoid are different than the values for the catenoid because we used the substitution $u = \widetilde{u} - \frac{\pi}{2}$ in Example 2.51). Make sure that the x, y, z-axes are positioned in the same directions. Then plot the following unit normals, $\mathbf{n}$, on each surface at the following points and observe that $\mathbf{n}$ points in the same direction as prescribed by Theorem 2.55:

(a) at $\left(\frac{\pi}{3}, 0\right)$,   (b) at $\left(\frac{\pi}{4}, -\frac{\pi}{2}\right)$,   (c) at $\left(\frac{\pi}{6}, \frac{\pi}{2}\right)$.
    Try it out!

2.4. Weierstrass Representation

At the end of the last section, we saw that we could apply complex analysis to minimal surface theory to define conjugate surfaces. In this chapter we will again use complex analysis to learn more about minimal surfaces. First, we will use inherent properties of an isothermal paramatrization to give us a necessary and sufficient condition for a surface to be minimal. This condition is very useful for finding and classifying minimal surfaces, as we will see. First, we state a theorem which will lead to the desired condition.

THEOREM 2.58. If the parametrization $\mathbf{x}$ is isothermal, then

$$\mathbf{x}_{uu} + \mathbf{x}_{vv} = 2EH\mathbf{n},$$

where E is a coefficient of the first fundamental form and H is the mean curvature.

EXERCISE 2.59 (Proof of Theorem 2.58). The set $\{\mathbf{x}_u, \mathbf{x}_v, \mathbf{n}\}$ forms a basis for $\mathbb{R}^3$. Assume $F = 0$. Then the vector $\mathbf{x}_{uu}$ can be expressed in terms of these bases vectors.

141

That is,
$$\mathbf{x}_{uu} = \Gamma^u_{uu}\mathbf{x}_u + \Gamma^v_{uu}\mathbf{x}_v + e\mathbf{n},$$
where the coefficients, Γ^u_{uu} and Γ^v_{uu}, are known as Christoffel symbols and e comes from the coefficient of the second fundamental form. That is, $e = \mathbf{n} \cdot \mathbf{x}_{uu}$.

(a) Show that $\Gamma^u_{uu} = \frac{E_u}{2E}$ and $\Gamma^v_{uu} = -\frac{E_v}{2G}$ by taking the inner product of $\mathbf{x}_{uu}$ with each of the basis vectors. In a similar manner, it can be shown that
$$\mathbf{x}_{vv} = -\frac{G_u}{2E}\mathbf{x}_u + \frac{G_v}{2G}\mathbf{x}_v + g\mathbf{n}.$$

(b) Use the mean curvature equation (5) and the results from (a) to show that if the parametrization $\mathbf{x}$ is isothermal, then
$$\mathbf{x}_{uu} + \mathbf{x}_{vv} = 2EH\mathbf{n}.$$

Try it out!

Now, where do we go from here? Recall that for a minimal surface, $H \equiv 0$. So Theorem 2.58 tells us that $\mathbf{x}_{uu} + \mathbf{x}_{vv} \equiv 0$. But what does this last equation represent? It is Laplace's equation and relates to harmonic functions. Recall that $\varphi(u, v)$ is a real-valued *harmonic function* if $\varphi_{uu} + \varphi_{vv} = 0$ (for example, $\varphi(u, v) = u^2 - v^2$ is harmonic). This leads to the condition for a surface to be minimal that we referred to in the beginning of this section. The corollary below provides the final piece of the puzzle that will allow us to explicitly describe minimal surfaces using the Weierstrass representation, which we will derive in this section.

COROLLARY 2.60. A surface M with an isothermal parametrization $\mathbf{x}(u, v) = \big(x_1(u, v), x_2(u, v), x_3(u, v)\big)$ is minimal if and only if x_1, x_2, x_3 are harmonic.

Make sure you understand the significance of this result. First, we need an isothermal parametrization for our surface, but this is not a difficulty because of Theorem 2.45. Then this result tells us we will have a minimal surface if and only if the coordinate functions of that parametrization are harmonic functions. This will provide us another way to create and to prove a surface is minimal.

PROOF. ($\Rightarrow$) If M is minimal, then $H = 0$ and so by Theorem 2.58 $\mathbf{x}_{uu} + \mathbf{x}_{vv} = 0$, and hence the coordinate functions are harmonic. ($\Leftarrow$) Suppose x_1, x_2, x_3 are harmonic. Then $\mathbf{x}_{uu} + \mathbf{x}_{vv} = 0$. So by Theorem 2.58 we have that $2(\mathbf{x}_u \cdot \mathbf{x}_u)H\mathbf{n} = 0$. But $\mathbf{n} \neq 0$ and $E = \mathbf{x}_u \cdot \mathbf{x}_u \neq 0$. Hence, $H = 0$ and M is minimal. $\square$

EXERCISE 2.61. Given the parametrization for the Enneper surface
$$\mathbf{x}(u, v) = \left(u - \frac{1}{3}u^3 + uv^2, v - \frac{1}{3}v^3 + u^2v, u^2 - v^2\right),$$
use Corollary 2.60 to prove that the Enneper surface is a minimal surface.
Try it out!

The importance of Corollary 2.60 is not in proving specific surfaces are minimal. Instead, it lies in establishing a general formula that will guarantee any surface created by it will be minimal. This formula is known as the Weierstrass representation for minimal surfaces. This is neat, because it will provide us with a simple way to construct a lot of examples of minimal surfaces using functions from complex analysis. After stating the Weierstrass representation in Theorem 2.66 we will use the rest of this section to create more minimal surfaces. However, it turns out that not all minimal surfaces are of equal interest. So in section 2.5 of this chapter we leave behind the idea of creating arbitrary minimal surfaces and instead explore properties that make certain minimal surfaces more interesting.

Now we will derive this important formula, the Weierstrass Representation, and bring in the connection with complex analysis. Suppose M is a minimal surface with an isothermal parametrization $\mathbf{x}(u, v)$. Let $z = u + iv$ be a point in the complex plane. Recall that $\bar{z} = u - iv$ is the conjugate of z. Using these representations for z and $\bar{z}$ we can solve for u, v in terms of $z, \bar{z}$ to get

$$u = \frac{z + \bar{z}}{2} \text{ and } v = \frac{z - \bar{z}}{2i}.$$

Then the parametrization of the minimal surface M can be written in terms of the complex variables z and $\bar{z}$ as:

$$\mathbf{x}(z, \bar{z}) = \big(x_1(z, \bar{z}), x_2(z, \bar{z}), x_3(z, \bar{z})\big).$$

EXERCISE 2.62. Let $f(u, v) = x(u, v) + iy(u, v)$ be a complex function. Using the notation $u = \frac{z + \bar{z}}{2}$ and $v = \frac{z - \bar{z}}{2i}$, we can express f in terms of z and $\bar{z}$ instead of u and v. That is, we have the function $f(z, \bar{z})$. In this exercise you will prove the neat result that f is analytic if and only if f can be written in terms of $z = u + iv$ alone without using $\bar{z} = u - iv$.

(a) Using the chain rule, derive the following formulas:

$$\frac{\partial f}{\partial z} = \frac{1}{2}\left(\frac{\partial x}{\partial u} + \frac{\partial y}{\partial v}\right) + \frac{i}{2}\left(\frac{\partial y}{\partial u} - \frac{\partial x}{\partial v}\right),$$

$$\frac{\partial f}{\partial \bar{z}} = \frac{1}{2}\left(\frac{\partial x}{\partial u} - \frac{\partial y}{\partial v}\right) + \frac{i}{2}\left(\frac{\partial y}{\partial u} + \frac{\partial x}{\partial v}\right).$$

(b) Show that f is analytic $\iff \frac{\partial f}{\partial \bar{z}} = 0$.

Try it out!

EXAMPLE 2.63. The function $f_1(z) = z^2$ is analytic, because $\frac{\partial f_1}{\partial \bar{z}}(z^2) = 0$. However, $f_2(z) = |z|^2 = z\bar{z}$ is not analytic, because $\frac{\partial f_1}{\partial \bar{z}} = z \neq 0$.

EXERCISE 2.64. Prove that

(10)
$$4\left(\frac{\partial}{\partial z}\left(\frac{\partial f}{\partial \bar{z}}\right)\right) = f_{uu} + f_{vv}.$$

Try it out!

The next theorem expands upon Corollary 2.60 to establish the Weierstrass representation for minimal surfaces.

THEOREM 2.65. Let M be a surface with parametrization $\mathbf{x} = (x_1, x_2, x_3)$ and let $\phi = (\varphi_1, \varphi_2, \varphi_3)$, where $\varphi_k = \frac{\partial x_k}{\partial z}$. Then $\mathbf{x}$ is isothermal $\iff \phi^2 = (\varphi_1)^2 + (\varphi_2)^2 + (\varphi_3)^2 = 0$. If $\mathbf{x}$ is isothermal, then M is minimal $\iff$ each φ_k is analytic.

Before we prove Theorem 2.65, let's look at applying it to a specific example to help us better understand what the theorem is saying. Suppose we have the parametrization $\mathbf{x} = (x_1, x_2, x_3) = (z - \frac{1}{3}z^3, -i(z + \frac{1}{3}z^3), z^2)$. Then $\varphi_1 = \frac{\partial x_1}{\partial z} = 1 - z^2$, $\varphi_2 = \frac{\partial x_2}{\partial z} = -i(1 + z^2)$, and $\varphi_3 = \frac{\partial x_3}{\partial z} = 2z$. Notice that $\phi^2 = [1 - z^2]^2 + [-i(1 + z^2)]^2 + [2z]^2 = 0$. Thus, by the theorem, the parametrization $\mathbf{x}$ is isothermal. Also, each φ_k is a polynomial and hence analytic. So $\mathbf{x}$ is a parametrization of a minimal surface (in fact, it is Enneper's surface). Make sure you understand how this example relates to Theorem 2.65 before you read the following proof of the theorem.

PROOF. Applying the complex differential operator $\frac{\partial f}{\partial z}$ from Exercise 2.62 to this situation and then squaring the terms, we have $(\varphi_k)^2 = \left(\frac{\partial x_k}{\partial z}\right)^2 = \left[\frac{1}{2}\left(\frac{\partial x_k}{\partial u} - i\frac{\partial x_k}{\partial v}\right)\right]^2 = \frac{1}{4}\left[\left(\frac{\partial x_k}{\partial u}\right)^2 - \left(\frac{\partial x_k}{\partial v}\right)^2 - 2i\frac{\partial x_k}{\partial u}\frac{\partial x_k}{\partial v}\right]$. Also, recall that $\mathbf{x_u} \cdot \mathbf{x_u} = \left(\frac{\partial x_1}{\partial u}\right)^2 + \left(\frac{\partial x_2}{\partial u}\right)^2 + \left(\frac{\partial x_3}{\partial u}\right)^2 = \sum_{k=1}^{3}\left(\frac{\partial x_k}{\partial u}\right)^2$ and similarly $\mathbf{x_v} \cdot \mathbf{x_v} = \sum_{k=1}^{3}\left(\frac{\partial x_k}{\partial v}\right)^2$. Hence,

$$
\begin{aligned}
\phi^2 &= (\varphi_1)^2 + (\varphi_2)^2 + (\varphi_3)^2 \\
&= \frac{1}{4}\left[\sum_{k=1}^{3}\left(\frac{\partial x_k}{\partial u}\right)^2 - \sum_{k=1}^{3}\left(\frac{\partial x_k}{\partial v}\right)^2 - 2i\sum_{k=1}^{3}\frac{\partial x_k}{\partial u}\frac{\partial x_k}{\partial v}\right] \\
&= \frac{1}{4}\left(\mathbf{x}_u \cdot \mathbf{x}_u - \mathbf{x}_v \cdot \mathbf{x}_v - 2i(\mathbf{x}_u \cdot \mathbf{x}_v)\right) \\
&= \frac{1}{4}(E - G - 2iF).
\end{aligned}
$$

Thus, $\mathbf{x}$ is isothermal $\iff E = G, F = 0 \iff \phi^2 = 0$.

Now suppose that $\mathbf{x}$ is isothermal. By Corollary 2.60, it suffices to show that for each k, x_k is harmonic $\iff \varphi_k$ is analytic. Using eq (10) and Exercise 2.62 this follows because

$$
\frac{\partial^2 x_k}{\partial u \partial u} + \frac{\partial^2 x_k}{\partial v \partial v} = 4\left(\frac{\partial}{\partial \bar{z}}\left(\frac{\partial x_k}{\partial z}\right)\right) = 4\left(\frac{\partial}{\partial \bar{z}}\left(\varphi_k\right)\right) = 0.
$$

$\square$

144

Note that if $\mathbf{x}$ is isothermal

$$|\phi|^2 = \left|\frac{\partial x_1}{\partial z}\right|^2 + \left|\frac{\partial x_2}{\partial z}\right|^2 + \left|\frac{\partial x_3}{\partial z}\right|^2 = \frac{1}{4}\left(\sum_{k=1}^{3}\left(\frac{\partial x_k}{\partial u}\right)^2 + \sum_{k=1}^{3}\left(\frac{\partial x_k}{\partial v}\right)^2\right)$$

$$= \frac{1}{4}\left(\mathbf{x}_u \cdot \mathbf{x}_u + \mathbf{x}_v \cdot \mathbf{x}_v\right) = \frac{1}{4}(E + G) = \frac{E}{2}.$$

So if $|\phi|^2 = 0$, then all the coefficients of the first fundamental form are zero and M degenerates to a point. Similarly, we want $|\phi|^2$ to be finite.

Finally, we need to solve $\varphi_k = \frac{\partial x_k}{\partial z}$ for x_k since the parametrization of the surface is given as $\mathbf{x} = (x_1, x_2, x_3)$. The difficulty is that x_k is a function of two variables, z and $\bar{z}$, and we want to have a representation in which we only have to integrate with respect to one variable. To overcome this difficulty, we will use some ideas about differentials (see [26] for a nice introduction to differentials). First, since x_k is also a function of the two variables u and v, we can write

$$(11) \qquad dx_k = \frac{\partial x_k}{\partial u}du + \frac{\partial x_k}{\partial v}dv.$$

Also, $dz = du + idv$. Using Exercise 2.62 we have

$$\varphi_k dz = \frac{\partial x_k}{\partial z}dz = \frac{1}{2}\left(\frac{\partial x_k}{\partial u} - i\frac{\partial x_k}{\partial v}\right)(du + idv)$$

$$= \frac{1}{2}\left[\frac{\partial x_k}{\partial u}du + \frac{\partial x_k}{\partial v}dv + i\left(\frac{\partial x_k}{\partial u}dv - \frac{\partial x_k}{\partial v}du\right)\right],$$

$$\overline{\varphi_k dz} = \overline{\varphi_k}\,\overline{dz} = \frac{\overline{\partial x_k}}{\partial z}\overline{dz} = \frac{1}{2}\left(\frac{\partial x_k}{\partial u} + i\frac{\partial x_k}{\partial v}\right)(du - idv)$$

$$= \frac{1}{2}\left[\frac{\partial x_k}{\partial u}du + \frac{\partial x_k}{\partial v}dv - i\left(\frac{\partial x_k}{\partial u}dv - \frac{\partial x_k}{\partial v}du\right)\right].$$

Adding these two equations yields

$$(12) \qquad \frac{\partial x_k}{\partial u}du + \frac{\partial x_k}{\partial v}dv = \varphi_k dz + \overline{\varphi_k dz} = 2\,\text{Re}\{\varphi_k dz\}.$$

Combining eq (11) and eq (12), we have

$$dx_k = 2\,\text{Re}\{\varphi_k dz\}.$$

Therefore, $x_k = 2\,\text{Re}\int \varphi_k dz + c_k$. Since adding c_k just translates the image by a constant amount and multiplying each coordinate function by 2 just scales the the surface, these constants do not affect the geometric shape of the surface. Hence, we do not need these constants and we will let our coordinate function be

$$x_k = \text{Re}\int \varphi_k dz.$$

145

Summary: If we have analytic functions φ_k $(k = 1, 2, 3)$ such that

$$\phi^2 = 0 \qquad \text{and} \qquad |\phi|^2 \neq 0 \text{ and is finite,}$$

then the parametrization

(13) $$\mathbf{x} = \left(\operatorname{Re} \int \varphi_1(z)dz, \operatorname{Re} \int \varphi_2(z)dz, \operatorname{Re} \int \varphi_3(z)dz \right)$$

defines a minimal surface.

For example, consider

$$\begin{aligned} \varphi_1 &= p(1 + q^2) \\ \varphi_2 &= -ip(1 - q^2) \\ \varphi_3 &= -2ipq. \end{aligned}$$

Then

$$\begin{aligned} \phi^2 &= [p(1 + q^2)]^2 + [-ip(1 - q^2)]^2 + [-2ipq]^2 \\ &= [p^2 + 2p^2q^2 + p^2q^4] - [p^2 - 2p^2q^2 + p^2q^4] - [4p^2q^2] \\ &= 0, \end{aligned}$$

and

$$\begin{aligned} |\phi|^2 &= |p(1 + q^2)|^2 + |-ip(1 - q^2)|^2 + |-2ipq|^2 \\ &= |p|^2[(1 + q^2)(1 + \bar{q}^2) + (1 - q^2)(1 - \bar{q}^2) + 4q\bar{q}] \\ &= |p|^2[2(1 + 2q\bar{q} + q^2\bar{q}^2) \\ &= 4|p|^2(1 + |q|^2)|^2 \neq 0 \qquad \text{(note: if } p = 0, \text{ then } \varphi_k = 0 \text{ for all } k\text{).} \end{aligned}$$

Notice that p, pq^2, and pq have to be analytic in order for each φ_k to be analytic. If p is analytic with a zero of order $2m$ at z_0, then q can have a pole of order no larger than m at z_0. Recall that a function that is analytic in a domain D except possibly at poles is known as a *meromorphic* function in D. This leads to the following result.

THEOREM 2.66 (Weierstrass Representation (p,q)). Every regular minimal surface has a local isothermal parametric representation of the form

$$\begin{aligned} \mathbf{x} &= (x_1(z), x_2(z), x_3(z)) \\ &= \left(\operatorname{Re}\left\{ \int_a^z p(1 + q^2)dz \right\}, \right. \\ &\qquad \operatorname{Re}\left\{ \int_a^z -ip(1 - q^2)dz \right\}, \\ &\qquad \left. \operatorname{Re}\left\{ \int_a^z -2ipqdz \right\} \right), \end{aligned}$$

146

where p is an analytic function and q is a meromorphic function in some domain $\Omega \subset \mathbb{C}$, having the property that at each point where q has a pole of order m, p has a zero of order at least $2m$, and $a \in \Omega$ is a constant.

EXAMPLE 2.67. For $p(z) = 1, q(z) = iz$, we get

$$\mathbf{x} = \left(\operatorname{Re}\left\{ \int_0^z (1 - z^2)\, dz \right\}, \operatorname{Re}\left\{ \int_0^z -i(1 + z^2)\, dz \right\}, \operatorname{Re}\left\{ \int_0^z 2z\, dz \right\} \right)$$
$$= \left(\operatorname{Re}\left\{ z - \frac{1}{3}z^3 \right\}, \operatorname{Re}\left\{ -i\left(z + \frac{1}{3}z^3 \right) \right\}, \operatorname{Re}\left\{ z^2 \right\} \right).$$

Letting $z = u + iv$, this yields

$$\mathbf{x}(u, v) = \left(u - \frac{1}{3}u^3 + uv^2, v - \frac{1}{3}v^3 + u^2v, u^2 - v^2 \right)$$

which gives the Enneper surface.

You can use the applet, *MinSurfTool*, to graph an image of this surface using the functions p and q. After opening *MinSurfTool*, make sure that the W.E.(p,q) tab is on top. In the appropriate boxes, put $p(z) = 1$ and $q(z) = i * z$. Then click on the **Graph** button. Remember that you can increase the size of the image of the surface by clicking on the left button on the mouse, and you can decrease the size by clicking on the right mouse button. Also, you can rotate the surface by placing the cursor arrow on the image of the surface, then click on and hold the left button on the mouse as you move the cursor.
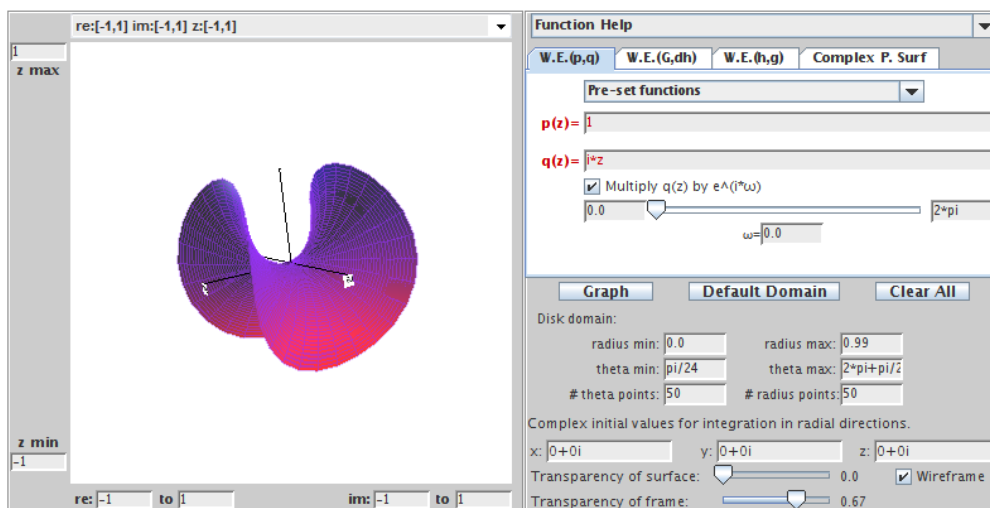

FIGURE 2.30. Enneper surface using $p(z) = 1$ and $q(z) = iz$.

EXAMPLE 2.68. Let $p(z) = 1$ and $q(z) = 1/z$ on the domain $\mathbb{C} - \{0\}$. Notice that q is meromorphic with a pole of order 1 at $z_0 = 0$ while p does not have a zero of

order 2 at $z_0 = 0$. This does not violate the conditions of Theorem 2.66, because the domain is $\mathbb{C} - \{0\}$. We will show that this generates a helicoid. Using Weierstrass Representation (p,q) and letting $z = u + iv$, we get $\mathbf{x}(u, v) = (x_1, x_2, x_3)$, where

$$x_1 = \operatorname{Re} \int_1^z \left(1 + \frac{1}{z^2}\right) dz = \operatorname{Re}\left(z - \frac{1}{z}\right) = u - \frac{u}{u^2 + v^2}$$

$$x_2 = \operatorname{Re} \int_1^z -i\left(1 - \frac{1}{z^2}\right) dz = \operatorname{Im}\left(z + \frac{1}{z}\right) = v - \frac{v}{u^2 + v^2}$$

$$x_3 = \operatorname{Re} \int_1^z -2i\frac{1}{z}\, dz = 2\operatorname{Im}(\log z) = 2 \arg z = 2 \arctan\left(\frac{v}{u}\right).$$

Notice that this parametrization is different than the following parametrization we have been using for the helicoid:

$$\widetilde{\mathbf{x}}(\widetilde{u}, \widetilde{v}) = (\widetilde{x}_1, \widetilde{x}_2, \widetilde{x}_3) = (a \sinh \widetilde{v} \cos \widetilde{u}, a \sinh \widetilde{v} \sin \widetilde{u}, a\widetilde{u}).$$

To show that $\mathbf{x}$ also gives an image of the helicoid, we will find a substitution that will change $\mathbf{x}$ into $\widetilde{\mathbf{x}}$. Note that

$$x_1^2 + x_2^2 = (u^2 + v^2) - 2 + \frac{1}{u^2 + v^2}$$

$$\widetilde{x}_1{}^2 + \widetilde{x}_2{}^2 = a^2 \sinh^2 \widetilde{v} = a^2 \left(\frac{e^{\widetilde{v}} - e^{-\widetilde{v}}}{2}\right)^2.$$

Equating the right hand side of these equations and letting $a = 2$, we get that

$$u^2 + v^2 = e^{2\widetilde{v}}.$$

Also, with $x_3 = \widetilde{x}_3$, we see that

$$\frac{v}{u} = \tan \widetilde{u}.$$

Now, using these last two equations we can solve for u and v to get

$$u = e^{\widetilde{v}} \cos \widetilde{u} \quad \text{and} \quad v = e^{\widetilde{v}} \sin \widetilde{u}.$$

If we substitute these values for u and v into $\mathbf{x}(u, v)$ we get the parametrization $\widetilde{\mathbf{x}}(\widetilde{u}, \widetilde{v})$ for the helicoid.

Using the W.E.(p,q) tab in *MinSurfTool*, we can get a graph of the helicoid by choosing from the **Pre-set functions** the values $p(z) = 1$ and $q(z) = 1/z$. Note that since the domain is $\mathbb{C} \setminus \{0\}$, the Disk domain: radius min: box is set to 0.2. Also, because of the singularity at $z = 0$, there needs to be a function entered into the x, y and z boxes in the Complex initial values for integration in radial direction section. These functions come from explicitly solving the integrals for $x_1(z)$, $x_2(z)$, and $x_3(z)$ in the Weierstrass representation when $p(z) = 1$ and $q(z) = 1/z$ and then substituting in $re^{i\theta}$ for z. If you use a choice from the **Pre-set functions**, these functions will be entered automatically. However, if you enter your own p and q values,

you will need to compute these functions and enter them into these boxes in order to get the correct minimal surface image.
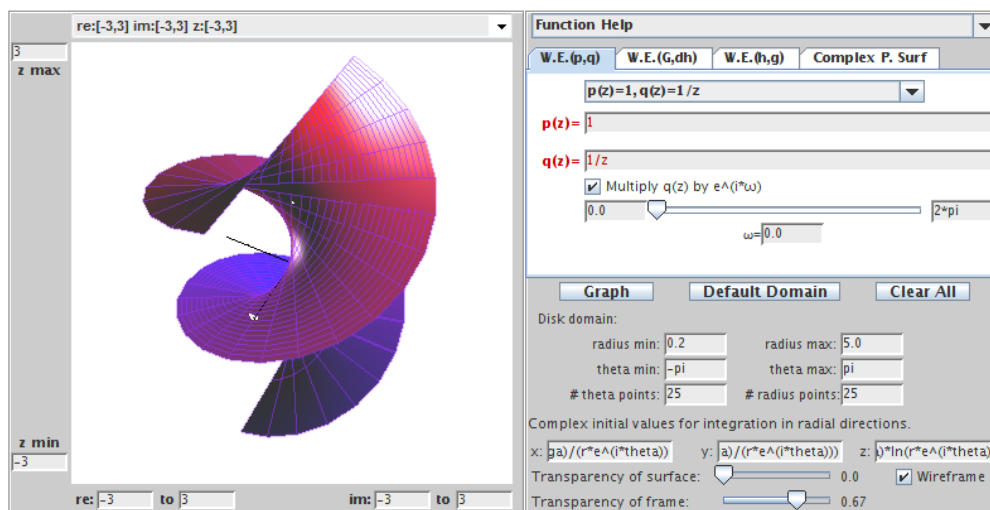

FIGURE 2.31. The helicoid using $p(z) = 1$ and $q(z) = \frac{1}{z}$.

EXERCISE 2.69. Show that the minimal surfaces generated by using $p(z) = 1$ and $q(z) = 0$ on the domain $\mathbb{C}$ in the Weierstrass representation is the plane.
Try it out!

EXERCISE 2.70. Show that the minimal surfaces generated by using $p(z) = 1$ and $q(z) = i/z$ on the domain $\mathbb{C} - \{0\}$ in the Weierstrass representation is the catenoid. Use the appropriate **Pre-set function** in the W.E.(p,q) tab of *MinSurfTool* to graph an image of this surface.
Try it out!

EXPLORATION 2.71. Enneper's surface can be constructed also with $p(z) = 1$ and $q(z) = z$. Recall that it has four leaves (two pointing up and two pointing down). The number of leaves can be increased.

(a) Using $p(z) = 1$ and $q(z) = z^2$ on the domain $\mathbb{C}$ in the Weierstrass representation gives the Enneper surface with six leaves (see Figure 2.32). Compute the parametrization $\mathbf{x}(u, v)$ for this surface.
(b) Conjecture the values of p and q for the Enneper surface with n leaves.
(c) Use *MinSurfTool* with the W.E.(p,q) tab to check your conjectured values of p and q for the Enneper surface with n leaves.

Try it out!

EXPLORATION 2.72. Use *MinSurfTool* with the W.E.(p,q) tab to graph an image of the surface generated by the **Pre-set functions** $p(z) = \frac{1}{1-z^4}$ and $q(z) = z$.
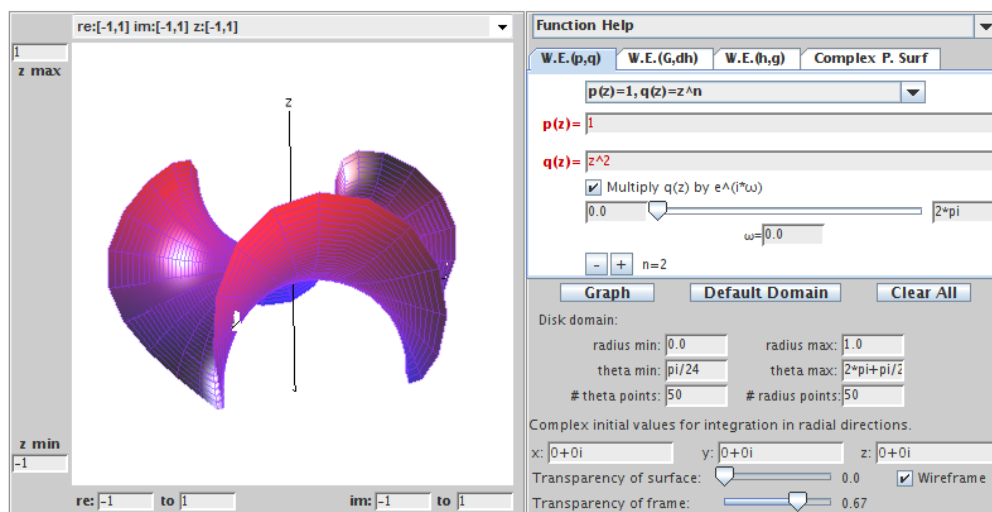
149

FIGURE 2.32. Enneper surface with 6 leaves using $p(z) = 1$ and $q(z) = z^2$.

(a) What minimal surface is this?
(b) Click on the box "Multiply $q(z)$ by e∧(i*ω)" and move the slider to generate a family of minimal surfaces. These surfaces are associated surfaces (see the paragraph after Definition 2.50). When $\theta = \frac{\pi}{2}$ you get the conjugate surface. In this case, what is the conjugate surface?
(c) Experiment with *MinSurfTool* to view the associated family and find the conjugate surface of various minimal surfaces discussed above.

Try it out!

EXPLORATION 2.73. Scherk's doubly periodic surface is generated with $p(z) = \frac{1}{1-z^4}$ and $q(z) = iz$. Use the appropriate **Pre-set function** in the W.E.(p,q) tab of *MinSurfTool* to graph an image of the surface generated by $p(z) = \frac{1}{1-z^{2n}}$ and $q(z) = iz^{n-1}$ for various values of $n = 2, 3, 4, \ldots$. Note the $-$ and $+$ boxes above the **Graph** button; these can be used to increase or decrease the value of n.

(a) What happens to the surface as n increases?
(b) Notice that the surface has leaves that alternate between going up and going down. How is n related to the number of leaves?
(c) What is the image of the projection of the surface onto the x_1x_2-plane for each n?
(d) Using the previous parts conjecture how many leaves the surface would have if $p(z) = \frac{1}{1-z^5}$. Why could such a surface not exist?

Try it out!

EXPLORATION 2.74. Use the appropriate **Pre-set function** in the `W.E.(p,q)` tab of *MinSurfTool* to graph an image of the surface generated by $p(z) = \frac{1}{(1-z^4)^2}$ and $q(z) = iz^3$. This surface is known as the 4-noid (see Figure 2.33).

(a) Try to create a 3-noid by changing the values of p and q and graphing the result in *MinSurfTool*.

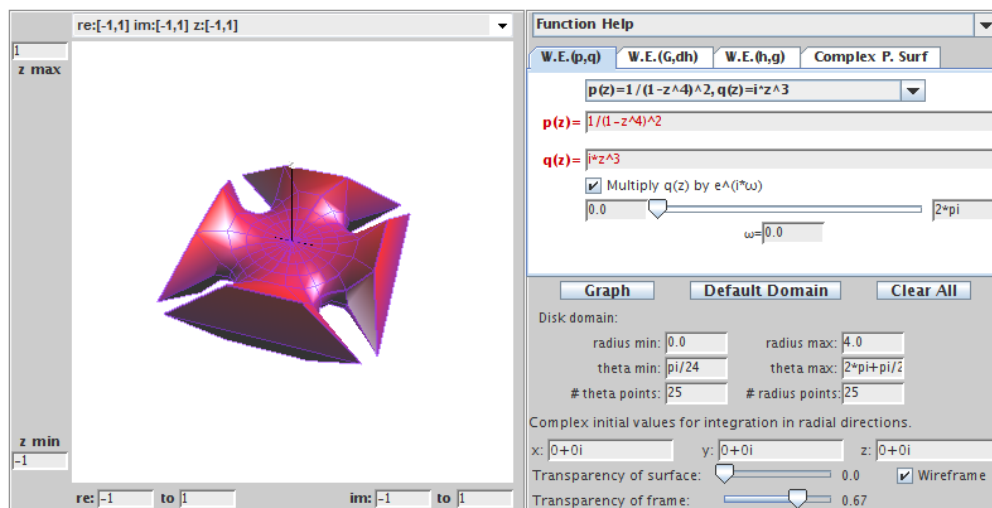(b) Conjecture the values of p and q that will generate an n-noid.

FIGURE 2.33. Image of the 4-noid minimal surface.

Try it out!

While the Weierstrass Representation will generate a minimal surface, there is no guarantee that the minimal surface will be embedded. Recall from the introduction of this chapter that we said that a surface is embedded if it has no self-intersections. The plane, the catenoid, the helicoid, and Scherk's doubly periodic surface are examples of embedded minimal surfaces. However, Enneper's surface is not embedded. In Exploration 2.11 you saw that Enneper's surface intersects itself when the domain contains a disk centered at the origin of radius $R \geq \sqrt{3}$.

EXPLORATION 2.75. Using *MinSurfTool* with the `W.E.(p,q)` tab, come up with three sets of functions p and q that create other minimal surfaces that are not embedded. Remember if the chosen p and q functions result in a singularity in the Weierstrass representation, then you will need to enter functions into the x, y and z boxes in the `Complex initial values for integration in radial direction` section. These functions come from explicitly solving the integrals for $x_1(z)$, $x_2(z)$, and $x_3(z)$ in the Weierstrass representation for the chosen $p(z)$ and $q(z)$ values and then substituting in $re^{i\theta}$ for z.

Try it out!

An important area in minimal surface theory is the study of complete (boundary-less) embedded minimal surfaces. The following theorem tells us that any minimal surface without boundary cannot be closed and bounded.

THEOREM 2.76. If M is a complete minimal surface in $\mathbb{R}^3$, then M is not compact.

PROOF. By Theorem 2.45, we can assume that M has an isothermal parametrization. Now, if M were compact, then each coordinate function would attain a maximum. Since the real part of an analytic function is harmonic we see from Theorem 2.65, the coordinate functions in this parametrization are harmonic. But harmonic functions attain their maximum on the boundary of the set. So, M must have a boundary which contradicts M being complete. □

2.5. The Gauss map, G, and height differential, dh

We can use other representations for $\phi = (\varphi_1, \varphi_2, \varphi_3)$ to form different Weierstrass representations as long as $\phi^2 = 0$ and $|\phi|^2 \neq 0$ (see the Summary on page 146). An important representation employs the functions known as the Gauss map, G, and the height differential, dh. This representation is useful, because the functions G and dh describe the geometry of the minimal surface. To develop this representation, we first need some background about the Gauss map.

Recall that the curvature of a unit speed curve, α, at a point s is $|\alpha''(s)|$. That is, the curvature of a curve is described by the rate of change of the tangent vector. Similarly, the curvature of a surface is related to the rate of change of the tangent plane. Since each tangent plane is essentially determined by its unit normal vector, $\mathbf{n}$, we can investigate the curvature of a surface by studying the variation of the unit normal vector. This is the idea behind the Gauss map.

DEFINITION 2.77. Let $M : \Omega \to \mathbb{R}^3$ be a surface with a chosen orientation (that is, a differentiable field of unit normal vectors $\mathbf{n}$). The *Gauss map*, $\mathbf{n}_p$, translates the unit normal on M at a point p to the unit vector at the origin pointing in the same direction as the unit normal and thus corresponds to a point on the unit sphere S^2.
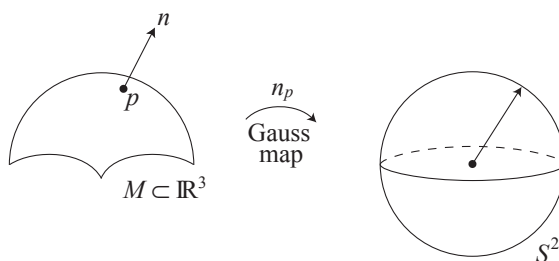

FIGURE 2.34. The Gauss map.

EXAMPLE 2.78. Let's determine the image of the Gauss map for the catenoid. A meridian is a curve formed by a vertical slice on the surface (see Exercise 2.15). Consider a meridian on the entire catenoid (remember that the image in Figure 2.35 is just part of a catenoid and that it actually extends on forever). The Gauss map, $\mathbf{n_p}$, of this meridian will be a meridian on S^2 from the north pole, $(0, 0, 1)$, to the south pole, $(0, 0, -1)$, that excludes these end points. Note that $(0, 0, 1)$ and $(0, 0, -1)$ are excluded, because no matter how far the catenoid extends, the unit normal $\mathbf{n}$ never points exactly straight up or exactly straight down. Now, since the catenoid is a surface of revolution if we revolve this meridian on S^2, we get that the image of the Gauss map for the catenoid is $S^2 \setminus \{(0, 0, 1), (0, 0, -1)\}$.
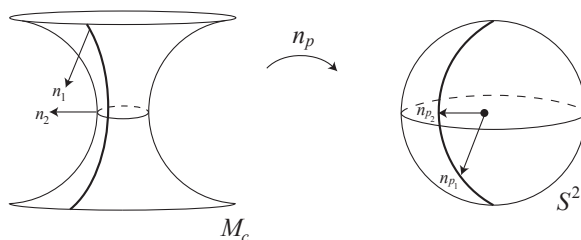

FIGURE 2.35. Image of a meridian on a catenoid under the Gauss map.

EXERCISE 2.79. Describe the image of the Gauss map for the following surfaces:
   (a) a right circular cylinder;
   (b) a torus;
   (c) Ennepers surface defined just on $\mathbb{D}$;
   (d) helicoid;
   (e) Scherk's doubly periodic surface.
   Try it out!

THEOREM 2.80 ([**22**]). Let M be a minimal surface with an isothermal parametrization. Then the Gauss map of M preserves angles.

While the Gauss map preserves angles, it reverses orientation. Such maps are known as *anticonformal*. To help visualize the fact that the Gauss maps reverses orientation, consider three points A, B, and C on a curved path near the neck of the catenoid (see Figure 2.36). Since A is above the neck of the catenoid, the outward pointing unit normal at A will be pointing downward and hence the Gauss map will put it below the equator on S^2 at the point A'. The point B is on the neck of the catenoid and so the outward pointing unit normal at B will be horizontal. So, the Gauss map will put it on the equator of S^2 at the point B'. Similarly, the normal at C will get mapped to C'. Thus, following the curve path from A to B to C in the positive direction on the catenoid gets sent by the Gauss map to a curve from A' to B' to C' in the negative direction on S^2. That is, we have an orientation-reversing map.
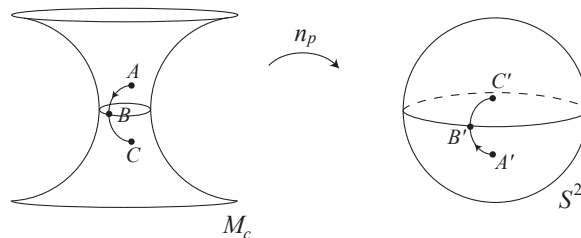
FIGURE 2.36. $\mathbf{n_p}$ is orientation reversing.

Since the Gauss map associates a point on M with a point on S^2, we can also associate it with a point in the complex plane $\mathbb{C}$ by using stereographic projection. Recall that stereographic projection, σ, takes a point on S^2 to a point in the extended complex plane, $\mathbb{C} \cup \infty$. To do this, we place the complex plane through the equator of the sphere and take a line connecting the north pole, $(0, 0, 1) \in S^2$, with the given point $(x_1, x_2, x_3) \in S^2$. This line will intersect the extended complex plane at some point, $z = x + iy$. In such a setting the unit sphere is known as the *Riemann sphere*.
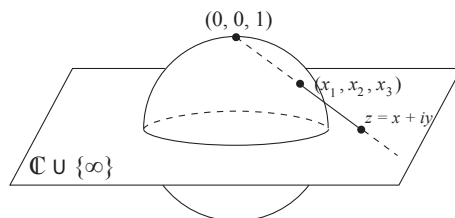

FIGURE 2.37. Stereographic projection.

EXERCISE 2.81. Describe the projections of the following sets on the Riemann sphere onto the extended complex plane:
  (a) meridians;
  (b) parallels;
  (c) circles;
  (d) circles that contain (i.e., touch) the point $(0, 0, 1)$;
  (e) antipodal points (i.e., diametrically opposite points).
  Try it out!

Finally, let $\bar{\sigma}$ be the projection of $(x_1, x_2, x_3) \in S^2$ to the point $x - iy \in \mathbb{C}$ given first by stereographic projection of (x_1, x_2, x_3) to $z = x + iy$ followed by reflection of $z = x + iy$ across the real axis to $\bar{z} = x - iy$. Note that $\bar{\sigma}$ is anticonformal.

Now, let $G : D \subset \mathbb{C} \to \mathbb{C}$ be the map defined by $G = \bar{\sigma} \circ \mathbf{n} \circ \mathbf{x}$. Note that G preserves angles since $\bar{\sigma}$, $\mathbf{n}$, and $\mathbf{x}$ preserves angles and so the composition preserves angles. Also,

G is orientation preserving, because both $\bar{\sigma}$ and $\mathbf{n}$ are orientation reversing and so their composition is orientation preserving. Thus, G is a meromorphic function. It is also called the Gauss map. The map G is of more interest to us, so for the remainder of the chapter "the Gauss map" will refer to this map.
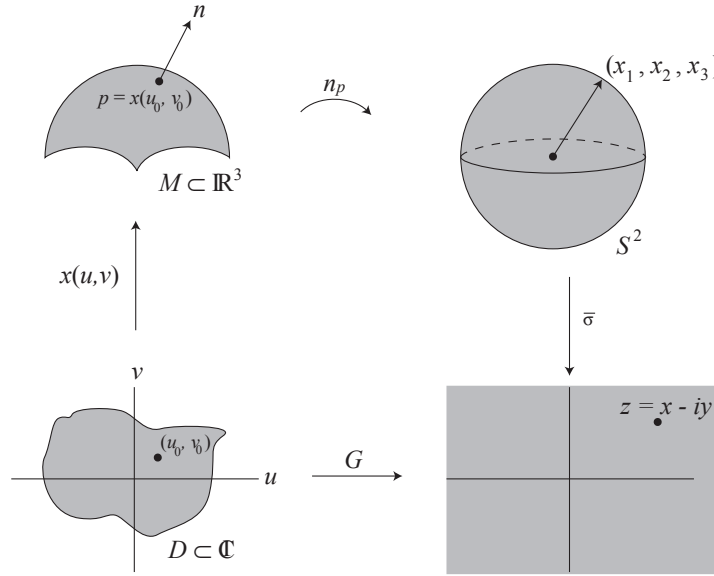

FIGURE 2.38. The map G.

EXAMPLE 2.82. Using the geometry of the Enneper surface, M_E, we can determine specific values of a Gauss map, G, on M_E even though we do not know what function G is. What is $G(0)$? Enneper's surface is formed by bending a disk into a saddle surface. The point $0 \in \mathbb{C}$ should get mapped to the point in the center of the Enneper surface. For simplicity sake, we will take the downward pointing normal $\mathbf{n}$. Hence, the unit normal at the center of M_E points straight down, and thus mapping it to S^2 under $\mathbf{n_p}$ gives the vector pointing at $(0, 0, -1)$ (see Figure 2.40). Taking the stereographic projection, σ, results in the point $z = 0 \in \mathbb{C}$ and reflecting this across the real axis does not change 0, so $\bar{\sigma}(0, 0, -1) = 0$. Hence, $G(0) = 0$.

Next, what is $G(r)$ when $r \in [0, 1]$? The points r get mapped under $\mathbf{x}$ to a curve moving upward along one of the upward pointing leaves of the Enneper surface. The corresponding downward pointing unit normal, n_r, stays in the $x_1 x_3$-half plane (where $x_1 \geq 0$) also moving upward (i.e., the x_3 value is increasing). As r approaches 1, n_r approaches being parallel to the x_1 axis (see Figure 2.41). Thus, mapping these unit normals to S^2, the curve $\{r \in \mathbb{D} : 0 \leq r < 1\}$ traces a meridian on S^2 from $(0, 0, -1)$ to $(1, 0, 0)$. The stereographic projection of this onto the complex plane gives $\{r \in \mathbb{D} : 0 \leq r < 1\}$ and reflecting this across the real axis does not change the values. Hence, $G(r) = r$, where $0 \leq r \leq 1$.
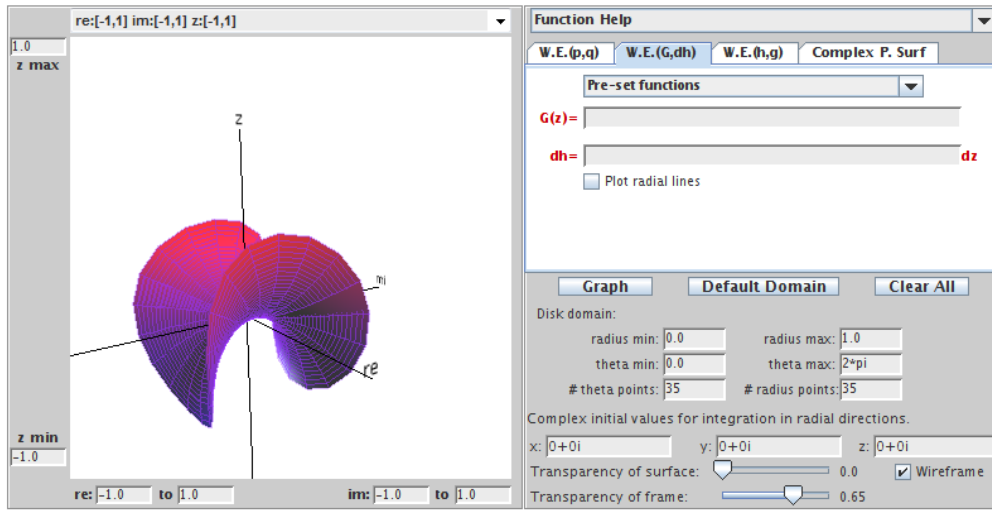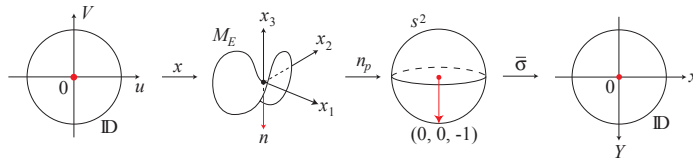
FIGURE 2.39. The Enneper surface.

FIGURE 2.40. $G(0) = 0$ for the Enneper surface.
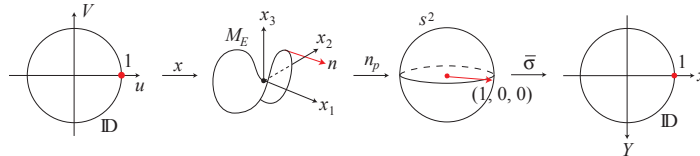
FIGURE 2.41. $G(1) = 1$ for the Enneper surface.

Finally, what is $G(e^{i\theta})$ for $0 \leq \theta \leq \frac{\pi}{2}$? If we restrict the domain of the Enneper surface to $\mathbb{D}$, these points get mapped to the edge of our image of the Enneper surface in a positive direction. At $\theta = 0$, the unit normal is pointing outward (i.e., away from the opposite leaf) and under the Gauss map, n_p, this corresponds to $(1, 0, 0) \in S^2$. As θ moves from 0 to $\frac{\pi}{2}$, the unit normal moves from pointing outward to pointing inward (i.e., toward the opposite leaf). So that at $\theta = \frac{\pi}{2}$, the unit normal is mapped under n_p to $(0, -1, 0) \in S^2$ which projects under σ to $-i \in \mathbb{C}$ (see Figure 2.42). Reflecting this across the real axis gives $\overline{\sigma}(0, -1, 0) = \overline{-i} = i$. A similar argument shows that the same thing happens for all $\theta \in [0, \frac{\pi}{2}]$. That is, $G(e^{i\theta}) = e^{i\theta}, (0 \leq \theta \leq \frac{\pi}{2})$.

156

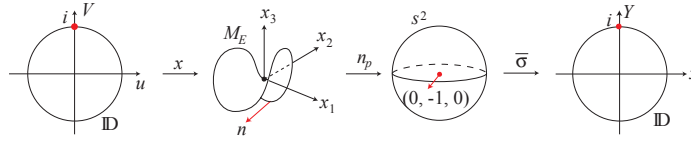FIGURE 2.42. $G(i) = i$ for the Enneper surface.

EXAMPLE 2.83. Let's determine some specific values for the Gauss map G for the singly periodic Scherk surface with six leaves, M_S. The domain used here is $\mathbb{D}$. Also, the leaves are centered at rays from the origin through each of the 6th roots of unity (i.e., $e^{i\pi k/3}, (k = 0, ..., 5)$) with the leaf centered at the positive real axis pointing upward and the subsequent leaves alternating between downward pointing and upward pointing (see Figure 2.43).
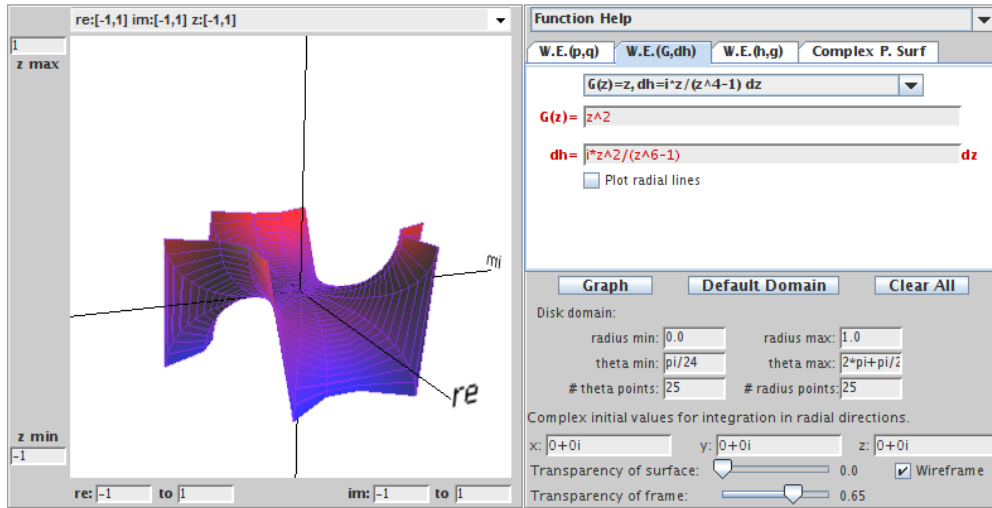
FIGURE 2.43. The singly periodic Scherk surface with six leaves.

As in the previous example we will use the downward pointing unit normal, and so $G(0) = 0$. Next, because the leaves are centered at the 6th roots of unity (i.e., $e^{i\pi k/3}, (k = 0, ..., 5)$), let's look at $G(e^{i\pi k/3})$, where $(k = 0, ..., 5)$. First, the point 1 gets mapped under $\mathbf{x}$ to the "edge" of M_S above the positive real axis. By looking at the graph of M_S, we see that this is in the middle of an upward pointing leaf, and the corresponding unit normal $\mathbf{n}$ lies above the positive real axis and pointing away from the origin. Also, it lies in a plane parallel to the horizontal x_1x_2-plane. Mapping this normal under $\mathbf{n}_p$ and then $\overline{\sigma}$, results in the point 1. Hence, $G(1) = 1$. Now, consider $G(e^{i\pi/3})$. The point $e^{i\pi/3}$ gets mapped to the "edge" of M_S above the line $re^{i\pi/3}, r > 0$. Since the leaves alternate between pointing upward and pointing downward, this is in the middle of a downward pointing leaf and points toward the

157

origin. Mapping this normal under $\mathbf{n}_p$ and then $\overline{\sigma}$, results in the point $\overline{e^{i4\pi/3}} = e^{i2\pi/3}$. Hence, $G(e^{i\pi/3}) = e^{i2\pi/3}$.

In a similar way, we get the following values:

$$G(1) = 1, \qquad G\left(e^{i\frac{\pi}{3}}\right) = e^{i\frac{2\pi}{3}}, \qquad\qquad G\left(e^{i\frac{2\pi}{3}}\right) = e^{i\frac{4\pi}{3}},$$

$$G(-1) = 1, \qquad G\left(e^{i\frac{4\pi}{3}}\right) = e^{i\frac{2\pi}{3}} = e^{i\frac{8\pi}{3}}, \qquad G\left(e^{i\frac{5\pi}{3}}\right) = e^{i\frac{4\pi}{3}} = e^{i\frac{10\pi}{3}}.$$

EXERCISE 2.84. A picture of the half catenoid on its side defined on $\mathbb{D}$ is shown in Figure 2.44 with the positive real axis on the right.
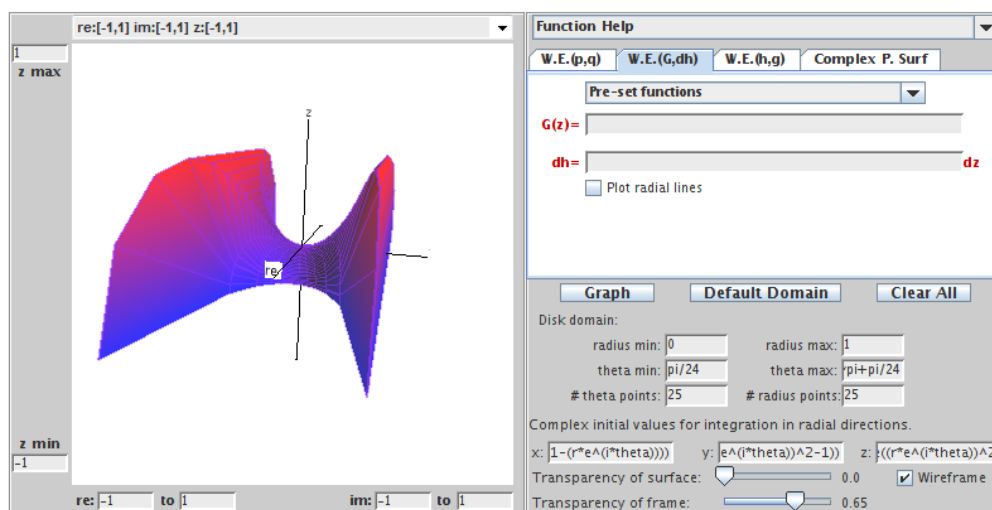
FIGURE 2.44. A view of the half catenoid on its side.

For this catenoid on its side determine:

(a) $G(0)$;　　(b) $G(1)$;　　(c) $G(-1)$;　　(d) $G(i)$;　　(e) $G(-i)$.

Try it out!

EXERCISE 2.85. For the 4-noid (see Figure 2.33), determine:

(a) $G(0)$;　　(b) $G(1)$;　　(c) $G(-1)$;　　(d) $G(i)$;　　(e) $G(-i)$.

Try it out!

We will use the Gauss map, G, to form another Weierstrass representation of a parametrized minimal surface. In doing so, we will also need the height differential,

dh, which is called such because it is locally (though not globally) the differential of the height coordinate. We will not get into the definition of differential forms; if you are interested in learning about differential forms, check out [**26**]. However, it is worth mentioning that at points where the Gauss map is vertical (i.e., $G = 0$ or $G = \infty$), the height function ought to have local minimums and maximums. Hence, dh ought to have a zero at these points (for example, see Figure 2.40).

THEOREM 2.86 (Weierstrass Representation (G,dh)). Every regular minimal surface has a local isothermal parametric representation of the form

$$(14) \qquad \mathbf{x} = \mathrm{Re} \int_a^z \left(\frac{1}{2}\left(\frac{1}{G} - G\right), \frac{i}{2}\left(\frac{1}{G} + G\right), 1 \right) dh,$$

where G is the Gauss map, dh is the height differential, and $a \in \Omega$ is a constant.

PROOF. From the Summary on page 146, we need

$$\phi^2 = 0 \quad \text{and} \quad |\phi|^2 \neq 0 \text{ and be finite.}$$

Comparing eq (14) and eq (13), we have that

$$\varphi_1 = \frac{1}{2}\left(\frac{1}{G} - G\right) dh, \quad \varphi_2 = \frac{i}{2}\left(\frac{1}{G} + G\right) dh, \quad \varphi_3 = dh.$$

In Exercise 2.87 you will show that

$$\phi^2 = 0 \quad \text{and} \quad |\phi|^2 \neq 0.$$

$\square$

EXERCISE 2.87. Prove that $\phi^2 = 0$ and $|\phi|^2 \neq 0$ in the proof of Theorem 2.86. ***Try it out!***

Note that

$$G = \frac{\varphi_1 + i\varphi_2}{-\varphi_3} \quad \text{and} \quad dh = \varphi_3.$$

One advantage of using this Weierstrass representation with the Gauss map and height differential is that the complex analytic properties of G and dh are related to the geometry of a minimal surface. We will discuss this in a bit, but first we will look at some examples. The following is a list of the Weierstrass data for some common minimal surfaces.

(a) The Enneper surface: $G(z) = z$ $dh = z\, dz$ on $\mathbb{C}$.
(b) The catenoid: $G(z) = z$ $dh = \frac{1}{z}\, dz$ on $\mathbb{C} \setminus \{0\}$.
(c) The helicoid: $G(z) = z$ $dh = \frac{i}{z}\, dz$ on $\mathbb{C} \setminus \{0\}$.
(d) Scherk's doubly periodic surface: $G(z) = z$ $dh = \frac{z}{z^4 - 1}\, dz$ on $\mathbb{D}$.
(e) Scherk's singly periodic surface: $G(z) = z$ $dh = \frac{iz}{z^4 - 1}\, dz$ on $\mathbb{D}$.
(f) Polynomial Enneper: $G(z) = p(z)$ $dh = p(z)\, dz$ on $\mathbb{C}$.
(g) Wavy plane: $G(z) = z$ $dh = dz$ on $\mathbb{C} \setminus \{0\}$.

EXAMPLE 2.88. For $G(z) = z^k$ and $dh = z^k\, dz$, where $k = 1, 2, \ldots$, we get

$$\mathbf{x} = \operatorname{Re} \int_0^z \left(\frac{1}{2}\left(\frac{1}{z^k} - z^k\right), \frac{i}{2}\left(\frac{1}{z^k} + z^k\right), 1\right) z^k\, dh$$

$$= \left(\operatorname{Re} \frac{1}{2}\left\{ z - \frac{1}{2k+1} z^{2k+1} \right\}, \operatorname{Re} \frac{1}{2}\left\{ -i\left(z + \frac{1}{2k+1} z^{2k+1}\right)\right\}, \operatorname{Re}\left\{ \frac{z^{k+1}}{k+1}\right\}\right).$$

This is the Enneper surface with $2k + 2$ leaves (see Exploration 2.71).

EXERCISE 2.89. You may have noticed that the Weierstrass data for the catenoid and the helicoid, which are conjugate surfaces (see Definition 2.50), have the same Gauss map, G, while the height diffferentials dh differ by a multiple of i. Prove that this is true for any conjugate surfaces.
Try it out!

EXERCISE 2.90. Let $G(z) = z^4$ and $dh = z^2\, dz$.
   (a) Using eq (14), compute the parametrization.
   (b) This minimal surface has a planar end (i.e., looks like a plane) and an Enneper end. To graph this surface, use the corresponding entry under the **Pre-set functions** in the **W.E. (G,dh)** tab of *MinSurfTool*.
Try it out!

The catenoid and the surface in Exercise 2.90 are examples of minimal surfaces with ends. Loosely, an *end* of a minimal surface is a piece that "goes on forever," or, more precisely, leaves all compact subsets of the minimal surface. Recall from Theorem 2.76 that all complete minimal surfaces in $\mathbb{R}^3$ are not compact, and hence they must possess at least one end.

EXERCISE 2.91. Determine the number of ends each of the following surfaces have:
(a) the catenoid;   (b) the plane;   (c) the helicoid;   (d) Enneper's surface.
Try it out!

Ends occur in a deleted neighborhood (i.e., a disk with the centered removed) centered at a singularity. Three common types of ends for minimal surfaces are: (1) Enneper ends; (2) catenoid ends; and (3) flat or planar ends. In discussing ends, we will need to represent ds, the metric (i.e., a way to measure distance) on a minimal surface, in terms of G and dh. Using $ds^2 = |\phi|^2$ and eq (14), we derive

(15) $$ds = \frac{1}{\sqrt{2}}\left(|G| + \frac{1}{|G|}\right)|dh|.$$

An Enneper end has $ds \sim |z^k| \cdot |dz|$, so the metric grows in the same manner as a polynomial as we approach the end. On the other hand, a catenoidal end and a planar end have $ds \sim |dz|$, so the metric becomes Euclidean. A catenoidal end differs from a planar end in that the residue of dh is logarithmic.

EXERCISE 2.92. Prove eq. (15).
Try it out!

EXAMPLE 2.93. We know that the catenoid has two catenoid ends, but let's show how we could prove this if we did not know what type of ends these are. Using the Weierstrass data, $G(z) = z$ and $dh = \frac{1}{z} \, dz$, for the catenoid, we have the parametrization

$$\mathbf{x}(z) = \left(\frac{1}{2} \operatorname{Re} \left(-\frac{1}{z} - z \right), \frac{i}{2} \operatorname{Re} \left(-\frac{1}{z} + z \right), \operatorname{Re} \left(\log z \right) \right).$$

So there is a singularity or pole of order 1 at 0. Also, there is a pole of order 1 at ∞. To see that there is a singularity at ∞, we replace z with $\frac{1}{w}$ and look at the limit as w goes to 0. Thus, the catenoid will have two ends (one at 0 and one at ∞). To determine what types of ends these are, we look at ds at these points. Note that

$$ds = \frac{1}{\sqrt{2}} \left(|z| + \frac{1}{|z|} \right) \frac{1}{|z|} |dz|.$$

As $z \to \infty$, $ds \sim |dz|$ and because x_3 is logarithmic, we have a catenoid end.

At $z = 0$, plugging in 0 does not work, so instead we let $w = \frac{1}{z}$ and consider $w \to \infty$. Note that in this case,

$$dh = \frac{1}{z} \, dz = w \left(\frac{1}{w} \right)' = w \left(-\frac{1}{w^2} \, dw \right) = -\frac{1}{w} \, dw.$$

Therefore,

$$ds = \frac{1}{\sqrt{2}} \left(|w| + \frac{1}{|w|} \right) \frac{1}{|w|} |dw|.$$

As $w \to \infty$, $ds \sim |dw|$ and again because x_3 is logarithmic, we have a catenoid end.

EXAMPLE 2.94. The surface in Exercise 2.90 has $G(z) = z^4$ and $dh = z^2 \, dz$, and the corresponding parametrization is

$$\mathbf{x}(z) = \left(\frac{1}{2} \operatorname{Re} \left(\frac{1}{7} z^7 - \frac{1}{z} \right), \frac{i}{2} \operatorname{Re} \left(\frac{1}{7} z^7 + \frac{1}{z} \right), \operatorname{Re} \left(\frac{1}{3} z^3 \right) \right).$$

Note there are singularities at 0 and at ∞ and

$$ds = \frac{1}{\sqrt{2}} \left(|z|^4 + \frac{1}{|z|^4} \right) |z|^2 |dz|.$$

As $z \to \infty$, $ds \sim |z|^6 \, |dz|$ and so we have an Enneper end.

At $z = 0$, we again let $w = \frac{1}{z}$ and consider $w \to \infty$. Then $G(w) = w^4$ and $dh = -\frac{1}{w^4} \, dw$. Hence,

$$ds = \left(|w|^4 + \frac{1}{|w|^4} \right) \frac{1}{|w|^4} |dw|.$$

As $w \to \infty$, $ds \sim |dw|$, but because there is no logarithmic term, we have a planar end.
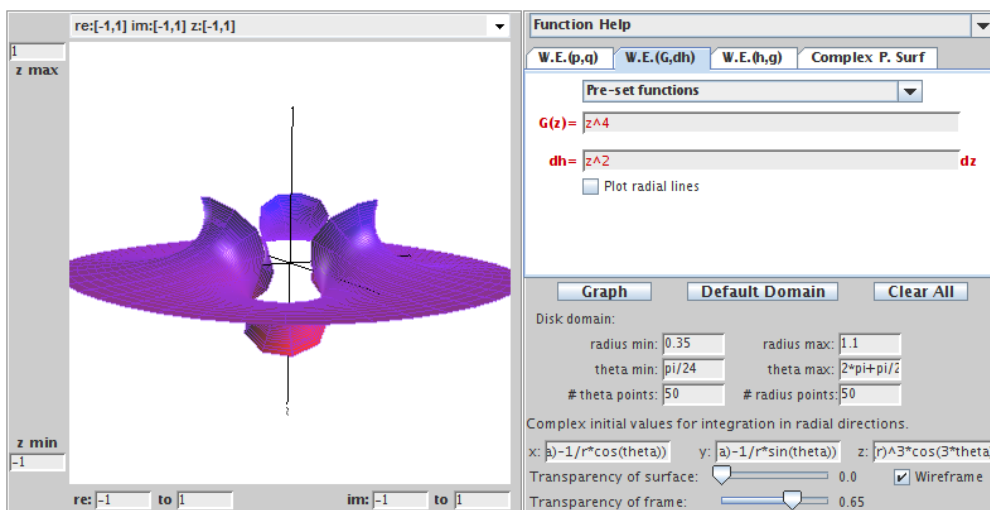
FIGURE 2.45. A minimal surface with Enneper and planar ends.

EXERCISE 2.95. Let $G(z) = \frac{z^2+3}{z^2-1}$ and $dh = \frac{z^2+3}{z^2-1} \, dz$. Show that this minimal surface has one planar end and two catenoid ends.
Try it out!

The Gauss map and height differential also tell us about two important types of curves on a minimal surface. These are known as the asymptotic lines and the curvature lines. To understand what these lines are, we will review the terms normal curvature and principal directions, both of which were discussed in Section 2. Let p be a point on a curve on a minimal surface M. The tangent vector $\mathbf{w}$ and normal vector $\mathbf{n}$ at p form a plane that intersects the surface in another curve, say α (see Figure 2.18). The normal curvature in the direction $\mathbf{w}$ is $\alpha'' \cdot \mathbf{n}$ and measures how much the surface bends toward $\mathbf{n}$ as you move in the direction of $\mathbf{w}$ at point p. An *asymptotic line* is a curve that is tangent to a direction in which the normal curvature is zero.

As we rotate the plane through the normal $\mathbf{n}$, we will get a set of curves on the surface each of which has a value for its curvature. The directions in which the normal curvature attains its absolute maximum and absolute minimum values are known as the principal directions. *Curvature lines* are curves that are always tangent to a principal direction.

A nice relationship between these lines and the Weierstrass data is:

$$\text{A curve } z(t) \text{ is an asymptotic line} \iff \frac{dG}{G}(z) \cdot dh(z) \in i\mathbb{R}.$$

$$\text{A curve } z(t) \text{ is a curvature line} \iff \frac{dG}{G}(z) \cdot dh(z) \in \mathbb{R}.$$

162

EXAMPLE 2.96. Let $G(z) = z$ and $dh(z) = dz$. This is a parametrization of the wavy plane. Computing the Weierstrass representation, we get the parametrization:

$$(x_1(z), x_2(z), x_3(z)) = \left(\operatorname{Re}\left\{ \frac{1}{2}\log(z) - \frac{1}{4}z^2 \right\}, \operatorname{Re}\left\{ \frac{i}{2}\log(z) + \frac{i}{4}z^2 \right\}, \operatorname{Re}\left\{ z \right\} \right).$$

We can plot an image of this surface using the corresponding function under **Pre-set functions** in the **W.E.(G,dh)** tab in *MinSurfTool*.
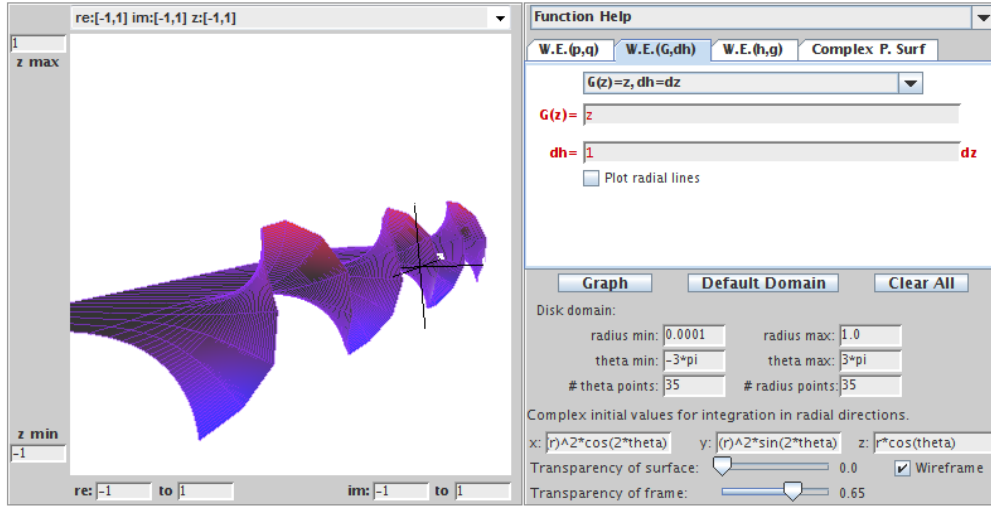
FIGURE 2.46. The wavy plane.

Now, for the wavy plane

$$\frac{dG}{G}(z) \cdot dh(z) = \frac{dz}{z} \cdot dz.$$

If we let $z = e^{i\theta}$ (since we let radius max=1), then $dz = ie^{i\theta}d\theta$ and

$$\left. \frac{dG}{G}(z) \cdot dh(z) \right|_{z=e^{i\theta}} = -e^{i\theta}(d\theta)^2.$$

So, from the equations above, we get that for $k \in \mathbb{Z}$: (1) the asymptotic lines occur when $\theta = \frac{\pi}{2} + k\pi$; and (2) the curvature lines occur when $\theta = k\pi$.

If we use *MinSurfTool* to plot the wavy plane with theta min $= -\frac{\pi}{2}$ and theta max $= \frac{\pi}{2}$, we see that these asymptotic lines lie in the x_1x_2-plane. Similarly, if we plot theta min $= 0$ and theta max $= \pi$, we see that these curvature lines are reflection lines through which the wavy plane can be reflected as if through a mirror to get a smooth continuation of the minimal surface.

EXERCISE 2.97. Using $G(z) = z$ and $dh = zdz$ for Enneper's surface, show that for $z = e^{i\theta}$ the asymptotic lines occur when $\theta = \frac{\pi}{4} + \frac{k\pi}{2}$, where $k \in \mathbb{R}$, and the

curvature lines occur when $\theta = \frac{k\pi}{2}$, where $k \in \mathbb{Z}$. Use *MinSurfTool* to plot these lines on Enneper's surface.

   Try it out!

   EXERCISE 2.98. Prove that for conjugate surfaces, the asymptotic lines (and curvature lines) of one surface are the curvature lines (and asymptotic lines) of the other surface.

   Try it out!

   As mentioned earlier, an advantage of using the Gauss map and height differential is that the complex analytic properties of G and dh are related to the geometry of a minimal surface. So, let's look at the complex analytic properties of G and dh. First, recall from the Summary on page 146, we need $|\phi|^2$ to be finite and nonzero. Because of eq (15), this leads to the following condition.

   PROPOSITION 2.99. At a nonsingular point, G has a zero or pole of order n if and only if dh has a zero of order n.

   Second, note that the integrals in the Weierstrass representation in eq (14) might depend upon the path of integration if the domain of G and dh is not simply connected. If the representation does not depend on the path of integration, then all closed paths γ in the domain satisfy the three conditions below:

$$\text{Re}\left(\frac{1}{2} \int_\gamma \left(\frac{1}{G} - G \right) dh \right) = 0;$$

$$\text{Re}\left(\frac{i}{2} \int_\gamma \left(\frac{1}{G} + G \right) dh \right) = 0;$$

$$\text{Re} \int_\gamma dh = 0.$$

If a curve has the property that one of these quantities is a nonzero constant, then the surface will have periodic behavior. Example 2.101 illustrates an example of this behavior.

   The three equations above can be reduced to the following two period conditions:

(16)
$$(i) \ \int_\gamma G \, dh = \overline{\int_\gamma \frac{1}{G} \, dh} \quad \text{(horizontal period condition)},$$

$$(ii) \ \text{Re} \int_\gamma dh = 0 \quad \text{(vertical period condition)}.$$

for all closed paths γ in the domain.

   EXERCISE 2.100. Show that the conditions

$$\text{Re}\left(\frac{1}{2} \int_\gamma \left(\frac{1}{G} - G \right) dh \right) = 0, \ \text{and} \ \text{Re}\left(\frac{i}{2} \int_\gamma \left(\frac{1}{G} + G \right) dh \right) = 0$$

are equivalent to the condition

$$\int_\gamma G \, dh = \overline{\int_\gamma \frac{1}{G} \, dh}.$$

Try it out!

These period conditions are useful in determining horizontal periods (e.g., Scherk's doubly periodic surface), vertical periods (e.g., Scherk's singly periodic surface) and in determining possible constant values in G and dh.

EXAMPLE 2.101. Consider Scherk's doubly periodic surface with the Weierstrass data $G(z) = z$ and $dh(z) = \frac{z}{z^4-1} dz$. The horizontal period condition is $\int_\gamma G \, dh = \overline{\int_\gamma \frac{1}{G} \, dh}$, for all closed paths γ in the domain. Note that both integrands are meromorphic with poles of order 1 at $\pm 1, \pm i$. So, the only paths that concern us are ones that enclose one, two, or three of these poles. A nice way to calculate these integrals along such paths is to use the Residue Theorem that states if γ is a simple closed positively-oriented contour and f is analytic inside and on γ except at the points $z_1, \ldots, z_n$ inside γ, then

$$\int_\gamma f(z) \, dz = 2\pi i \sum_{j=1}^n \mathrm{Res}(f, z_j).$$

Recall that for poles of order 1,

$$\mathrm{Res}(f, z_j) = \lim_{z \to z_j} (z - z_j) f(z).$$

Thus, for $\displaystyle\int_\gamma G \, dh$, we have

$$\mathrm{Res}(G \, dh, z_j) = \lim_{z \to z_j} \frac{z^3 - z_j z^2}{z^4 - 1} = \lim_{z \to z_j} \frac{3z^2 - 2z_j z}{4z^3} = \lim_{z \to z_j} \frac{3z^3 - 2z_j z^2}{4z^4} = \frac{z_j^3}{4}.$$

In particular,

$$\mathrm{Res}(G \, dh, 1) = \frac{1}{4} \qquad\qquad \mathrm{Res}(G \, dh, i) = \frac{-i}{4}$$

$$\mathrm{Res}(G \, dh, -1) = \frac{-1}{4} \qquad\qquad \mathrm{Res}(G \, dh, -i) = \frac{i}{4}.$$

Similarly, we can compute that

$$\mathrm{Res}\left(\frac{1}{G} \, dh, 1\right) = \frac{1}{4} \qquad\qquad \mathrm{Res}\left(\frac{1}{G} \, dh, i\right) = \frac{i}{4}$$

$$\mathrm{Res}\left(\frac{1}{G} \, dh, -1\right) = \frac{-1}{4} \qquad\qquad \mathrm{Res}\left(\frac{1}{G} \, dh, -i\right) = \frac{-i}{4}.$$

Now, if the path γ_1 just contains the pole at $z_1 = 1$, then the horizontal period conditions result in

$$\int_{\gamma_1} G \, dh = 2\pi i \operatorname{Res}(G \, dh, 1) = \frac{i\pi}{2}, \qquad \overline{\int_{\gamma_1} \frac{1}{G} \, dh} = \overline{2\pi i \operatorname{Res}\left(\frac{1}{G} \, dh, 1\right)} = \frac{-i\pi}{2}.$$

These integrals should be equal, which occurs if the minimal surface is periodic in the imaginary direction with period of π. Likewise, if we take a path γ_2 that just contains the pole $z_2 = i$, then we get

$$\int_{\gamma_2} G \, dh = \frac{\pi}{2}, \qquad \overline{\int_{\gamma_2} \frac{1}{G} \, dh} = \frac{-\pi}{2},$$

and the minimal surface is periodic in the real direction with period π. All other paths γ are covered by these two cases. Finally, if we look at the vertical period condition, we get that the condition is automatically true for all paths γ and so the minimal surface is not periodic in the vertical direction. This matches the image of Scherk's doubly periodic surface from *MinSurfTool*.

EXERCISE 2.102. Show that the period conditions given in eq. (16) result in Scherk's singly periodic surface being periodic in the vertical direction.
    Try it out!

Let's look at example of how all of this can help us use the geometry of a minimal surface to determine G and dh.

EXAMPLE 2.103. From the list of Weierstrass data on page 159, we know that $G(z) = z$ and $dh = z \, dz$ for the Enneper surface. However, we want to show how this Weierstrass data can be determined by using the geometric shape of the surface. First, let's determine a plausible candidate for G. To do this, we will make a guess based on the value of G at a few specific points. From Example 2.82, we know that $G(0) = 0$, $G(r) = r$ for $0 \le r \le 1$, and $G(e^{i\theta}) = e^{i\theta}$ for $0 \le \theta \le \frac{\pi}{2}$. Therefore, it seems plausible to let $G(z) = z$. Second, given this G, let's determine dh. Because eq (15) must be finite and the Enneper surface has no ends in $\mathbb{C}$, dh cannot have any poles in $\mathbb{C}$. However, from the sentence before Exercise 2.91, we know that Enneper's surface must have at least one end. This end corresponds to the point at infinity, $z = \infty$, and so dh has a pole at ∞. Thus, $dh = \rho z^n \, dz$, for some $n \in \mathbb{N}$ and $\rho \in \mathbb{C}$. Since $G(z) = z$ has a zero of order 1 at 0, by Proposition 2.99 dh must also have a zero of order 1 at 0 and no other zeros. Thus, $dh = \rho z \, dz$. For simplicity sake, we let $\rho = 1$. Finally, notice that the period conditions in eq (16) hold, because there are no poles in $\mathbb{C}$, and so every integral along any closed path γ will equal 0. Hence, the Weierstrass data

$$G(z) = z, \quad dh = z \, dz$$

generates a minimal surface.

EXAMPLE 2.104. Consider the singly periodic Scherk surface with six leaves, M_S (see Figure 2.43). Note that these leaves go off to infinity. Hence, we will have 6 poles. Because of symmetry, we will choose these poles to be at the 6th roots of unity (i.e., $e^{i\pi k/3}, (k = 0, ..., 5)$). This means that dh will have the term $z^6 - 1$ in its denominator. However, we will need to determine G first in order to know what should be in the numerator of dh. From the results in Example 2.83, it seems reasonable that $G(z) = z^2$. Since $G(z) = z^2$ has a zero of order 2 at 0, by Proposition 2.99 dh must also have a zero of order 2 at 0 and no other zeros. Thus, we so far have

$$G(z) = z^2, \qquad dh = \rho \frac{z^2}{z^6 - 1} \, dz,$$

where $\rho \in \mathbb{C}$. To determine possible values of ρ, consider the period conditions in eq (16). There are poles of order 1 at $e^{ik\pi/3}, k = 0, \ldots, 5$. We compute that

$$\text{Res}(G \, dh, z_j) = \frac{\rho z_j^5}{6}, \qquad \text{Res}\left(\frac{1}{G} \, dh, z_j\right) = \frac{\rho z_j}{6}.$$

Hence, if γ contains the pole z_j, then the horizontal period condition requires

$$\int_\gamma G \, dh = 2\pi i \, \text{Res}(G \, dh, z_j) = \frac{\rho \pi i z_j^5}{3}, \text{ and}$$

$$\overline{\int_\gamma \frac{1}{G} \, dh} = \overline{2\pi i \, \text{Res}\left(\frac{1}{G} \, dh, z_j\right)} = \frac{-\bar{\rho} \pi i \bar{z}_j}{3}$$

to be equal (since there is no periodicity of M_S in the horizontal direction). These integrals will be equal for these poles $z_j^5 = \bar{z}_j$ if $\rho = -\bar{\rho}$. That is, ρ is purely imaginary. Without loss of generality, we let $\rho = i$ and we check that the vertical period condition holds. Hence, we have that the Weierstrass data for singly periodic Scherk surface with six leaves is

$$G(z) = z^2, \qquad dh = \frac{iz^2}{z^6 - 1} \, dz.$$

EXERCISE 2.105. Let M be the Enneper surface with 8 leaves. Using the approach of Example 2.103 determine G and dh for this surface.
    Try it out!

EXERCISE 2.106. Let M be the 3-noid with ends symmetrically placed so that if the surface is rotated by $\frac{2\pi}{3}$ you will get the same image. Determine G and dh for this surface.
    Try it out!

SMALL PROJECT 2.107. Let M be a minimal surface that has 6 symmetrically-placed catenoidal ends with 4 ends along the side (like a 4-noid), 1 end on the top, and 1 end on the bottom. So, M will look the same if it is rotated horizontally by $\frac{\pi}{2}$ and if it is rotated vertically by $\frac{\pi}{2}$. Determine G and dh for this surface.
    Optional

SMALL PROJECT 2.108. For Scherk's singly periodic surface the four ends are symmetrically placed so that if the surface is rotated by $\frac{\pi}{2}$ you will get the same image. This is because the denominator of dh is $z^4 - 1$ which has zeros that are equally spaced on the unit circle. It is possible to create a variation of Scherk's singly periodic surface that has four ends with rotational symmetry of π. That is, if the ends are labelled $E_1, \ldots, E_4$, then E_2 will be closer to E_1 than to E_3 (and likewise, E_4 will be closer to E_3 than to E_1) and if the surface is rotated by π you will get the same image. Determine G and dh for this surface.

Optional

LARGE PROJECT 2.109. Describe and classify the possible minimal surfaces where one of the coordinates of the parametrization is fixed while the other two coordinates vary. For example, if x_3 is fixed to a specific function, what are the possible coordinate functions for x_1 and x_2? Try to generalize this approach as much as possible.

Optional

LARGE PROJECT 2.110. Describe and classify the possible minimal surfaces with $G(z) = z^m$ and $dh = z^n\ dz$, for all $n, m \in \mathbb{N}$ (see Example 2.88 and Example 2.94). There are several distinct cases to consider. Determine how to separate m, n into these distinct cases remembering to discuss types of surface, types of ends, lines of symmetry, etc.

Optional

2.6. Minimal Surfaces and Harmonic Univalent Mappings

In the Summary on page 146 before the first Weierstrass representation, we learned that each coordinate function of the parametrization $\mathbf{x}$ of a minimal surface had the form $x_k = \operatorname{Re} \int \varphi_k dz$ with φ_k being analytic. Since the real part of an analytic function is a harmonic function, we see that each x_k is harmonic. Also from this Summary, we have that $(\varphi_1)^2 + (\varphi_2)^2 + (\varphi_3)^2 = 0$. This means that if we know the functions φ_1 and φ_2, we can determine the function φ_3. So, another way to get a Weierstrass representation for minimal surfaces is to use two harmonic functions x_1 and x_2. In other words, we can investigate minimal surfaces by studying harmonic mappings in the complex plane. Such mappings are known as planar harmonic mappings and have been studied independently of minimal surfaces.

In this section we will develop another Weierstrass representation. In this case we will use planar harmonic mappings instead of p and q as in Section 2.4 or G and dh as in Section 2.5. What benefit do we obtain from this new approach? First, we can use results about planar harmonic mappings to prove results about minimal surfaces. Since these two areas of mathematics developed independently, this approach can be fruitful. Second, in the study of planar harmonic mappings, much of the work deals with $1 - 1$ functions. That is, f will be $1 - 1$ in G means that if $f(z_1) = f(z_2)$, then $z_1 = z_2$. Geometrically, this means that the image, $f(G)$, will not overlap or intersect

itself. When such functions are lifted from the complex plane into $\mathbb{R}^3$, the resulting surfaces are minimal graphs and hence have no self-intersections. This can be beneficial is establishing the embeddedness of the resulting minimal surfaces. For example, the $1 - 1$ planar harmonic function given by

$$f(z) = h(z) + \overline{g(z)} = \operatorname{Re}\left[\frac{i}{2}\log\left(\frac{i+z}{i-z}\right)\right] + i\operatorname{Im}\left[\frac{1}{2}\log\left(\frac{1+z}{1-z}\right)\right]$$

maps the unit disk onto a square region. This square region is the projection (i.e., shadow) of Scherk's doubly periodic surface onto the plane. In other words, we can
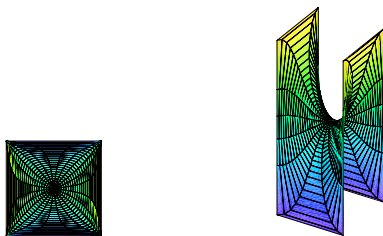

FIGURE 2.47. The image of $f(\mathbb{D})$ and Scherk's doubly periodic surface

lift f from $\mathbb{C}$ into $\mathbb{R}^3$ and to get Scherk's doubly periodic surface. Planar harmonic mappings that are $1 - 1$ are also known as harmonic univalent mappings. Harmonic univalent mappings can be studied on their own without bringing in minimal surfaces and such a study is the topic of Chapter 4 of this book.

EXERCISE 2.111.
  (a). Although all minimal graphs are embedded, the converse is not true. Give an example of an embedded minimal surface that is not a minimal graph.
  (b). Suppose you have a nonunivalent harmonic mapping. Why could it not be the projection of a minimal graph?
  Try it out!

Now that we have given an overview of this section, let's briefly discuss harmonic univalent mappings. A planar harmonic mapping is a function $f = u(x,y) + iv(x,y)$ where u and v are real harmonic functions. This concept is more general than that of an analytic function, because we do not require u and v to be harmonic conjugates. However, the following theorem allows us to relate a planar harmonic mapping to analytic functions. For our purposes, we will assume that the domain of f is the unit disk, $\mathbb{D}$.

THEOREM 2.112. Define a function $f = u + iv$, where u and v are real harmonic functions. If D is a simply-connected domain and $f : D \to \mathbb{C}$, then there exist analytic functions h and g such that $f = h + \overline{g}$.

EXERCISE 2.113.

(a) Show that $f(x, y) = u(x, y) + iv(x, y) = (x^3 - 3xy^2) + i(-3x^2y + y^3)$ is complex-valued harmonic by showing that u and v are real harmonic functions.

(b) Using $x = \frac{1}{2}(z + \overline{z})$ and $y = \frac{1}{2i}(z - \overline{z})$, rewrite $f(x, y) = (x^3 - 3xy^2) + i(-3x^2y + y^3)$ in terms of z and $\overline{z}$.

(c) Determine the analytic functions h and g such that $f = h + \overline{g}$.

Try it out!

EXAMPLE 2.114. In the previous exercise, we saw that the planar harmonic map $f : \mathbb{D} \to \mathbb{C}$ defined by

$$f(x, y) = u(x, y) + iv(x, y) = (x^3 - 3xy^2) + i(-3x^2y + y^3)$$

can be written as

$$f(z) = h(z) + \overline{g}(z) = z + \frac{1}{3}\overline{z}^3.$$

What is the image of $\mathbb{D}$ under f? It is a hypocycloid with 4 cusps. This fact can be computed by considering $f(e^{i\theta}) = u(\theta) + iv(\theta)$ and comparing the component functions, $u(\theta)$ and $v(\theta)$, to the parametrized equation for a hypocycloid with 4 cusps. To help us visualize the image, we can use the applet *ComplexTool*. To graph the image of $\mathbb{D}$ under the harmonic function $f(z) = z + \frac{1}{3}\overline{z}^3$, enter this function in *ComplexTool* in the form $z + 1/3\ \texttt{conj}\ (z \wedge 3)$ (see Figure 2.48). Remember this example; we will show that this function is related to a minimal graph.
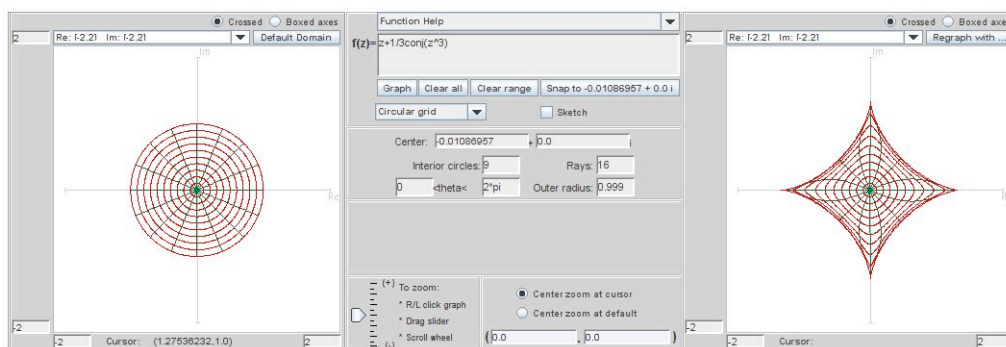
FIGURE 2.48. Image of $\mathbb{D}$ under the harmonic function $f(z) = z + \frac{1}{3}\overline{z}^3$

Note that the harmonic function $f(z) = h(z) + \overline{g}(z)$ can also be written in the form

(17) $$f(z) = \operatorname{Re}\left\{h(z) + g(z)\right\} + i\operatorname{Im}\left\{h(z) - g(z)\right\}.$$

This is because,

$$\operatorname{Re}\{h + g\} = \frac{1}{2}[(h + g) + \overline{(h + g)}] \quad \text{and} \quad \operatorname{Im}\{h - g\} = \frac{1}{2i}[(h - g) - \overline{(h - g)}].$$

Hence, in the previous example, $f(z) = z + \frac{1}{3}\bar{z}^3$ can also be written as $f(z) = \mathrm{Re}\left\{z + \frac{1}{3}z^3\right\} + i\,\mathrm{Im}\left\{z - \frac{1}{3}z^3\right\}$.

We are interested in harmonic functions that are $1-1$ or univalent, because this is one necessary condition in order to lift the harmonic mapping to a minimal graph. One theorem that establishes univalency requires the following background material.

DEFINITION 2.115. The *dilatation* of $f = h + \bar{g}$ is $\omega(z) = g'(z)/h'(z)$.

THEOREM 2.116 (Lewy [**11**]). $f = h + \bar{g}$ is locally univalent and orientation-preserving if and only if $|g'(z)/h'(z)| < 1$, for all $z \in \mathbb{D}$.

See Chapter 4 of this book for more background on the dilatation and on orientation-preserving.

EXERCISE 2.117. Show that if $z \in \mathbb{D}$, then $|\omega(z)| < 1$ for:

(a) $\omega_1(z) = e^{i\theta}z$, where $\theta \in \mathbb{R}$;

(b) $\omega_2(z) = z^n$, where $n = 1, 2, 3, \ldots$;

(c) $\omega_3(z) = \dfrac{z + a}{1 + \bar{a}z}$, where $|a| < 1$;

(d) $\omega_4(z)$ being the composition of any of the functions ω above.

Try it out!

Creating nontrivial examples of harmonic univalent mappings that lift to minimal graphs is not easy. However, one way to do this is to use the shearing technique of Clunie and Sheil-Small. Before we proceed, we need to discuss a certain type of domain.

DEFINITION 2.118. A domain Ω is *convex in the direction of the real axis* (or convex in the horizontal direction, CHD) if every line parallel to the real axis has a connected intersection with Ω.

CHD   not CHD

THEOREM 2.119 (Clunie and Sheil-Small). A harmonic function $f = h + \bar{g}$ locally univalent in $\mathbb{D}$ is a univalent mapping of $\mathbb{D}$ onto a CHD domain $\Longleftrightarrow h - g$ is an analytic univalent mapping of $\mathbb{D}$ onto a CHD domain.

REMARK 2.120. This technique is known as the "shear" method or "shearing" a function. In our situation, suppose $F = h - g$ is an analytic univalent function convex in the real direction. Then the corresponding harmonic shear is

$$f = h + \bar{g} = h - g + g + \bar{g} = h - g + 2\operatorname{Re}\{g\}.$$

So, the harmonic shear differs from the analytic function by adding a real function to it. Geometrically, you can think of this as taking F, the original analytic univalent function convex in the real direction, and cutting it up into thin horizontal slices which are then translated and/or scaled in a continuous way to form the corresponding harmonic function, f. This is why the method is called "shearing." Since F is univalent and convex in the real direction and we are only adding a continuous real function to it, the univalency is preserved.

EXAMPLE 2.121. Let

(18)
$$h(z) - g(z) = \frac{1}{2}\log\left(\frac{1+z}{1-z}\right)$$

which is an analytic function that maps $\mathbb{D}$ onto a horizontal strip convex in the direction of the real axis (see Figure 2.49).
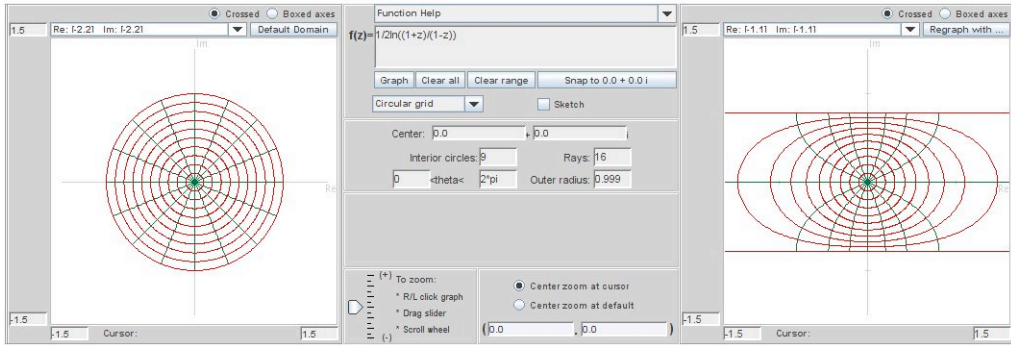
FIGURE 2.49. Image of $\mathbb{D}$ under the analytic function $\frac{1}{2}\log\left(\frac{1+z}{1-z}\right)$

Let

$$\omega(z) = g'(z)/h'(z) = -z^2.$$

Applying the shearing method from Theorem 2.119 with the substitution $g'(z) = -z^2 h'(z)$, we have

$$h'(z) - g'(z) = \frac{1}{1-z^2} \Rightarrow h'(z) + z^2 h'(z) = \frac{1}{1-z^2}$$

$$\Rightarrow h'(z) = \frac{1}{1-z^4} = \frac{1}{4}\left[\frac{1}{1+z} + \frac{1}{1-z} + \frac{1}{i+z} + \frac{1}{i-z}\right].$$

172

Integrating $h'(z)$ and normalizing so that $h(0) = 0$, yields

(19)
$$h(z) = \frac{1}{4} \log \left(\frac{1+z}{1-z} \right) + \frac{i}{4} \log \left(\frac{i+z}{i-z} \right).$$

We can use this same method to solve for normalized $g(z)$, where $g(0) = 0$. Note that we can also find $g(z)$ by using eqs. (56) and (57). Either way, we get

$$g(z) = -\frac{1}{4} \log \left(\frac{1+z}{1-z} \right) + \frac{i}{4} \log \left(\frac{i+z}{i-z} \right).$$

So

$$f(z) = h(z) + \overline{g(z)} = \mathrm{Re} \left[\frac{i}{2} \log \left(\frac{i+z}{i-z} \right) \right] + i \, \mathrm{Im} \left[\frac{1}{2} \log \left(\frac{1+z}{1-z} \right) \right].$$

What is $f(\mathbb{D})$? Notice that

$$f(z) = \left[-\frac{1}{2} \arg \left(\frac{i+z}{i-z} \right) \right] + i \left[\frac{1}{2} \arg \left(\frac{1+z}{1-z} \right) \right] = u + iv.$$

Let $z = e^{i\theta} \in \partial \mathbb{D}$. Then

$$\frac{i+z}{i-z} = \frac{i + e^{i\theta}}{i - e^{i\theta}} \frac{-i - e^{-i\theta}}{-i - e^{-i\theta}} = \frac{1 - i(e^{i\theta} + e^{-i\theta}) - 1}{1 + i(e^{i\theta} - e^{-i\theta}) + 1} = -i \frac{\cos \theta}{1 - \sin \theta}.$$

Thus,

$$u = -\frac{1}{2} \arg \left(\frac{i+z}{i-z} \right) \bigg|_{z=e^{i\theta}} = \begin{cases} \frac{\pi}{4} & \text{if } \cos \theta > 0, \\ -\frac{\pi}{4} & \text{if } \cos \theta < 0. \end{cases}$$

Likewise, we can show that

$$v = \begin{cases} \frac{\pi}{4} & \text{if } \sin \theta > 0, \\ -\frac{\pi}{4} & \text{if } \sin \theta < 0. \end{cases}$$

Therefore, we have that $z = e^{i\theta} \in \partial \mathbb{D}$ is mapped to

$$u + iv = \begin{cases} z_1 = \frac{\pi}{2\sqrt{2}} e^{i\frac{\pi}{4}} = \frac{\pi}{4} + i\frac{\pi}{4} & \text{if } \theta \in (0, \frac{\pi}{2}), \\ z_2 = \frac{\pi}{2\sqrt{2}} e^{i\frac{3\pi}{4}} = -\frac{\pi}{4} + i\frac{\pi}{4} & \text{if } \theta \in (\frac{\pi}{2}, \pi), \\ z_3 = \frac{\pi}{2\sqrt{2}} e^{i\frac{5\pi}{4}} = -\frac{\pi}{4} - i\frac{\pi}{4} & \text{if } \theta \in (\pi, \frac{3\pi}{2}), \\ z_4 = \frac{\pi}{2\sqrt{2}} e^{i\frac{7\pi}{4}} = \frac{\pi}{4} - i\frac{\pi}{4} & \text{if } \theta \in (\frac{3\pi}{2}, 2\pi). \end{cases}$$

Thus, this harmonic function maps $\mathbb{D}$ onto the interior of the region bounded by a square with vertices at z_1, z_2, z_3 and z_4.

EXERCISE 2.122. Verify that shearing $h(z) - g(z) = z - \frac{1}{3}z^3$ with $\omega(z) = z^2$ yields $f(z) = z + \frac{1}{3}\overline{z}^3$ from Example 2.114.
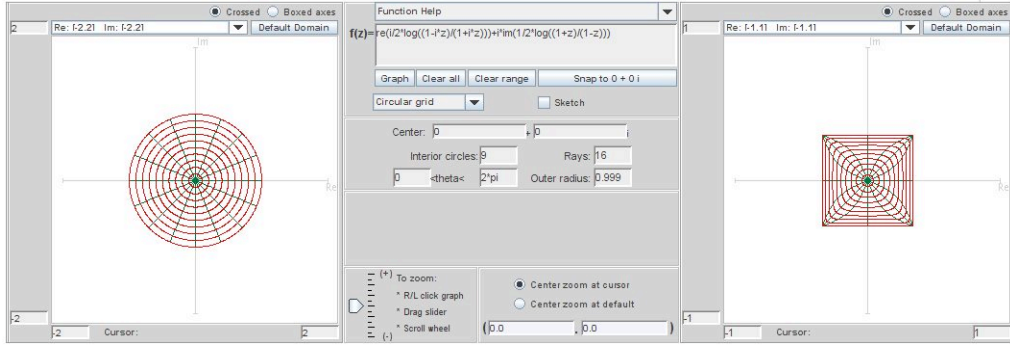 **Try it out!**

173

FIGURE 2.50. Image of $\mathbb{D}$ under $f(z) = \operatorname{Re}\left[\dfrac{i}{2}\log\left(\dfrac{i+z}{i-z}\right)\right] + i\operatorname{Im}\left[\dfrac{1}{2}\log\left(\dfrac{1+z}{1-z}\right)\right]$.

To actually find the minimal graph that is associated with specific types of harmonic univalent mappings, we need to develop the appropriate Weierstrass representation as outlined in eq (13). Recall that it must satisfy the properties $\phi^2 = 0$ and $|\phi|^2 \neq 0$, and we want it to use planar harmonic mappings. A natural choice is to consider

$$x_1 = \operatorname{Re}(h + g) = \operatorname{Re}\int (h' + g')\,dz = \operatorname{Re}\int \varphi_1\,dz$$

$$x_2 = \operatorname{Im}(h - g) = \operatorname{Re}\int -i(h' - g')\,dz = \operatorname{Re}\int \varphi_2\,dz$$

$$x_3 = \operatorname{Re}\int \varphi_3\,dz$$

and then solve for φ_3.

EXERCISE 2.123. Derive that $\varphi_3 = 2ih'\sqrt{g'/h'} = 2i\sqrt{g'h'}$.
Try it out!

We need φ_3 to be analytic and so we require the dilatation $\omega = g'/h'$ to be a perfect square.

THEOREM 2.124 (Weierstrass Representation (h,g)). If $f = h + \overline{g}$ is a sense-preserving harmonic univalent mapping of $\mathbb{D}$ onto some domain $\Omega \in \mathbb{C}$ with dilatation $\omega = q^2$ for some function q analytic in $\mathbb{D}$, then the isothermal parametrization

$$\mathbf{x}(u, v) = (x_1, x_2, x_3)$$
$$= \left(\operatorname{Re}\{h(z) + g(z)\}, \operatorname{Im}\{h(z) - g(z)\}, 2\operatorname{Im}\left\{\int_0^z \sqrt{g'(\zeta)h'(\zeta)}d\zeta\right\} \right)$$

defines a minimal graph whose projection onto the complex plane is f. Conversely, if a minimal graph $\mathbf{x}(u, v) = \{(u, v, F(u, v)) : u + iv \in \Omega\}$ is parametrized by sense-preserving isothermal parameters $z = x + iy \in \mathbb{D}$, then the projection onto its base

174

plane defines a harmonic univalent mapping $f(z) = u + iv = \text{Re}\{h(z) + g(z)\} + i\,\text{Im}\{h(z) - g(z)\}$ of $\mathbb{D}$ onto Ω whose dilatation is the square of an analytic function.

Summary: Let $f = h + \bar{g}$ defined on $\mathbb{D}$ be a harmonic univalent mapping such that the dilatation $\omega = g'/h'$ is the square of an analytic function and $|\omega(z)| < 1$ for all $z \in \mathbb{D}$. Then f lifts to a minimal graph using the Weierstrass formula given in Theorem 2.124.

EXAMPLE 2.125. Recall from Example 2.114 the harmonic univalent mapping

$$f(z) = z + \frac{1}{3}\bar{z}^3 = \text{Re}\left(z + \frac{1}{3}z^3\right) + i\,\text{Im}\left(z - \frac{1}{3}z^3\right).$$

Note that $h(z) = z$ and $g(z) = \frac{1}{3}z^3$. Also, $\omega(z) = z^2$ which is the square of an analytic function. Hence this harmonic mapping lifts to a minimal graph. We compute that

$$x_3 = 2\,\text{Im}\int_0^z \sqrt{h'(\zeta)g'(\zeta)}\,d\zeta = \text{Im}\left(z^2\right).$$

This yields a parametrization of a surface that is the conjugate of the Enneper's surface given in Example 2.67:

$$\mathbf{x} = \left(\text{Re}\left\{z + \frac{1}{3}z^3\right\}, \text{Im}\left\{z - \frac{1}{3}z^3\right\}, \text{Im}\left\{z^2\right\}\right)$$

and hence yields Enneper's surface. Note that the projection of the Enneper surface onto the x_1x_2-plane is the image of $\mathbb{D}$ under the harmonic mapping f. Also, while Enneper's surface is not a graph over $\mathbb{C}$, it is a graph over $\mathbb{D}$ as this result proves. You can see this by using *MinSurfTool* with the **W.E. (h,g)** tab. Enter in the functions $h(z) = z$, $g(z) = \frac{1}{3}z^3$, and $\sqrt{h' \cdot g'} = z$. Make sure the `Surface and projection` button is clicked (this will allow you to see both the minimal surface and its projection that is related to the harmonic mapping). Also, use the Disk domain for the unit disk (i.e., radius min: 0; radius max: 1; theta min: 0; theta max: 2 pi). The minimal surface is colored purple while the $f(\mathbb{D})$ is colored green (see Figure 2.51). As you move the image so that it is viewed from the top, the projection of the minimal surface matches the image of $f(\mathbb{D})$ (see Figure 2.52).

EXPLORATION 2.126. In the Weierstrass representation (h,g), we require that $\omega = g'/h'$ be the square of an analytic function. This is necessary because $\varphi_3 = ih'\sqrt{g'/h'}$, and if g'/h' were not the square of an analytic function, then there would be two branches of the square root. Geometrically, we can see that this is necessary. Use *MinSurfTool* with the **W.E. (h,g)** tab to graph the following images and describe why the geometry of those functions $f = h + \bar{g}$ in the left column do lift to a minimal graph

175
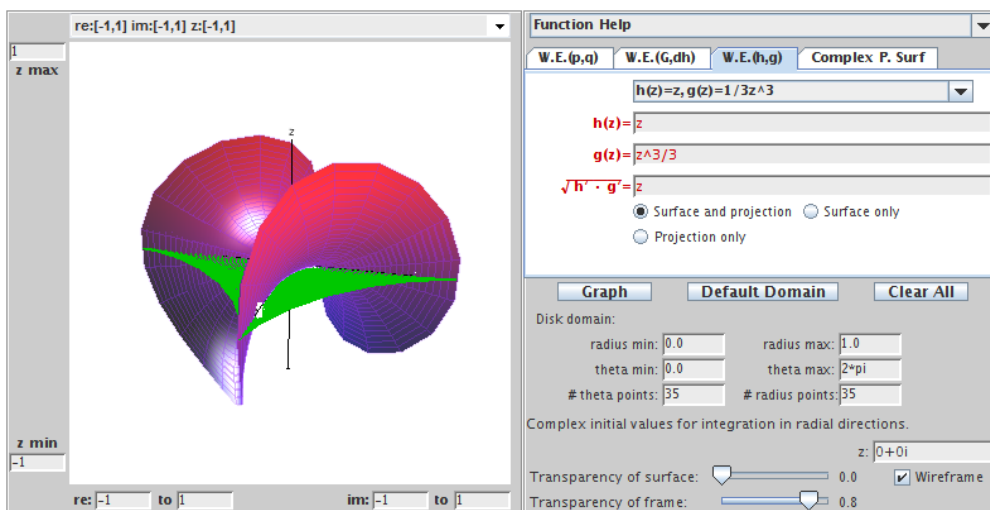
FIGURE 2.51. Side view of the Enneper surface and the image of the unit disk under the harmonic map.
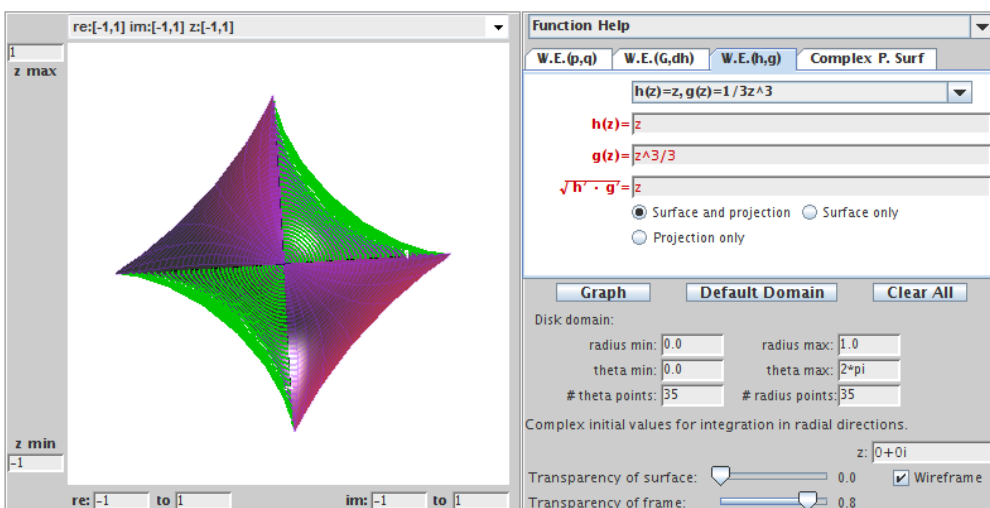
FIGURE 2.52. The projection of the Enneper surface is the image of the unit disk under the harmonic map.

while those in the right column do not. To enter $\sqrt{z}$ into the applet type `sqrt(z)`.

(a) $z + \frac{1}{3}\bar{z}^3$;
 (note: $\omega = z^2$ and $\sqrt{h' \cdot g'} = z$)

(b) $z + \frac{1}{2}\bar{z}^2$;
 (note: $\omega = z$ and $\sqrt{h' \cdot g'} = \sqrt{z}$)

(c) $z - \frac{1}{5}\bar{z}^5$;
 (note: $\omega = -z^4$ and $\sqrt{h' \cdot g'} = iz^2$)

(d) $z - \frac{1}{4}\bar{z}^4$.
 (note: $\omega = -z^5$ and $\sqrt{h' \cdot g'} = i\sqrt{z^5}$)

176

EXAMPLE 2.127. Consider the harmonic univalent mapping from Example 2.121 given by

$$f(z) = h(z) + \overline{g(z)} = \mathrm{Re}\left[\frac{i}{2}\log\left(\frac{i+z}{i-z}\right)\right] + i\,\mathrm{Im}\left[\frac{1}{2}\log\left(\frac{1+z}{1-z}\right)\right].$$

Because $\omega(z) = -z^2$ is the square of an analytic function, we can lift this harmonic mapping to a minimal graph. We compute that

$$x_3 = 2\,\mathrm{Im}\int_0^z \frac{iz}{1-z^4}\,d\zeta = \frac{1}{2}\,\mathrm{Im}\left\{i\log\left(\frac{1+z^2}{1-z^2}\right)\right\}.$$

This yields a parametrization of Scherk's doubly periodic minimal surface:

$$\mathbf{x} = \left(\mathrm{Re}\left[\frac{i}{2}\log\left(\frac{i+z}{i-z}\right)\right], \mathrm{Im}\left[\frac{1}{2}\log\left(\frac{1+z}{1-z}\right)\right], \frac{1}{2}\,\mathrm{Im}\left[i\log\left(\frac{1+z^2}{1-z^2}\right)\right]\right).$$

We can use **Pre-set functions** from the **W.E. (h,g)** tab on *MinSurfTool* to plot the minimal graph and the image of the unit disk under the planar harmonic mapping, where

$$h(z) = \frac{1}{4}\log\left(\frac{1+z}{1-z}\right) + \frac{i}{4}\log\left(\frac{i+z}{i-z}\right),$$

$$g(z) = -\frac{1}{4}\log\left(\frac{1+z}{1-z}\right) + \frac{i}{4}\log\left(\frac{i+z}{i-z}\right).$$

Because of the singularities at $\pm 1, \pm i$, the radius max is set to 0.999. Notice that the projection of Scherk's doubly periodic surface onto the $x_1 x_2$-plane is a square which is the image of $\mathbb{D}$ under the harmonic mapping f.

FIGURE 2.53. Side view of Scherk's doubly periodic surface and the image of the unit disk under the harmonic map.

177

EXPLORATION 2.128. Use the **Pre-set functions** from the **W.E. (h,g)** tab on *MinSurfTool* to plot the minimal graphs associated with the given functions h and g for the planar harmonic mappings. Determine which minimal surfaces these are.

(a) $h(z) = z$, $g(z) = \frac{1}{2n+1} z^{2n+1} (n = 1, 2, 3, \ldots)$;

(b) $h(z) = z$, $g(z) = \frac{1}{z}$;

(c) $h(z) = z$, $g(z) = -\frac{1}{z}$;

(d) $h(z) = \frac{1}{5} z^5$, $g(z) = -\frac{1}{z}$;

(e) $h(z) = \frac{1}{4} \log \left(\frac{i+z}{i-z} \right) - \frac{i}{4} \log \left(\frac{1+z}{1-z} \right)$, $g(z) = \frac{1}{4} \log \left(\frac{i+z}{i-z} \right) + \frac{i}{4} \log \left(\frac{1+z}{1-z} \right)$.

Try it out!

If we start with a harmonic univalent mapping that has a perfect square dilatation we can use Theorem 2.124 to find the parametrization of the corresponding minimal graph. In Example 2.125, we saw that the minimal graph corresponding to the harmonic univalent map $f(z) = z + \frac{1}{3}\bar{z}^3$ is Enneper's surface, and in Example 2.127 the minimal graph corresponding to the harmonic univalent map $f(z) = \text{Re}[\frac{i}{2} \log(\frac{i+z}{i-z})] + i \,\text{Im}[\frac{1}{2} \log(\frac{1+z}{1-z})]$ is Scherk's doubly periodic surface. There have been several research papers in the field of harmonic univalent mappings that have used Theorem 2.124 to create minimal graphs from harmonic univalent mappings (e.g., [**8**], [**9**], [**10**], [**11**], [**17**]). However, many of them have not identified the specific minimal graph created, because it is often not easy to do so as it was in Examples 2.125 and 2.127.

Question: Given a harmonic univalent mapping we can use Theorem 2.124 to find the parametrization of a minimal graph. Can we determine which minimal graph this is?

The following examples will show how can determine the minimal graph when the parametrization is not a standard parametrization for a known minimal surface.

EXAMPLE 2.129. The analytic function,

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

is an important and interesting function in complex analysis. It maps $\mathbb{D}$ onto the right half-plane $\{z \in \mathbb{C} \mid \text{Re}\{z\} > 0\}$. If we shear

$$h(z) - g(z) = \frac{z}{1-z} \quad \text{with} \quad \omega(z) = g'(z)/h'(z) = z^2,$$

then we get the harmonic univalent mapping $f = h + \bar{g}$, where

$$h = \frac{1}{8} \log \left(\frac{z+1}{z-1} \right) + \frac{3z - 2z^2}{4(1-z)^2} \quad \text{and} \quad g = \frac{1}{8} \log \left(\frac{z+1}{z-1} \right) - \frac{z - 2z^2}{4(1-z)^2}.$$

The image of $\mathbb{D}$ under $f = h + \overline{g}$ is shown in Figure 2.54.

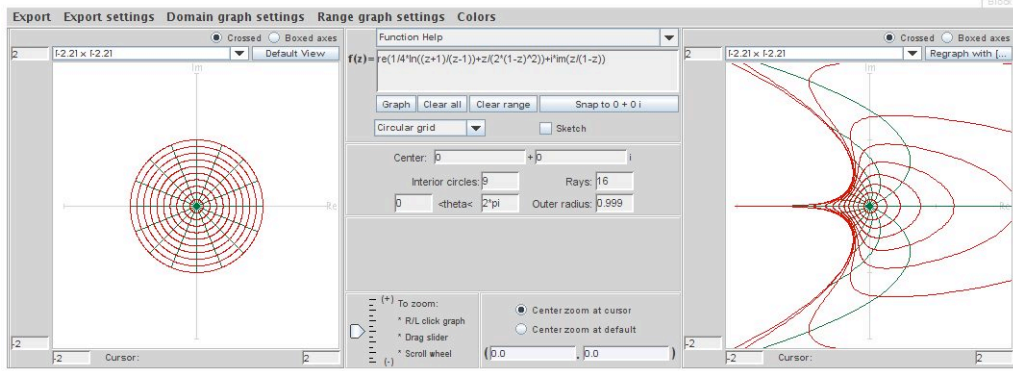

FIGURE 2.54. Image of $\mathbb{D}$ under $f = \text{Re}\left(\frac{1}{4}\log\left(\frac{z+1}{z-1}\right) + \frac{z}{2(1-z)^2}\right) + \text{Im}\left(\frac{z}{1-z}\right)$.

Because $\omega(z) = z^2$, we can use Theorem 2.124 to find the parametrization of the corresponding minimal graph:

$$\mathbf{x} = \left(\text{Re}\left\{\frac{1}{4}\log\left(\frac{z+1}{z-1}\right) + \frac{z}{2(1-z)^2}\right\}, \text{Im}\left\{\frac{z}{1-z}\right\}, \text{Im}\left\{\frac{1}{4}\log\left(\frac{z+1}{z-1}\right) - \frac{z}{2(1-z)^2}\right\}\right).$$

What minimal graph is this? It is difficult to tell. The coordinate functions do not look like a standard parametrization for a known minimal surface. Let us try to use a substitution to rewrite this parametrization into a form that allows us to identify the minimal graph. If we use a Möbius transformation for the substitution, it will not affect the geometry of the minimal graph. Letting $z \mapsto \frac{\hat{z}+1}{\hat{z}-1}$, we get the parametrization:

$$\hat{\mathbf{x}} = \left(\frac{1}{4}\text{Re}\left\{\log(\hat{z}) + \frac{1}{2}\hat{z}^2 - \frac{1}{2}\right\}, -\frac{1}{2}\text{Im}\left\{\hat{z}\right\}, -\frac{1}{4}\text{Im}\left\{\log(\hat{z}) - \frac{1}{2}\hat{z}^2 + \frac{1}{2}\right\}\right).$$

This is useful, because it simplifies the log terms in x_1 and x_3. Also, notice that by switching the coordinate functions, and factoring out $\frac{1}{2}$ we have:

$$(20) \quad \widetilde{\mathbf{x}} = \left(-\frac{1}{2}\left[\frac{1}{2}\text{Im}\left\{\log(\widetilde{z}) - \frac{1}{2}\widetilde{z}^2\right\}\right], \frac{1}{2}\left[\frac{1}{2}\text{Re}\left\{\log(\widetilde{z}) + \frac{1}{2}\widetilde{z}^2\right\}\right], -\frac{1}{2}\left[\text{Im}\left\{\widetilde{z}\right\}\right]\right),$$

and this looks close to the standard parametrization for the wavy plane. In Figure 2.55 we have used the **Complex P. Surf** tab in *MinSurfTool* to graph the image of the surface with this parametrization.

In fact, the coordinate functions above correspond to the conjugate surface of the wavy plane scaled by $\frac{1}{2}$. This is clear given the actual coordinates of the wavy plane
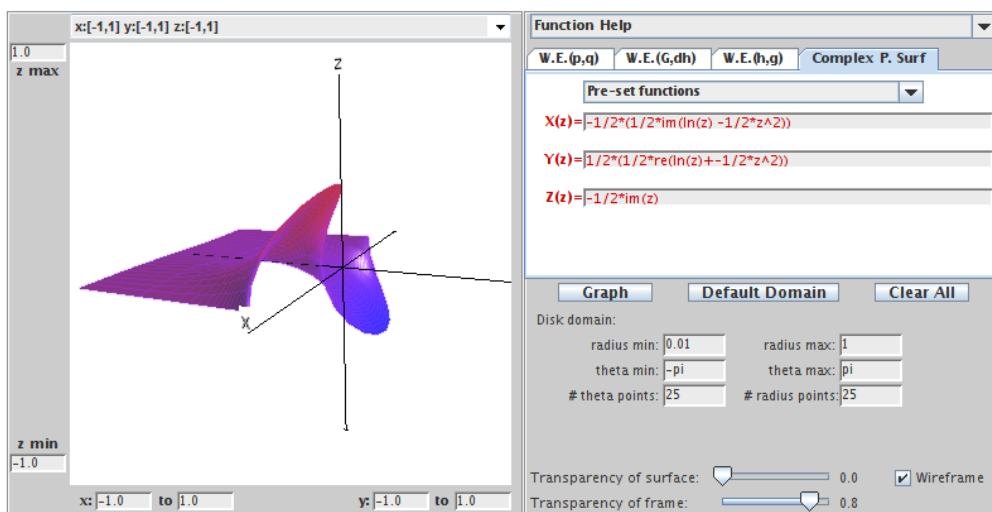
FIGURE 2.55. The conjugate of the wavy plane.

below:

$$\mathbf{W} = \left(\frac{1}{2} \operatorname{Re} \left\{ \log (z) - \frac{1}{2} z^2 \right\}, -\frac{1}{2} \operatorname{Im} \left\{ \log (z) + \frac{1}{2} z^2 \right\}, \operatorname{Re} \{z\} \right).$$

Since the wavy plane is its own conjugate surface, this means that it is accurate to describe our original minimal graph as the wavy plane.

How does this match with the image in Figure 2.54 of $\mathbb{D}$ under the original harmonic univalent map? In deriving the familiar parametrization of the wavy plane, we applied the transformation $z \mapsto \frac{\hat{z}+1}{\hat{z}-1}$, switched the coordinate functions, and took the conjugate function. These changes are equivalent to altering the original domain $\mathbb{D}$, so we are looking at a piece of the wavy plane coming from a different domain G. The projection of the wavy plane coming from G is shown in Figure 2.56. Notice the similarity between the image in Figure 2.54 and the image in Figure 2.56.

EXERCISE 2.130. An important function in complex analysis is the Koebe function given by $\frac{z}{(1-z)^2}$. By shearing

$$h(z) - g(z) = \frac{z}{(1 - z)^2} \quad \text{with} \quad \omega(z) = z^2,$$

we derive the harmonic univalent mapping $f = h + \overline{g}$, where

$$h = \frac{z - z^2 + \frac{1}{3} z^3}{(1 - z)^3} \quad \text{and} \quad g = \frac{\frac{1}{3} z^3}{(1 - z)^3}.$$

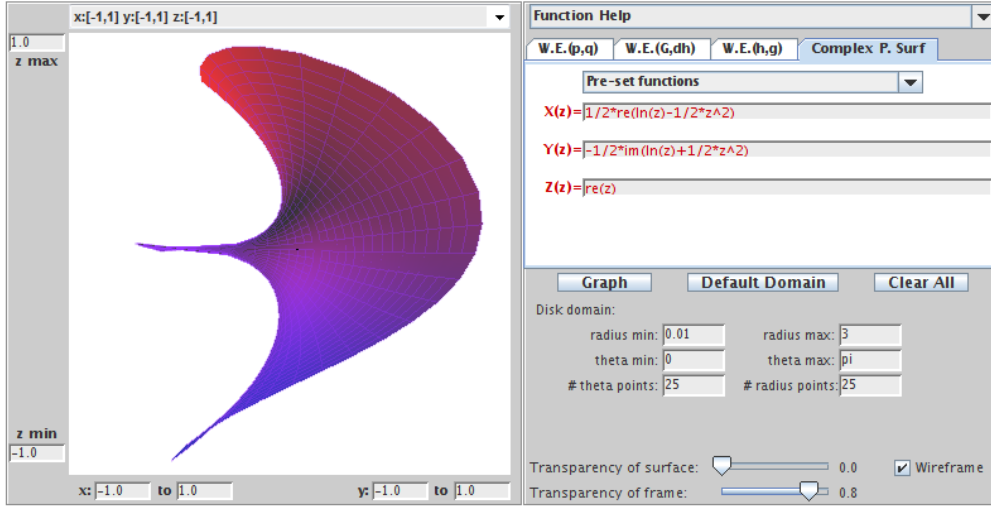(a) Use Theorem 2.124 to find the parametrization of the minimal graph that f lifts to.

FIGURE 2.56. The wavy plane mapped from domain G.

   (b) Use the approach of Example 2.129 to show analytically that this minimal
       graph is Enneper's surface.
   (c) Use *ComplexTool* to graph the image of $\mathbb{D}$ under f and *MinSurfTool* to sketch
       the corresponding minimal graph.

Try it out!

EXAMPLE 2.131. By shearing $h(z) - g(z) = \frac{1}{2} \log\left(\frac{1+z}{1-z}\right)$ with $\omega(z) = g'(z)/h'(z) = m^2 z^2$, where $|m| \leq 1$, it was shown in [9] that the harmonic function $f = h + \overline{g}$ is univalent, where

$$h(z) = \frac{1}{2(1-m^2)} \log\left(\frac{1+z}{1-z}\right) + \frac{m}{2(m^2-1)} \log\left(\frac{1+mz}{1-mz}\right)$$

$$g(z) = \frac{m^2}{2(1-m^2)} \log\left(\frac{1+z}{1-z}\right) + \frac{m}{2(m^2-1)} \log\left(\frac{1+mz}{1-mz}\right).$$

When $m = e^{i\frac{\pi}{2}}$, the function f is the same as in Example 2.121 and the image of $\mathbb{D}$ under $f = h + \overline{g}$ is a square.

In fact, for every m such that $|m| = 1$, the image of $\mathbb{D}$ under $f = h + \overline{g}$ is a parallelogram.

Since $\omega(z) = g'(z)/h'(z) = m^2 z^2$, we can lift f to a minimal graph. We compute that

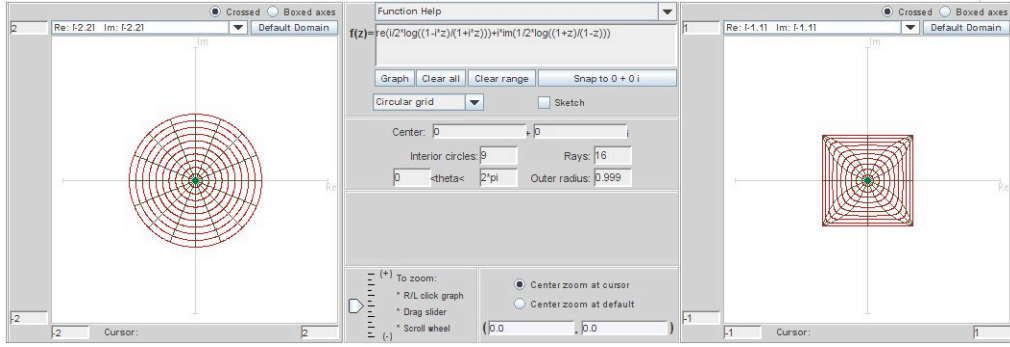$$x_3 = \text{Im}\left\{\frac{m}{1-m^2} \log\left(\frac{1-m^2 z^2}{1-z^2}\right)\right\}.$$

181

FIGURE 2.57. Image of $\mathbb{D}$ under $f = h + \overline{g}$ when $m = e^{i\frac{\pi}{2}}$.
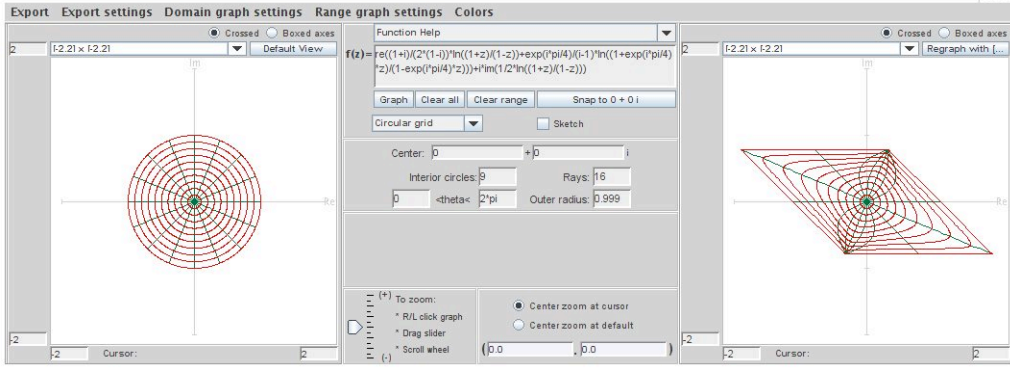

FIGURE 2.58. Image of $\mathbb{D}$ under $f = h + \overline{g}$ when $m = e^{i\frac{\pi}{4}}$.

Hence, the corresponding parametrization of the minimal graph is

$$\mathbf{x} = \left(\mathrm{Re} \left\{ \frac{1 + m^2}{2(1 - m^2)} \log \left(\frac{1 + z}{1 - z} \right) + \frac{m}{(m^2 - 1)} \log \left(\frac{1 + mz}{1 - mz} \right) \right\}, \right.$$

$$\left. \mathrm{Im} \left\{ \frac{1}{2} \log \left(\frac{1 + z}{1 - z} \right) \right\}, \mathrm{Im} \left\{ \frac{m}{1 - m^2} \log \left(\frac{1 - m^2 z^2}{1 - z^2} \right) \right\} \right).$$

When $m = e^{i\frac{\pi}{2}}$, the minimal graph is Scherk's doubly periodic surface (see Figure 2.59).

For $m = e^{i\theta}, (0 < \theta < \frac{\pi}{2})$, the minimal graphs are slanted Scherk's surfaces (see Figure 2.60).

What is the minimal graph when $m = 1$? In the limit (i.e., $\theta = 0$) we have the equation

$$\mathbf{x} = \left(\mathrm{Re} \left\{ \frac{z}{1 - z^2} \right\}, \mathrm{Re} \left\{ -\frac{i}{2} \log \left(\frac{1 + z}{1 - z} \right) \right\}, \mathrm{Re} \left\{ \frac{-iz^2}{1 - z^2} \right\} \right).$$

182
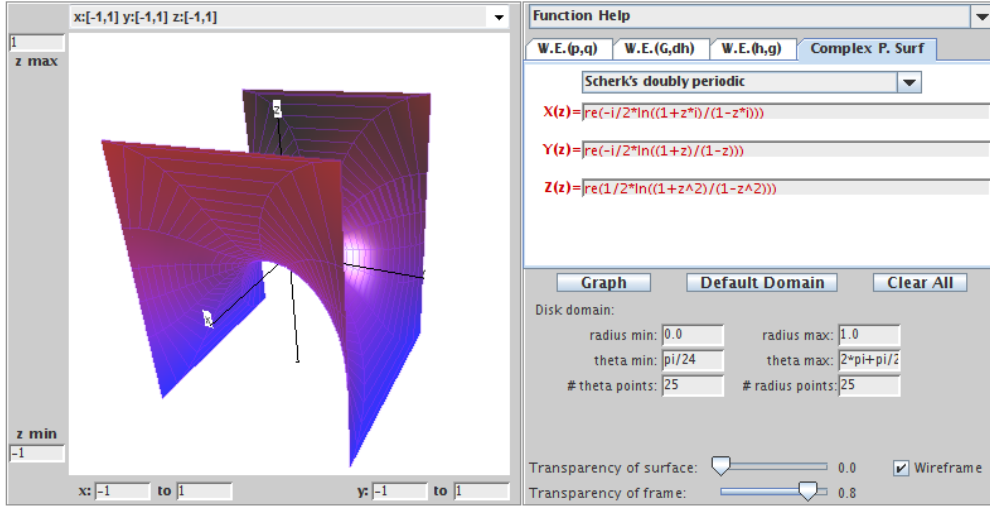
FIGURE 2.59. Scherk's doubly periodic square surface ($m = e^{i\frac{\pi}{2}}$).
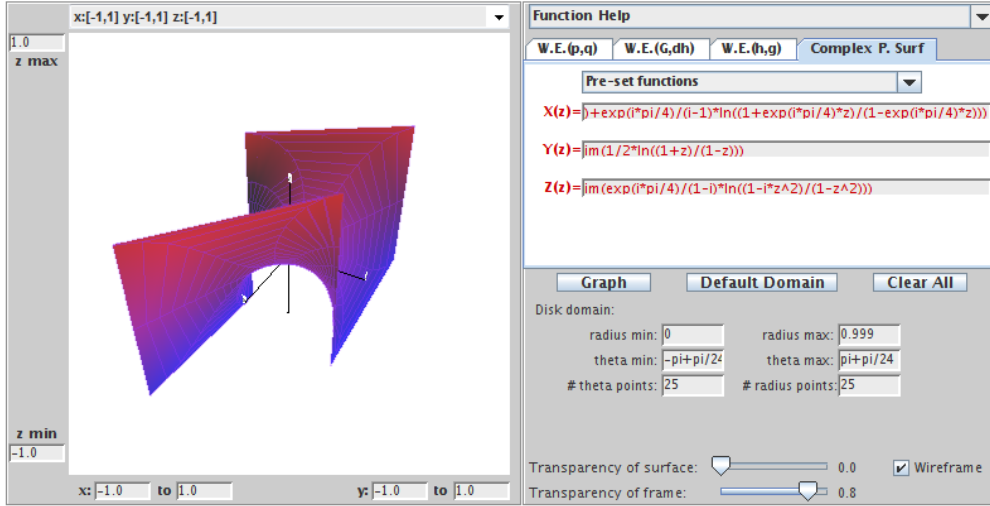
FIGURE 2.60. A Scherk's doubly periodic parallelogram surface ($m = e^{i\frac{\pi}{4}}$).

Again, this parametrization does not look like a standard parametrization of a known minimal surface. However, we can use a substitution to rewrite this parametrization into a form that allows us to identify the minimal graph. Using the substitution $z \longmapsto \frac{e^z-1}{e^z+1}$ and the fact that $\operatorname{Re}\left\{\dfrac{-iz^2}{1-z^2}\right\} = \operatorname{Re}\left\{\dfrac{1}{2i}\dfrac{1+z^2}{1-z^2}\right\}$, this equation is equivalent to

$$X = \left(\frac{1}{2}\sinh u \cos v, \frac{1}{2}v, \frac{1}{2}\sinh u \sin v\right),$$

183

which is an equation of a helicoid. Thus, we have a family of minimal surfaces that transform Scherk's doubly periodic surface continuously into the helicoid. This is neat!
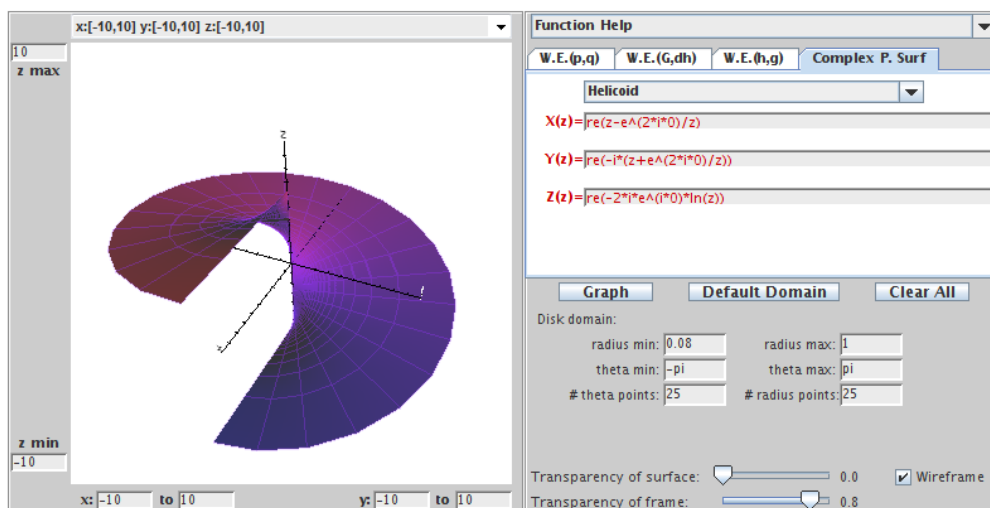
FIGURE 2.61. Side view of helicoid that is the limit function of the slanted Scherk's surfaces.

EXERCISE 2.132. Show that $\mathrm{Re}\left\{\dfrac{-iz^2}{1-z^2}\right\} = \mathrm{Re}\left\{\dfrac{1}{2i}\dfrac{1+z^2}{1-z^2}\right\}$.

Try it out!

EXERCISE 2.133. Consider the harmonic univalent map $f(z) = h(z) + \overline{g}(z)$, where

$$h = \frac{1}{4}\log\left(\frac{1+z}{1-z}\right) + \frac{\frac{1}{2}z}{1-z^2} \quad \text{and} \quad g = \frac{1}{4}\log\left(\frac{1+z}{1-z}\right) - \frac{\frac{1}{2}z}{1-z^2}.$$

(a) Use Theorem 2.124 to find the parametrization of the minimal graph that f lifts to.
(b) Use *ComplexTool* to graph the image of $\mathbb{D}$ under f and *MinSurfTool* with tab **W.E.(h,g)** to sketch the corresponding minimal graph.
(c) Use the approach of Example 2.131 to show analytically that this minimal graph is the catenoid.

Try it out!

LARGE PROJECT 2.134. The analytic function, $F(z) = z$, maps the unit disk, $\mathbb{D}$, onto itself. Shear $h(z) - g(z) = z$ with various perfect-square dilatations, ω, that satisfy the condition $|\omega| < 1$ for all $z \in \mathbb{D}$ (e.g., $\omega = z^{2n}(n \in \mathbb{N}), \omega = e^{i\theta}z^2(\theta \in \mathbb{R}), \omega = \left(\frac{z-a}{1-\overline{a}z}\right)^2(|a| < 1)$). Determine the corresponding minimal graphs.
Optional

184

LARGE PROJECT 2.135. The analytic function, $F(z) = \frac{z}{1-z}$, maps the unit disk, $\mathbb{D}$, onto a right half-plane and is an important function. Shear $h(z) + g(z) = \frac{z}{1-z}$ with various perfect-square dilatations, ω, that satisfy the condition $|\omega| < 1$ for all $z \in \mathbb{D}$ (e.g., $\omega = z^{2n}(n \in \mathbb{N}), \omega = e^{i\theta}z^2(\theta \in \mathbb{R}), \omega = \left(\frac{z-a}{1-\bar{a}z}\right)^2(|a| < 1)$). Determine the corresponding minimal graphs.

Optional

2.7. Conclusion

We have presented an introduction to minimal surfaces and described a few topics that students can explore using the exercises, the exploratory problems, and the projects along with the applets. For a deeper and thorough explanation of differential geometry consult [**7**], [**18**], or [**21**] for beginners, and [**3**] for intermediates. Also, you should consider Spivak's five volume work [**24**]. Oprea's book [**22**] is a very nice source for an introduction to minimal surfaces. For more background on minimal surfaces we recommend [**25**], [**14**], [**15**], [**6**], [**23**], and [**20**].

2.8. Additional Exercises

Differential Geometry

EXPLORATION 2.136. An oblique cylinder can be parametrized by

$$\mathbf{x}(u, v) = (\cos u, \sin u + v \cos \theta, v \sin \theta),$$

where $\theta \in \left(0, \frac{\pi}{2}\right)$ is a fixed value. Use *DiffGeomTool* to explore what happens to the oblique cylinder as θ varies between 0 and $\frac{\pi}{2}$.

EXERCISE 2.137. Use *DiffGeomTool* to graph the surface parametrized by

$$\mathbf{x}(u, v) = \left(\cos u \left(1 + v \sin \left(\frac{1}{2} u \right) \right), \sin u \left(1 + v \sin \left(\frac{1}{2} u \right) \right), v \cos \left(\frac{1}{2} u \right) \right),$$

where $-\pi < u < \pi$, $-\frac{1}{2} < v < \frac{1}{2}$. This surface is known as the Möbius strip and is nonorientable; that is, the normal vector can change from pointing outward to pointing inward as it travels along a closed path on the surface. You can see this in *DiffGeomTool* by clicking on the **Normal vector** box and setting the `Point locator: (u,v)=` to $(\pi - 0.1, 0)$. Next, change this u coordinate to each of the following values: $u = \pi - 0.1 - 1$, $u = \pi - 0.1 - 2$, $u = \pi - 0.1 - 3$, $u = \pi - 0.1 - 4$, $u = \pi - 0.1 - 5$, and $u = \pi - 0.1 - 6$. As you do so, observe that $\mathbf{n}$ will make nearly a complete path along a closed curve but it will change the direction it is pointing from where it started to where it ended.

EXERCISE 2.138. Describe the u-parameter and v-parameter curves on the Enneper surface.

EXERCISE 2.139. In Exploration 2.11(c), you proved the largest value of r for which the Enneper surface has no self-intersections assuming that the intersection occurs on the x_3-axis. In this exercise, prove the same result without assuming the intersection occurs on the x_3-axis.

EXERCISE 2.140. Compute the coefficients of the first and the second fundamental forms for the Enneper surface whose parametrization is

$$\mathbf{x}(u, v) = \left(u - \frac{1}{3} u^3 + uv^2, v - \frac{1}{3} v^3 + u^2 v, u^2 - v^2 \right).$$

EXERCISE 2.141. A *CMC (Constant Mean Curvature) surface* is a surface that has the same mean curvature everywhere on the surface. Minimal surfaces are a subset of CMC surfaces. Using *DiffGeomTool* sketch the following surfaces and determine which are CMC surfaces:

(a). $\mathbf{x}(u, v) = (u - v, u + v, 2(u^2 + v^2))$, where $-1 < u < 1$, $-1 < v < 1$;

(b). $\mathbf{x}(u, v) = (\cos u, \sin u, v)$, where $-\pi < u < \pi$, $-2 < v < 2$;

(c). $\mathbf{x}(u, v) = \left((2 + \cos v) \cos u, (2 + \cos v) \sin u, \sin v \right)$, where $0 < u, v < 2\pi$;

(d). $\mathbf{x}(u, v) = (\sqrt{1 - u^2} \cos v, \sqrt{1 - u^2} \sin v, u)$, where $-1 < u < 1$, $-\pi < v < \pi$;

Minimal Surfaces

EXERCISE 2.142. Use eq (5) to show that the Enneper surface parametrized by

$$\mathbf{x}(u,v) = \left(u - \frac{1}{3}u^3 + uv^2, v - \frac{1}{3}v^3 + u^2v, u^2 - v^2 \right)$$

is a minimal surface.

EXERCISE 2.143. Prove Theorem 2.39 in the special case that the surface of revolution has the parametrization

$$\mathbf{x}(u,v) = \left(f(v) \cos u, f(v) \sin u, v \right).$$

EXERCISE 2.144. An oblique cylinder is a cylinder whose side forms an angle θ with the x_1x_2-plane, where $0 < \theta \le \frac{\pi}{2}$. For a fixed θ it can be parametrized by

$$\mathbf{x}(u,v) = \left(\cos u, \sin u + v \cos \theta, v \sin \theta \right).$$

Determine the values of θ for which $\mathbf{x}$ is isothermal.

EXERCISE 2.145. Show that the parametrization

$$\mathbf{x}(u,v) = \left(\arctan\left(\frac{2u}{1-(u^2+v^2)} \right), \arctan\left(\frac{-2v}{1-(u^2+v^2)} \right), \right.$$
$$\left. \frac{1}{2}\ln\left(\frac{(u^2-v^2+1)^2 + 4u^2v^2}{(u^2-v^2-1)^2 + 4u^2v^2} \right) \right)$$

is an isothermal parametrization of Scherk's doubly periodic surface (that is, show that it is isothermal and that there is transformation that maps this parametrization to the parametrization given in Exercise 2.43(b) for Scherk's doubly periodic surface).

EXERCISE 2.146. Let X be a minimal surface that is not a plane, and let Y be its conjugate minimal surface. Is it possible that the plane is one of the associated surfaces for X and Y? If so, describe the geometry of the minimal surface X. If not, explain why.

Weierstrass Representation

EXERCISE 2.147. In Example 2.67, show the details in going from the Enneper surface parametrization

$$\mathbf{x} = \left(\operatorname{Re}\left\{ z - \frac{1}{3}z^3 \right\}, \operatorname{Re}\left\{ -i\left(z + \frac{1}{3}z^3 \right) \right\}, \operatorname{Re}\left\{ z^2 \right\} \right)$$

to the parametrization

$$\mathbf{x}(u,v) = \left(u - \frac{1}{3}u^3 + uv^2, v - \frac{1}{3}v^3 + vu^2, u^2 - v^2 \right)$$

that is also for the Enneper surface.

EXERCISE 2.148. Compute the parametrization for the minimal surfaces generated by using $p(z) = \frac{1}{2z}$ and $q(z) = iz$ on the domain $\mathbb{C} - \{0\}$ in the Weierstrass representation. Use *MinSurfTool* with the W.E.(p,q) tab to graph an image of this surface which is known as the *wavy plane*. [Use radius min=0.001, radius max=1.3, theta min=-pi, theta max=pi with initial values $x = \text{Re}(1/2 * \log(z) - 1/4 * z^2)$, $y = \text{Im}(1/2 * \log(z) + 1/4 * z^2)$, and $z = \text{Re}(z)$.]

EXERCISE 2.149. Compute the parametrization for the minimal surfaces generated by using $p(z) = z^2$ and $q(z) = \frac{i}{z^2}$ on the domain $\mathbb{C} - \{0\}$ in the Weierstrass representation. Use *MinSurfTool* with the W.E.(p,q) tab to graph an image of this surface which is known as *Richmond's surface*. [Use radius min=0.1, radius max=1, theta min=pi/24, theta max=2pi+pi/24 with initial values $x = \text{Re}(1/3 * z^3 + 1/z)$, $y = \text{Im}(1/3 * z^3 - 1/z)$, and $z = \text{Re}(2 * z)$.]

EXERCISE 2.150. Compute the parametrization for the minimal surfaces generated by using $p(z) = \frac{(z+1)^2}{z^4}$ and $q(z) = \frac{z^2(z-1)}{z+1}$ on the domain $\mathbb{D} - \{0\}$ in the Weierstrass representation. Use *MinSurfTool* with the W.E.(p,q) tab to graph an image of this surface which is known as the *wavy plane*. [Use radius min=0.1, radius max=0.9, theta min=pi/24, theta max=2pi+pi/24 with initial values $x =$, $y =$, and $z =$.]

The Gauss map, G, and height differential, dh

EXERCISE 2.151. Show that if $z = x + iy$ is the projection of the point (x_1, x_2, x_3) on the Riemann sphere onto to complex plane, then

$$x = \frac{x_1}{1 - x_3}, \quad y = \frac{x_2}{1 - x_3}.$$

EXERCISE 2.152. For Scherk's doubly periodic surface find:

(a) $G(0)$;   (b) $G(1)$;   (c) $G(-1)$;   (d) $G(i)$;   (e) $G(-i)$.

EXERCISE 2.153. The Weierstrass data for a 4-noid are

$$G(z) = z^3 \quad \text{and} \quad dh = \frac{z^3}{(z^4 - 1)^2} \, dz.$$

Show that the ends of the 4-noid are catenoid ends.

EXERCISE 2.154. Determine the asymptotic and curvature lines for Scherk's doubly periodic surface with $G(z) = z$ and $dh(z) = \frac{iz}{z^4-1}dz$.

EXERCISE 2.155. Determine the period conditions for the wavy plane with $G(z) = z$ and $dh(z) = dz$.

EXERCISE 2.156. Let M be the Scherk doubly periodic surface with 6 ends. Using the approach of Example 2.104 determine G and dh for this surface.

Minimal Surfaces and Harmonic Univalent Mappings

EXERCISE 2.157. Prove that if $f = u + iv$ is harmonic in a simply-connected domain G, then $f = h + \bar{g}$, where h and g are analytic.

EXERCISE 2.158. Prove that the representations $f(z) = h(z) + \bar{g}(z)$ and $f(z) = \mathrm{Re}\left\{h(z) + g(z)\right\} + i\,\mathrm{Im}\left\{h(z) - g(z)\right\}$ are equivalent.

EXERCISE 2.159. Shear $h(z) - g(z) = \frac{z}{1-z}$ with $\omega(z) = z^2$ to get the harmonic univalent function $f = h + \bar{g}$ given in Example 2.129, where

$$h = \frac{1}{8}\log\left(\frac{z+1}{z-1}\right) + \frac{3z - 2z^2}{4(1-z)^2} \quad \text{and} \quad g = \frac{1}{8}\log\left(\frac{z+1}{z-1}\right) - \frac{z - 2z^2}{4(1-z)^2}.$$

EXERCISE 2.160. Shear $h(z) - g(z) = \frac{z}{(1-z)^2}$ with $\omega(z) = z^2$ to get the harmonic univalent function $f = h + \bar{g}$ given in Exercise 2.130, where

$$h = \frac{1}{8}\log\left(\frac{z+1}{z-1}\right) + \frac{3z - 2z^2}{4(1-z)^2} \quad \text{and} \quad g = \frac{1}{8}\log\left(\frac{z+1}{z-1}\right) - \frac{z - 2z^2}{4(1-z)^2}.$$

EXERCISE 2.161. Show that the parametrization:

$$\mathbf{x} = \left(\mathrm{Re}\left[\frac{i}{2}\log\left(\frac{i+z}{i-z}\right)\right], \mathrm{Im}\left[\frac{1}{2}\log\left(\frac{1+z}{1-z}\right)\right], \frac{1}{2}\mathrm{Im}\left[i\log\left(\frac{1+z^2}{1-z^2}\right)\right]\right)$$

is equivalent to the parametrization in Exercise 2.145 that gives Scherk's doubly periodic minimal surface.

EXERCISE 2.162. Consider the harmonic univalent map $f(z) = h(z) + \bar{g}(z)$, where

$$h = \frac{3}{16}\log\left(\frac{1+z}{1-z}\right) - \frac{3i}{16}\log\left(\frac{1+iz}{1-iz}\right) + \frac{1}{4}\frac{z}{1-z^4}$$

$$g = -\frac{3}{16}\log\left(\frac{1+z}{1-z}\right) - \frac{3i}{16}\log\left(\frac{1+iz}{1-iz}\right) + \frac{1}{4}\frac{z^3}{1-z^4}.$$

(a) Use Theorem 2.124 to find the parametrization of the minimal graph that f lifts to.
(b) Use *ComplexTool* to graph the image of $\mathbb{D}$ under f and *MinSurfTool* with tab **W.E.(h,g)** to sketch the corresponding minimal graph.
(c) This minimal surface has 4 helicoidial ends and is not embedded (i.e., it does have self-intersections). However, Theorem 2.124 states that this surface should be a minimal graph and hence have no self-intersections. Explain why this is not a contradiction.

LARGE PROJECT 2.163. The analytic function, $F(z) = \frac{z}{(1-z)^2}$, maps the unit disk, $\mathbb{D}$, onto $\mathbb{C} \setminus (-\infty, -\frac{1}{4})$ and is an important function. Shear $h(z) - g(z) = \frac{z}{(1-z)^2}$ with various perfect-square dilatations, ω, that satisfy the condition $|\omega| < 1$ for all $z \in \mathbb{D}$ (e.g., $\omega = z^{2n}(n = \mathbb{N})\omega = e^{i\theta}z^2, (\theta \in \mathbb{R}), \omega = \left(\frac{z-a}{1-\bar{a}z}\right)(|a| < 1)$). Determine the corresponding minimal graphs.

Bibliography

[1] F. J. Almgren and J. Taylor, The geometry of soap films and soap bubbles, *Sci. Am.* **235** (1976), 8293.

[2] L. Bers, *Riemann Surfaces*, New York Univ., Institute of Mathematical Sciences, New York, 1957-1958.

[3] W. Boothby, *An introduction to differentiable manifolds and Riemannian geometry*, Second edition. Pure and Applied Mathematics, vol 120. Academic Press, Inc., Orlando, FL, 1986.

[4] J. Clunie and T. Sheil-Small, Harmonic univalent functions, *Ann. Acad. Sci. Fenn. Ser. A.I Math.* **9** (1984), 3-25.

[5] R. Courant and H. Robbins, *What Is Mathematics?*, Oxford University Press, 1941.

[6] U. Dierkes, S. Hildebrandt, A. Kster, and O. Wohlrab, *Minimal surfaces I*, Grundlehren der Mathematischen Wissenschaften, 295, Springer-Verlag, Berlin, 1992.

[7] M. do Carmo, *Differential geometry of curves and surfaces*, translated from the Portuguese, Prentice-Hall, Inc., Englewood Cliffs, N.J., 1976.

[8] M. Dorff, Minimal graphs in $\mathbb{R}^3$ over convex domains, *Proc. Amer. Math. Soc.* **132** (2004), no. 2, 491-498.

[9] M. Dorff and J. Szynal, Harmonic shears of elliptic integrals, *Rocky Mountain J. Math.* **35** (2005), no. 2, 485–499.

[10] K. Driver and P. Duren, Harmonic shears of regular polygons by hypergeometric functions, *J. Math. Anal. App.* **239** (1999), 72-84.

[11] P. Duren and W. Thygerson, Harmonic mappings related to Scherk's saddle-tower minimal surfaces, *Rocky Mountain J. Math.* **30** (2000), no. 2, 555-564.

[12] W. Hengartner and G. Schober, On schlicht mappings to domains convex in one direction, *Comment. Math. Helv.* **45** (1970), 303-314.

[13] S. Hildebrandt and A. Tromba, *Mathematics and Optimal Form*, Scientific American Library, 1985.

[14] H. Karcher, Construction of minimal surfaces, *Survey in Geometry*, Univ. of Tokyo, 1989.

[15] H. Karcher, Introduction to the complex analysis of minimal surfaces, Lecture notes given at the NCTS, Taiwan, 2003.

[16] H. Lewy, On the non-vanishing of the Jacobian in certain one-to-one mappings, *Bull. Amer. Math. Soc.* **42** (1936), no. 10, 689-692.

[17] J. McDougall and L. Schaubroeck, Minimal surfaces over stars, *J. Math. Anal. Appl.* **340** (2008), no. 1, 721-738.

[18] R. Millman and G. Parker, *Elements of differential geometry*, Prentice-Hall Inc., Englewood Cliffs, N. J., 1977.

[19] F. Morgan, Minimal surfaces, crystals, shortest networks, and undergraduate research, *Math. Intel.*, **14** (1992), 37-44.

[20] J. Nitsche, *Lectures on Minimal Surfaces*, vol. 1, Cambridge U. Press, 1989.

[21] J. Oprea, *Differential geometry and its applications.*, 2nd ed., Classroom Resource Materials Series. Math. Assoc. of America, Washington, DC, 2007.

[22] J. Oprea, *The mathematics of soap films: explorations with Maple*, Student Mathematical Library, vol 10, Amer. Math. Soc., Providence, RI, 2000.

[23] R. Osserman, *A survey of minimal surfaces*, 2nd ed. Dover Publications, Inc., New York, 1986.

[24] M. Spivak, *A Comprehensive Introduction to Differential Geometry*, vols 1-5, 3rd ed., Publish or Perish, Inc., Houston, TX, 2005.

[25] M. Weber, *Classical minimal surfaces in Euclidean space by examples*, Preliminary notes for the Clay Institute Summer School on minimal surfaces, MSRI, Berkeley, Calif., 2001.

[26] S. Weintraub, *Differential Forms: A Complement to Vector Calculus*, Academic Press, Inc., San Diego, CA, 1997.

CHAPTER 3

Applications to Flow Problems

Michael Brilleslyper (text), Jim Rolf (applets)

3.1. Introduction

This chapter grew out of a series of lectures prepared for an undergraduate mathematical physics course. At the time, the goal was to show the students some applications of complex function theory that connected to familiar topics from calculus and physics. The idea to look at the two dimensional flows of ideal fluids was a natural fit. Many common ideas from vector calculus are used in the development of the subject and there are numerous applications of the methods that are developed. Modeling ideal fluid flow is a standard application of *conformal mappings* and is readily found in most undergraduate complex analysis texts (see [**3**] or [**5**]). However, in preparing the notes it became apparent that there was a need for a unified treatment that included a variety of applications and extensions such as including sources and sinks in the flow, or accounting for the role played by sources or sinks at infinity. Additionally, the emphasis in many texts is on the analytic aspects of the subject and not on the geometric or visual aspects of the flows. This chapter combines all this material and more. It is self-contained and relies only on basic results from vector calculus and a standard first course in complex variables. To permit discovery and experimentation, we created the easy-to-use applet *FlowTool* that can be accessed online at http://www.jimrolf.com/explorationsInComplexVariables/chapter3.html:

- *FlowTool* plots the streamlines for the flow of an ideal fluid. It permits the user to select the number and location of sources or sinks on the boundary or in the interior of the region and then allows the strength to be dynamically varied. The applet shows steady state fluid flow in four preset regions: the entire plane, the half plane, the quadrant, and the strip.

We note that the applet is limited to certain pre-set regions and particular types of sources. To account for the wide variety of problems that may be encountered in practice, we also make use of powerful computer algebra systems such as *Mathematica*.

Vector fields arise naturally in many applications. They are used to model physical phenomena such as the velocity field of a fluid flowing in a region or the electric force field generated by a collection of charges. The geometry of the region, along with any

sources or sinks in the field, determines the nature of the resulting vector field. Finding descriptions of these fields is the focus of this chapter. It is also of interest to find the integral curves of such a vector field. These are curves that are everywhere tangent to the vector field. Integral curves have different names depending on the context including *flow lines*, *stream lines*, or *lines of force.*

In this chapter we study physical situations in which complex function theory can be used to help solve the problem of finding particular instances of such vector fields and their integral curves. We focus mainly on using techniques of complex analysis to solve flow problems. To motivate our discussion and to provide a road map for where the chapter is headed, we start with an example. We omit many of the details here hoping the reader will be intrigued enough to press on with the material and learn how to solve this type of problem.

EXAMPLE 3.1. Imagine an infinitely long, very shallow, channel in which fluid flows. Because the channel is very shallow we assume this is a two dimensional flow (real flows are three dimensional, but we assume the flow is identical on all parallel planes). Orient the channel on the complex plane so that one edge runs along the real axis and the other along the horizontal line $\text{Im } z = \pi$ (thus our channel has width π). Now we assume that fluid is being pumped into the channel and drained from the channel at various points along the edges. For the time being we assume that all pumps and drains operate at the same constant rate. Our goal is to describe the velocity of the fluid in the channel. We assume that the flow is in a *steady state,* meaning the velocity at a point of the domain does not change with time. For this scenario, let the fluid be pumped into the channel at equal rates at $z = 0$ and $z = 2$, and suppose fluid is drained from the channel at the same rate at the point $z = -2 + \pi i$. Note that there is more fluid being pumped into the channel than there is being drained from the channel. Figure 3.1 illustrates the situation. The curves shown are the integral curves of the underlying vector field and they represent the path that a drop of dye would follow if placed into the flow. A critical assumption in this model is that we are dealing with an *ideal fluid.* This means that the fluid is incompressible, non viscous, and there is no loss of energy due to friction between the walls of the channel and the fluid. Such fluids do not really exist, but they frequently provide good approximations to real physical situations and they have certain mathematical properties that lend themselves to analysis and modeling.

The plot in figure3.1 was generated using the computer algebra system *Mathematica.* Software packages such as *Mathematica, Maple,* and *Matlab* are very powerful and can be useful in generating complicated graphics. However, they may not be easy to use or to adapt to a particular problem. In order to facilitate your understanding of this chapter we have provided the *FlowTool* applet. This applet allows you to experiment with complicated flows for a fixed number of domains. We will make use of this applet throughout the chapter and in numerous exercises. We also provide the basic syntax for generating these types of plots in *Mathematica.*
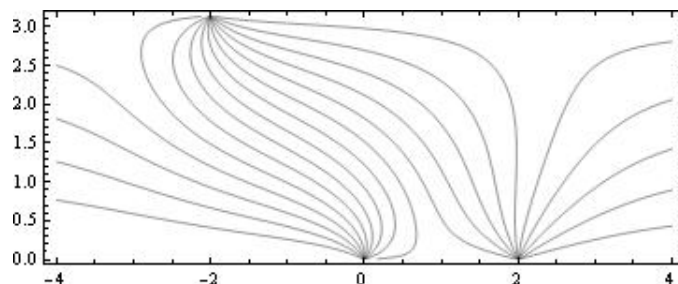
FIGURE 3.1. Flow Lines for an Ideal Fluid in a Channel with Sources and Sinks

Before moving on you should open the *FlowTool* applet and follow the instructions below to generate your own version of Figure 3.1. Open the applet and select the strip domain from the drop down menu. Uncheck the box marked Equipotential lines (we only wish to view the flow lines in this example). Use the mouse to place a source at $z = 0$, $z = 2$, and $z = -2 + \pi i$ by clicking once near each of the three locations. Notice that a dialog box with a slider opens in the right hand panel each time a source is placed on the boundary of the channel. The sliders are used to change the strength of a source. If the slider is moved to the left, the strength will eventually become negative, signifying that the source has changed to a sink. Move the slider to create a sink of strength -1 at the location $z = -2 + \pi i$. The other two locations are sources of strength 1. Figure 3.2 shows a graph similar to what you should see on the applet. You should feel free to experiment further by adding or deleting sources and using the sliders to vary the relative strengths. Observe how the flow lines run parallel to the edges of the channel, indicating that the boundary of the region acts like a frictionless barrier to the flow.

There is a lot of interesting mathematics behind the plot of the flow lines in figures 3.1 and 3.2. This chapter follows a path that develops all the key ideas needed to solve flow problems in various regions with combinations of sources and sinks along the boundary. In addition, later sections in the chapter extend the ideas and methods to more complicated settings.

Section 3.2 starts with a review of basic ideas from vector calculus including the important notions of the divergence and curl of a vector field and the definition of a harmonic function. This material should be familiar to most students.

Section 3.3 provides a review of the main results needed from a standard course in complex variables. This section also makes the critical connection between planar vector fields and complex functions. We also introduce the Polya field and give a mathematical description of sources and sinks.

Section 3.4 introduces the complex potential function which is the main tool for constructing ideal fluid flows. We make several key connections between the complex potential function and the underlying vector field. This material leads directly to the
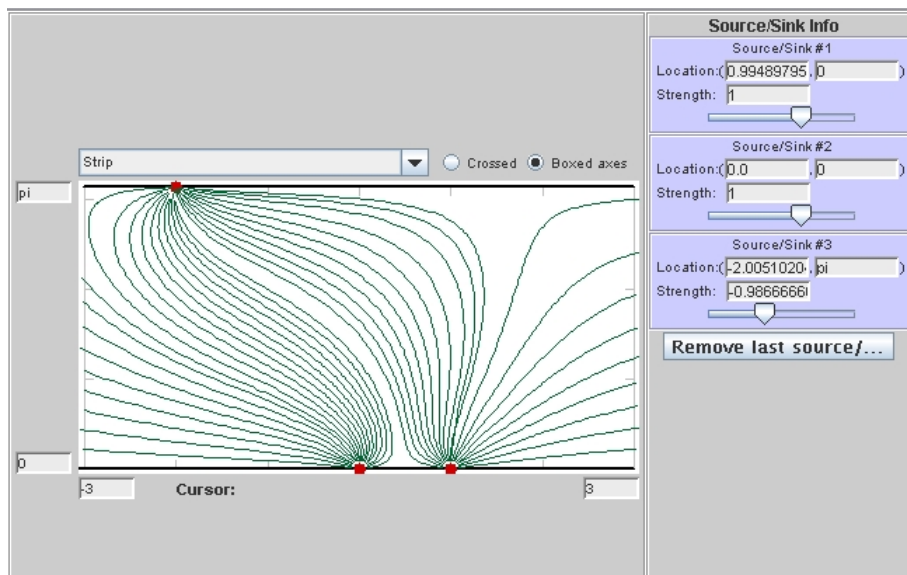
FIGURE 3.2. Flow Lines in a Channel with Sources and Sinks generated with FlowTool Applet.

construction of uniform flows in various regions in Section 3.5. We then continue the development in Section 3.6 by allowing sources and sinks along the boundary of the region. We then take a careful look at the uniform flow in a channel in Section 3.7. This development sheds light on the important role played by sources or sinks at infinity. Section 3.8 puts all the previous material together by discussing flows in any region with various combinations of sources or sinks along the boundary.

The final five sections focus on various applications or extensions of the material. These topics include interval sources, sources or sinks in the interior of regions, flows inside disks, dipoles, and steady state temperature problems.

3.2. Background and Fundamental Results

Several fundamental concepts from multivariable calculus are required for this material. We briefly review the main ideas here. The reader wishing to obtain more background regarding vector fields and their associated operations should consult any standard calculus text such as [**1**].

We represent a 2-dimensional or planar vector field in Cartesian coordinates using two real-valued functions of position: $\vec{F}(x, y) =< P(x, y), Q(x, y) >$. Simple examples include constant vector fields such as $\vec{F}(x, y) =< 3, 4 >$ and the field tangent to concentric circles about the origin given by $\vec{G}(x, y) =< -y, x >$. Other examples include slope fields for first order differential equations or the magnetic field in a plane perpendicular to a wire with a current flowing through it. Vector fields are represented graphically in the plane by drawing arrows indicating the direction of the

195

field at selected points. The magnitude of the field at a point is given by
$$|\vec{F}| = \sqrt{P(x,\,y)^2 + Q(x,\,y)^2}$$
and the direction θ satisfies
$$\tan\theta = \frac{Q(x,\,y)}{P(x,\,y)}$$
for $P(x,\,y) \neq 0$. If $Q(x,\,y) \neq 0$ and $P(x,\,y) = 0$, then the direction is $\pm\frac{\pi}{2}$ at $(x,\,y)$, the choice being determined by the sign of $Q(x,\,y)$. The vector $< 0, 0 >$ has no defined direction. It is usually impossible to draw vectors with their true magnitude, so graphical representations frequently scale the length of the vectors down to a more manageable size. Figure 3.3 shows a constant vector field, while Figure 3.4 shows the magnetic field generated by a current running perpendicular to the plane through the origin. We also show an integral curve of the field.
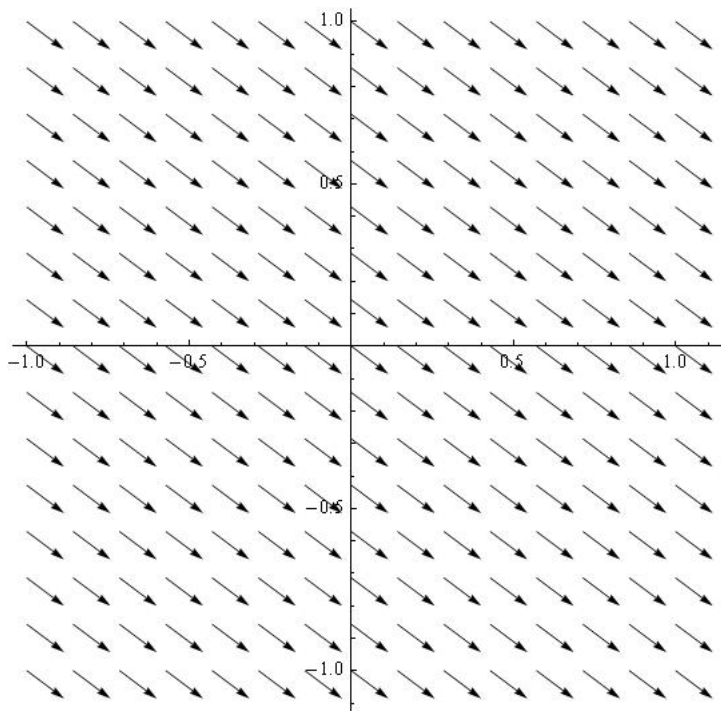

FIGURE 3.3. The constant vector field given by $\vec{F}(x,\,y) = < 4, -3 >$.

The *curl* and *divergence* of $\vec{F}$ are defined respectively in terms of the vector differential operator $\nabla = < \partial_x,\, \partial_y >$ as follows:

(21)
$$\nabla \times \vec{F} = (Q_x - P_y)\hat{k} \text{ (Curl)}$$

(22)
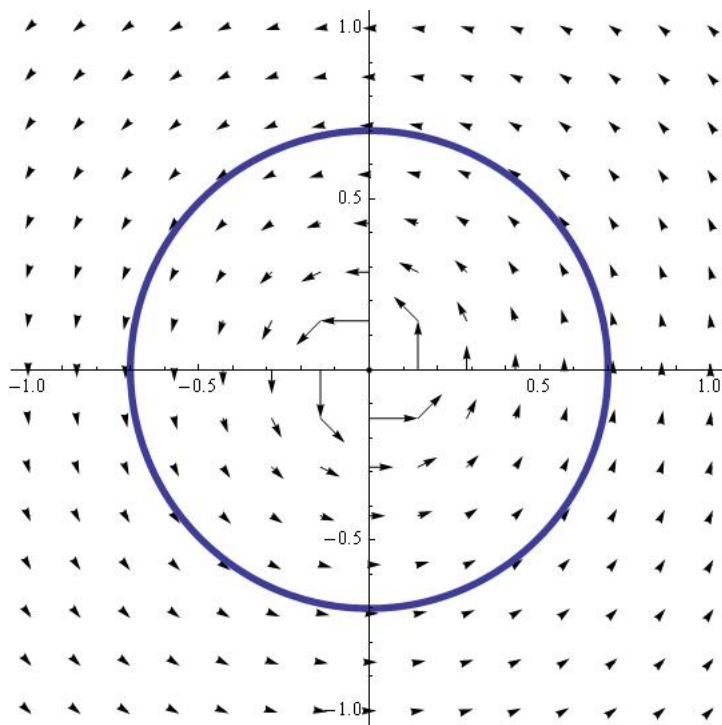$$\nabla \cdot \vec{F} = P_x + Q_y \text{ (Divergence)}$$

FIGURE 3.4. Magnetic field generated by a current in a wire. The circle represents an integral curve of the field.

The curl is really defined on 3-dimensional vector fields and yields another vector field orthogonal to the original field. However, in the case of 2-dimensional planar fields we simply assume the $\hat{k}$-component of the vector field $\vec{F}$ is zero, implying that the curl will always point in the $\hat{k}$-direction. Thus, it is sufficient to simply compute the scalar component of $\nabla \times \vec{F}$, namely $Q_x - P_y$.

We say that a vector field is *irrotational* at a point if the curl is zero. Similarly, we say the vector field is *incompressible* at a point if the divergence is zero. The standard physical description of a fluid with zero curl is that an infinitesimally small paddle wheel placed horizontally into the flow would not rotate. That is the flow contains no vortices. The physical description of the divergence being zero is the idea of the fluid being incompressible. One consequence of this is that the amount of fluid entering some region must be equal to the amount of fluid leaving the same region.

In this chapter we study vector fields that are *both* irrotational and incompressible throughout some region. Whereas this may seem restrictive, it actually encompasses a wide range of physically meaningful phenomena such as the electric field in a region free from charges or the velocity field of an ideal fluid flowing in a region. In addition, the gravity force field generated by a collection of masses also falls in this category.

197

The requirement that both the curl and the divergence be zero imposes conditions on the component functions of the vector field $\vec{F}(x,\,y) =< P(x,\,y),\,Q(x,\,y) >$. From Equations (21) and (22) we obtain a pair of partial differential equations relating the components of $\vec{F}$:

$$(23) \qquad\qquad\qquad \begin{aligned} P_x &= -Q_y \\ P_y &= Q_x \end{aligned}$$

Note the similarity of these equations to the Cauchy-Riemann equations satisfied by analytic functions. Indeed, if $f(z) = u(x,\,y) + iv(x,\,y)$ is an anlytic function, then the real and imaginary parts satisfy the Cauchy-Riemann equations:

$$(24) \qquad\qquad\qquad \begin{aligned} u_x &= v_y \\ u_y &= -v_x \end{aligned}$$

EXERCISE 3.2. Determine if the following vector fields are incompressible and/or irrotational:

(1) $\vec{F}(x, y) =< ax + by,\, cx + dy >$, where $a, b, c,$ and d are real
(2) $\vec{G}(x, y) =< e^x \cos y,\, e^x \sin y >$

Try it out!

Recall that given a vector field $\vec{F}(x,\,y)$ and a path C,, the line integral of $\vec{F}$ along C is denoted by $\int_C \vec{F} \cdot d\vec{R}$ and measures the work done by the field on a particle traversing the path C. The line integral is computed by finding a parameterization of C, given by $r(t) = (x(t),\,y(t))$, where $a \leq t \leq b$, and computing $\int_a^b \vec{F}(r(t)) \cdot r'(t)\,dt$.

A vector field $\vec{F}$ is said to be *conservative* in a simply connected domain D if the line integral $\int_C \vec{F} \cdot d\vec{R}$ between any two fixed points in D is independent of the path C chosen for the integration (provided the path lies entirely in D). There are numerous equivalent conditions that guarantee a field is conservative. See [**1**] for a complete discussion.

THEOREM 3.3. Let $\vec{F}$ be a vector field with component functions that are continuous and have continuous first order partial derivativesthroughout a simply connected region D. The following statements are equivalent:

(1) $\vec{F}$ is conservative.
(2) There exists a differentiable *potential* function ϕ such that $\nabla \phi = \vec{F}$, where $\nabla \phi =< \phi_x,\,\phi_y > .$
(3) $\int_C \vec{F} \cdot d\vec{R} = 0$ for every closed loop C in D.
(4) $\nabla \times \vec{F} = 0$.

We make extensive use of the result that a curl-free vector field has a potential function. The condition $\nabla\phi = \vec{F}$ implies the system of partial differential equations

(25)
$$\phi_x = P(x, y)$$
$$\phi_y = Q(x, y)$$

be satisfied. These equations allow us to construct potential functions through partial integration. The function ϕ is unique up to an additive constant (why?) and is called the *real potential function* of $\vec{F}$.

Another key definition that plays an important role in our work is that of a *harmonic function*. These functions are solutions of Laplace's equation. Harmonic functions play a critical role in much of applied mathematics. They arise frequently as steady-state solutions to various physical problems. Harmonic functions and their properties are closely tied to the theory of analytic functions in complex analysis. In solving flow problems, harmonic functions play a central role.

DEFINITION 3.4. Let $u : D \subseteq \mathbb{R}^2 \to \mathbb{R}$ have continuous 2nd order partial derivatives. If u satisfies Laplace's Equation $\Delta u = \nabla \cdot \nabla u = u_{xx} + u_{yy} = 0$, then u is said to be a harmonic function in D.

The following exercises explore the mechanics of working with vector fields, potential functions, and harmonic functions. In particular, Exercise 3.6 shows that the potential function for an irrotational and incompressible field must be harmonic.

EXERCISE 3.5. Let $\vec{F}(x, y) =< x^3 - 3xy^2, y^3 - 3x^2y >$. Compute both the curl and divergence of $\vec{F}$. If $\vec{F}$ is conservative, then find a potential function for it. If $\vec{F}$ is also incompressible, then show that the potential function is harmonic. (Optional) Use a program such as Mathematica or Matlab (or some other utility) to graph both $\vec{F}$ and several level curves of the potential function. What key geometrical observation connects the direction of $\vec{F}$ with the tangents to the level curves of the potential function? ***Try it out!***

EXERCISE 3.6. Suppose that $\vec{F}$ is both irrotational and incompressible. Let ϕ be a real potential function of $\vec{F}$. Show that ϕ is a *harmonic function*. ***Try it out!***

The result of Exercise 3.6 is particularly important. Note that it is the irrotational feature of the vector field that implies the existence of the potential function and the incompressibility then implies the potential function is harmonic.

EXERCISE 3.7. Consider an attracting force at the origin of the xy-plane whose magnitude at a point is inversely proportional to the square of the distance from the origin to the point. Determine a formula for the vector field representing this force field and determine if the field is both irrotational and incompressible in any region not containing the origin. If possible, find a potential function for the vector field. Finally, change the field by assuming the force of attraction at a point is inversely proportional

to the distance to the point (as opposed to the square of the distance). Now compute the curl and divergence of this field. How do your answers compare? ***Try it out!***

3.3. Complex Functions and Vector Fields

In this section we make several basic connections between vector fields and complex functions. A vector field being both irrotational and incompressible is closely related to the concept of an analytic function.

Any complex function of the complex variable $z = x + iy$ can be expressed in terms of its real and imaginary parts as $f(z) = u(x, y) + iv(x, y)$, where u and v are real-valued functions of x and y. This is accomplished by setting $z = x + iy$ and separating into real and imaginary parts. Recall that the function f is said to be *analytic* at $z = z_0$ if the complex derivative $f'(z)$ exists for every z in a neighborhood of z_0. An immediate and far-reaching consequence of this definition is that the functions u and v must satisfy the Cauchy-Riemann equations:

(26)
$$u_x = v_y$$
$$u_y = -v_x$$

We recall a basic result about analytic functions:

THEOREM 3.8. Let $f(z) = u(x, y) + iv(x, y)$ be analytic in a domain D, then the functions u and v are both harmonic in D.

This result is an easy consequence of the Cauchy-Riemann equations and the reader is strongly encouraged to prove the result in the following exercise.

EXERCISE 3.9. Prove Theorem 3.8. ***Try it out!***

An equally important result is that every harmonic function on a simply connected domain is the real part of an analytic function. For a complete discussion, see [**5**].

THEOREM 3.10. Let $u(x, y)$ be a harmonic function defined on a simply connected domain D. Then, there exists another harmonic function $v(x, y)$ defined on D called the *harmonic conjugate* of u, such that $f(z) = u(x, y) + iv(x, y)$ is analytic on D. The harmonic conjugate is not unique.

EXERCISE 3.11. Find a harmonic conjugate for $u(x, y) = 3x^2 - 2y - 3y^2$. ***Try it out!***

EXERCISE 3.12. Let $u(x, y) = \frac{1}{2}\mathrm{Log}(x^2 + y^2)$. Show that u is harmonic on the punctured plane. Next, omit any infinite ray emenating from the origin and construct a harmonic conjugate for u on the plane minus the ray. ***Try it out!***

Next we establish a correspondence between the set of complex functions and the set of planar vector fields. Indeed, since both complex functions and planar vector

fields can be represented by a pair of real-valued functions, any complex function may be thought of as a vector field and visa-versa. We note the following correspondence:

(27) $\qquad V(z) = P(x,y) + iQ(x,y) \iff \vec{F}(x,y) = <P(x,y),\ Q(x,y)>$

Initially the correspondence is only a renaming of one object into another with no obvious value. However, looking at complex functions as vector fields sometimes yields insights that are not apparent in the more traditional view of complex functions as mapping regions from one copy of the complex plane to another copy. When plotting the vector field of a complex function, the magnitude of the vector field at z is $|V(z)|$ and the direction is $\arg(V(z))$. For example, Figure 3.5 shows the vector field corresponding to the complex function $V(z) = z^2$
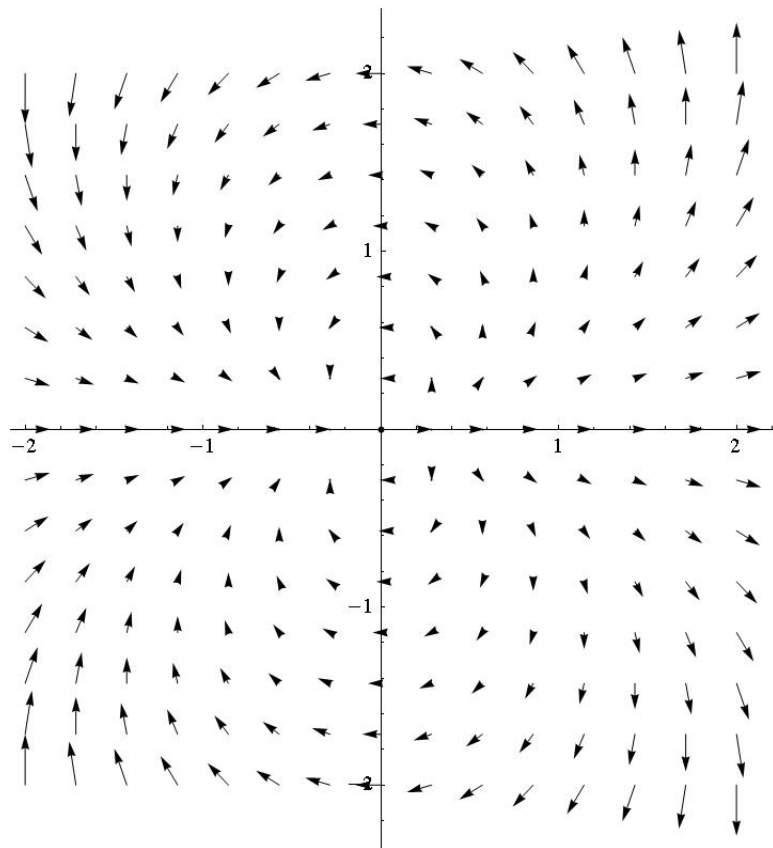

FIGURE 3.5. Vector field representation of the function $V(z) = z^2$

For the remainder of this chapter we will use the notation $V(z)$ to refer to both complex functions and their corresponding vector field representation.

The correspondence given by (27), though particularly simple, is not the most useful vector field representation of a complex function. For reasons that will become clear, we instead associate a complex function with its *Polya vector field*.

DEFINITION 3.13. Let $V(z) = u(x, y) + i\, v(x, y)$ be a complex function. The vector field given by the conjugate $\overline{V(z)} = u(x, y) - i\, v(x, y)$ is called the *Polya vector field* of $V(z)$.

It is important to note then when plotting $\overline{V(z)}$, that we still attach the arrow to the point z, not $\bar{z}$.

The Polya vector field of $V(z) = z^2$ is $\overline{V(z)} = \bar{z}^2$ and is shown in Figure 3.6. Contrast this field with the one shown in Figure 3.5. We will see the value of the Polya field representation when we look at the vector field corresponding to an analytic function.
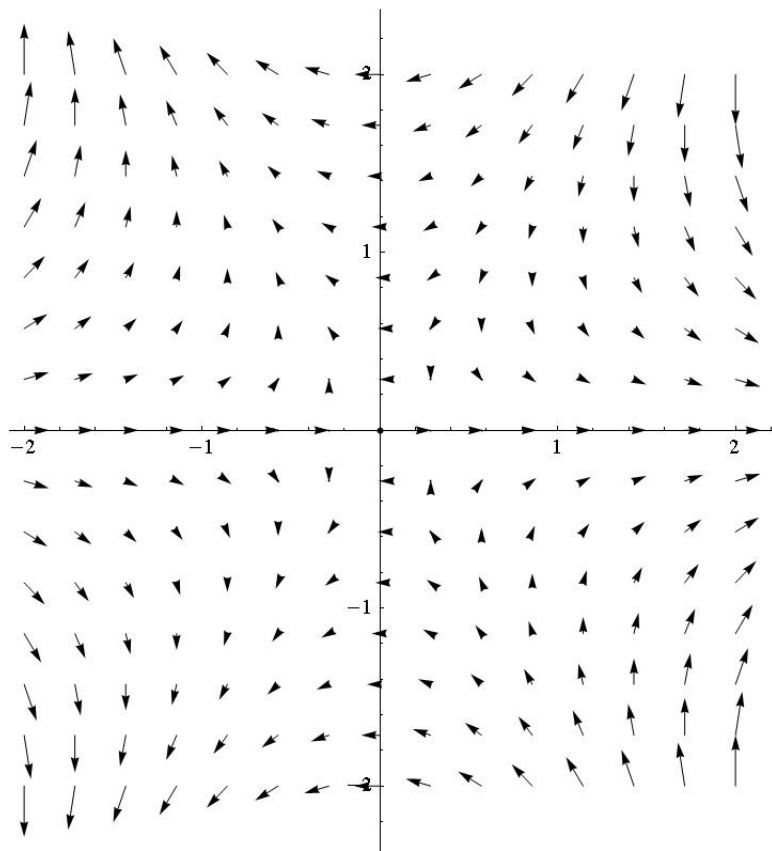

FIGURE 3.6. The irrotational, incompressible Polya field $\overline{V(z)} = \bar{z}^2$

The following theorem shows that if a complex function is analytic, then its Polya field is both irrotational and incompressible. The proof is a straightforward calculation involving the Cauchy-Riemann equations and is left as an exercise.

THEOREM 3.14. Let $V(z)$ be a complex function which is twice differentiable on a domain D. $V(z)$ is analytic on D if and only if its Polya vector field $\overline{V(z)}$ is irrotational and incompressible on D.

EXERCISE 3.15. Prove Theorem 3.14. ***Try it out!***

The previous Theorem 3.14 establishes a bijection between the set of irrotational, incompressible planar vector fields and the set of analytic functions. It is precisely this connection that allows us to use the techniques of complex analysis to solve flow problems.

EXERCISE 3.16. Find the component functions of each irrotational and incompressible vector field corresponding to the following analytic functions. Verify directly that the resulting vector fields have both curl and divergence equal to zero.

(1) $V(z) = z^2$
(2) $V(z) = e^z$
(3) $V(z) = \frac{1}{z}$

Try it out!

EXERCISE 3.17. Verify that the following vector field is irrotational and incompressible and then find the corresponding analytic function.

$$\overline{V(z)} = (x^2 - y^2 - 2x) + i(-2xy + 2y).$$

Try it out!

The third example in Exercise 3.16 deserves some additional comments. The function $V(z) = 1/z$ is not defined at $z = 0$, but is analytic at every point in any neighborhood of $z = 0$. Such a point is called a *singular point* or a *singularity* of the function. The corresponding Polya vector field $\overline{V(z)}$ is also undefined at $z = 0$. However, graphing the vector field reveals that the field lines emanate radially from the origin; see Figure 3.7. We call such a point a *source*. Conversely, if the field lines went towards the origin, we would refer to the origin as a *sink*. This leads to the following definition.

DEFINITION 3.18. Let $V(z)$ be an analytic function except at $z = 0$. If $V(z)$ has a pole of order 1 at $z = 0$, then zero is either a source or sink of the corresponding Polya field $\overline{V(z)}$. The sign of the coefficient of the z^{-1}-term of the Laurent expansion of $V(z)$ about $z = 0$ determines whether the point is a source or a sink.

Sources and sinks have physical interpretations in terms of some quantity being created or destroyed. In the case of flow problems we interpret a source as a location where fluid is pumped into the region, whereas a sink acts like a drain, removing fluid from the region. In the case of an electric field, sources and sinks correspond to positive and negative charges respectively. For gravitational fields, point masses correspond to sinks of various strengths (depending on the mass). Allowing vector fields to have sources and sinks greatly extends the examples to which we may apply our methods.
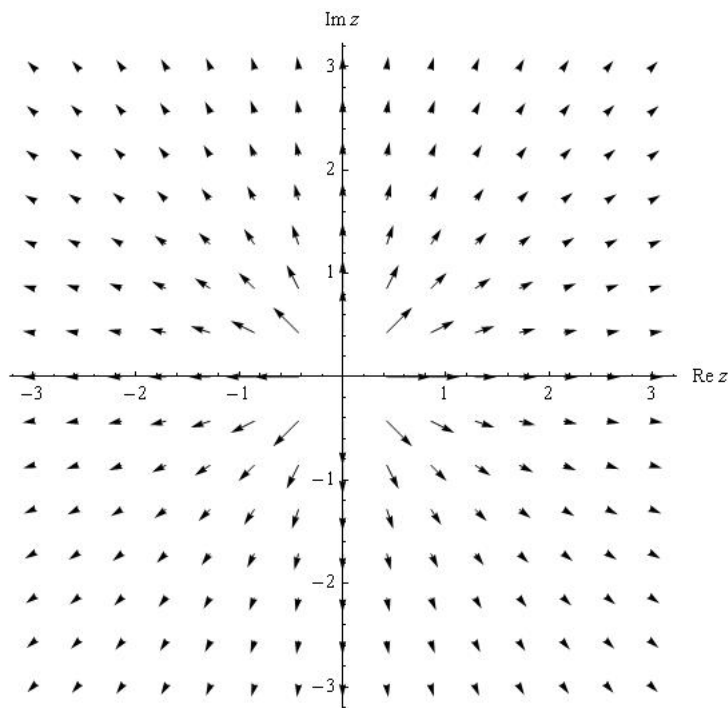
FIGURE 3.7. Polya vector field of the analytic function $f(z) = \frac{1}{z}$

EXERCISE 3.19. Let $V(z) = 1/z$. Show that the Polya field $\overline{V(z)}$ corresponds to the familiar inverse square law from physics. That is, show that the vector field emanates radially from the origin and that the magnitude of the field at a point in the plane is inversely proportional to the distance from the point to the origin. ***Try it out!***

If we desire to have a vector field with a source at some point $z = z_1$ rather than $z = 0$ we only need to translate the corresponding analytic function to obtain $V(z) = 1/(z - z_1)$. A sink is obtained by changing the sign of the function. The relative strength of a source or sink is changed by multiplying $V(z)$ by a real scalar. Finally, to account for more than one source or sink we use the *Principle of Superposition.* This principle states that if several components act to generate a vector field, then the field is obtained as the sum of the fields generated from each component separately. This is nothing more than the linearity of vector addition, but the consequences are far-reaching: complicated systems may be analyzed by studying the simpler components from which they are generated. For example, Figure 3.8 shows the field generated by a sources at $z = 0$ and $z = 1$ and sink of twice the strength of either source located at $z = 1 + 2i$. The analytic function corresponding to this field is given by

(28)
$$V(z) = \frac{1}{z} + \frac{1}{z - 1} - \frac{2}{z - (1 + 2i)}.$$

204
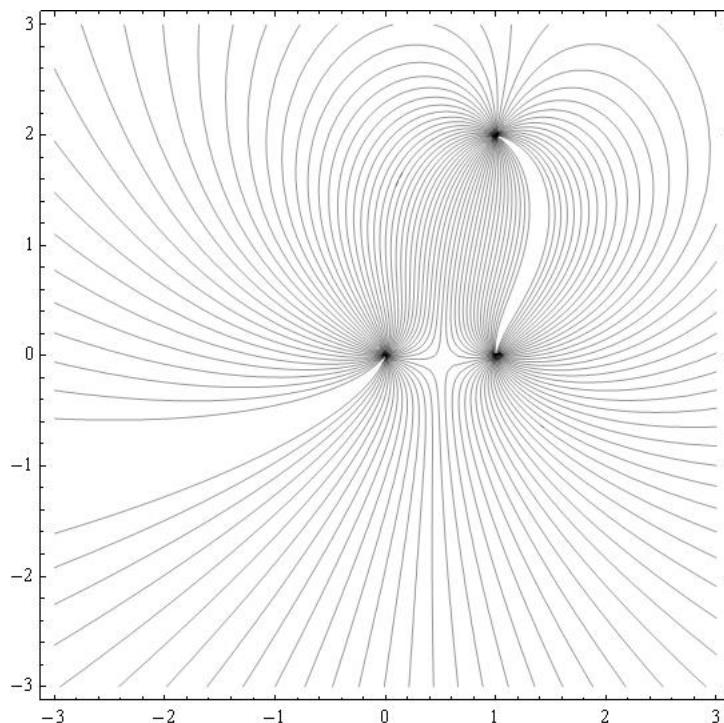
FIGURE 3.8. Vector field with two point sources and one sink. Integral curves are shown for clarity rather than arrows.

From a qualitative standpoint only the relative strengths of any sources or sinks matters in the analysis of the situation. Nevertheless, we compute the actual strength of a source or sink for completeness. We define the strength of a source to be the flux of the vector field over any simple closed loop C that encloses the source, but no other singularities of the function.

DEFINITION 3.20. Let $\vec{F}(x, y) =< P(x, y), Q(x, y) >$ be an irrotational, incompressible vector field with a source at $z = z_0$. The strength of the source is defined to be

$$(29) \qquad \int_C P\, dy - Q\, dx,$$

where C is a loop traversed in a counter clockwise direction containing z_0 and no other singularities of the field. If the strength is negative, then the source is actually a sink.

Equation (29) is the familiar formula for computing the flux of a vector field.

There is an interesting relationship between the flux and circulation of a vector field and the complex integral of the associated Polya field. Indeed, the strength of the source can be determined directly from the complex integral $\int_C \overline{V(z)}\, dz$ as shown in the following important exercise.

205

EXERCISE 3.21. Let $V(z) = u(x, y) + i\,v(x, y)$ be a vector field Let C be a simple loop not passing through any singular points of the field. Expand the integral $\int_C \overline{V(z)}\,dz$ in real and imaginary parts. (Hint: let $dz = dx + i\,dy$.) Show that

$$\int_C \overline{V(z)}\,dz = \int_C u\,dx + v\,dy + i\int_C u\,dy - v\,dx.$$

This computation shows that the flux of $V(z)$ outward across C is obtained from the imaginary part of the integral $\int_C \overline{V(z)}\,dz$. Note as a consequence that the circulation of $V(z)$ is given by the real part of the integral. ***Try it out!***

The above exercise gives a useful characterization and interpretation of complex integration in terms of the familiar concepts of work and flux. An immediate consequence is that if $\overline{V(z)}$ is irrotational and incompressible both inside and on C, then both the work along C and the flux across C must both be zero. This implies that the function $V(z)$ is analytic (as was already established). An exceptional discussion of Polya vector fields and their relationship to complex integration can be found in Needham [**2**].

In particular, if $V(z)$ is analytic everywhere except for a finite number of simple poles located at $z_i, i = 1, 2, \ldots, n$, then the strength of the source or sink of $\overline{V(z)}$ located at z_k is given by the imaginary part of $\int_C V(z)\,dz$, where C is a simple closed loop traversed in the counter clockwise direction such that z_k is contained in the interior of C and none of the other z_i are inside C.

Let C be a simple closed curve containing z_k, but no other z_i for $i \neq k$. The function $V(z)$ can be represented in the form

$$\frac{g(z)}{z - z_k},$$

where $g(z)$ is analytic inside and on C. By Cauchy's Integral Formula, $\int_C V(z)\,dz = 2\pi i g(z_k)$.

EXERCISE 3.22. In constructing vector fields with sources or sinks, the most typical case is where $V(z)$ has the form

$$V(z) = \sum_{j=1}^{n} \frac{a_j}{z - z_j},$$

where $a_j \in \mathbb{R}$. Show that for the corresponding Polya field that the strength of the source or sink located at $z = z_k$ is simply $2\pi a_k$. ***Try it out!***

EXERCISE 3.23. Let $a \in \mathbb{R}$. Show that the vector field $\overline{V(z)} = a/(\bar{z} - \overline{z_0})$ has a source of strength $2\pi a$ at $z = z_0$ by directly computing the imaginary part of $\int_C V(z)\,dz$, where C is a circle centered at z_0. Do this by using a parameterization of C as opposed to Cauchy's Integral Formula. ***Try it out!***

EXERCISE 3.24. Consider the Polya field of the function

$$V(z) = \frac{\sin z}{z^2}.$$

This vector field has a simple pole at the origin. What is the strength of the source? **Try it out!**

3.4. Complex Potential Functions

At this point, we have established a one-to-one correspondence via complex conjugation between irrotational, incompressible vector fields in the plane with a finite collections of point sources and sinks and complex functions which are analytic except at finitely many points where they have poles of order one.

Having established this correspondence we now proceed to construct the *complex potential function* of the underlying vector field. Let $V(z)$ be analytic on a simply connected domain so that $\overline{V(z)}$ is irrotational and incompressible, hence $\overline{V(z)}$ has a potential function ϕ that is harmonic. Thus, by Theorem 3.10 we know there exists another harmonic function ψ such that the complex function $\Omega(z) = \phi(x, y) + i\,\psi(x, y)$ is analytic. Note that ψ is only determined up to an additive constant. The function Ω is called a complex potential function of $\overline{V(z)}$. The function Ω tells us a great deal about the underlying vector field.

DEFINITION 3.25. Given an irrotational, incompressible vector field $\overline{V(z)}$ on a simply connected domain, a *complex potential function* of $\overline{V(z)}$ is the *analytic* function $\Omega(z) = \phi(x, y) + i\,\psi(x, y)$, where ϕ is the real potential function of $\overline{V(z)}$ and ψ is a harmonic conjugate of ϕ.

It should be noted that a globally defined complex potential function may not exist. For example, $\overline{V(z)} = 1/\bar{z}$ is irrotational and incompressible on the punctured plane, but a complex potential for this field involves choosing a branch of the complex logarithm function which is not analytic on any neighborhood containing the origin.

As we already know, the level curves of ϕ (the real potential function) are orthogonal to the direction of $\overline{V(z)}$. Thus these curves form the *equipotential* lines of the field. In the case of a velocity field of an ideal fluid, these curves represent points where the velocity is constant. Whereas in the case of an electric field, the level curves of ϕ represent curves of constant electrostatic potential. So, what do the level curves of ψ represent? The answer is found in the following standard result from complex variables. We restate the result here.

THEOREM 3.26. Let $\Omega(z) = \phi(x, y) + i\,\psi(x, y)$ be analytic at $z_0 = x_0 + i\,y_0$ and suppose that $\Omega'(z_0) \neq 0$. Then the tangents to the level curves of ϕ and ψ are orthogonal at the point (x_0, y_0).

For a proof see Zill and Shanahan [**3**].

Theorem 3.26 implies that the level curves of ψ are parallel to the underlying vector field $\overline{V(z)}$. Thus, these level curves are the integral curves of the field. They are, in fact, exactly the curves that were sketched by the *FlowTool* applet in the opening example to this chapter. The function ψ is often referred to as the *stream function* in the case of a flow problem. The level curves of ψ are called the *stream lines.*

The vector field can be obtained from the complex potential function according to the following exercise.

EXERCISE 3.27. Let $\overline{V(z)}$ be an irrotational and incompressible vector field with analytic complex potential function $\Omega(z)$. Show that $\Omega'(z) = V(z)$. (Hint: Write $\Omega(z)$ in terms of its real and imaginary parts and differentiate using the Cauchy-Riemann equations.) ***Try it out!***

An equivalent statement to the preceding exercise is that

(30) $$\Omega(z) = \int V(z)\, dz.$$

This formulation is useful in the following exercise:

EXERCISE 3.28. Let $\overline{V(z)} = 3\bar{z}^2 - 4\bar{z}$ be an irrotational, incompressible vector field.
  (1) Determine a complex potential function on $\mathbb{C}$ using Equation (30).
  (2) Find the real and imaginary parts of the complex potential function to easily find the real potential function and the stream function.
Try it out!

We end this section with an exercise that illustrates a mapping property of the complex potential function.

EXERCISE 3.29. Let Ω be the complex potential for an irrotational, incompressible vector field. Show that when viewed as a mapping of the z-plane to the w-plane, Ω maps equipotential curves to vertical lines and flow lines to horizontal lines. ***Try it out!***

3.5. Uniform Flows in the Plane and other Regions

We now begin the process of building up examples that allow us to solve a wide variety of problems in different regions of the plane. We start by considering very simple flows in the entire plane. Consider the function $\Omega(z) = z$. Thinking of Ω as a complex potential function, the natural question is to determine the underlying vector field. The following exercise asks you to determine this field.

EXERCISE 3.30. Show that $\Omega(z) = z$ is the complex potential function for a uniform flow to the right. Be sure to explicitly compute the vector field $V(z)$. See Exercise 3.27. Also show directly that the level curves of the imaginary part of Ω are stream lines for the flow. ***Try it out!***

EXERCISE 3.31. How does the answer to Exercise 3.30 change if the complex potential function is $\Omega(z) = (2+3i)z$? Show that the stream lines are given by the family of linear equations $3x + 2y = c$. See Figure 3.9. ***Try it out!***
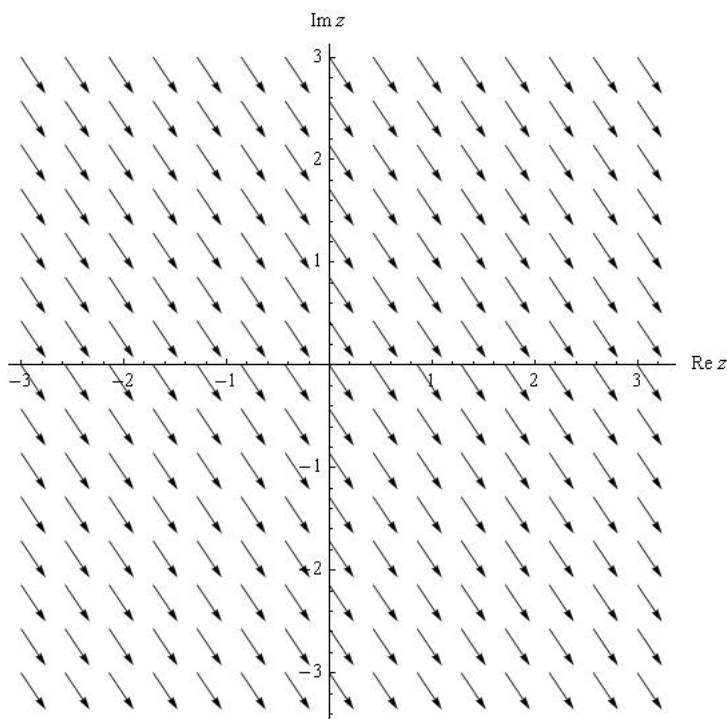
FIGURE 3.9. Uniform Flow with complex potential $\Omega(z) = (2 + 3i)z$

Our goal is to solve flow problems in various regions such as sectors, strips, and disks. However, to this point we have only studied flows in the entire complex plane. The connection lies in the theory of conformal mappings. The subject of conformal mappings is central to any course in complex analysis. In brief, a conformal mapping is a one-to-one complex function mapping some region in the z-plane to some region in the w-plane in such a way that angles and orientation are preserved. A key result is that an analytic function is conformal at all points where its derivative does not equal zero. The critical property of conformal mappings for our development is the following theorem:

THEOREM 3.32. Let $\Omega(w)$ be analytic on a domain $D' \in \mathbb{C}$, and let $f : D \to D'$ be conformal for some domain $D \in \mathbb{C}$. Then the composition $\widetilde{\Omega}(z) = (\Omega \circ f)(z)$ is analytic on the domain D.

For a proof see Zill and Shanahan [**3**].

The significance of Theorem 3.32 is that it allows a given problem to be translated to a simpler domain, solved, and then translated back to the original domain. Unfortunately, it is usually impossible to conformally map a given region to the entire complex plane. Instead, we make use of a deep and powerful theorem from complex analysis.

THEOREM 3.33 (Riemann Mapping Theorem). Every simply connected domain in the complex plane, with the exception of the entire plane, can be conformally mapped to the upper half plane, $\mathbb{H} = \{z \mid \operatorname{Im} z > 0\}$.

The Riemann Mapping Theorem gives no indication of how to find the needed conformal mapping, just that one exists. Nevertheless, there are extensive tables of conformal mappings that give specific instances of the needed mappings for a large collection regions. A good table of conformal mappings can be found in [**3**]. Despite being only an existence result, Theorem 3.33 indicates that it is valuable to know how to solve problems on the upper half plane $\mathbb{H} = \{z \mid \operatorname{Im} z > 0\}$ in $\mathbb{C}$. Indeed, we will translate many different problems to equivalent problems in $\mathbb{H}$ in order to solve them.

Suppose an ideal fluid is flowing from left to right in $\mathbb{H}$. The real axis in this case acts as a boundary for the flow. That is the real axis acts as a streamline for the flow. It is clear that a constant vector field such as $\overline{V(z)} = 1$ is such a flow. It is easy to see the complex potential of this field is $\Omega(z) = z$. Thus, the identity function may be regarded as the complex potential for a uniform flow to the right in $\mathbb{H}$. Our goal is to use this fact, combined with the theory of conformal mappings to find the complex potential for flows in other regions and in regions that have sources and sinks.

Our first example deals with a uniform flow in a quadrant, often called *flow around a corner*.

EXAMPLE 3.34. Imagine if a vertical barrier were inserted along the imaginary axis into the uniform flow in $\mathbb{H}$ described above. The result is likely to look like the flow shown in Figure 3.10. The goal is to find the complex potential function for this flow. The key is Theorem 3.32. We seek a conformal mapping that maps the region shown in Figure 3.10 to $\mathbb{H}$ with the property that the boundaries of the region are mapped to the real axis. This last part is important because the boundaries of the region are always streamlines for the flow (that is the flow is parallel to the boundaries and there is no friction). If we let $h(z) = -z^2$ then this function will map the 2nd quadrant to $\mathbb{H}$. Observe that the negative real axis is mapped to the negative real axis and the imaginary axis is mapped to the positive real axis. We know that the complex potential for the uniform flow in $\mathbb{H}$ is given by $\Omega(z) = z$, so we set $\widetilde{\Omega}(z) = \Omega(h(z)) = h(z) = -z^2$. Note that $\widetilde{\Omega}(z) = -x^2 + y^2 - 2xyi = -z^2$ and that the underlying vector field is $\overline{\widetilde{\Omega}'(z)} = -2\bar{z}$. Plotting the streamlines gives Figure 3.10. The streamlines are given by the level curves of the imaginary part of $\widetilde{\Omega}$, namely $-2xy = c$ for different values of the constant c.
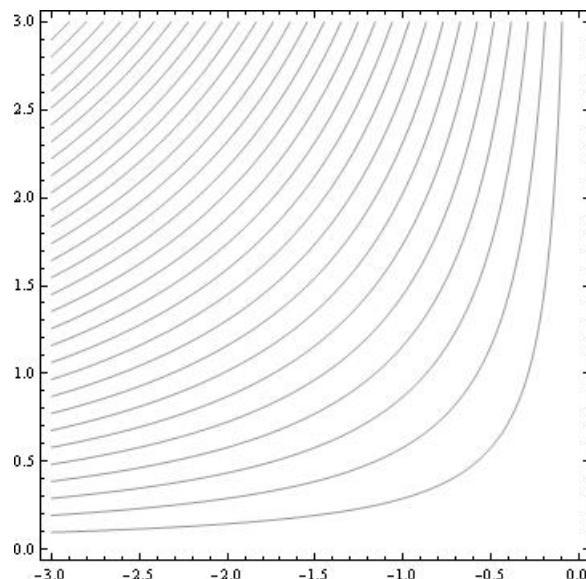
FIGURE 3.10. Uniform Flow around a Corner

EXERCISE 3.35. Let R be the region given in polar coordinates by

$$R = \{(r, \theta) \,|\, r \geq 0 \text{ and } 0 \leq \theta \leq \frac{\pi}{4}\}.$$

Find the complex potential function for a uniform (no sources or sinks) irrotational, incompressible flow in R. ***Try it out!***

The discussion above along with Exercise 3.35 demonstrates the technique of using conformal mappings to solve uniform flow problems. Essentially any conformal mapping from a region D onto $\mathbb{H}$ is the complex potential for the velocity field of an ideal fluid flowing in D.

Another example showing the power of conformal mappings is *flow around a cylinder.*

EXAMPLE 3.36. Consider the function $h(z) = z + 1/z$. This function is a conformal mapping from $\mathbb{H}$ minus the upper half of the unit disk onto $\mathbb{H}$. More precisely, h maps $\{z \,|\, |z| \geq 1 \text{ and } \operatorname{Im} z > 0\}$ onto $\mathbb{H}$. The streamlines are shown in Figure 3.11 and are given explicitly by the family of curves

$$y - \frac{y}{x^2 + y^2} = c.$$

3.6. Sources and Sinks

Having seen how to solve for uniform flows in various regions, we now turn our attention to regions that have sources or sinks at various locations on the boundary of
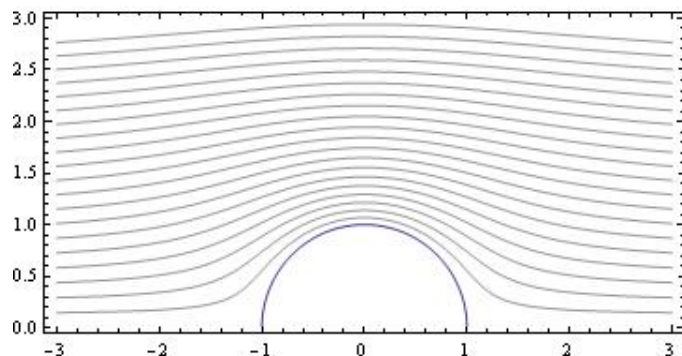
FIGURE 3.11. Streamlines around a Cylinder

the region. Recall the example of the channel from the introduction. Before tackling problems involving channels or other regions, we will first see how to solve the problem in $\mathbb{H}$. Here we will allow any number of sources or sinks of various strength along the real axis.

Let $a \in \mathbb{R}$. We have already seen that $\overline{V(z)} = 1/(\bar{z} - a)$ is an irrotational, incompressible vector field with a source at $z = a$ of strength 2π. However, it is important to note that we are only interested in the part of $V(z)$ that lies in $\mathbb{H}$. Since this is exactly half of the field, it makes sense to define the *effective strength* of the source on the boundary to be half of the strength. In this case, the effective strength of the source is π. As we have already mentioned, only the relative strengths of the sources or sinks matter with regard to the qualitative behavior of the flow. However, if we wish to balance a source on the boundary with a sink in the interior of the region, then the notion of effective strength comes into play.

Before proceeding, it is important to revisit the idea of a source in vector field. The following exploration shows how physical principles lead to our complex function representation of a source or sink.

EXPLORATION 3.37. Suppose that we have an irrotational and incompressible vector field defined on the punctured plane with a source of strength S at $z = 0$. Furthermore, assume that that the flow lines are directed radially away from the origin. We wish to determine a formula for the field. Let $\vec{F}(x, y)$ denote the vector field at all points other than the origin. We assumed the direction of $\vec{F}$ is always directly away from the origin, so the only thing to determine is the magnitude of $\vec{F}$ at each fixed distance from the origin. Since $\vec{F}$ is incompressible, the amount of fluid crossing into any closed region must equal the amount leaving the region. We apply this principle to circles centered at the origin. We interpret the source having strength S to mean that there are S units of fluid entering the region per unit time. Now consider the circle $x^2 + y^2 = R^2$. There must be S units of fluid crossing this circle per unit time. On the other hand, the amount crossing the circle must be equal to the magnitude of the

vector field multiplied by the length of the circle. Setting these two quantities equal to each other gives

$$(31) \qquad S = |\vec{F}(x,y)| \cdot 2\pi R.$$

Solve this equation for the magnitude of the field and combine the result with the fact that the direction is radial to find an explicit formula for the field.

Now show that the complex representation of the vector field is given by $\overline{V(z)} = S/2\pi\bar{z}$. Hint: $1/\bar{z} = z/|z|^2$.

Immediate consequences of exploration (3.37) are the following results that we have already established:

(1) If the source is located at $z = a$ instead of $z = 0$, then the complex representation of the vector field is given by $\overline{V(z)} = 1/(\bar{z} - \bar{a})$.
(2) If a sink is desired at $z = a$, then the function is given by $\overline{V(z)} = -1/(\bar{z} - \bar{a})$
(3) The strength of a source or sink is changed simply by multiplying $V(z)$ by a real number.

EXERCISE 3.38. Find the complex representation of the irrotational, incompressible vector field in $\mathbb{H}$ if there is a source of effective strength π at $z = 3$ and a sink of effective strength 2π at $z = -2$. Hint: The vector field at any point is the sum of the vector fields determined from each individual source or sink. Use the FlowTool applet to view the solution. ***Try it out!***

Recall that if $\overline{V(z)} = 1/(\bar{z} - \bar{a})$, then the complex potential function on $\mathbb{H}$ can be obtained from $V(z)$ by using Equation (30):

$$(32) \qquad \Omega(z) = \int V(z)\, dz = \text{Log}(z - a).$$

Thus, we see that the complex logarithm is involved whenever there are sources and sinks along the boundary. It is important to remember that the complex logarithm extends the real logarithm function. Observe that $\Omega(z)$ is analytic for $\text{Im } z > 0$, but it is not analytic in any domain containing $z = a$.

We recall that

$$\text{Log}(z - a) = \text{Log}|z - a| + i \, \text{arg}(z - a).$$

Thus, the stream lines are given by the family of equations

$$\arctan\left(\frac{y}{x - a}\right) = c.$$

This in turn is easily manipulated into the family of rays emanating from $z = a$ given by $y = (\tan c)(x - a)$. Observe that the values $c = 0$ and $c = \pi$ correspond to the two stream lines that run along the boundary of $\mathbb{H}$ in the positive and negative directions from the source at $z = a$.

The proceeding development allows us to write down the complex potential for an irrotational, incompressible vector field in $\mathbb{H}$ with any combination of sources and sinks along the boundary. We state the complete result in the following theorem.

THEOREM 3.39. Let $\overline{V(z)}$ be the irrotational, incompressible vector field in $\mathbb{H}$ generated by a finte collection of simple sources or sinks of various strengths. Assume there are sources located along the boundary of $\mathbb{H}$ at $z = a_i$, $i = 1, \ldots, n$, $a_i \in \mathbb{R}$, with corresponding effective strengths, $S_i \pi$, $i = 1, \ldots, n$ (note, if $S_j < 0$, then there is a sink at a_j). Then the complex potential of $\overline{V(z)}$ is given by

$$\Omega(z) = \sum_{j=1}^{n} S_j \mathrm{Log}(z - a_j).$$

The following exercise asks you to experiment with the *FlowTool* Applet to develop some intuition regarding flows in $\mathbb{H}$. Feel free to expand beyond the suggestions in the exercise and to experiment with a variety of situations. In each case, the result should make physical sense.

EXERCISE 3.40. Use the *FlowTool* applet to view the flow in $\mathbb{H}$ for the following situations:
  (1) A source of strength 2π at $z = -1$ and a sink of strength 2π at $z = 1$.
  (2) A source of strength 2π at $z = -1$ and a source of strength 2π at $z = 1$.
  (3) A source of strength 4π at $z = -1$ and a sink of strength 2π at $z = 1$.
  (4) Sources of strength 2π at $z = -3$ and $z = 0$ and a sink of strength 4π at $z = 2$. ***Try it out!***

You should observe in Exercise 3.40 that in some cases all the fluid emanating from the sources is taken in by the sinks, but in other cases some of the fluid either escapes to infinity or seems to emanate from infinity. If we think of infinity as a point on the boundary of $\mathbb{H}$, then we are led to the notion that there must be a sink or source at infinity.

In order to further examine the behavior at infinity, we revisit a flow in the entire plane with a finite collection of sources or sinks. Specifically, consider the vector field

$$\overline{V(z)} = \sum_{j=1}^{n} \frac{S_j}{2\pi} \frac{1}{\bar{z} - \bar{z}_j}.$$

Let C be a simple closed contour that encloses all the singularities of $V(z)$ and consider $\int_C V(z)\,dz$. By the residue theorem,

$$\mathrm{Im} \int_C V(z)\,dz = \sum_{j=1}^{n} S_j,$$

which gives the total flux across C. Now, $\int_C V(z)\,dz$ can also be viewed as a line integral "around infinity" traversed in the clockwise direction. When viewed in this

way, the value of the integral is the opposite sign of the total flux outward across C. For example, if $V(z)$ has a source of strength 3 and a sink of strength 1 at points in the plane, then there must a sink of strength 2 at infinity. Observe that if the net flux in the finite plane is zero, then the sum of the sinks and sources at infinity must be zero as well. The idea that infinity can be both a source and a sink will be explored further as we progress through the chapter. In particular, Section 3.7 investigates flow in an infinite channel, where the concept of infinity as a simultaneous source and sink plays a critical role.

3.7. Flow in a Channel

Here we return to the example of a uniform flow in an infinitely long channel described in the introduction. However, before dealing with any sources or sinks along the boundary, we first investigate the case of uniform flow to the right in the channel. Assume the channel is oriented with one edge along the real axis and the other edge along the line $\operatorname{Im} z = \pi$. Since the edges of the channel are horizontal, the flow is simply the restriction of the uniform flow to the right in $\mathbb{H}$ restricted to the channel. Hence the complex potential is $\Omega(z) = z$.

On the other hand, the function $f(z) = e^z$ is a conformal mapping of the strip onto $\mathbb{H}$ such that boundary edges of the strip are mapped to the real axis. Hence, proceeding as in the case of the quadrant (flow around a corner), and using Theorem 3.32, we see that $\widetilde{\Omega}(z) = \Omega(f(z)) = e^z$ should be the complex potential function for the uniform flow in the strip. However, note that $\operatorname{Im} \widetilde{\Omega}(z) = e^x * \sin y$. Plotting the level curves of this function gives the streamlines shown in Figure 3.12. The streamlines should simply go from left to right parallel to the edges of the strip. What went wrong?
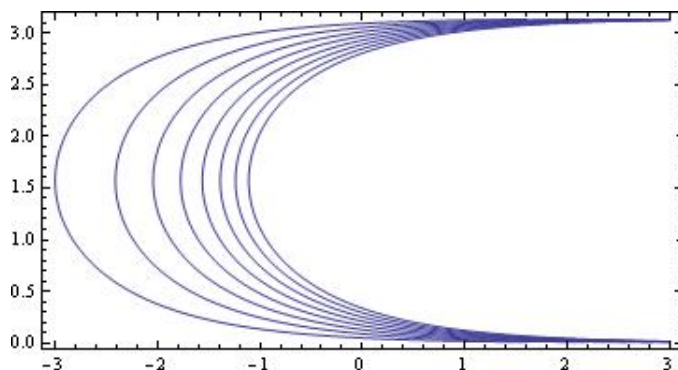

FIGURE 3.12. Incorrect uniform flow in a channel

The incorrect flow occurred from a failure to account for the behavior of infinity under the conformal mapping. Indeed, we may think of the flow in the channel as having a source at the "left end" of the strip and a sink of equal strength at the "right

end" of the strip. Now observe that points in the strip with very large negative real parts are mapped near zero. In fact, for all z with $0 \leq y \leq \pi$, we have

$$\lim_{x \to -\infty} e^{x+iy} = 0.$$

Thus, under the conformal mapping e^z, we expect to see a source at zero in $\mathbb{H}$. On the other hand, as $x \to \infty$, the values of e^z also approach ∞ and the sink is mapped to infinity in $\mathbb{H}$. This means that our complex potential in $\mathbb{H}$ has a source at zero. The complex potential for this flow in $\mathbb{H}$ is $\Omega(z) = \mathrm{Log}(z)$. Thus, the complex potential for the uniform flow in the strip is given by Theorem 3.32 as $\widetilde{\Omega}(z) = \Omega(e^z) = \mathrm{Log}(e^z) = z$ as expected.

With an understanding of how sources and sinks at infinity must be accounted for, it is easy to incorporate sources and sinks along the boundary. The following exercise asks you to find the complex potential for various flows in a channel. You should note in each case that the sum of all sources and sinks must be zero. Use the *FlowTool* applet to see the streamlines for each different flow and to note the behavior at infinity.

EXERCISE 3.41. Consider an infinitely long channel of width π having its lower edge along the real axis. Suppose there is a source of strength 6π at $z = 0$ and a sink of strength 6π at $-3 + \pi i$. Find the complex potential for this flow. Use the FlowTool applet to visualize the streamlines of the flow. Does any of the flow "escape" to infinity? Why or why not? ***Try it out!***

EXERCISE 3.42. Consider a similar problem to Exercise 3.41, but now suppose that the sink at $-3 + \pi i$ has strength 2π. Carefully describe the behavior at infinity. Find the complex potential and use the FlowTool applet to plot the streamlines. Does the result make sense from a physical perspective? ***Try it out!***

3.8. Flows in Other Regions

The methods used to solve flow problems in a strip apply equally well to any other region with sources or sinks along the boundary. All that is required is a conformal mapping from the region in question to the upper half-plane $\mathbb{H}$, such that the boundary of the region is mapped to the real axis. It is then a matter of determining where the sources and sinks are mapped and accounting for any behavior at infinity. It is then straightforward to find the complex potential for the transformed problem in $\mathbb{H}$ and then compose the result with the conformal mapping to obtain the solution.

EXAMPLE 3.43. As an example of these techniques, consider the region $R = \mathbb{H} - \{z \,|\, |z| \leq 1\}$. Imagine a uniform flow in R going left to right, along with a source of strength 2π at $z = -1$ and a sink of strength 2π at $z = 1$. We find the complex potential for this flow in R as follows:

(1) The mapping $f(z) = z + 1/z$ maps R conformally to $\mathbb{H}$.
(2) Observe that $f(-1) = -2$ and $f(1) = 2$.
(3) The complex potential in $\mathbb{H}$ is $\Omega(z) = \mathrm{Log}(z+2) - \mathrm{Log}(z-2)$.

(4) Composing Ω with f gives the result: $\widetilde{\Omega}(z) = \mathrm{Log}(z+1/z+2)-\mathrm{Log}(z+1/z-2)$.

(5) The flow lines are given by the level curves of $\mathrm{Im}\,\widetilde{\Omega}$.

The flow lines are shown in Figure 3.13.
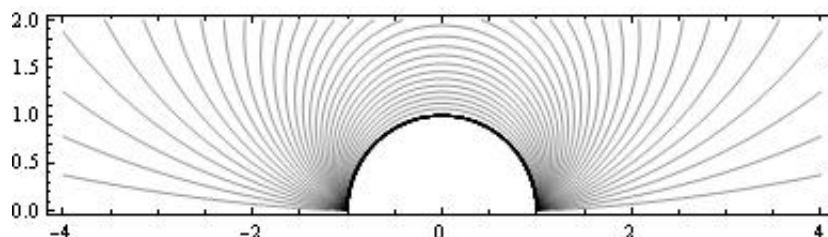
FIGURE 3.13. Flow around cylinder with sources and sinks

Apply the methods outlined in the example above in the following exercise.

EXERCISE 3.44. Let $R = \{z \,|\, 0 \le \arg(z) \le \pi/3\,\}$. Assume there is a source of strength 4π at $z = 0$ and a sink of strength 2π at $z = 2$. Find complex potential for the flow in R. The stream lines are shown in Figure 3.14
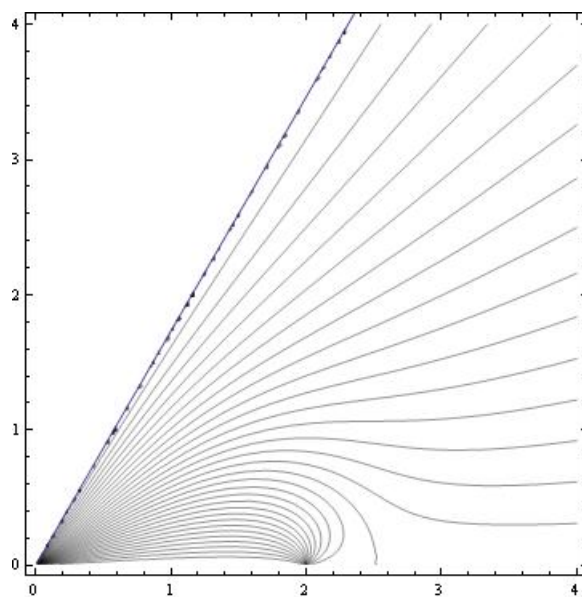
FIGURE 3.14. Stream lines for Exercise 3.44

It should be noted that even with a sophisticated graphing program it may not be easy to obtain a sketch of the flow lines. Many of the plots in this chapter have been generated using *Mathematica*. The *Mathematica* code used to generate Figure 3.14 is shown below.

```
>F[z_]=4*Log[z^3]-2*Log[z^3-8]
>a=ContourPlot[If[y<Sqrt[3]*x, Im[F[x+I*y]]], {x, 0, 4}, {y, 0, 4},
   ContourShading->False, Contours->30]
>b=Plot[Sqrt[3]*x, {x, 0, 4}]
>Show[a, b]
```

In the *Mathematica* code, observe the use of the "If" statement to only plot stream lines in the region of interest. The second plot, labeled "b" is technically not needed, but it makes the boundary of the region much clearer in the final plot.

A few more comments regarding the use of *Mathematica* are in order. In order to make a contour plot of the real or imaginary part of a complex function, it is necessary to replace z by $x + i\,y$ in the plot command. In addition, if the various branch cuts for complex logarithm functions may result in a plot with some breaks in the flow lines that does not look quite right. Sometimes it is possible to fix this problem by combining all the logarithm terms. For example in the code fragment above, the function could have been defined as

```
F[z_]=Log[z^12/(z^3-8)^2]
```

The *FlowTool* applet provides the first quadrant as one of the domains available for study. In the absence of any sources or sinks on the boundary, the uniform flow in this region is simply the "flow around a corner" that we have already mentioned. By including sources or sinks some some very interesting flows can be seen.

EXPLORATION 3.45. Develop a general procedure for finding the complex potential for a flow in the first quadrant with various sources or sinks on the boundary. Use your method to find the complex potentials for the scenarios given below. In each case, use the *FlowTool* applet to investigate the flow.

   (1) Sources of equal strength at $z = 1$ and $z = i$.
   (2) A source and a sink of equal strengths at $z = 1$ and $z = i$.
   (3) Sources of equal strength at $z = 1$ and $z = i$, and a sink of double strength at the origin.

3.9. Flows inside the Disk

The methods developed thus far can be applied equally well to fluid flow inside a disk. The following set of exercises explores various scenarios. Obtaining the graphical output for these flows can be challenging. We present the *Mathematica* code to generate one of the plots.

EXERCISE 3.46. Let $\mathbb{D} = \{z \in \mathbb{C} \,|\, |z| < 1\}$. Show that

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

is a conformal mapping from the unit disk $\mathbb{D}$ to the upper half plane $\mathbb{H}$. Determine the images of the points ± 1 and $\pm i$. Use the conformal mapping to find a uniform flow inside $\mathbb{D}$. The flow lines are shown in Figure 3.15. Explain why there appears to be both a source and a sink at $z = -1$. ***Try it out!***
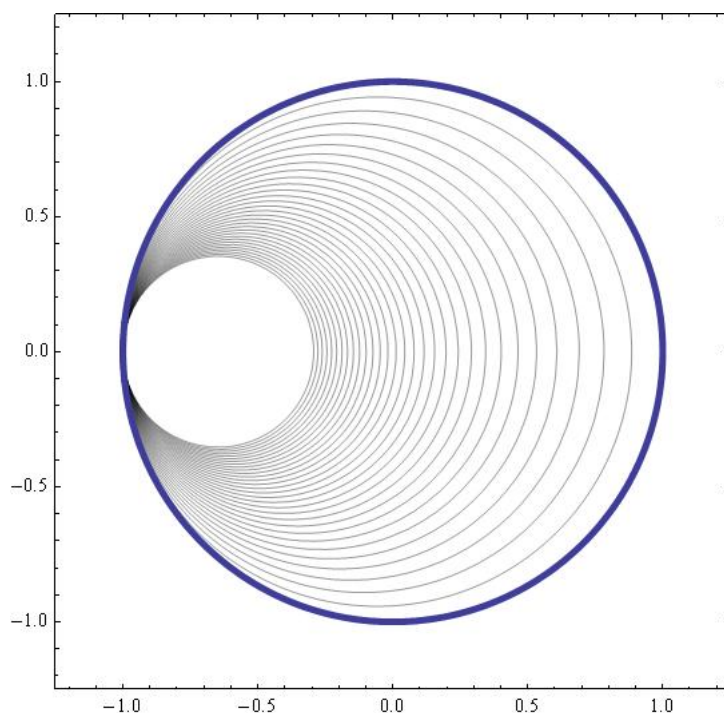
FIGURE 3.15. Uniform flow in a disk. Observe the behavior at $z = -1$.

The *Mathematica* code for Figure 3.15 is given below.

```
>g[z_] = I*(1 - z)/(1 + z)
>M[x_, y_] = If[x^2 + y^2 <= 1, Im[g[x + I y]]]
>a = ContourPlot[M[x, y], {x, -1.2, 1.2}, {y, -1.2, 1.2},
  ContourShading -> False, Contours -> 30]
>b = ParametricPlot[{Cos[t], Sin[t]}, {t, 0, 2*Pi},
  PlotStyle -> Thickness[.01]]
>Show[a, b]
```

EXPLORATION 3.47. This exploration considers various flows inside a disk.
  (1) Find the velocity field of the ideal fluid flow inside $\mathbb{D}$ if there is a source of strength 2π located at $z = 1$. Note that the total strength of the sources in a problem must be balanced by a sink or sinks of equivalent strength. Where is the sink in this problem? Can you control the location of the sink?

(2) Find the complex potential for the flow in $\mathbb{D}$ if there is a sink of strength 2π at $z = i$.

(3) Consider the case where there are equal strength sources on the boundary of $\mathbb{D}$ located at $z = 1$ and $z = -1$. Consider the location of the resulting sink. Why does this occur?

3.10. Interval Sources and Sinks

In this section we extend some of the methods already developed. Instead of looking at point sources or sinks, we consider the case of an entire interval of sources or sinks. This will allow us to model different phenomenon such as flow through a levy or the electric field generated by a line of charges. We present the material as a series of exercises designed to lead to the key results. The reader is expected to complete these exercises rather than simply using the results.

We begin by examining the flow lines generated by a uniformly distributed source of total strength 2π located along the interval $a \le x \le b$.

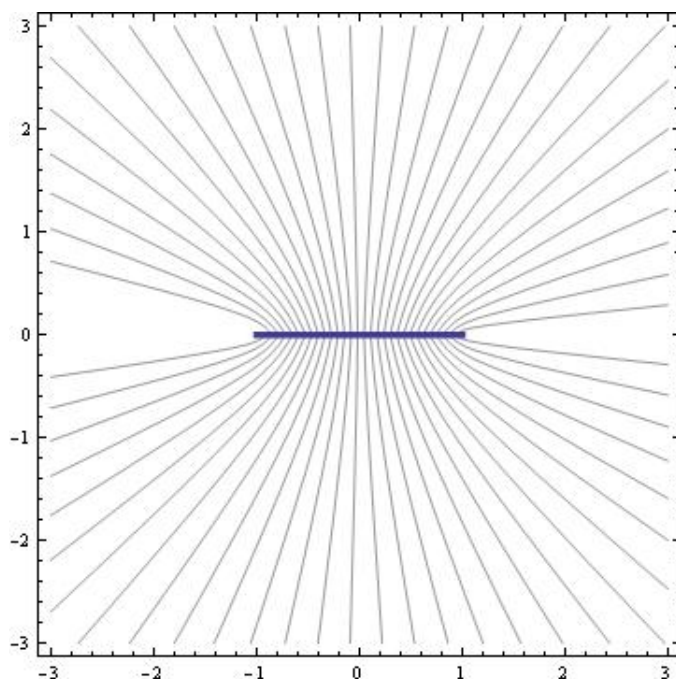We expect to see a vector field as in Figure 3.16.

FIGURE 3.16. Vector field with a uniform interval source

The next set of exercises shows that the complex potential for an interval source can be derived as a limiting process of a collection of point sources as the number of points approaches infinity.

EXERCISE 3.48. Let $a = x_0, x_1, \ldots, x_n = b$ be $n+1$ equally spaced points in the interval $[a, b]$. Assume there is a source of strength $2\pi/(n+1)$ at each x_j. Find the complex potential $\Omega(z)$ for the flow in $\mathbb{H}$ for this collection of point sources. ***Try it out!***

Now the idea at this point is to let $n \to \infty$ and to recognize the resulting limit as a definite integral.

EXERCISE 3.49. Let $\Delta x_j = x_{j+1} - x_j$, $j = 0, 1, \ldots, n$. Express the result of the previous exercise as a Riemann sum on the interval $[a, b]$. Show the limit of the Riemann sum leads to the definite integral

$$(33) \qquad \frac{1}{b-a} \int_a^b \mathrm{Log}(z-x)\, dx.$$

Try it out!

EXERCISE 3.50. Obtain the definition of the complex potential for a uniform interval source by using integration by parts to show the that the integral 33 can be evaluated to the expression:

$$(34) \qquad \Omega_a^b(z) = \frac{b-z}{b-a}\mathrm{Log}(z-b) + \frac{z-a}{b-a}\mathrm{Log}(z-a) - 1$$

Try it out!

EXERCISE 3.51. Use Equation (34) to find the complex potential in $\mathbb{H}$ for a flow that has a uniform source of strength 2π along the interval $[0, 3]$ and a uniform sink of strength 2π along the interval $[-2, -1]$. The graph appears in Figure 3.17. ***Try it out!***
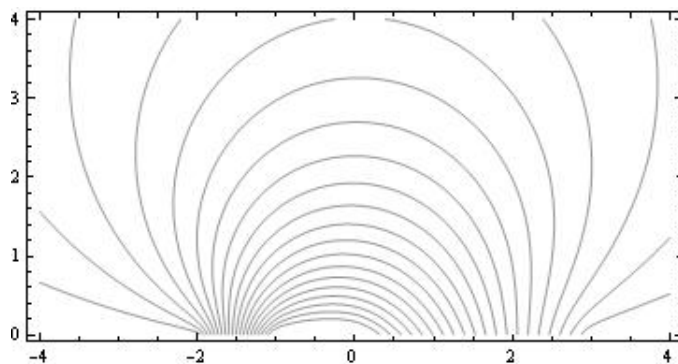

FIGURE 3.17. Flow lines with interval source/sink

Next we seek to extend the notion of interval sources or sinks to the boundaries of other regions. Some care must be taken to be sure the interval sources or sinks are still uniform as we now show. Consider the case of the first quadrant R in $\mathbb{C}$

with interval sources of equal strength located along the intervals $[1, 3]$ and $[i, 3i]$. We proceed as in previous sections. The conformal mapping $f(z) = z^2$ maps R to $\mathbb{H}$. Next we determine the behavior of the key intervals under the conformal mapping. The interval $[1, 3]$ is mapped to $[1, 9]$ and $[i, 3i]$ is mapped to $[-9, -1]$. Now, we know how to find the complex potential in $\mathbb{H}$ from Equation (34). Indeed we see that $\Omega(z) = \Omega_1^9(z) + \Omega_{-9}^{-1}(z)$. We then find the complex potential in R by composition: $\widetilde{\Omega}(z) = \Omega(z^2)$. Unfortunately this result is not quite correct. The problem is that that sources should be uniformly distributed across the intervals in R. When we apply the conformal mapping the resulting intervals in $\mathbb{H}$ are no longer uniformly distributed. The key to fixing this problem is described next.

To understand what is happening, focus on the interval $[1, 3]$ under the mapping $f(z) = z^2$. Clearly the interval is mapped to $[1, 9]$. Notice that the first half $[1, 2]$ is mapped to $[1, 4]$ and the second half $[2, 3]$ is mapped to $[4, 9]$. Hence the density of points is less in the second half of the interval. Thus, when solving for the complex potential in $\mathbb{H}$ we cannot treat the intervals as though the source is uniformly distributed. Recall that Equation (34) was derived under the assumption that the source was uniformly distributed along the interval. We must instead use Equation (33) and take into account the non-uniform distribution. Thus instead of computing the integral $\int_1^9 \mathrm{Log}(z - x)\, dx$ we compute $\int_1^3 \mathrm{Log}(z - x^2)\, dx$.

EXERCISE 3.52. Find a complex potential for the flow in the first quadrant generated by uniform interval sources along the intervals $[1, 3]$ and $[i, 3i]$. Plot the streamlines.

EXERCISE 3.53. Find the complex potential for an ideal flow in the region
$$R = \{re^{i\theta} \mid r \geq 0, \quad 0 \leq \theta \leq \frac{\pi}{3}\}$$
with a uniform source of strength 2π located along the interval $[2, 4]$. **Try it out!**

EXERCISE 3.54. Find the complex potential for an ideal flow in the infinite channel
$$R = \{z \mid 0 \leq \mathrm{Im}\, z \leq 2\}$$
with a uniform source of strength 2π located on the boundary along the interval $[1 + 2i, 4 + 2i]$. **Try it out!**

A natural extension of the previous material on interval sources is to consider intervals with a non-uniform density. This small project was inspired by Potter [**4**].

SMALL PROJECT 3.55. Consider a function $\lambda : [a, b] \subset \mathbb{R} \to \mathbb{R}$ with the property that
$$\int_a^b \lambda(x)\, dx = S,$$
where S is the total strength of the generalized source on $[a, b]$. We think of $\lambda(x)$ as giving the source density at $x \in [a, b]$. The goal is to find the complex potential for

this interval source with variable density. Observe that if $\lambda(x) = S/(b-a)$, for all x in $[a, b]$, then we obtain the case of a uniformly distributed source along the interval.

Rather than attempting to directly find the complex potential, it is better to first find the underlying vector field $\overline{V(z)}$. Begin by subdividing the interval into n equal subintervals, where the nth subinterval is $[x_i, x_{i+1}]$. Now, consider the vector field $\overline{V_i(z)}$ having a source at x_i with strength

$$S_i = \frac{b-a}{n}\lambda(x_i).$$

Show by summing the individual vector fields and taking a limit as $n \to \infty$ that the vector field is given by

(35)
$$\overline{V(z)} = \int_a^b \frac{\lambda(x)}{\overline{z} - x}\, dx.$$

From Equation (35) it follows that the complex potential is

(36)
$$\Omega(z) = \int \left(\int_a^b \frac{\lambda(x)}{z - x}\, dx \right) dz$$

Consider the specific example of $\lambda(x) = x$ on the interval $[-1, 2]$. Observe that the source density takes on both positive and negative values on the interval, with the positive density increasing towards the right end of the interval. The strength of this variable density source is

$$\int_{-1}^2 x\, dx = \frac{3}{2}.$$

Compute $V(z)$ by using the substitution $w = z - x$ in the integral. Use the result to find the complex potential and plot the flow lines. The result is shown in Figure 3.18.

Note that if λ is a polynomial, then a formula for $V(z)$ can be found explicitly using the same substitution as above. The reader is encouraged to find and plot the complex potential for the case where $\lambda(x) = x^2$ on $[0, 1]$.

3.11. Steady State Temperature Problems

A common problem in applied mathematics is to find the steady state temperature in a region of the plane given a fixed temperature distribution along the boundary of the region. In the case of the upper half plane $\mathbb{H}$, the problem takes the following form: Given a piecewise continuous, bounded function f defined on the real axis (i.e., the temperature), find a function T in $\mathbb{H}$ such that $\Delta T = 0$ and T "agrees" with f along the boundary. Note that solution involves finding a harmonic function in $\mathbb{H}$. We also point out that the we are seeking solutions that are physically meaningful. For example, if the boundary values are identically zero, then both $T_1(x, y) = 0$ and $T_2(x, y) = e^x \sin y$ are solutions in $\mathbb{H}$. However, T_2 is unbounded which does not correspond to a real temperature distribution.
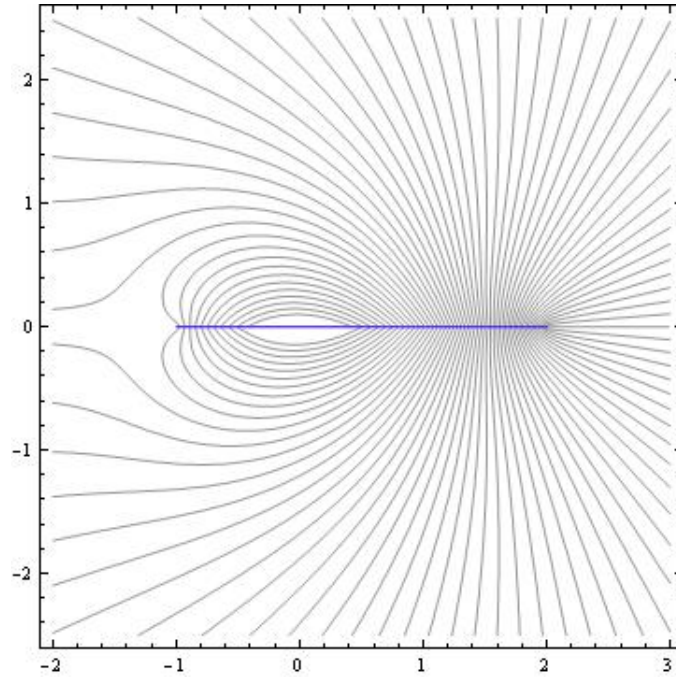
FIGURE 3.18. Flow lines for variable density interval source

Every time we construct a complex potential function in $\mathbb{H}$, both the real and imaginary parts are harmonic. To see if our methods are useful in this setting, we discuss how to manipulate the boundary values in an example.

We begin with the simple case of the complex potential for an ideal flow in $\mathbb{H}$ with a single point source of strength 2π at $z = 0$. We know that the complex potential is given by $\Omega(z) = \operatorname{Log} z$. Writing this in terms of its real and imaginary parts gives

$$(37) \qquad \Omega(z) = \operatorname{Log} |z| + i \operatorname{Arg}(z).$$

Now observe the values of $\operatorname{Arg}(z)$ along the real axis. When $x > 0$, we have $\operatorname{Arg}(x) = 0$, and when $x < 0$ we have $\operatorname{Arg}(z) = \pi$. The key observation is that the argument function is piecewise constant along the real axis.

Consider the steady state temperature problem where the distribution along the boundary is given by

$$(38) \qquad f(x) = \begin{cases} 0 & \text{if } x < 0 \\ 100 & \text{if } x \geq 0. \end{cases}$$

Starting from Equation (37), we see that multiplying $\operatorname{Im}\Omega(z)$ by $100/\pi$ gives a harmonic function in $\mathbb{H}$ satisfying the boundary condition

224

(39) $$\frac{100}{\pi}\operatorname{Arg}(x) = \begin{cases} 100 & \text{if } x < 0 \\ 0 & \text{if } x \geq 0. \end{cases}$$

This is almost correct. To get the boundary values switched around to the correct intervals we need some properties of the argument function. The following exercise leads to the correct result.

EXERCISE 3.56. If $\operatorname{Im} z > 0$, then show that reflecting $\operatorname{Arg}(z)$ across the y-axis gives the function $\operatorname{Arg}(-\bar{z})$. Next, show that

$$\operatorname{Arg}(-\bar{z}) = -\operatorname{Arg}(-z).$$

Try it out!

Using the results of Exercise 3.56, we see that

$$g(x, y) = -\frac{100}{\pi}\operatorname{Arg}(-z)$$

is harmonic and satisfies the correct boundary conditions. The graph of $g(x, y)$ is shown in Figure 3.19. Note that $g(x, y)$ is the imaginary part of the complex potential

$$\Omega(z) = -\frac{100}{\pi}\operatorname{Log}(-z).$$

The streamlines correspond to isotherms–curves of constant temperature.
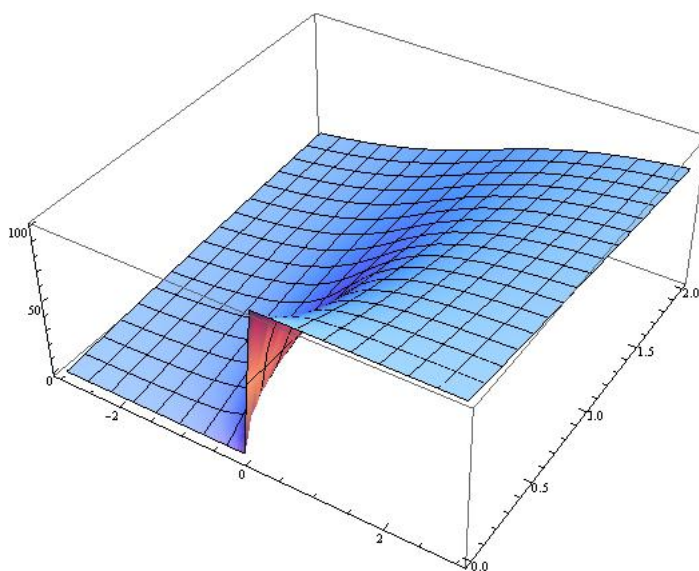

FIGURE 3.19. Steady state temperature distribution with piece-wise constant boundary values

The next few exercises explore slight variations on this problem. We then finish the chapter by extending to the case where segments of the boundary are not insulated (i.e., there are segments that have non-constant temperatures) and see how our methods involving intervals sources can be applied.

EXERCISE 3.57. Solve the steady state temperature problem in $\mathbb{H}$ where the boundary temperatures are given below. (Hint: This is very similar to the above example–simply start with a slightly different complex potential function.)

$$(40) \qquad f(x) = \begin{cases} 100 & \text{if } x < 2 \\ 0 & \text{if } x \geq 2. \end{cases}$$

Try it out!

EXERCISE 3.58. Solve the steady state temperature problem in $\mathbb{H}$ where the boundary temperatures are given below. (Hint: Add a constant to the function, but pay attention to the effect it has on the other boundary value.)

$$(41) \qquad f(x) = \begin{cases} 100 & \text{if } x < 2 \\ 50 & \text{if } x \geq 2. \end{cases}$$

Try it out!

The types of boundary conditions we have been considering can easily be extended to more segments. The following exercise leads to the algorithm for solving the general problem of this type.

EXERCISE 3.59. Solve the steady state temperature problem in $\mathbb{H}$ where the boundary temperatures are given below.

$$(42) \qquad f(x) = \begin{cases} 100 & \text{if } x < -3 \\ 50 & \text{if } -3 < x < 2 \\ 25 & \text{if } x \geq 2. \end{cases}$$

Make a contour plot of your solution and observe that the contours represent curves of constant temperature. *Try it out!*

The general problem of finding the temperature distribution in the half-plane given piece-wise constant boundary conditions is a standard application in complex analysis texts; see [**3**]. The problem usually takes the following form; see [**3**].

EXPLORATION 3.60. Solve:

$$T_{xx} + T_{yy} = 0$$

Subject to:

$$T(x,0) = \begin{cases} k_0 & \text{if } -\infty < x < x_1 \\ k_1 & \text{if } x_1 < x < x_2 \\ \vdots & \vdots \\ k_n & \text{if } x_n < x < \infty \end{cases}$$

Using the techniques developed in the exercises, derive the general solution to this problem.

Next we consider a slightly more general problem where segments of the boundary are not kept at constant temperature. For example, suppose we have the following temperature distribution along the boundary:

(43)
$$f(x) = \begin{cases} 100 & \text{if } x < 0 \\ 0 & \text{if } x > 1. \end{cases}$$

Observe that no temperature is specified on the interval $(0, 1)$. If this segment of the boundary is not insulated then we expect the temperature to change linearly from 100 degrees to 0 degrees (this can be seen by solving the one-dimensional heat flow problem on the interval $[0, 1]$ with the endpoints held at 100 and 0 degrees respectively).

We now investigate how our earlier techniques can be applied to this type of problem. Let us attempt to find the steady state temperature distribution in $\mathbb{H}$ for the boundary condition given in Equation (43). We begin by writing down the complex potential for an ideal flow with an interval source uniformly distributed along the interval $[0, 1]$. Recall from Equation (34) that the complex potential is given by

(44)
$$\Omega_0^1(z) = (1 - z)\text{Log}(z - 1) + z\text{Log}(z) - 1.$$

We know that $\text{Im}\,\Omega_0^1(z)$ must be constant along the intervals $x < 0$ and $x > 1$. We determine these values by choosing a test value from each interval. Thus,

$$\text{Im}\,\Omega_0^1(-1) = \text{Im}(2\text{Log}(2) - \text{Log}(-1) - 1) = -\pi,$$

and

$$\text{Im}\,\Omega_0^1(2) = \text{Im}(-\text{Log}(1) + 2\text{Log}(2) - 1) = 0.$$

Hence multiplying our complex potential by $-100/\pi$ gives the correct temperatures on the two intervals that are held constant. But, what about the interval $(0, 1)$? In figure 3.20, we show the graph of $\text{Im}\,\Omega_0^1(x)$, for $-3 < x < 3$ and observe the desired linear behavior on the interval $(0, 1)$.

The graph of the solution on $\mathbb{H}$ is shown in Figure 3.21. Notice how the surface representing the temperature matches up with the given boundary values.

The following small project asks you to generalize the previous example to a general problem of the same type.

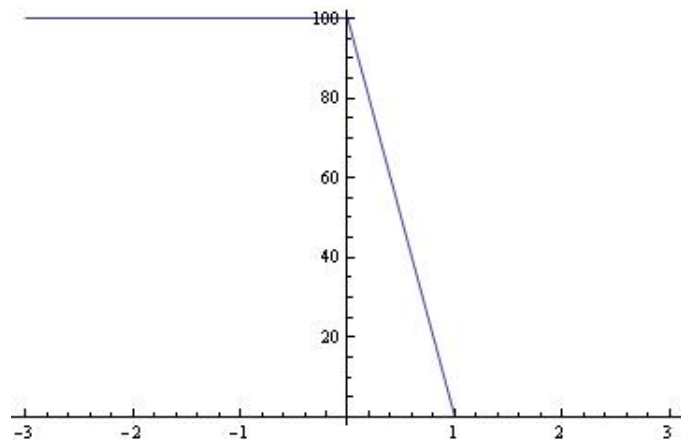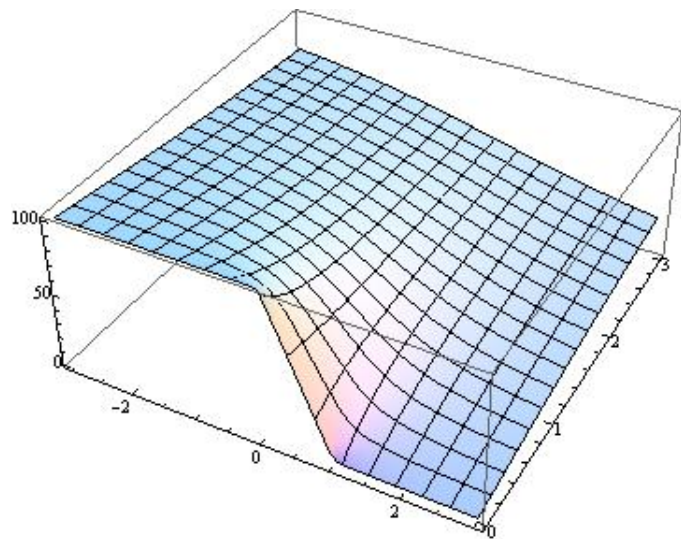FIGURE 3.20. Temperature along boundary

FIGURE 3.21. Solution to heat equation

SMALL PROJECT 3.61. Find a general formula for the steady-state temperature distribution $T(x, y)$ in the half-plane with the following boundary data:

$$T(x, 0) = \begin{cases} k_1 & \text{if } -\infty < x < x_1 \\ k_2 & \text{if } x_2 < x < x_3 \\ \vdots & \vdots \\ k_n & \text{if } x_n < x < \infty \end{cases}$$

228

We assume $x_1 < x_2 < \ldots < x_n$ and that $T(x, 0)$ is linear on the intervals in between the x_i locations.

3.12. Flows with Source and Sinks not on the Boundary

In this section we extend the ideas of this chapter to a wider array of the applications. Consider the standard example of the upper half plane $\mathbb{H}$, but with a source located at $z = i$, as opposed to being on the boundary. In this case, we expect the flow to look like an ordinary source near $z = i$, but since it is constrained to stay in $\mathbb{H}$ the real axis must deflect the flow so that it runs parallel to the boundary. Figure 3.22 shows the flow lines we should expect.
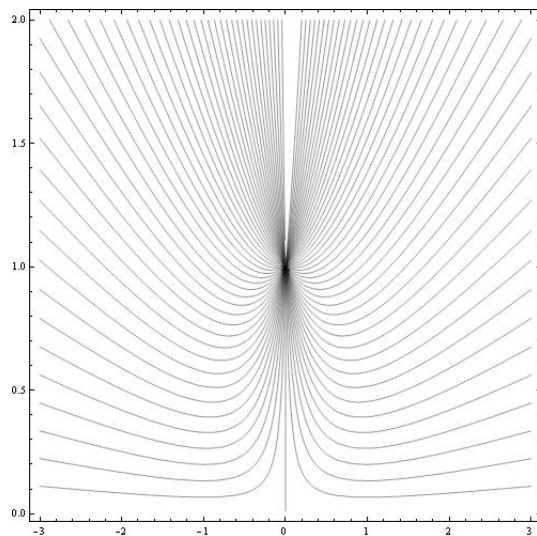
FIGURE 3.22. Flow in $\mathbb{H}$ with a single source at $z = i$.

The key idea in obtaining the complex potential that gives Figure 3.22 is to balance the source at $z = i$ with a source of equal strength at $z = -i$. By the Principle of Superposition, the vertical components of the underlying vector fields will sum to zero along the real axis. The resulting complex potential is $\Omega(z) = \text{Log}(z - i) + \text{Log}(z + i)$ or equivalently as $\Omega(z) = \text{Log}(z^2 + 1)$. Note that Ω is not analytic in on any punctured neighborhood of i. Thus, Ω is not formally a complex potential in all of $\mathbb{H}$. However, by choosing different branches of the complex logarithm, it always possible to construct an analytic complex potential in a simply connected domain not containing i.

This balancing sources or sinks across the boundaries, combined with our earlier work can be used to deal with sources or sinks in the interior of any region. Needham [**2**] refers to this approach as the *Method of Images*. The following exercise finds the complex potential for a flow in the first quadrant with a source located at the interior point $z = 1 + 2i$.

EXERCISE 3.62. Find the flow of an ideal fluid in the first quadrant with a single source located at $z = 1 + 2i$. Hint: Use the conformal mapping $f(z) = z^2$ to map the first quadrant to $\mathbb{H}$. Determine the location of source by computing $f(1 + 2i)$. Find the complex potential in $\mathbb{H}$ by balancing the source with another source symmetrically located across the real axis. Compose the result with $f(z)$ to obtain the complex potential of the flow in the first quadrant. The result gives the flow shown in Figure 3.23. ***Try it out!***
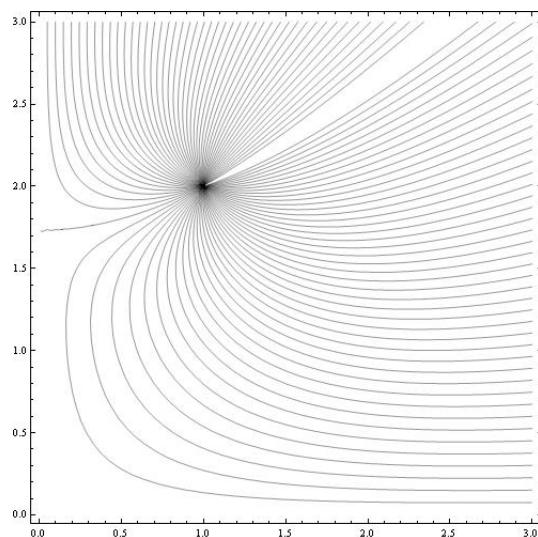
FIGURE 3.23. Flow in the first quadrant with a single source at $z = 1 + 2i$

When dealing with sources and sinks both on the boundary and in the interior the notion of *effective strength* comes into play. Indeed, a source on the boundary must have twice the strength as one in the interior in order to have the same effective strength.

EXERCISE 3.63. Find the complex potential for a uniform flow in $\mathbb{H}$ with a source at $z = 0$ and a sink of equal magnitude at $z = i$. Answer: $\Omega(z) = \mathrm{Log}(z^2/(z^2 + 1))$ The streamlines are shown in Figure 3.24 ***Try it out!***

The techniques developed thus far allow us to combine sources and sinks of various relative strengths both on the boundary and in the interior of some region. The following exercise shows that the complex potential for a very complicated flow can be systematically built up from simpler pieces.

EXERCISE 3.64. Let R be the infinite strip defined by $0 \leq \mathrm{Im}\, z \leq \pi$. Suppose there is a source of strength 2π at $z = 1$, a sink of strength 2π at $z = 4$ and a source of strength 2π in the middle of the strip at $z = \pi i/2$. Construct the complex potential of the velocity field. The streamlines are shown in Figure 3.25. ***Try it out!***
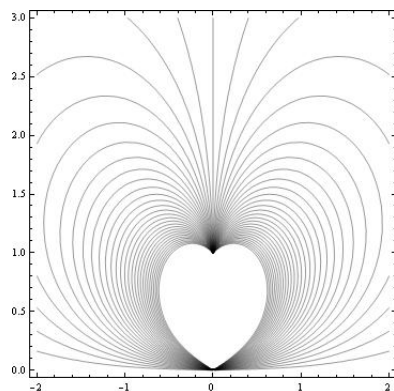
FIGURE 3.24. Balancing a source on the boundary with a sink in the interior
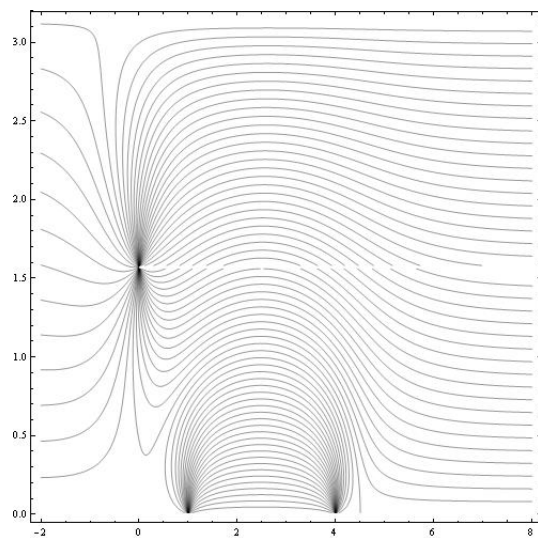
FIGURE 3.25. Streamlines for the flow in Exercise 3.64

3.13. Vector Fields with Other Types of Singularities

Another application arises when we interpret ideal flows as electric field lines generated by various positive and negative charges. In this context we imagine the plane as a copper plate with charges at various locations. A *dipole* is obtained when a positive and negative charge of equal strength are separated by a small distance. Consider the case of a source of strength 1 located at $z = \epsilon$ and a sink of strength -1 located at $z = -\epsilon$. The Polya vector field for the corresponding electric field is

$$\overline{f(z)} = \frac{1}{\bar{z} - \epsilon} - \frac{1}{\bar{z} + \epsilon}.$$

231

Observe that as ϵ approaches zero, the field vanishes. In order to prevent the field from vanishing, we need to increase the strengths of the source and sink inversely to the distance between them. keep the field strength constant while the distance between the charges approaches zero.

EXERCISE 3.65. Consider the Polya vector field

$$\overline{f(z)} = \frac{1}{2\epsilon}\left(\frac{1}{\bar{z}-\epsilon} - \frac{1}{\bar{z}+\epsilon}\right).$$

Compute the limiting field as ϵ goes to zero and then compute the complex potential for the field. Hint: Recall that $\Omega'(z) = f(z)$. ***Try it out!***

As shown in Exercise 3.65, the vector field generated by the dipole is $\overline{f(z)} = 1/\bar{z}^2$. Observe the complex potential is given by $\Omega(z) = -1/z$. The electric field lines are given by the level curves of $\mathrm{Im}\,\Omega(z)$ and are shown in Figure 3.26.

EXERCISE 3.66. Show that the electric field lines for the dipole are circles with their centers along the imaginary axis.
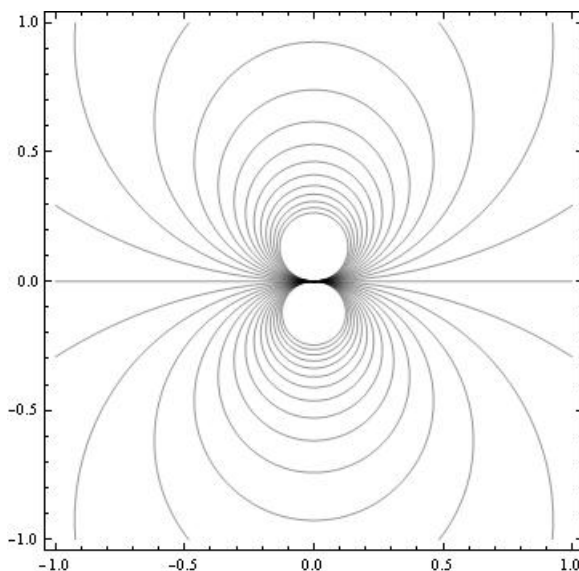
FIGURE 3.26. Electric field lines for a dipole.

Observe the orientation of the two small loops in Figure 3.26. The particular orientation of the dipole depends on the direction from which the source and the sink approach each other. In this case Polya field was $1/\bar{z}^2$ and the *dipole moment* was 1. Other dipole moments are obtained by considering the field $d/\bar{z}^2$, where d is any complex number. For example if $d = 1+i$, then we obtain the dipole field shown in Figure 3.27.
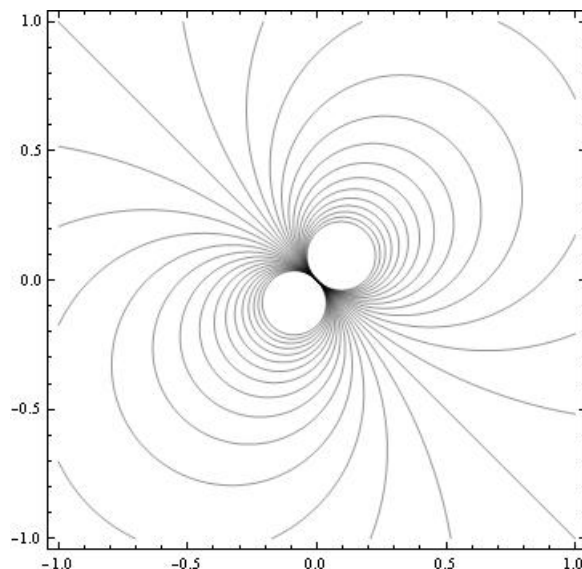
FIGURE 3.27. Electric field lines for a dipole with dipole moment equal to $1 + i$.

EXPLORATION 3.67. Investigate the general behavior of the dipole whose field is given by $\overline{f(z)} = (a + bi)/\bar{z}^2$.

It is also possible to consider the problem of multiple charges approaching each other to obtain *multipoles*. As an example, consider the complex potential function $\Omega(z) = 1/z^2$. The electric field lines are shown in Figure 3.28.

EXERCISE 3.68. What charges are converging to give the electric field lines shown in Figure 3.28?

An interesting problem involves looking at sources or sinks and multipoles at the same location. For example, consider the Polya field $\overline{V(z)} = 1/\bar{z}^2 + 1/\bar{z}$. The complex potential function is $\Omega(z) = -1/z + \text{Log } z$. As an exercise, you should plot the electric field lines both near zero and on a larger scale. Note how the dipole dominates the behavior near zero and the source dominates the behavior far from zero.

We close this chapter with a brief discussion that ties together a loose end: namely the nature of sources and sinks at infinity. Since we now understand the idea of a multipole, we are in a position to correctly state the nature of a flow at infinity.

Consider the most basic example of the uniform flow to the right in the entire plane $\mathbb{C}$. Since fluid appears from the left and disappears to the right it seems reasonable to say that there is both a source and a sink at infinity. Of course we now recognize this phenomenon as a dipole.

To study the behavior of a complex function at infinity, it is standard to replace z with $1/z$ and study the behavior of the resulting function at $z = 0$. For the case of the uniform flow, the complex potential is $\Omega(z) = z$. The behavior at infinity is
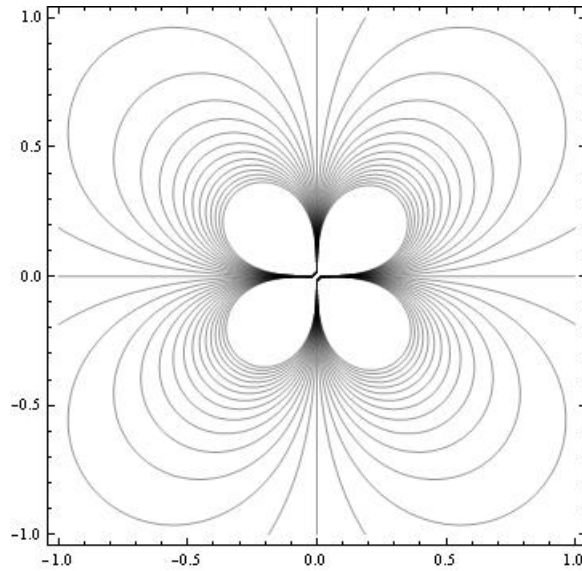
FIGURE 3.28. Electric field lines for a multipole.

determined by examining $\Omega(1/z) = 1/z$ at $z = 0$. Thus we have a potential function of $1/z$. Differentiating and taking the conjugate to obtain the underlying vector field, we get $1/\bar{z}^2$ which we recognize as a dipole.

More generally, for any vector field where the net sum of the sources and sinks is, say N, the strength of the source or sink at infinity is $-N$. If $N = 0$, then there is a multipole at infinity.

Bibliography

[1] James Stewart, *Multivariable Calculus,* 6th *ed.,* Thompson Brooks/Cole, 2008.

[2] Tristan Needham, *Visual Complex Analysis,* Oxford University Press, Oxford UK, 1997.

[3] Dennis G. Zill, Shanahan, P. D., *A First Course in Complex Analysis with Applications,* 2nd *ed.,* Jones and Bartlett, Boston-Toronto-London-Singapore, 2009.

[4] Harrison Potter, *On Conformal Mappings and Vector Fields*, Senior Thesis, Marietta College, Marietta, Ohio, 2008.

[5] Ruel V. Churchill, Brown, J. W., *Complex Variables and Applications,* 4th *ed.,* McGraw-Hill Book Company, New York, 1984.

CHAPTER 4

Anamorphosis, Mapping Problems, and Harmonic Univalent Functions

MICHAEL DORFF

4.1. Introduction

Complex-valued analytic functions have many very nice properties that are not necessarily possessed by real-valued functions. For example, we say a complex-valued function is analytic if you can differentiate it one time. It turns out that if a complex-valued function f is analytic, then you can differentiate it infinitely many times. In addition, complex-valued analytic functions can always be represented as a Taylor series, and they are conformal (that is, they preserve angles when $f' \neq 0$). Such properties are not true for real-valued functions that can be differentiated one time. Why does an analytic function have these properties? If $f = u + iv$ is an analytic function, then its real part, $u(x, y)$, and its imaginary part, $v(x, y)$, satisfy Laplace's equation and thus are both harmonic. Also, u and v satisfy the Cauchy-Riemann equations and are therefore harmonic conjugates of each other. In this chapter we discuss some ideas and problems related to a collection of univalent (i.e., $1 - 1$) complex-valued functions, $f = u + iv$, where u and v satisfy Laplace's equation but not necessarily the Cauchy-Riemann equations. This collection of functions are known as *harmonic univalent functions* or mappings and contain the collection of analytic univalent functions as a subset. Analytic univalent functions have been studied since the early 1900's, and there are thousands of research papers written on the subject. The study of harmonic univalent mappings is a fairly recent area of research. So, it is natural to consider the properties of analytic univalent functions as a starting point for our study of harmonic univalent mappings. A general theme will be "What properties of analytic univalent functions are still true for this larger class of harmonic univalent functions?"

Section 4.2 discusses the idea of determining the image of domains in $\mathbb{C}$ under a collection of complex-valued functions known as Möbius maps and introduces the applet *ComplexTool* to as an aid to visualize these images. Section 4.3 presents some background about the family of univalent analytic functions. Section 4.4 introduces the fundamentals of harmonic univalent functions. The study of harmonic univalent functions from the perspective of univalent complex-valued analytic functions is a new

area of research. Finding examples of such functions is not easy, but a very useful method of doing so is discussed in Section 4.5. Sections 4.2 – 4.5 should be read first. After that, the remaining sections can be read in any order and are independent of each other. There are three applets used in this chapter:

- *ComplexTool* is used to plot the image of domains in $\mathbb{C}$ under complex-valued functions.
- *ShearTool* is used to plot the image of domains in C under a complex-valued harmonic function that is formed by shearing an analytic function and a dilatation; the user enters the corresponding analytic function and dilatation without having to solve explicitly for the harmonic function.
- *LinComboTool* is used to plot and explore the convex combination of complex-valued harmonic polygonal maps.

They can be accessed online at
http://www.jimrolf.com/explorationsInComplexVariables/chapter4.html. Each section contains examples, exercises, and explorations that involve using the applets. You should do all of the exercises and explorations many of which present functions and concepts that will be used later in the chapter (there are additional exercises at the end of the chapter). In the study of harmonic univalent functions, there are many open problems. Some of these are specifically mentioned. In addition, there are short projects and long projects that are suitable as research problems for undergraduates to explore.

The goal of this chapter is not to give a comprehensive or step-by-step approach to this topic, but rather to get the reader engaged with the general notions, questions, and techniques of the area – but even more so, to encourage the reader to actively pose as well as pursue their own questions. To better understand the nature and purpose of this text, the reader should be sure to read the Introduction before proceeding. The study of harmonic univalent functions has many interesting problems that can be investigated by undergraduates through the use of computers and the applets. I anticipate that some students will explore the ideas in this chapter and that this will lead them to prove some new results in the field.

4.2. Anamorphosis and Möbius Maps

Numb3rs is a U.S. television show that ran from 2005 to 2010. It deals with a group of FBI special agents trying to solve crimes. One of the FBI agents is Don Eppes who has a mathematical genius brother, Charlie, who helps the FBI solve their cases. In the episode "Jack of All Trades" (season 5, episode 4), the FBI is trying to catch a thief who has eluded them for two years. In fact, they do not even have an accurate image of his face. At one point in this episode, the thief escapes by stealing a bus at night and drives away. Before he is able to escape, one agent takes a photo of the thief with a flash camera. Unfortunately, the flash from the camera reflects off the bus window obscuring the face of the thief. However, there is a metal cylindrical thermos

in the photo that displays a distorted image of the thief. Using mathematics, Charlie recreates the image of the thief's face giving the FBI information that soon leads to capturing the thief.

This method of producing a distorted image that appears normal by viewing it through a curved mirror or from an unusual angle is known as anamorphosis. Anamorphosis was studied by painters in the 15-16th centuries as they were trying to understand perspective. Two famous paintings that display anamorphosis are Jan van Eck's "The Arnolfini Marriage" and Hans Holbein's "The Ambassadors." Van Eyck's painting is thought to depict the Italian merchant Giovanni Arnolfini and his wife. On the wall behind them is a curved mirror reflecting a distorted image of the scene the viewer sees in the painting. In Holbein's painting there is a distorted shape lying diagonally

FIGURE 4.1. Van Eyck's "The Arnolfini Marriage"

at the bottom. This shape is a skull that can be seen more clearly when the viewer is looking at the painting from a certain angle.

In modern times anamorphosis has continued to be used. In the sketch "Mysterious Island" by the Hungarian artist István Orosz the viewer sees the image of a seashore with hills and the sun in the background, two men walking, and a ship being tossed in the sea. But if a cylindrical mirror is placed on top of the sun, the image of the author Jules Verne appears on the cylinder. And in 2010 the group *preventable.ca* was able to cause drivers to slow down while driving near a school in British Columbia, Canada. They did this by putting an anamorphosis image on the highway. The image is not clear until the driver reached a certain point on the highway when suddenly it appears that a young girl has darted into the road chasing after a pink ball. The idea was that in the same way that the anamorphosis image could suddenly appear to the driver, so could a child suddenly run in front of the driver.

This idea of how an image is distorted is used in complex analysis when trying to visualize the image that results when applying a domain to a complex-valued function. Such problems are called mapping problems. We will begin this chapter by exploring

FIGURE 4.2. Holbein's "The Ambassadors"

FIGURE 4.3. Orosz's "Mysterious Island"

FIGURE 4.4. Anamorphosis image of a child chasing a ball into the street.

an important collection of maps known as Möbius tranformations. First, let us review some ideas about mapping domains under complex-valued functions. If we have a one-dimensional real-valued function, such as $f(x) = x^2$, it is useful to represent the

graph of the function, because it can tell us about some of the properties of f (i.e., zeros, $1-1$, increasing, etc.). We want to do the same thing for complex-valued functions. However, we would need a 4-dimensional graph (two dimensions for the domain and two dimensions for the range). We can represent some of the properties of a complex-valued function by looking at specific sets in the domain and seeing where the complex-valued function $w = f(z)$ takes them.

We can use the accompanying applet *ComplexTool* to graph the image of certain domains under complex-valued functions or maps. To do so, open *ComplexTool* (see Figure 4.5). Suppose we want to find the image of the unit disk under the map
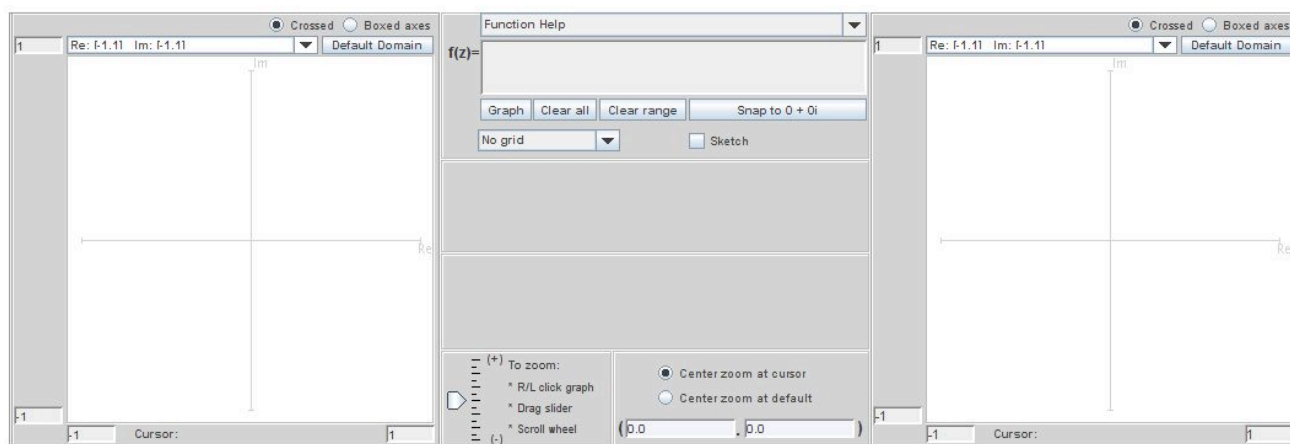
FIGURE 4.5. The applet *ComplexTool*

$f(z) = z + 2 - i$. In the middle section near the top there is a box that has **f(z)=** before it. In this box, enter $z + 2 - i$. Below this, there is a window that states `No grid`. Click on the down arrow ▼ and choose the option `Circular grid`; an image of a circular grid should appear on the left (we will call this the z-plane). Next, click on the button `Graph` which is in the middle section below the function you entered earlier. The image of the circular grid should appear on the right (we will call this the w-plane). To reduce the size of the image, click on the down arrow ▼ above the image and chose a different size, such as `Re: [-3,3] Im: [-3,3]`. Also, you can move the axes so that the image is centered by positioning the cursor over the image, clicking on the left mouse button, and dragging the image to the left (see Figure 4.6).

EXPLORATION 4.1. You can graph the halfdisk, $S = \{z||z| < 1, \operatorname{Im} z > 0\}$, by using a circular disk but restricting the the θ value to be between 0 and pi in the middle panel of the applet. For each of the following functions determine the image of S. Make a conjecture how a domain is transformed under the maps $f(z) = z + A$,
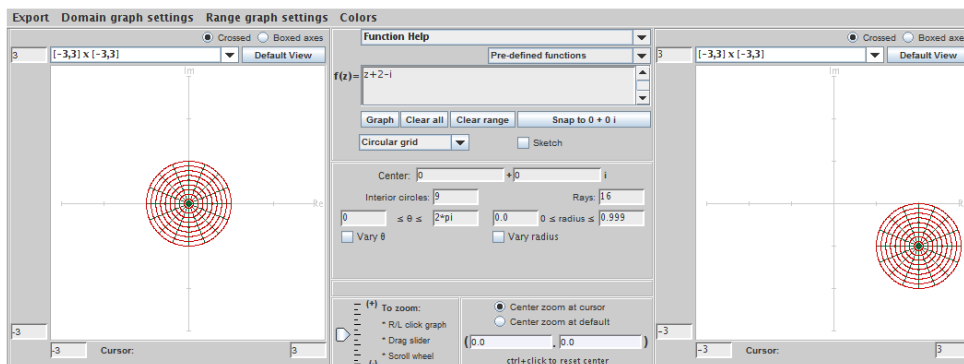
FIGURE 4.6. The image of the unit disk under the map $f(z) = z + 2 - i$.

$f(z) = Bz$, and $f(z) = e^{i\theta}z$, where $A = a_1 + ia_2 \in \mathbb{C}$, $B > 0$, and $\theta \in \mathbb{R}$.

(a) $f(z) = z - 1$;   (b) $f(z) = z + i$;   (c) $f(z) = z - 1 + i$;

(d) $f(z) = 0.5z$ (e) $f(z) = 2z$;   (f) $f(z) = 2.5z$;

(g) $f(z) = e^{i\pi/4}z$ (h) $f(z) = e^{i\pi/2}z$;   (i) $f(z) = e^{-i\pi/4}z$.

Try it out!

The previous exercise helps us see the following characteristics of these maps:

- if $A = a_1 + ia_2 \in \mathbb{C}$, then the map $f(z) = z + A$ moves the image domain a_1 units horizontally and a_2 units vertically;
- if $B > 0$, then the map $f(z) = Bz$ scales (i.e., expands or shrinks) the domain by a factor of B; and
- if $\theta \in \mathbb{R}$, then the map $f(z) = e^{i\theta}z$ rotates the domain about the origin in a counterclockwise direction by a factor of θ.

EXERCISE 4.2. Justify that these three functions $f(z) = z + A$, $f(z) = Bz$, and $f(z) = e^{i\theta}z$ map circles onto circles and lines onto lines.

Try it out!

If we think of a line as a circle with infinite radius, then we can say that these functions preserve circles. That is, they map circles onto circles. Another function we are interested in is is the inversion map $f(z) = \frac{1}{z}$. If we write $z = re^{i\theta}$, then

$$f(z) = \frac{1}{z} = \frac{1}{re^{i\theta}} = \frac{1}{r}e^{-i\theta}.$$

Thus, the function $f(z) = \frac{1}{z}$ scales the domain by a factor of $\frac{1}{r}$ and reflects it across the real axis because of the $e^{-i\theta}$.

241

EXPLORATION 4.3. Open *ComplexTool*. We want to explore the image of the unit circle under $f(z) = \frac{1}{z}$. To do this, choose the option `Circular grid`. In the middle panel below this, select `Interior circles:` to be 1 and `Rays:` to be 0. This should give the domain in the z-plane to just be the unit circle. Next, click on the down arrow ▼ above the image and chose `Re: [-3,3] Im: [-3,3]` for the z-plane. Do this also for the w-plane. Left click on the `Graph` button to produce the image of the unit circle in the right box. Now, left click on the domain circle and drag it around. Notice what happens to the image as you do this.

(a.) As you slowly move the center of the domain circle away from the origin in the z-plane, what happens to the size of the image circle in the w-plane?

(b.) What geometric shape does the image take when the domain circle intersects the origin?

(c.) What happens to the size of the image circle as the domain circle moves off the screen?

(d.) Can you move the domain circle to a position so that the image is neither a circle nor a line?

(e.) Explain all your observation in parts (a.)-(d.) in terms of the fact that the function $f(z) = \frac{1}{z}$ scales the domain by a factor of $\frac{1}{r}$ and reflects it across the real axis

Try it out!
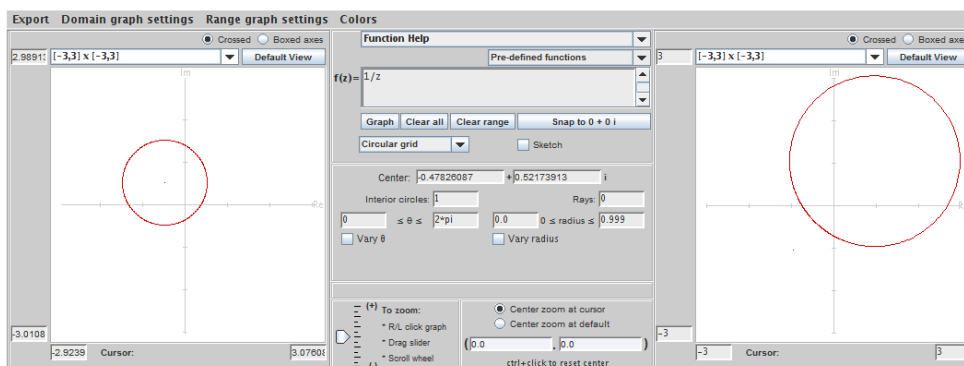
FIGURE 4.7. The image of a circle not intersecting the origin under the map $f(z) = \frac{1}{z}$.

It is true that the inversion map $f(z) = \frac{1}{z}$ maps circles to either circles or lines. Let's look at the image of vertical lines under $f(z) = \frac{1}{z}$. First, consider the line L_1 that starts at the bottom of the imaginary axis and travels upward on that axis. This line can be described as $z = 0 + iy$, where y varies from $-\infty$ to ∞. Then the image of L_1 is

$$\frac{1}{z} = \frac{1}{iy} = i\frac{-1}{y}.$$
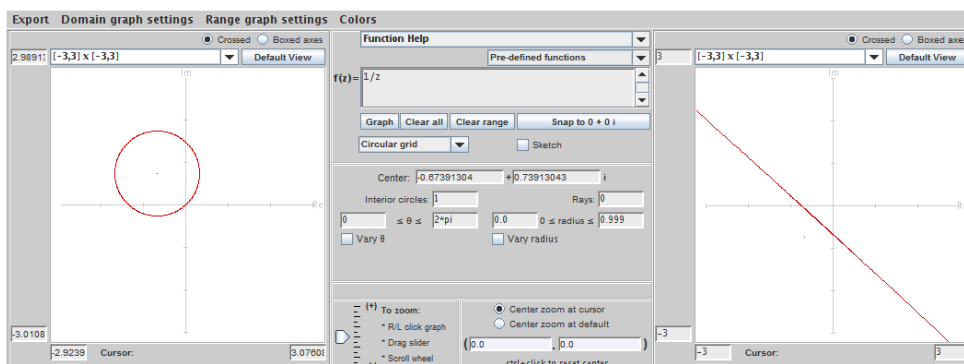
FIGURE 4.8. The image of a circle intersecting the origin under the map $f(z) = \frac{1}{z}$.

Because the real part is zero in this image, L_1 will get mapped into the imaginary axis. As y starts at $-\infty$ and increases, the image will start at 0 and move upward. When y passes through the point 0, the image will wrap around from the top of the imaginary axis and go to the bottom of the imaginary axis. As y continues up alone the positive imaginary axis, the image will continue up from the bottom of the imaginary axis. You can use *ComplexTool* to visualize this. Open *ComplexTool* and enter the function $f(z) = 1/z$. Chose the window size of `Re: [-3,3] Im: [-3,3]` for both the z-plane and the w-plane. Click on the box next to the `Sketch` feature so that a checkmark appears. Then click on the `Graph` button. Now place the cursor arrow on the imaginary axis in the z-plane, hold down the left mouse button, and drag the cursor upward. As you do so, the resulting image in the w-plane should appear (see Figure 4.9).
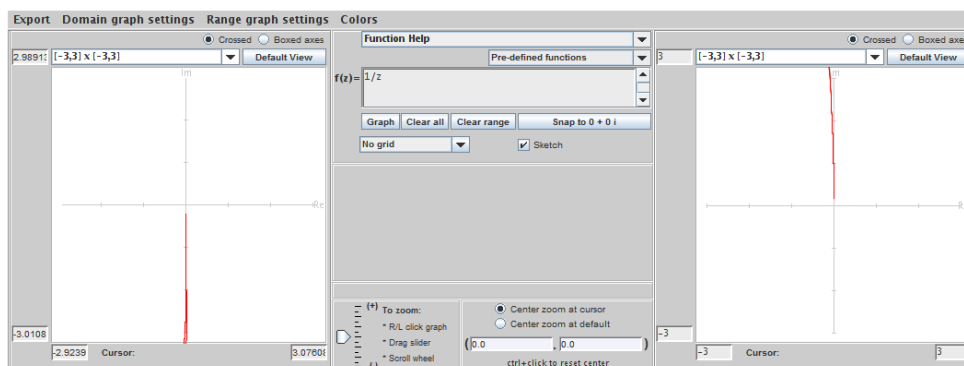
FIGURE 4.9. The image of a vertical line on the imaginary axis under the map $f(z) = \frac{1}{z}$.

Now, let's consider the image of the vertical line with $\operatorname{Re} z = \frac{1}{2}$. In the z-plane, let $z = x + iy$. So, in general,

$$\frac{1}{z} = \frac{\bar{z}}{z\bar{z}} = \frac{1}{|z|^2}\bar{z} = \frac{x}{x^2 + y^2} - i\frac{y}{x^2 + y^2}.$$

If we let $w = u + iv$ in the image domain, then we have that

$$u = \frac{x}{x^2 + y^2} \quad \text{and} \quad v = \frac{-y}{x^2 + y^2}.$$

In the z-plane, this vertical line with $\operatorname{Re} z = \frac{1}{2}$ can be described by $z = x + iy = \frac{1}{2} + iy$. Combining these equations, we have that the image of this vertical line under the inversion map is

$$\frac{1}{z} = u + iv = \frac{\frac{1}{2}}{\frac{1}{4} + y^2} + i\frac{-y}{\frac{1}{4} + y^2} = \frac{2}{4y^2 + 1} + i\frac{-4y}{4y^2 + 1}.$$

Now, what is the geometric shape described by u and v? It turns out that it is a circle centered at $u = 1$ of radius 1 (see Figure 4.10). To see this, notice that such a circle can be described by the equation $(u - 1)^2 + v^2 = 1$. Now,

$$\begin{aligned}
(u - 1)^2 + v^2 &= \left(\frac{2}{4y^2 + 1} - 1\right)^2 + \left(\frac{-4y}{4y^2 + 1}\right)^2 \\
&= \frac{(-4y^2 + 1)^2 + (-4y)^2}{(4y^2 + 1)^2} \\
&= \frac{16y^4 + 8y^2 + 1}{(4y^2 + 1)^2} \\
&= 1.
\end{aligned}$$
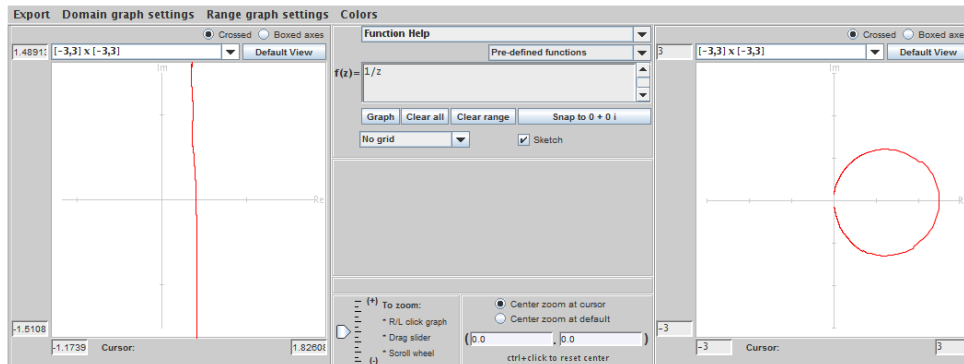
FIGURE 4.10. The image of a vertical line with $\operatorname{Re} z = \frac{1}{2}$ under the map $f(z) = \frac{1}{z}$.

EXERCISE 4.4. Using the approach above, show that $f(z) = \frac{1}{z}$ maps the vertical line $\operatorname{Re} z = c$ onto the circle $\left(u - \frac{1}{2c}\right)^2 + v^2 = \left(\frac{1}{2c}\right)^2$.
Try it out!

The inversion map $f(z) = \frac{1}{z}$ takes circles and lines onto circles and lines. Specifically,

- a circle not intersecting the origin is mapped onto a circle not intersecting the origin.
- a line intersecting the origin is mapped onto a line intersecting the origin.
- a circle intersecting the origin is mapped onto a line not intersecting the origin, and vice-versa.

The four maps we have discussed so far are special cases of a type of mappings known as Möbius transformations,

$$M(z) = \frac{Az + B}{Cz + D},$$

where $A, B, C, D \in \mathbb{C}$ and $AD \neq BC$.

EXERCISE 4.5. What happens to $M(z) = \frac{Az+B}{Cz+D}$ if $AD = BC$?
Try it out!

If $A = 1$, $C = 0$, and $D = 1$, we have $M(z) = z + B$, a translation map. If $B = 0$, $C = 0$, and $D = 1$, we have $M(z) = Az$ which is a scaling, a rotation, or both. If $A = 0$, $B = 1$, $C = 1$, and $D = 0$, we have $M(z) = \frac{1}{z}$, the inversion map. Also, note that

$$\frac{Az + B}{Cz + D} = \frac{\frac{A}{C}(Cz + D) - \frac{AD}{C} + B}{Cz + D} = \frac{A}{C} + \frac{B - \frac{AD}{C}}{Cz + D}.$$

So if we let

$$f_1(z) = Cz + D, \quad f_2(z) = \frac{1}{z}, \quad \text{and} \quad f_3(z) = \left(B - \frac{AD}{C}\right)z + \frac{A}{C},$$

then the general Möbius transformation $M(z) = \frac{Az+B}{Cz+D}$ can be expressed as $(f_3 \circ f_2 \circ f_1)(z)$. We know how the maps f_1, f_2, and f_3 affect a given domain, and since these functions map circles/lines to circles/lines, a Möbius transformation will do this too. By including the point at ∞ in our domain and range, the Möbius transformation is a $1 - 1$ and onto function. Thus, its inverse function exists. This inverse function is

$$M^{-1}(z) = \frac{Dz - B}{-Cz + A},$$

and we see that the inverse of any Möbius transformation is a Möbius transformation.

EXERCISE 4.6. Starting with any Möbius transformation M, compute that M^{-1} is $\frac{Dz-B}{-Cz+A}$.
Try it out!

Also, every circle is uniquely determined by three distinct points. These facts allow us to determine the image of a circle under a Möbius transformation. Specifically, to find the image of a circle under M, take any three points on the circle, find their image under M, and then determine the circle formed by those three image points.

EXAMPLE 4.7. Let's find the image of the unit circle $\{z \mid |z| = 1\}$ under the map
$$M(z) = \frac{z}{1-z}.$$
First, choose three points on the unit circle. We will choose the points i, -1, and $-i$. Now
$$M(i) = -\frac{1}{2} + \frac{1}{2}i, \quad M(-1) = -\frac{1}{2}, \quad \text{and} \quad M(-i) = -\frac{1}{2} - \frac{1}{2}i.$$
The three image points $-\frac{1}{2} + \frac{1}{2}i$, $-\frac{1}{2}$, and $-\frac{1}{2} - \frac{1}{2}i$ form a line (note that this also makes sense since the point 1 on the unit circle is mapped to $M(1) = \infty$). Specifically, it is the vertical line $\{z \mid \text{Re}\{z\} = -\frac{1}{2}\}$. Hence, the unit circle is mapped under M to this vertical line.

Question: What is the image of the unit disk $\{z \mid |z| < 1\}$ under the map $\frac{z}{1-z}$? Möbius maps will send the interior of the domain into the interior of the image. Since $M(0) = 0$, we know that the unit disk gets mapped onto the right half-plane $\{w \mid \text{Re}\{w\} > -\frac{1}{2}\}$.
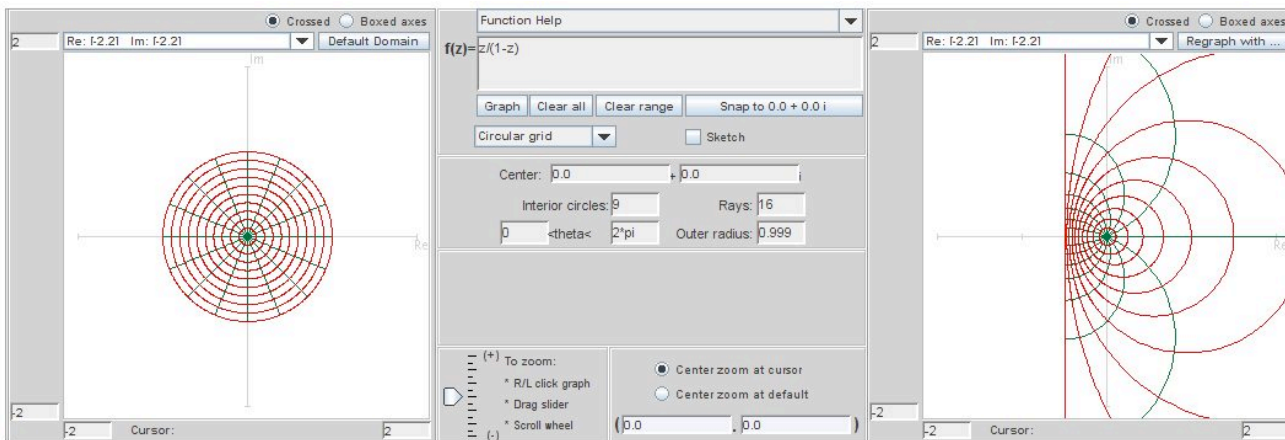
FIGURE 4.11. The image of the unit disk under the Möbius map $M(z) = \frac{z}{1-z}$.

EXERCISE 4.8. Using the approach in Example 4.7 analytically determine the image of the unit disk $\{z \mid |z| < 1\}$ under the transformation
$$M(z) = \frac{z-i}{z}.$$
Use *ComplexTool* to check your answer.
    Try it out!

Suppose that we want to map a given circle/line C_1 to a specific circle/line C_2. How do we construct a function that will do this? Recall that three points uniquely determine a circle and Möbius transformations send circles/lines to circles/lines. So, choose three points z_1, z_2, and z_3 on C_1 and three points w_1, w_2, and w_3 on C_2. Then a Möbius transformation $M(z)$ satisfying

$$M(z_1) = w_1, \quad M(z_2) = w_2, \quad \text{and} \quad M(z_3) = w_3$$

will map C_1 onto C_2.

We can construct $M(z)$ by using the cross-ratio formula

(45)
$$\frac{(w - w_1)(w_2 - w_3)}{(w - w_3)(w_2 - w_1)} = \frac{(z - z_1)(z_2 - z_3)}{(z - z_3)(z_2 - z_1)}.$$

If one of the terms z_i or w_i is ∞ (i.e., one of the circles C_1 or C_2 is actually a line), then we omit the expressions containing that term. For example, if $w_1 = \infty$, then eq (45) becomes

$$\frac{w_2 - w_3}{w - w_3} = \frac{(z - z_1)(z_2 - z_3)}{(z - z_3)(z_2 - z_1)}.$$

Let's see how we can use eq (45) to find a Möbius transformation that maps z_1 to w_1, z_2 to w_2, and z_3 to w_3.

EXAMPLE 4.9. Suppose we want to determine a Möbius transformation that maps the vertical line, $l = \{z \mid \text{Re}\{z\} = -\frac{1}{2}\}$, onto the unit circle, C. First, we pick three points on l. We will chose

$$z_1 = -\frac{1}{2}, \quad z_2 = -\frac{1}{2} + i\frac{1}{2}, \quad \text{and} \quad z_3 = \infty.$$

Next, we pick three points on C. Let's chose

$$w_1 = -1, \quad w_2 = i, \quad \text{and} \quad w_3 = 1.$$

Using eq (45), we have

$$\frac{z + \frac{1}{2}}{(-\frac{1}{2} + i\frac{1}{2}) + \frac{1}{2}} = \frac{(w + 1)(i - 1)}{(w - 1)(i + 1)}.$$

Simplifying and solving for w yields

$$w = \frac{z}{1 + z}.$$

Thus, the Möbius transformation $M(z) = \frac{z}{1+z}$ maps l onto C. Note that this transformation is the inverse of the transformation used in Example 4.7.

EXERCISE 4.10. In Example 4.9, work out the details that show

$$\frac{z + \frac{1}{2}}{(-\frac{1}{2} + i\frac{1}{2}) + \frac{1}{2}} = \frac{(w + 1)(i - 1)}{(w - 1)(i + 1)} \quad \text{simplifies to} \quad w = \frac{z}{1 + z}.$$

Try it out!

In Example 4.9, we found a Möbius transformation M that maps the line l onto the circle C. As an afterthought we see that M maps the right half-plane $D = \{z \mid \mathrm{Re}\{z\} > -\frac{1}{2}\}$ to the unit disk, because $M(0) = 0$. Of course, it could have happened that we had constructed a Möbius transformation that maps right half-plane to the exterior part of the unit circle instead of the unit disk. What can we do at the very beginning to guarantee that the Möbius transformation we construct will map our initial domain onto the region we want? Notice that we could have chosen switch the values of w_2 and w_3 so that $w_1 = -1$, $w_2 = 1$, and $w_3 = i$, and constructed a different Möbius transformation that maps l to C. In this case, the right half-plane D would have mapped onto the exterior of the unit circle. To guarantee that the Möbius transformation we construct will map our initial domain onto the region we want, we need to chose the order of the points w_1, w_2, and w_3 correctly. How do we do this? Notice that as we travel along l from the point $z_1 = -\frac{1}{2}$ to $z_2 = -\frac{1}{2} + i\frac{1}{2}$ to $z_3 = \infty$, the right half-plane D is on the right-side of the line l. In order to guarantee that M will map D onto the unit disk, we need to chose the order of the points w_1, w_2, and w_3 so that the unit disk is also on the right-side of the unit circle C. This does happen in Example 4.9, because we go from $w_1 = -1$ to $w_2 = i$ to $w_3 = 1$.

EXERCISE 4.11. Determine a Möbius transformation that maps the unit disk $\{z \mid |z| < 1\}$ onto the upper half-plane $\{z \mid \mathrm{Im}\{z\} > 0\}$. Use *ComplexTool* to check your answer. ***Try it out!***

There are more ideas we could discuss related to Möbius transformations, but we need to go on to other topics. However, before we do this, we want to mention that there is a very nice short video that illustrates some of the connections between Möbius transformations and motions of the sphere. The video is called "Moebius Transformations Revealed" by Douglas Arnold and Jonathan Rogness, and it can be easily searched for on the internet.

4.3. The Family S of Analytic, Normalized, Univalent Functions

We will be discussing mapping problems in complex analysis. This deals with the properties of a collection of functions that map one domain onto certain other image domains. Before we get into these properties, we need some background material.

DEFINITION 4.12.
  (1) Throughout this chapter, let $G \subset \mathbb{C}$ be a simply-connected domain.
  (2) Let $\mathbb{D} = \{z : |z| < 1\}$, the unit disk.
  (3) A function f is *univalent* in G if f is one-to-one in G. That is, if $f(z_1) = f(z_2)$, then $z_1 = z_2$, $\forall z_1, z_2 \in G$.

Univalent analytic functions are nice, because they guarantee the existence of an inverse function that is analytic.

EXAMPLE 4.13. Suppose we want to prove that $f(z) = (1+z)^2$ is univalent in $\mathbb{D}$. A standard argument for that is to let $z_1, z_2 \in \mathbb{D}$ and suppose $f(z_1) = f(z_2)$. This means that

$$f(z_1) = f(z_2) \Rightarrow (1+z_1)^2 = (1+z_2)^2$$
$$\Rightarrow 1 + 2z_1 + z_1^2 = 1 + 2z_2 + z_2^2$$
$$\Rightarrow z_1^2 - z_2^2 + 2(z_1 - z_2) = 0$$
$$\Rightarrow (z_1 - z_2)(z_1 + z_2 + 2) = 0.$$

Since $|z_1|, |z_2| < 1$, we know that $z_1 + z_2 + 2 \neq 0$. Hence, we must have $z_1 - z_2 = 0$. So, $z_1 = z_2$.

The image of $\mathbb{D}$ under the map $f(z) = (1+z)^2$ is shown in Figure 4.12.
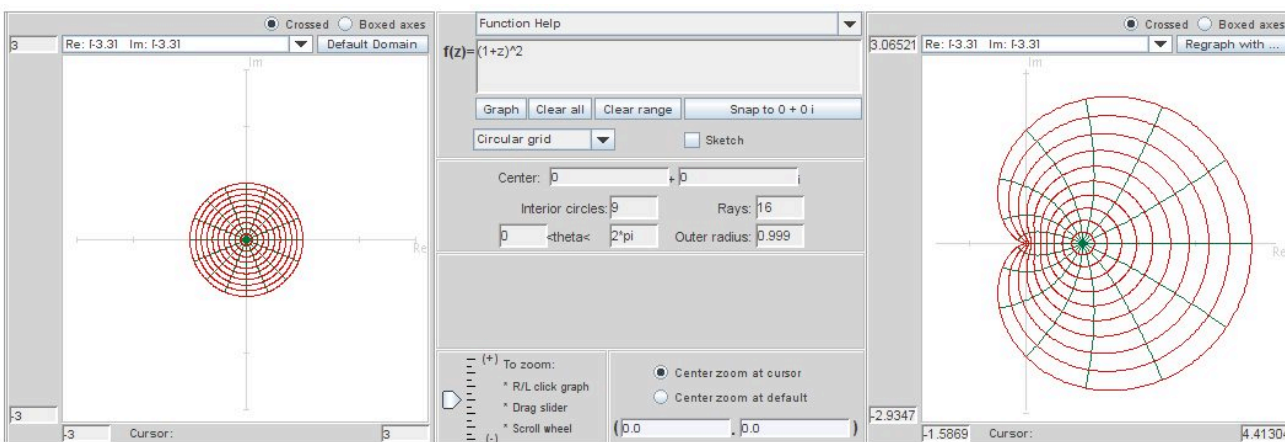
FIGURE 4.12. The image of the unit disk under the map $f(z) = (1+z)^2$.

EXPLORATION 4.14. From Example 4.13, we know that $f(z) = (1+z)^2$ is univalent on $\mathbb{D}$, while it can be shown that $f(z) = (1+z)^4$ is not univalent on $\mathbb{D}$ (see Exploration 4.15). Use *ComplexTool* to graph the image of $\mathbb{D}$ under the analytic function $f(z) = (1+z)^2$ and then under the analytic function $f(z) = (1+z)^4$. What aspects of these two images suggest that a function is univalent or is not univalent? Explore this idea by plotting the following further functions in *ComplexTool* and conjecture which of them is univalent:

(a) $g_1(z) = z - z^2$;

(b) $g_2(z) = z - \frac{1}{2}z^2$;

(c) $g_3(z) = 2z - z^2$;

(d) $g_4(z) = z + \frac{3}{4}z^2$;

(e) $g_5(z) = \frac{z}{1-z}$;

(f) $g_6(z) = \frac{z^2}{1-z}$;

(g) $g_7(z) = \frac{z}{(1-z)^2}$.

249

Try it out!

EXPLORATION 4.15. Prove that $f(z) = (1 + z)^4$ is not univalent in $\mathbb{D}$.

One way to do this is to find two distinct points $z_1, z_2 \in \mathbb{D}$ such that $f(z_1) = f(z_2)$. You can use *ComplexTool* to help you find z_1 and z_2. Plot the image of $\mathbb{D}$ under the map $f(z) = (1+z)^4$. Note that you can increase the size of the image by clicking on the left button on the mouse; to decrease the size, click on the right mouse button. Next, check the `Sketch` box in the top middle section. The `Sketch` command allows you to draw a shape in the original domain on the left and see the image under the function of that shape on the right. For example, draw a line along the imaginary axis from 0 to i and then draw a line along the imaginary axis from 0 to $-i$; you should see that the two image curves meet at $f(i) = f(-i)$. Compute $f(i)$ and $f(-i)$ to prove that this is true. However, this does not prove that f is not univalent in $\mathbb{D}$ since $i, -i \notin \mathbb{D}$. Use the `Sketch` feature of *ComplexTool* to help you find two points $z_1, z_2 \in \mathbb{D}$ such that $f(z_1) = f(z_2)$. To delete the shapes you have drawn, use the `Clear all` button and then regraph your image. [Hint: using *ComplexTool* find two distinct lines in the original domain that get mapped to the line on the real axis from -4 to 0 in the image domain; parametrize these two lines in such a way that for each t value, the image of these parametrized lines under f give the same image point.]

Try it out!

In Exploration 4.14, you may have noticed that the image of $\mathbb{D}$ under the function $g_2(z) = z - \frac{1}{2}z^2$ is similar to the image of $\mathbb{D}$ under the function $g_3(z) = 2z - z^2$. This is because $g_3 = 2g_2$. We want to avoid such repetitions. To do so, we will *normalize* all functions in the family of analytic univalent functions defined on $\mathbb{D}$. In doing this, first suppose f_1 is univalent and analytic in the simply connected domain $G \neq \mathbb{C}$. The Riemann Mapping Theorem can be stated in the following form:

THEOREM 4.16. **(Riemann Mapping Theorem)** Let $a \in G$. Then there exists a unique function $f_2 : G \to \mathbb{C}$ such that

(1) $f_2(a) = 0$ and $f_2'(a) > 0$;
(2) f_2 is univalent;
(3) $f_2(G) = \mathbb{D}$.

Thus $f_3 = f_1 \circ f_2^{-1} : \mathbb{D} \to f_1(G)$ with f_3 being univalent and analytic. So when studying mappings of simply-connected domains, we can simplify matters by letting $\mathbb{D}$ be our domain. Let $f_3 : \mathbb{D} \to \mathbb{C}$ be univalent and analytic. Since f_3 is analytic, it has a power series about the origin:

$$f_3(z) = \alpha_0 + \alpha_1 z + \alpha_2 z^2 + \alpha_3 z^3 + \cdots$$

that is convergent in $\mathbb{D}$. Notice that adding a constant to f_3 merely translates the image domain and does not effect the univalency. Hence

$$f_4(z) = f_3(z) - \alpha_0 = \alpha_1 z + \alpha_2 z^2 + \alpha_3 z^3 + \cdots$$

is also univalent and analytic in $\mathbb{D}$. Next, note that $\alpha_1 \neq 0$ because f_4 being univalent implies $f_4'(z) \neq 0$ (for all $z \in \mathbb{D}$); but $f_4'(0) = \alpha_1$. So consider

$$f_5(z) = \frac{1}{\alpha_1} f_4(z) = z + \frac{\alpha_2}{\alpha_1} z^2 + \frac{\alpha_3}{\alpha_1} z^3 + \cdots .$$

Recall that multiplying f_4 by $\frac{1}{\alpha_1}$ merely rotates and/or stretches (or shrinks) the image domain. Hence f_5 is still univalent and analytic in $\mathbb{D}$. These steps have "normalized" our original function f_3, so that $f'(0) = 1$, and $f(0) = 0$.

DEFINITION 4.17. The family of analytic, normalized, univalent functions is denoted by S (from the German word "schlicht" which means "simple" or "plain"); that is,

$$S = \{f : \mathbb{D} \to \mathbb{C} \mid f \text{ is analytic and univalent with } f(0) = 0, f'(0) = 1\}.$$

Thus $f \in S$ implies $f(z) = z + a_2 z^2 + a_3 z^3 + \cdots .$

EXERCISE 4.18. Show that $f(z) = z + a_2 z^2$ is univalent in $\mathbb{D} \Longleftrightarrow |a_2| \leq \frac{1}{2}$.
Try it out!

EXAMPLE 4.19. (**Polynomial map**)
Consider the function $g_2(z) = z - \frac{1}{2}z^2 \in S$. In Exploration 4.14, you graphed $g_2(\mathbb{D})$. While computer images are helpful, they can be misleading and sometimes inaccurate. So, it is important for us to be able to determine such images analytically. How can we determine $g_2(\mathbb{D})$ analytically? Consider the image of the boundary of $\mathbb{D}$.

$$\begin{aligned}
w = g_2(e^{i\theta}) &= e^{i\theta} - \frac{1}{2}e^{2i\theta} \\
&= (\cos\theta + i\sin\theta) - \frac{1}{2}(\cos 2\theta + i\sin 2\theta) \\
&= \left(\cos\theta - \frac{1}{2}\cos 2\theta\right) + i\left(\sin\theta - \frac{1}{2}\sin 2\theta\right) \\
&= u + iv
\end{aligned}$$

Thus, $g_2(\partial\mathbb{D})$ is parametrized by

$$\begin{aligned}
u(\theta) &= \cos\theta - \frac{1}{2}\cos 2\theta \\
v(\theta) &= \sin\theta - \frac{1}{2}\sin 2\theta.
\end{aligned}$$

What is this image? It is a cardiod, which is also known as an epicycloid with one cusp (see Figure 4.13).

DEFINITION 4.20. An *epicycloid* is the path traced out by a point p on a circle of radius b rolling on the outside of a circle of radius a:

$$x(\theta) = (a+b)\cos\theta - b\cos\left(\left(\frac{a}{b}+1\right)\theta\right)$$

$$y(\theta) = (a+b)\sin\theta - b\sin\left(\left(\frac{a}{b}+1\right)\theta\right).$$
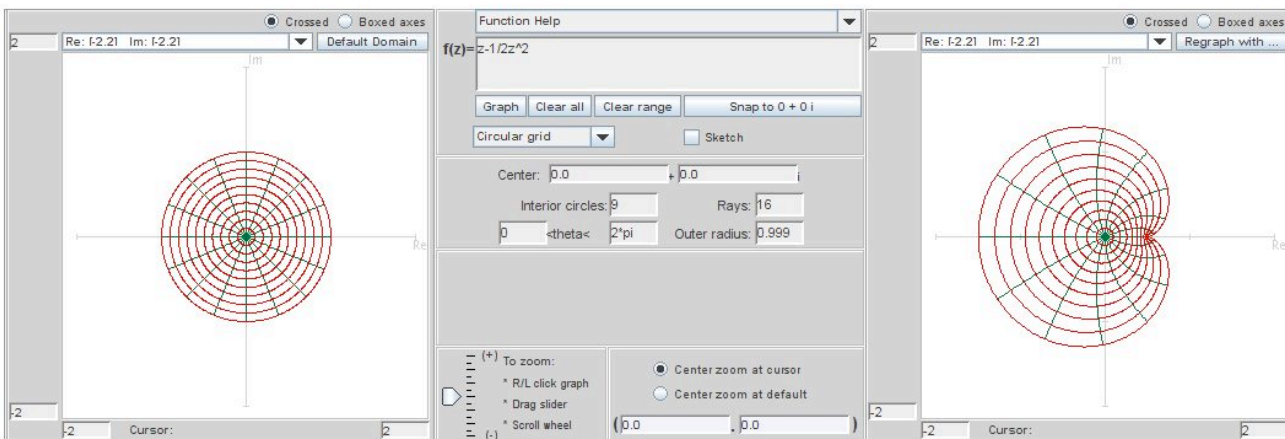
FIGURE 4.13. The image of the unit disk under the map $z - \frac{1}{2}z^2$.

EXPLORATION 4.21. In Exercise 4.18, you showed that $f(z) = z + a_2 z^2$ is univalent in $\mathbb{D} \iff |a_2| \leq \frac{1}{2}$. We want to make a conjecture about the generalization of this result. Use *ComplexTool* to graph $f(z) = z + a_3 z^3$ for various values of a_3. What do you conjecture is the bound on a_3 for which f is univalent on $\mathbb{D}$? Do the same for $f(z) = z + a_4 z^4$, $f(z) = z + a_5 z^5$, etc. What do you conjecture is the bound on a_n for which $f(z) = z + a_n z^n$ is univalent on $\mathbb{D}$? What do you conjecture $f(\mathbb{D})$ is when $a_n = -\frac{1}{n}$?
Try it out!

Let's look at determining $f(\mathbb{D})$ analytically for a few important examples that were included in Exploration 4.14.

EXAMPLE 4.22. **(Right half-plane map)**
Consider the function

$$f_r(z) = \frac{z}{1-z} \in S.$$

Since $\dfrac{1}{1-z} = \sum_{n=0}^{\infty} z^n$, we can multiply by z to get:

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

252

Recall that this is the Möbius transformation that maps $\mathbb{D}$ onto the right half-plane whose boundary is the line $\text{Re}\{w\} = -\frac{1}{2}$ (see Figure 4.14).
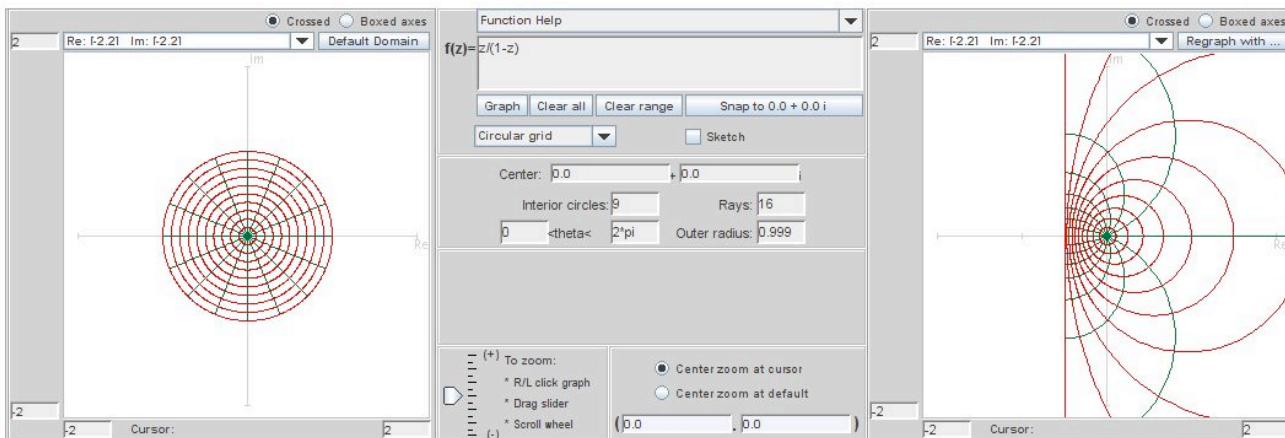
FIGURE 4.14. The image of the unit disk under the analytic right half-plane map in S.

EXAMPLE 4.23. (**Koebe map**)
Next, consider the function

$$f_k(z) = \frac{z}{(1-z)^2} \in S.$$

We can compute the power series for f_k by differentiation the series for $\frac{1}{1-z}$ and then multiplying by z:

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

Notice, for this function, $a_n = n$ for all n. We will now show that the image of $\mathbb{D}$ under f_k is a slit domain. That is, it is a domain consisting of the entire complex plane except that a ray (a.k.a. a slit) is cut out of it. To determine $f_k(\mathbb{D})$, consider the following sequence of maps:

$$u_1(z) = \frac{1+z}{1-z}, \quad u_2(z) = z^2, \quad u_3(z) = \frac{1}{4}[z-1].$$

Now,

$$u_3 \circ u_2 \circ u_1(z) = \frac{1}{4}\left[\left(\frac{1+z}{1-z}\right)^2 - 1\right] = \frac{z}{(1-z)^2}.$$

Note that u_1 is the Möbius transformation that maps $\mathbb{D}$ onto the right half-plane whose boundary is the imaginary axis. Also, u_2 is the squaring function, while u_3 translates the image one space to the left and then multiplies it by a factor of $\frac{1}{4}$.

Thus the image $\mathbb{D}$ is the entire complex plane except for a slit along the negative real axis from $w = \infty$ to $w = -\frac{1}{4}$ (see Figure 4.15). The function $f_k(z) = \frac{z}{(1-z)^2}$ is known as the Koebe function.
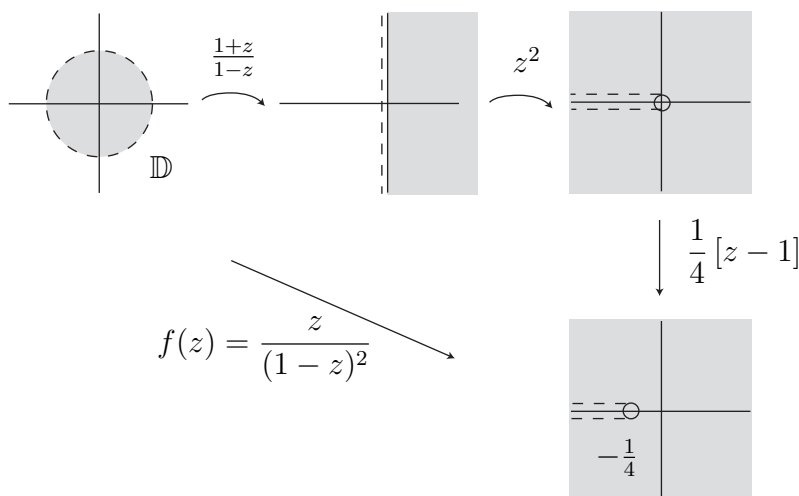
FIGURE 4.15. The image of the unit disk under the Koebe function.

EXPLORATION 4.24. It is difficult to interpret the image of $\mathbb{D}$ under the Koebe function using *ComplexTool*. One way to help understand the image is to use increasing values that approach 1 for the radius of circles in $\mathbb{D}$. We can do this by using the left box in the "$\leq$ radius $\leq$" feature in the center panel of the applet. Graph $\mathbb{D}$ under the map $\frac{z}{(1-z)^2}$ several times using the values of 0.8, 0.85, 0.9, 0.95, and 0.999 in the left box in the "$\leq$ radius $\leq$" feature.
Try it out!

Suppose we have an analytic function f with a Taylor series representation $f(z) = z + a_2 z^2 + a_3 z^3 + \cdots$. One question to ask is for what values of a_n is f in the family of schlicht functions? Consider the case in which all $a_n = 0$ except possibly a_2. So, $f(z) = z + a_2 z^2$. By Exercise 4.18, we know that $|a_2| \leq \frac{1}{2}$ if and only if $f \in S$. The function $f(z) = z - \frac{1}{2}z^2$ from Example 4.19 is an extremal function. An extremal function is a function that is on the boundary between those that satisfy a condition and those that do not. Here, $f(z) = z - \frac{1}{2}z^2$ is extremal, because if we increase $|a_2| = \left|-\frac{1}{2}\right|$ even just a little bit, then $f(z) = a + a_2 z^2$ is no longer schlicht. In general, how large can $|a_n|$ be and $f(z) = z + a_2 z^2 + a_3 z^3 + \cdots$ still be schlicht? Recall, for the

254

Koebe function,

$$\frac{z}{(1-z)^2} = \sum_{n=1}^{\infty} n z^n,$$

and so in this case we have that $a_n = n$. This lead Bieberbach to make his famous conjecture in 1916.

Bieberbach Conjecture For $f \in S$, $|a_n| \leq n$, for all n. In particular, $|a_2| \leq 2$.

Because the image of $\mathbb{D}$ under the Koebe function covers all of $\mathbb{C}$ except a slit along the real axis, it seems plausible that the the Bieberbach Conjecture is true with the Koebe function being extremal. It is true. However, it was not until 1984 that deBranges proved it.

We say an inequality is *sharp* if it is impossible to improve it (that is, we cannot decrease the upper bound or increase the lower bound). We can show that an inequality is sharp by finding a function with the desired properties and for which the inequality becomes equality. Such a function for which equality holds is known as an *extremal* function. Note that for deBranges Theorem (or Bieberbach Conjecture), the Koebe function is extremal.

There is another case in which the Koebe function is extremal. We know that if $f \in S$, then $f(\mathbb{D})$ is not the entire complex plane. That is, there is some point $a \in \mathbb{C}$ such that $a \notin f(\mathbb{D})$. This leads to the question of how small can $|a|$ be for such an $a \in f(\mathbb{D})$? For example, if $f(z) = z$, then $|a| = 1$; if $f(z) = \frac{z}{1-z}$, the right half-plane mapping in Example 4.22, then $|a| = \frac{1}{2}$. The answer to the question is that for all $f \in S$, $|a| \geq \frac{1}{4}$. This is known as the Koebe $\frac{1}{4}$ Theorem. The Koebe function is extremal in this case also, because $|a| = \frac{1}{4}$ for the Koebe function. For these reasons and others, the Koebe function is very important in the study of schlicht functions.

EXPLORATION 4.25. Consider the function $f(z) = \dfrac{z - tz^2}{(1-z)^2}$, where $0 \leq t \leq 1$. From what you have read so far, what is $f(\mathbb{D})$ when $t = 0$? What is $f(\mathbb{D})$ when $t = 1$? Using your answers to these two questions and not using *ComplexTool* yet, make a conjecture of what $f(\mathbb{D})$ is, when $0 < t < 1$. Now, use *ComplexTool* to modify or strengthen your conjecture. Using *ComplexTool*, what happens to $f(\mathbb{D})$ for $t > 1$? For $t = ik, 0 \leq k \leq 1$?

Try it out!

The family S has been studied extensively for many years. Here are a few facts about normalized, analytic univalent functions that will be used later in our discussion:

(1) (uniqueness in the Riemann Mapping Theorem) Let $G \neq \mathbb{C}$ be a specific simply-connected domain with $a \in G$. Because of the Riemann Mapping Theorem, the map $f \in S$ that maps $\mathbb{D}$ onto G with $f(0) = a$ and $f'(0) > 0$ is unique.

(2) (deBranges Theorem) For $f \in S$, $|a_n| \leq n$, for all n.

(3) (Koebe $\frac{1}{4}$ Theorem) The range of every function in class S contains the disk $G = \{w : |w| < \frac{1}{4}\}$.

    Remark: This is a consequence of the fact that $|a_2| \leq 2$ which was proved by Bieberbach in 1916.

(4) Let $f \in S$. Then $f(\mathbb{D})$ omits some value on each circle $\{w : |w| = R\}$ where $R \geq 1$. In particular, there is no function $f \in S$ for which $f(\mathbb{D})$ contains $\partial\mathbb{D}$, the unit circle.

4.4. The Family S_H of Normalized, Harmonic, Univalent Functions

About the same time that deBranges proved the Bierbach Conjecture, Clunie and Sheil-Small studied a family, S_H, of complex-valued *harmonic* functions that contained S as a proper subset and considered some of the same properties on S_H that had been investigated in S.

Recall that function $\phi(x, y)$ is harmonic if and only if $\phi_{xx} + \phi_{yy} = 0$.

DEFINITION 4.26. A continuous function $f = u + iv$ defined in G is a *complex-valued, harmonic function* in G if u and v are real harmonic (but not necessarily harmonic conjugates) in G.

EXAMPLE 4.27. The function
$$f(x, y) = u(x, y) + iv(x, y) = (x^2 - y^2) + i2xy$$
is complex-valued, harmonic because
$$u_{xx} + u_{yy} = 2 - 2 = 0$$
$$v_{xx} + v_{yy} = 0 + 0 = 0.$$

EXERCISE 4.28. Show that
$$f(x, y) = u(x, y) + iv(x, y) = \left(x + \frac{1}{2}x^2 - \frac{1}{2}y^2\right) + i(y - xy)$$
is complex-valued harmonic.

Try it out!

Although harmonic functions are more general than analytic functions, some familiar theorems about analytic functions have an equivalent form for harmonic functions. These include the mean-value theorem, the maximum-modulus theorem, Liouville's Theorem, and the Argument Principle. However, by considering all harmonic functions instead of just the subclass of analytic functions we can sometimes get more information as is in the case in the study of minimal surfaces (see Chapter 2 on "Minimal Surfaces").

One way of thinking of a function $f(x,y) = u(x,y) + iv(x,y)$ as being analytic is that f can be expressed solely in terms of $z = x + iy$ without using $\bar{z} = x - iy$. Hence, the function $f = z^2$ is analytic while $f = z\bar{z}$ is not. To explore this idea, let's say that $\zeta := z = x + iy$ and $\xi := \bar{z} = x - iy$. Then, we can "formally" write $x = \frac{1}{2}(\zeta + \xi)$ and $y = \frac{1}{2i}(\zeta - \xi)$. Using the chain rule with the function $f(x(\zeta, \xi), y(\zeta, \xi))$ and since $\zeta = z$ and $\xi = \bar{z}$, we can show that

$$(46) \qquad \frac{\partial f}{\partial z} = \frac{1}{2}\left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y}\right) + \frac{i}{2}\left(\frac{\partial v}{\partial x} - \frac{\partial u}{\partial y}\right)$$

$$(47) \qquad \frac{\partial f}{\partial \bar{z}} = \frac{1}{2}\left(\frac{\partial u}{\partial x} - \frac{\partial v}{\partial y}\right) + \frac{i}{2}\left(\frac{\partial u}{\partial y} + \frac{\partial v}{\partial x}\right).$$

EXERCISE 4.29.

(a) Derive eqs (46) and (47).
(b) Use these equations and the Cauchy-Riemann equations to prove that $f(x,y) = u(x,y) + iv(x,y)$ is analytic $\iff \dfrac{\partial f}{\partial \bar{z}} = 0$.

Try it out!

EXERCISE 4.30.

(a) Using $x = \frac{1}{2}(z + \bar{z})$ and $y = \frac{1}{2i}(z - \bar{z})$, rewrite

$$f(x,y) = u(x,y) + iv(x,y) = \left(x + \frac{1}{2}x^2 - \frac{1}{2}y^2\right) + i(y - xy)$$

in terms of z and $\bar{z}$.
(b) Use Exercise 4.29 to determine if f is analytic.
(c) Show that all analytic functions are complex-valued harmonic, but not all complex-valued harmonic functions are analytic.

Try it out!

The next theorem tells us that a complex-valued harmonic function defined on $\mathbb{D}$ is related to analytic functions.

THEOREM 4.31. If $f = u + iv$ is harmonic in a simply-connected domain G, then $f = h + \bar{g}$, where h and g are analytic.

PROOF. Recall that if u and v are real harmonic on a simply-connected domain, then there exists analytic functions K and L such that $u = \operatorname{Re} K$ and $v = \operatorname{Im} L$. Hence,

$$f = u + iv = \operatorname{Re} K + i \operatorname{Im} L = \frac{K + \bar{K}}{2} + i\frac{L - \bar{L}}{2i} = \frac{K + L}{2} + \frac{\overline{K - L}}{2} = h + \bar{g}.$$

$\square$

EXERCISE 4.32. Let

$$f(x, y) = u(x, y) + iv(x, y) = \left(x + \frac{1}{2}x^2 - \frac{1}{2}y^2\right) + i(y - xy)$$

be defined on $\mathbb{D}$. In Exercise 4.28 you showed that f is harmonic. Find analytic functions h and g such that $f = h + \overline{g}$.
Try it out!

We can use the applet *ComplexTool* to graph complex-valued harmonic functions. For example, to graph the image of $\mathbb{D}$ under the harmonic function $f(z) = z + \frac{1}{2}\overline{z}^2$, enter this function in *ComplexTool* in the form $z + 1/2 \ \texttt{conj} \ (z \wedge 2)$ (see Figure 4.16).
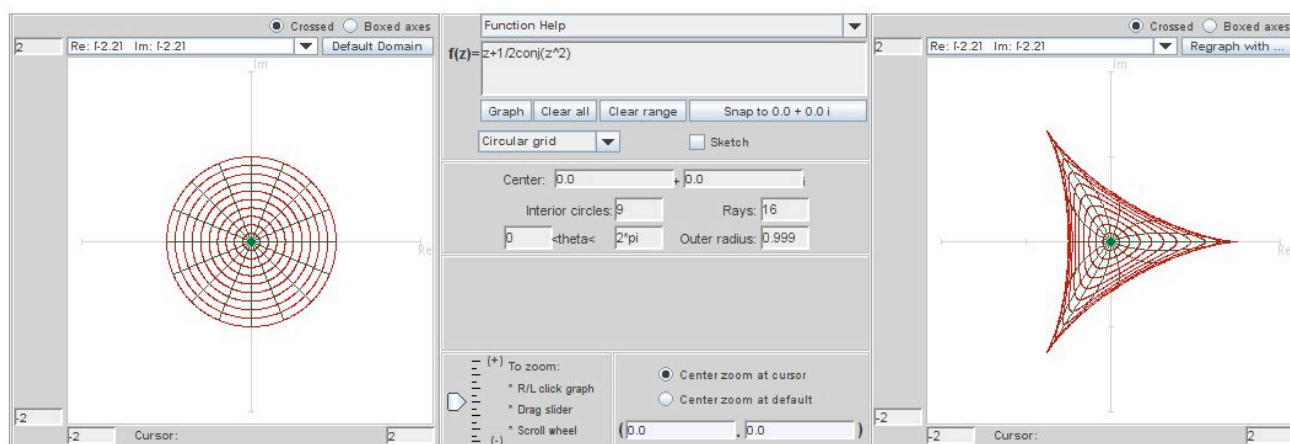
FIGURE 4.16. Image of $\mathbb{D}$ under the harmonic function $f(z) = z + \frac{1}{2}\overline{z}^2$

Note that the harmonic function $f(z) = h(z) + \overline{g}(z)$ can also be written in the form

(48) $$f(z) = \operatorname{Re}\left\{h(z) + g(z)\right\} + i\operatorname{Im}\left\{h(z) - g(z)\right\}.$$

Hence, in Example 4.32, $f(z) = z + \frac{1}{2}\overline{z}^2$ can also be written as $f(z) = \operatorname{Re}\left\{z + \frac{1}{2}z^2\right\} + i\operatorname{Im}\left\{z - \frac{1}{2}z^2\right\}$. In *ComplexTool* you can also enter the harmonic function in this form. To do so, you would type in $\texttt{re}(z + 1/2z \wedge 2) + \texttt{i*im}\,(z - 1/2z \wedge 2)$.

EXERCISE 4.33. Prove that the representations $f(z) = h(z) + \overline{g}(z)$ and $f(z) = \operatorname{Re}\left\{h(z) + g(z)\right\} + i\operatorname{Im}\left\{h(z) - g(z)\right\}$ are equivalent.
Try it out!

EXPLORATION 4.34. Graph the image of $\mathbb{D}$ under the following harmonic maps. Describe characteristics that appear to be different for harmonic mappings as compare

to analytic mappings.

(a) $f_1(z) = z + \frac{1}{3}\overline{z}^3$;

(b) $f_2(z) = \mathrm{Re}\left(\frac{z}{1-z}\right) + i\,\mathrm{Im}\left(\frac{z}{(1-z)^2}\right)$;

(c) $f_3(z) = \frac{z}{1-z} - \frac{1}{2}e^{\frac{\overline{z}+1}{\overline{z}-1}}$;

(d) $f_4(z) = \mathrm{Re}\left(\frac{i}{\sqrt{3}}\ln\left(\frac{1+e^{-i\frac{\pi}{3}}z}{1+e^{i\frac{\pi}{3}}z}\right)\right) + i\,\mathrm{Im}\left(\frac{1}{3}\ln\left(\frac{1+z+z^2}{1-2z+z^2}\right)\right)$;

(e) $f_5(z) = z + 2\ln(z+1) + (\overline{z}+1)e^{\frac{\overline{z}-1}{\overline{z}+1}}$.

Try it out!

Since $f = h + \overline{g}$, where h and g are analytic, f has the following series representation:

$$f(z) = \sum_{n=0}^{\infty} a_n z^n + \sum_{n=1}^{\infty} b_n \overline{z}^n.$$

Hence, we may normalize the harmonic univalent functions in a way similar to the normalized analytic univalent functions.

DEFINITION 4.35. Let S_H be the family of complex-valued harmonic, univalent mappings that are normalized on the unit disk. Specifically,

$S_H = \{f : \mathbb{D} \to \mathbb{C} \mid f$ is harmonic, univalent

$\qquad\qquad$ with $f(0) = a_0 = 0$, $f_z(0) = a_1 = 1\}$.

$$S_H^O = \{f \in S_H \mid f_{\overline{z}}(0) = b_1 = 0\}.$$

Thus, $S \subset S_H^O \subset S_H$.

Let's look at some significant examples.

EXAMPLE 4.36. **Harmonic polynomial map**
Consider

$$f(z) = h(z) + \overline{g(z)} = z + \frac{1}{2}\overline{z}^2.$$

In the next section we will prove that f is univalent and hence in S_H^O. But for now we will assume this and look at the image of $\mathbb{D}$ under f (see Figure 4.16). The function f maps $\mathbb{D}$ onto the interior of the region bounded by a hypocycloid with 3 cusps.

DEFINITION 4.37. A *hypocycloid* is the curve produced by a fixed point p on a small circle of radius b rolling the inside of a larger circle of radius a:

$$x(\theta) = (a - b)\cos\theta + b\cos\left(\left(\frac{a}{b} - 1\right)\theta\right)$$

$$y(\theta) = (a - b)\sin\theta + b\sin\left(\left(\frac{a}{b} - 1\right)\theta\right).$$

In Example 4.53, we will see that the function f is a shearing of the function $F(z) = z - 1/2z^2$ which maps $\mathbb{D}$ to an epicycloid (see Example 4.19).

EXPLORATION 4.38.

(a) Use *ComplexTool* to plot the image of $\mathbb{D}$ under the analytic polynomial map $F(z) = z - frac12e^{it\frac{\pi}{6}}z^2$ for $t = 0, 1, 2, 3, 4, 5, 6$. Describe what happens to the image as t varies.

(b) Use *ComplexTool* to plot the image of $\mathbb{D}$ under the harmonic polynomial map $f(z) = z + 1/2e^{it\frac{\pi}{6}}\overline{z}^2$ for $t = 0, 1, 2, 3, 4, 5, 6$. Describe what happens to the image as t varies.

(c) What differences do you notice between the images in (a) and (b) as t increases? Explain why it is reasonable for this difference to occur.

Try it out!

SMALL PROJECT 4.39.

(a) Use *ComplexTool* to plot the images of the following polynomials:

Harmonic Functions	Analytic Functions
(i) $f_1(z) = z + \frac{1}{3}\overline{z}^3$;	(ii) $F_1(z) = z + \frac{1}{3}z^3$;
(iii) $f_2(z) = z + \frac{1}{4}z^2 + \frac{1}{4}\overline{z}^2 + \frac{1}{3}\overline{z}^3$;	(iv) $F_2(z) = z + \frac{1}{2}z^2 + \frac{1}{3}z^3$;
(v) $f_3(z) = z + \frac{1}{6}\overline{z}^2 + \frac{1}{6}\overline{z}^4$;	(vi) $F_3(z) = z + \frac{1}{6}z^6$;
(vii) $f_4(z) = z + \frac{1}{6}z^2 + \frac{1}{6}\overline{z}^4$.	

(b) Write a list of similarities and differences between the images of $\mathbb{D}$ under harmonic and the analytic functions.

(c) Some questions to consider are: If the polynomial has three or more terms then how large can their coefficients be in modulus to guarantee univalency? If the polynomial has three terms, what difference does it make if the last two terms are $\overline{z}^2$ and $\overline{z}^3$ instead of z^2 and $\overline{z}^3$?

(d) Plot your own examples of harmonic and analytic polynomials and see if the properties in your list from (b) are still valid.

Optional

EXAMPLE 4.40. **Harmonic right half-plane map**

Consider

$$f(z) = h(z) + \overline{g(z)} = \frac{z - \frac{1}{2}z^2}{(1-z)^2} - \frac{\frac{1}{2}\overline{z}^2}{(1-\overline{z})^2}$$

$$= \operatorname{Re}(h(z) - g(z)) + i\operatorname{Im}(h(z) - g(z)) = \operatorname{Re}\left(\frac{z}{1-z}\right) + i\operatorname{Im}\left(\frac{z}{(1-z)^2}\right).$$

We will prove that f is univalent in the next section. The image of $\mathbb{D}$ under f using *ComplexTool* is shown in Figure 4.17.
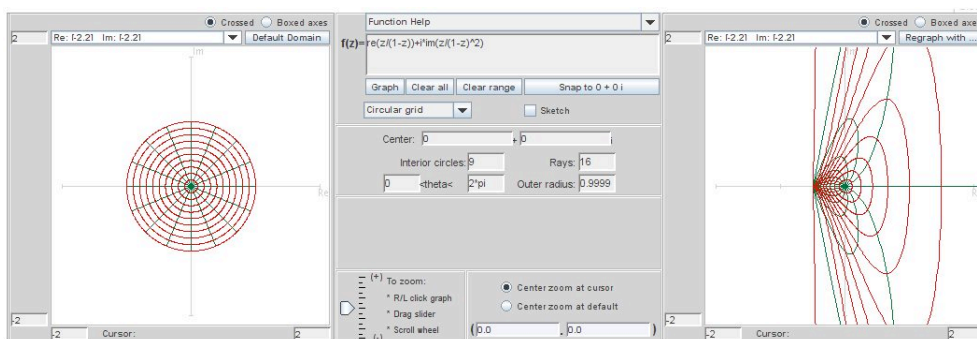
FIGURE 4.17. Image of $\mathbb{D}$ under $f(z) = \operatorname{Re}\left(\frac{z}{1-z}\right) + i\operatorname{Im}\left(\frac{z}{(1-z)^2}\right)$

It turns out that the image of $\mathbb{D}$ under the harmonic map f is the right half-plane $\{w \in \mathbb{C} \mid \operatorname{Re}\{w\} \geq -\frac{1}{2}\}$. This is the same region as the image of $\mathbb{D}$ under the analytic map $\frac{z}{1-z}$ although the boundary behavior is different.

EXPLORATION 4.41.

(a) Use *ComplexTool* to plot $\mathbb{D}$ under the analytic right half-plane map $\frac{z}{1-z}$. Use the `Sketch` box to draw radial lines from the origin to the boundary of $\mathbb{D}$ in the original domain. What are the images of points on the unit circle under this analytic map?

(b) Use *ComplexTool* to plot $\mathbb{D}$ under the harmonic right half-plane map $\frac{z - \frac{1}{2}z^2}{(1-z)^2} -$ $\frac{\frac{1}{2}\overline{z}^2}{(1-\overline{z})^2}$. Use the `Sketch` box to draw radial lines from the origin to the boundary of $\mathbb{D}$ in the original domain. What is the image of points on the unit circle under this analytic map?

(c) Using (a) and (b), describe how the boundary behavior is different between the analytic right half-plane map and this harmonic right half-plane map?

Try it out!

We are interested in harmonic univalent functions. However, results about univalent functions are difficult to obtain. So it is sometimes useful to just consider locally univalent functions as opposed to globally univalent (a.k.a univalent). In general, locally univalent functions are functions that are locally $1 - 1$. This means that there is a small neighborhood around a point $z_0 \in \mathbb{D}$ such as a small disk in $\mathbb{D}$ centered at z_0, and the function is $1 - 1$ for all points z in that neighborhood. Note that a locally univalent function is not guaranteed to be univalent in $\mathbb{D}$.

EXPLORATION 4.42. Earlier in Exploration 4.15, you proved that $f(z) = (1+z)^4$ is not univalent in $\mathbb{D}$. This can be shown by noticing that the points $-\frac{1}{2} + \frac{1}{2}i$ and $-\frac{1}{2} - \frac{1}{2}i$ in $\mathbb{D}$ both get mapped to $-\frac{1}{4}$ under $f(z) = (1+z)^4$. However, $f(z) = (1+z)^4$ is locally univalent in $\mathbb{D}$. We won't prove this. Instead, we just want to get a sense of what local univalence means. To do this, use *ComplexTool* to graph $f(z) = (1+z)^4$ on the sector $\{z = re^{i\theta} \in \mathbb{D} \mid \frac{\pi}{2} < \theta < \frac{3\pi}{2}\}$. You can do this by entering these values for θ into the left and right boxes for the "$\leq \theta \leq$" feature in the center panel of the applet. Notice that the point $z_0 = -\frac{1}{2} + \frac{1}{2}i$ is in this sector. Also, notice that the image indicates that $f(z) = (1+z)^4$ is not univalent in this sector. This does not mean that f is not locally univalent. Remember, local univalence means that at z_0 there is <u>some</u> neighborhood in which the function is $1 - 1$ for all points z in that neighborhood. Use *ComplexTool* to determine a neighborhood of the points z_0 in which f is $1 - 1$. Remember that you can increase the size of the image by clicking on the left button on the mouse and decrease the size by clicking on the right mouse button.

(a) $z_0 = -\frac{1}{2} + \frac{1}{2}i$;   (b) $z_0 = -\frac{1}{2} - \frac{1}{2}i$;

(c) $z_0 = -\frac{5}{6} + \frac{1}{6}i$;   (d) $z_0 = -0.9 + 0.1i$.

Try it out!

Let's look at the idea of local univalence for for a moment

DEFINITION 4.43. A function $f = h + \overline{g}$ is *locally univalent* on G if $J_f \neq 0$ on G, where J_f is the Jacobian of $f = u + iv$:

$$J_f = \det \begin{vmatrix} u_x & u_y \\ v_x & v_y \end{vmatrix}$$

$$= \det \begin{vmatrix} (\text{Re } h)_x + (\text{Re } g)_x & (\text{Re } h)_y + (\text{Re } g)_y \\ (\text{Im } h)_x - (\text{Im } g)_x & (\text{Im } h)_y - (\text{Im } g)_y \end{vmatrix}.$$

262

For analytic functions F, the Cauchy-Riemann equations yield $(\text{Re } F)_y = -(\text{Im } F)_x$ and $(\text{Im } F)_y = (\text{Re } F)_x$. Hence we have

$$= \det \begin{vmatrix} (\text{Re } h)_x + (\text{Re } g)_x & -(\text{Im } h)_x - (\text{Im } g)_x \\ (\text{Im } h)_x - (\text{Im } g)_x & (\text{Re } h)_x - (\text{Re } g)_x \end{vmatrix}$$

$$= (\text{Re } h)_x^2 - (\text{Re } g)_x^2 + (\text{Im } h)_x^2 - (\text{Im } g)_x^2$$

$$= |h'|^2 - |g'|^2.$$

Thus, we want $|h'|^2 - |g'|^2 \neq 0$.

Besides local univalence, another important property of these functions is sense-preserving. What is sense-preserving? A continuous function f is *sense-preserving* (a.k.a. *orientation-preserving*) if it preserves orientation. Consider the following example. Let f_1, f_2 be defined on the punctured disk $\mathbb{D} - \{0\}$ by

$$f_1(z) = \frac{1}{z} \qquad \text{and} \qquad f_2(z) = \overline{z}.$$

Both functions map the unit circle, $\partial \mathbb{D}$, onto itself. In particular, both functions map the points $A = 1$, $B = e^{i\frac{\pi}{4}}$, and $C = i$ to the points $A' = 1$, $B' = e^{-i\frac{\pi}{4}}$, and $C' = -i$, respectively (see Figure 4.18). Also, notice that in the domain as we travel along the unit circle in a counterclockwise direction (i.e., going from A to B to C), the region to the left of this path is $\mathbb{D}$. We will call this region the left hand side domain (LHS) and the region to the right of the path we will call the right hand side domain (RHS). So, in this case, $\mathbb{C} - \mathbb{D}$ is RHS. Now, where does $\mathbb{D}$, a LHS, get mapped under these two functions? Note that as we travel along $\partial \mathbb{D}$ in a counterclockwise direction in the domain set, the image curve under both functions will be $\partial \mathbb{D}$ traversed in a clockwise direction. So, in the image, $\mathbb{D}$ is now RHS while $\mathbb{C} - \mathbb{D}$ is LHS. The function f_1 maps the point $\frac{1}{2} \in \mathbb{D}$ to $2 \in \mathbb{C} - \mathbb{D}$ and so f_1 maps the LHS onto the LHS. Functions that map the LHS onto the LHS are sense-preserving. On the other hand, f_2 maps the point $\frac{1}{2} \in \mathbb{D}$ to $\frac{1}{2} \in \mathbb{D}$ and so f_2 maps the LHS onto the RHS. Functions that map the LHS onto the RHS are sense-reversing.
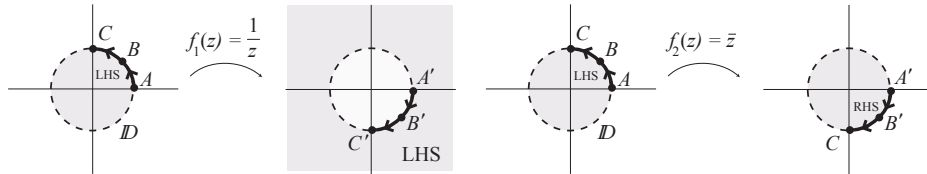
FIGURE 4.18. f_1 is sense-preserving, while f_2 is sense-reversing.

More exactly, as we travel counterclockwise along any simple closed contour $\gamma \in G$, there exists a left hand side domain (LHS) and a right hand side domain (RHS). Consider the image curve $f(\gamma)$. The function f is *sense-preserving* if the original LHS

domain with regard to γ is mapped to the LHS domain with regard to $f(\gamma)$. The function f is *sense-reversing* if the original LHS domain with regard to γ is mapped to the RHS domain with regard to $f(\gamma)$. It turns out that all analytic functions are sense-preserving, while some complex-valued harmonic functions are sense-preserving and some are sense-reserving.

EXPLORATION 4.44. For each of the following harmonic functions, use *ComplexTool* to make a conjecture if the function is: (a) locally univalent; and (b) sense-preserving [Hint: for sense-preserving, you can use the Sketch feature to draw a counterclockwise curve on the unit circle and see its image under the function]

$$\text{(a) } z + 2\overline{z}; \qquad\qquad \text{(b) } z + \tfrac{1}{2}\overline{z}$$

$$\text{(c) } z + 2\overline{z}^2; \qquad\qquad \text{(d) } z + \tfrac{1}{2}\overline{z}^2$$

$$\text{(e) } 2z^2 + \overline{z}; \qquad\qquad \text{(f) } \tfrac{1}{2}z^2 + \overline{z}$$

Try it out!

Now we will need the following important definition.

DEFINITION 4.45. $\omega(z) = g'(z)/h'(z)$ is known as the *dilatation* of $f = h + \overline{g}$.

There is a connection between the dilatation of a harmonic function and its locally univalent and sense-preserving nature.

THEOREM 4.46 (Lewy). The function $f = h + \overline{g}$ is locally univalent and sense-preserving $\iff |\omega(z)| < 1$ for all $z \in G$.

EXERCISE 4.47. Compute $\omega(z)$ for each of the functions in Exploration 4.44. Then use your results from Exploration 4.44 to verify that Lewy's Theorem for the functions in that Exploration.

Try it out!

EXERCISE 4.48. Show that for $z \in \mathbb{D}$, $|\omega_k(z)| < 1$ for:

(a) $\omega_1(z) = e^{i\theta}z$, where $\theta \in \mathbb{R}$;

(b) $\omega_2(z) = z^n$, where $n = 1, 2, 3, \ldots$;

(c) $\omega_3(z) = \dfrac{z + a}{1 + \overline{a}z}$, where $|a| < 1$;

(d) $\omega_4(z)$ being the composition of any of the functions ω above.

Try it out!

REMARK 4.49. Let $f = h + \overline{g}$ be a sense-preserving harmonic map with dilatation $\omega = g'/h'$. If $|\omega(z)| = 1$ for all z in an arc γ of $\partial\mathbb{D}$, then the image of γ under f is either:

- a concave arc (i.e., there are points in $f(\mathbb{D})$ such that the line connecting these points goes outside of $f(\mathbb{D})$) (see Example 4.36); or
- a stationary point (see Example 4.40).

We discuss the dilatation more in Section 4.6.

4.5. The Shearing Technique

Finding examples of univalent harmonic mappings that are not analytic is not easy. One very useful way to construct new examples of univalent harmonic mappings was provided by Clunie and Sheil-Small. It is known as the *shearing technique*. Basically, the shearing technique starts with a given analytic function F and a given dilatation ω both of which have certain properties. Then, we can find a <u>univalent</u> harmonic function $f = h + \overline{g}$ by writing F as $F = h - g$ and ω as $\omega = g'/h'$, and explicitly solving for h and g. Before we proceed, we need to discuss a needed property of F.

DEFINITION 4.50. A domain Ω is convex in the direction $e^{i\varphi}$ if for every $a \in \mathbb{C}$ the set $\Omega \cap \{a + te^{i\varphi} : t \in \mathbb{R}\}$ is either connected or empty. In particular, a domain is convex in the horizontal direction (CHD) if every line parallel to the real axis has a connected intersection with Ω.

CHD   not CHD
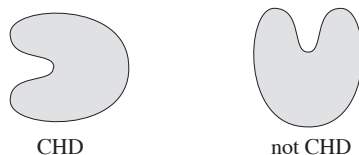
EXERCISE 4.51. For which values of $n = 1, 2, 3, \ldots$ do the following functions map $\mathbb{D}$ onto a CHD domain?

(1) $f(z) = z^n$,
(2) $f(z) = z - \frac{1}{n}z^n$ (see Example 4.19 and Definition 4.20),
(3) $f(z) = \frac{z}{(1-z)^n}$ (see Examples 4.22 and 4.23 to get you started).

Try it out!

Next, we have a theorem by Clunie and Sheil-Small that forms the basis of the shearing technique.

THEOREM 4.52. Let $f = h + \overline{g}$ be a harmonic function that is locally univalent in $\mathbb{D}$ (i.e., $|\omega(z)| < 1$ for all $z \in \mathbb{D}$). The function $F = h - g$ is an analytic univalent mapping of $\mathbb{D}$ onto a CHD domain $\Longleftrightarrow f = h + \overline{g}$ is a univalent mapping of $\mathbb{D}$ onto a CHD domain.

Before we prove Theorem 4.52, let's look at an example.

EXAMPLE 4.53. **Harmonic polynomial map**

Recall that the shearing technique starts with a given univalent analytic function F that maps $\mathbb{D}$ onto a CHD domain and a given dilatation ω where $|\omega(z)| < 1$ for all $z \in \mathbb{D}$. Then, we can construct a univalent harmonic function $f = h + \overline{g}$ by writing F as $F = h - g$ and ω as $\omega = g'/h'$, and explicitly solving for h and g.

In Example 4.36, we claimed that the harmonic polynomial $f(z) = z + \frac{1}{2}\overline{z}^2$ (see Figure 4.19) is univalent and is related to the analytic function $F(z) = z - \frac{1}{2}z^2$ (see Figure 4.20). We can use Theorem 4.52 to show this.
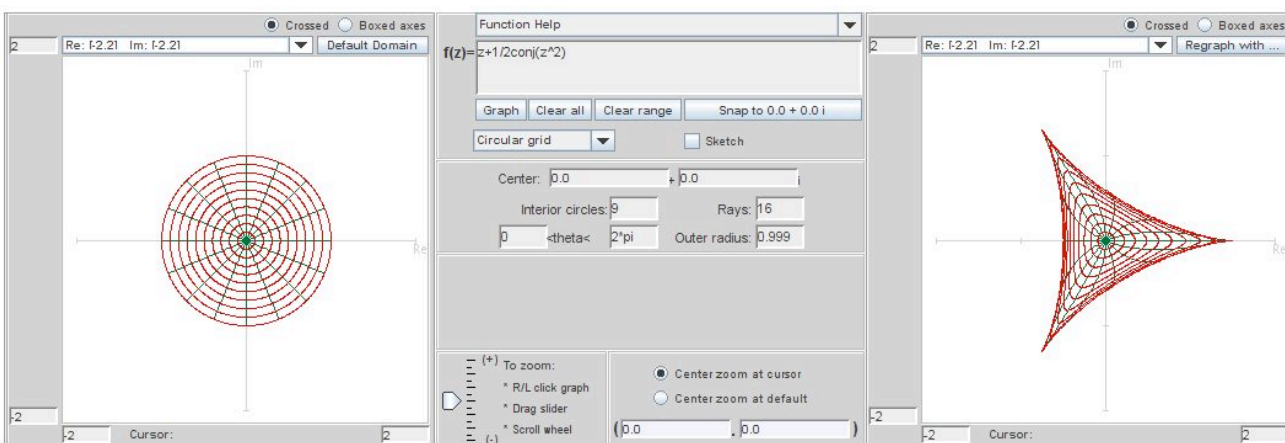
FIGURE 4.19. Image of $\mathbb{D}$ under the harmonic map $f(z) = z + \frac{1}{2}\overline{z}^2$.
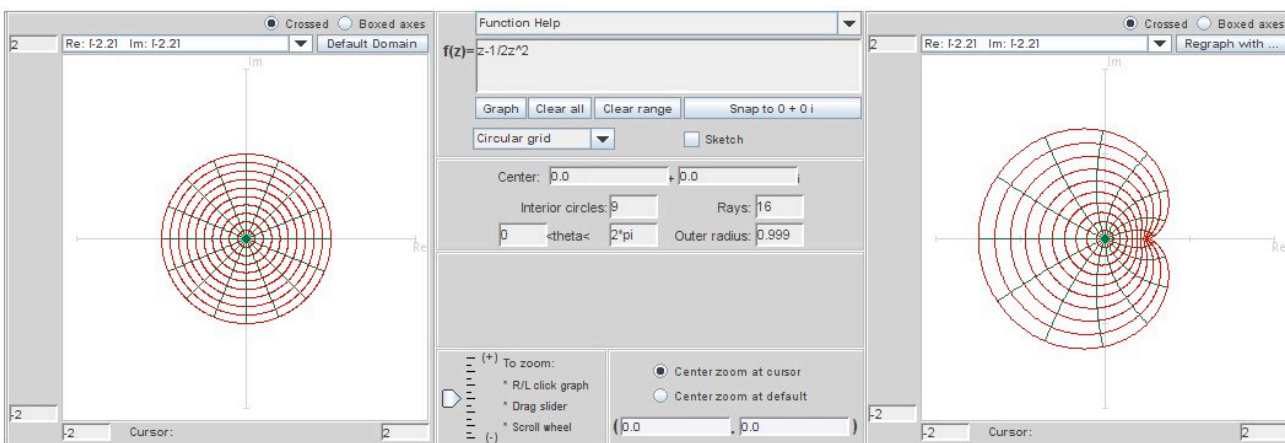
FIGURE 4.20. Image of $\mathbb{D}$ under the analytic map $F(z) = z - \frac{1}{2}z^2$.

Let's start with the analytic univalent function

$$F(z) = h(z) - g(z) = z - \frac{1}{2}z^2.$$

Recall that F maps $\partial\mathbb{D}$ to an epicycloid with 1 cusp (see Example 4.19) which is a CHD domain. Next, choose $\omega(z) = g'(z)/h'(z) = z$. Hence, we can apply the shearing technique and solve for h and g:

$$h'(z) - g'(z) = 1 - z \Rightarrow h'(z) - zh'(z) = 1 - z$$
$$\Rightarrow h'(z) = 1$$
$$\Rightarrow h(z) = z.$$

Since $g'(z) = zh'(z) = z$, we also have $g(z) = \frac{1}{2}z^2$. Notice that h and g are normalized; that is, $h(0) = 0$ and $g(0) = 0$. So, the corresponding harmonic univalent function is

$$f(z) = h(z) + \overline{g(z)} = z + \frac{1}{2}\overline{z}^2 \in S_H^O$$

which was derived from shearing the analytic univalent function

$$F(z) = z - \frac{1}{2}z^2 \in S$$

with the dilatation $\omega(z) = z$.

REMARK 4.54. This technique is known as "shearing." The word *shear* means to cut (as in "shearing a sheep to get its wool"). As in Theorem 4.52, suppose $F = h - g$ is an analytic univalent function convex in the horizontal direction. Then the corresponding harmonic shear is

$$f = h + \overline{g} = h - g + g + \overline{g} = h - g + 2\operatorname{Re}\{g\}.$$

So, the harmonic shear differs from the analytic function by adding a real function to it. Geometrically, you can think of this as taking the image $F(\mathbb{D})$, that is convex in the horizontal direction and is formed from the analytic univalent function F, and cutting it up into thin horizontal slices which are then translated and/or scaled in a continuous way to form the image $f(\mathbb{D})$, that is also convex in the horizontal direction and is formed from the harmonic univalent function f (see Figure 4.21). This is why the method is called "shearing."

EXERCISE 4.55. Let $f = h + \overline{g}$ with $h(z) - g(z) = z$ and $\omega(z) = z$. Compute h and g explicitly so that $f \in S_H^O$ and use *ComplexTool* to sketch $f(\mathbb{D})$.

EXERCISE 4.56. Let $f = h + \overline{g}$ with $h(z) - g(z) = z - \frac{1}{3}z^3$ and $\omega(z) = z^2$. Compute h and g explicitly so that $f \in S_H^O$ and use *ComplexTool* to sketch $f(\mathbb{D})$.

For exploring shears of functions, the applet *ShearTool* (see Figure 4.22) has an advantage over *ComplexTool*, because it easily allows you to see the image of $\mathbb{D}$ under a shear of $h(z) - g(z)$ without having to explicitly compute the resulting harmonic
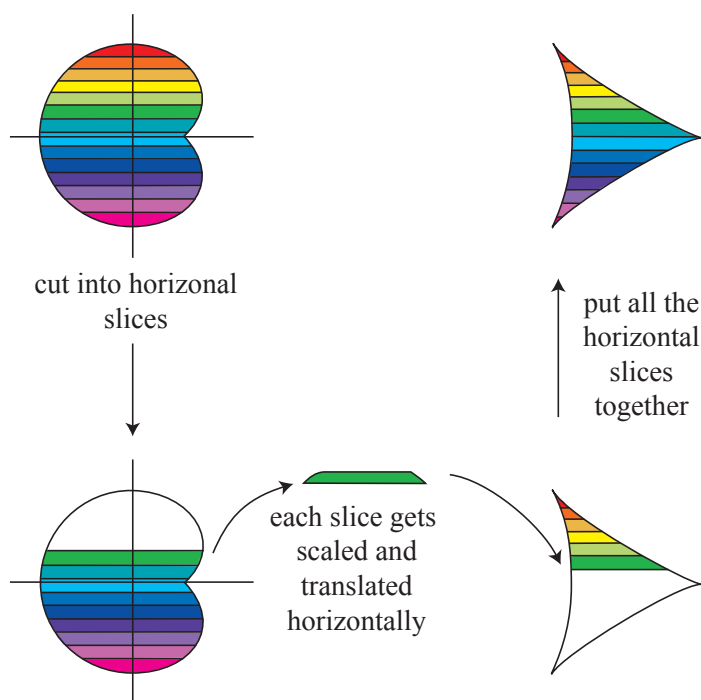
FIGURE 4.21. Shearing an analytic function to preserve univalency

function $f = h + \overline{g}$. When using *ShearTool*, enter an analytic function that is convex in the horizontal direction in **h-g=** box in the upper left section and enter the dilatation in the ω box below it. The click on the `Graph` button.

In Example 4.53, we sheared $h(z) - g(z) = z - \frac{1}{2}z^2$ with $\omega(z) = z$. Entering these two functions into *ShearTool* we get the image of a hypocycloid with 3 cusps (see Figure 4.23).

EXPLORATION 4.57.

(a.) Use *ShearTool* to graph the image of $\mathbb{D}$ under $f = h + \overline{g}$ where $h(z) - g(z) = z$ and $\omega(z) = -z$. Note the slight difference between this and Exercise 4.55.

(b.) Use *ShearTool* to graph the image of $\mathbb{D}$ under $f = h + \overline{g}$ where $h(z) - g(z) = z - \frac{1}{3}z^3$ and $\omega(z) = -z^2$. Note the slight difference between this and Exercise 4.56.

Try it out!

To prove Theorem 4.52, we will use the following lemma

LEMMA 4.58. Let $\Omega \subset \mathbb{C}$ be a CHD domain and let ρ be a real-valued continuous function in Ω. Then the map $\Psi(w) = w + \rho(w)$ is one-to-one in $\Omega \iff \Psi$ is locally one-to-one. If Ψ is one-to-one, then its range is a CHD domain.
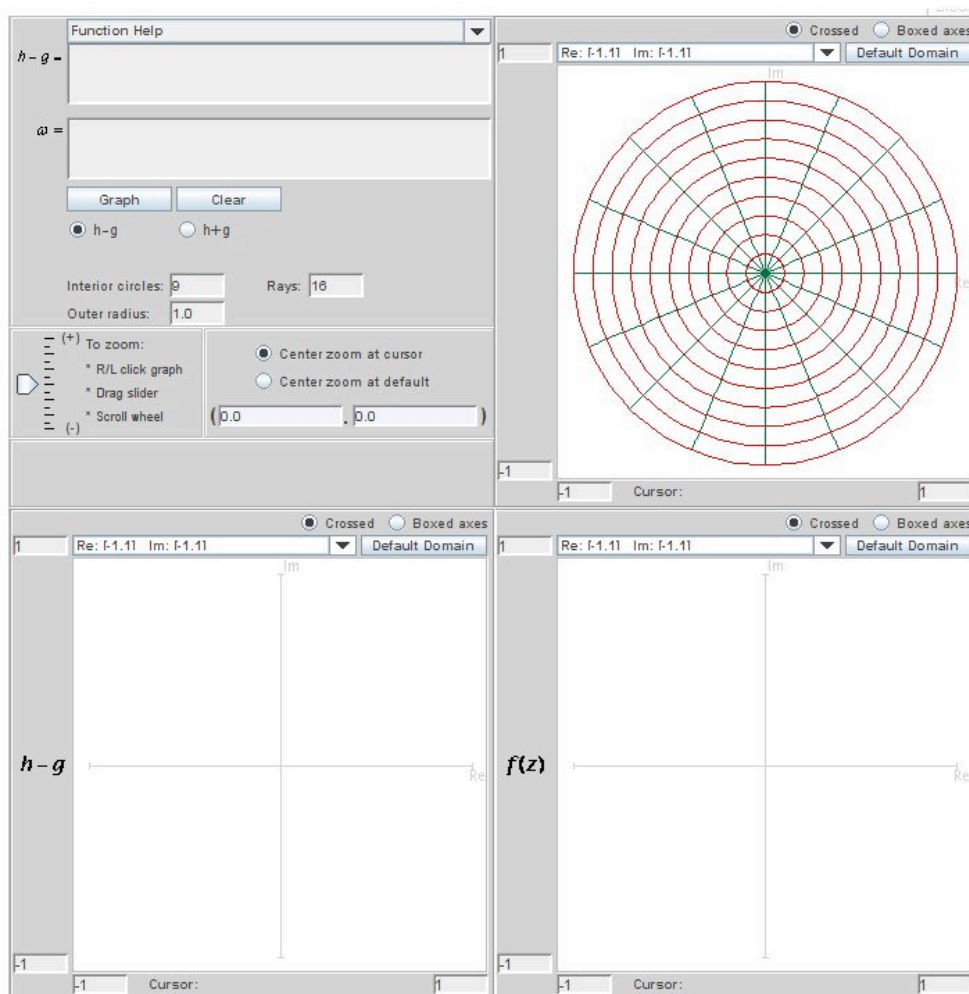
FIGURE 4.22. The applet *ShearTool*

PROOF OF LEMMA. ($\Rightarrow$) Trivial.

($\Leftarrow$) Suppose the map $\Psi(w) = w + \rho(w)$ is not $1-1$. Then, there are distinct points $w_1 = u_1 + iv_1$, $w_2 = u_2 + iv_2 \in \Omega$ such that $\Psi(w_1) = \Psi(w_2)$. Since ρ is real-valued, we have that $\mathrm{Im}(\Psi(w_1)) = \mathrm{Im}(w_1 + \rho(w_1)) = \mathrm{Im}(w_1) = v_1$. Similarly, $\mathrm{Im}(\Psi(w_2)) = v_2$. Since $\Psi(w_1) = \Psi(w_2)$, we must have $v_1 = v_2$. Let's say $k = v_1 = v_2$ for some constant $k \in \mathbb{R}$. Now consider the map $\Phi : \widetilde{\Omega} \subset \mathbb{R} \to \mathbb{R}$ defined by $\Phi(u) = u + \rho(u + ik)$. Note that because of the assumption that Ψ is not $1-1$, we have

$$\Phi(u_1) = u_1 + \rho(u_1 + k) = \mathrm{Re}\{\Psi(w_1)\} = \mathrm{Re}\{\Psi(w_2)\} = u_2 + \rho(u_2 + k) = \Phi(u_2).$$

Thus, Φ is not a strictly monotonic function. Let's say u_0 is the point at which the monotonicity of Φ changes. Because Φ maps into $\mathbb{R}$, the function Φ cannot be locally

269
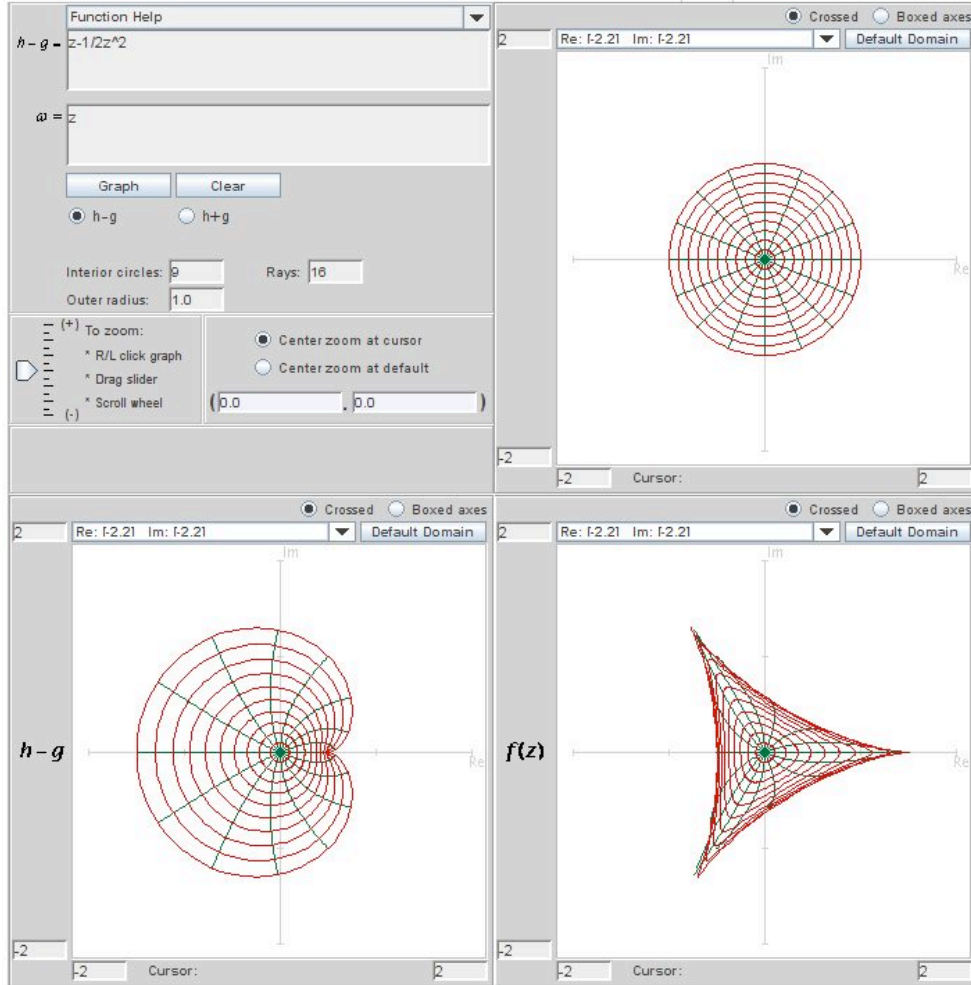
FIGURE 4.23. The image of $\mathbb{D}$ when shearing $h(z) - g(z) = z - \frac{1}{2}z^2$ with $\omega(z) = z$.

$1 - 1$ in any neighborhood of u_0. Hence, Φ is not locally $1 - 1$, and consequently, Ψ is not locally $1 - 1$.

Geometrically, Ψ acts as a "shear" in the horizontal direction and hence its range is CHD. $\qquad\square$

PROOF OF THEOREM 4.52. ($\Rightarrow$) Assume $f = h + \overline{g}$ is $1-1$ and $\Omega = f(\mathbb{D})$ is CHD. Note that $f = h - g + g + \overline{g} = h - g + 2\operatorname{Re}\{g\}$. Then the function

$$(h - g) \circ f^{-1}(w) = (f - 2\operatorname{Re}\{g\}) \circ f^{-1}(w) = w - 2\operatorname{Re}\{g(f^{-1}(w))\} = w + p(w)$$

may be defined in Ω, where p is real-valued and continuous. Since f is locally $1-1$, $|g'| < |h'| \iff g'(z) \neq h'(z)$, $\forall z \in \mathbb{D}$. Hence $h - g$ is locally $1-1$ in $\mathbb{D}$, and thus

270

$w \to w + p(w)$ is also locally $1-1$ on Ω since it is the composition of locally $1-1$ functions. By Lemma 4.58, $w \to w + p(w)$ is in fact univalent and its range is CHD. Hence, $(h-g)(z) = [w + p(w)] \circ f(z)$ is univalent being the composition of univalent functions, and its image is CHD.
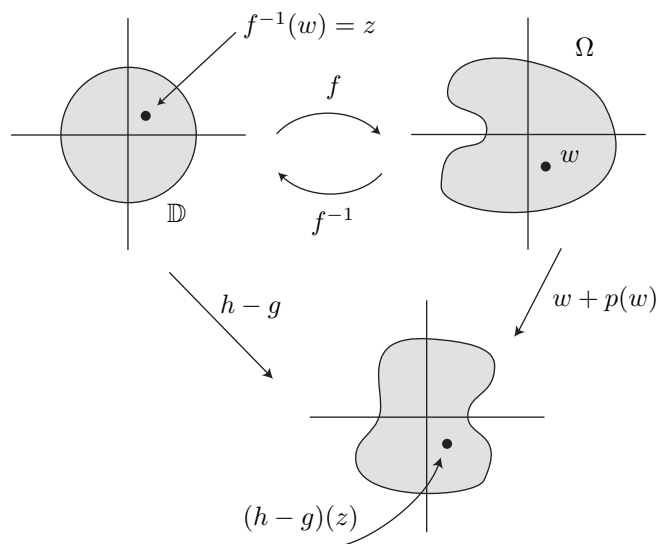
($\Leftarrow$) Now assume that $F = h - g$ is univalent on $\mathbb{D}$ and that $\Omega = F(\mathbb{D})$ is CHD. Then $f = F + 2\mathrm{Re}\{g\}$, and

$$f(F^{-1}(w)) = w + 2\mathrm{Re}\{g(F^{-1}(w))\}$$
$$= w + q(w)$$

271
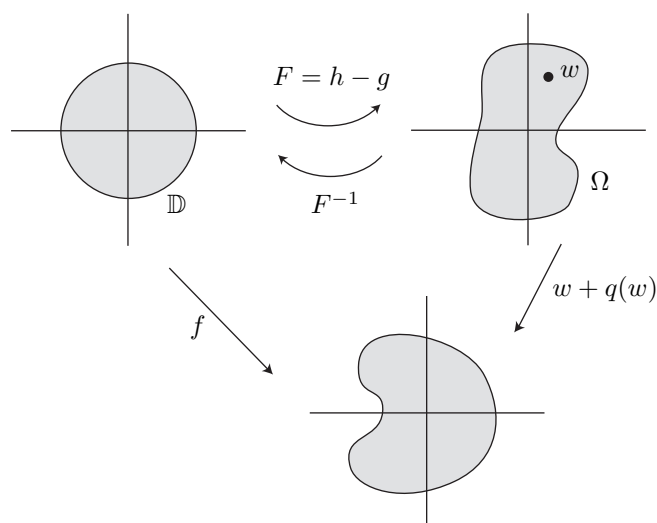
is locally 1-1 (being the composition of locally 1-1 functions) in Ω. By Lemma 4.58, $f \circ F^{-1}$ is univalent in Ω and has a range that is CHD.

$\square$

EXAMPLE 4.59. **Harmonic Koebe map**

Since the Koebe function is an important function in S (see Example 4.23), let's see what happens when it is sheared with a standard dilatation. In particular, let

(49)
$$h(z) - g(z) = \frac{z}{(1-z)^2}$$

and let

$$\omega(z) = g'(z)/h'(z) = z.$$

Apply the shearing technique, we have

$$h'(z) - g'(z) = \frac{1+z}{(1-z)^3} \Rightarrow h'(z) - zh'(z) = \frac{1+z}{(1-z)^3}$$

$$\Rightarrow h'(z) = \frac{1+z}{(1-z)^4}.$$

Integrating $h'(z)$ and normalizing so that $h(0) = 0$, yields

(50)
$$h(z) = \frac{z - \frac{1}{2}z^2 + \frac{1}{6}z^3}{(1-z)^3}.$$

We can use this same method to solve for normalized $g(z)$, where $g(0) = 0$. However, this time we will find $g(z)$ by using eqs. (49) and (50).

$$g(z) = h(z) - \frac{z}{(1-z)^2} = \frac{\frac{1}{2}z^2 + \frac{1}{6}z^3}{(1-z)^3}.$$

So

$$f(z) = h(z) + \overline{g(z)} = \mathrm{Re}\left(\frac{z + \frac{1}{3}z^3}{(1-z)^3}\right) + i\,\mathrm{Im}\left(\frac{z}{(1-z)^2}\right) \in S_H^O.$$

What is the image of $\mathbb{D}$ under f? It turns out that $f(\mathbb{D})$ is similar to the image of $\mathbb{D}$ under the analytic Koebe function (see Figure 4.15) with the slit on the negative real axis except in this case the tip of the slit is at $-\frac{1}{6}$ instead of $-\frac{1}{4}$. Recall in Example 4.23 we used the transformation $\frac{1+z}{1-z} = w = u + iv$. Let's use this transformation again (in general, it is useful when there are horizontal slit domains). Note that since $z \in \mathbb{D}$, $w = \frac{1+z}{1-z}$ is the right half-plane $\{w = u + iv \in \mathbb{C} \mid \mathrm{Re}\,w = u > 0, -\infty < v < \infty\}$. Then $z = \frac{w-1}{w+1}$. Substituting this into $h(z)$ and $g(z)$ and simplifying, we get:

$$h\left(\frac{w-1}{w+1}\right) = \frac{1}{8}\left[\frac{2}{3}w^3 + w^2 - \frac{5}{3}\right]$$

$$g\left(\frac{w-1}{w+1}\right) = \frac{1}{8}\left[\frac{2}{3}w^3 - w^2 + \frac{1}{3}\right].$$

272

Recall that $f = \mathrm{Re}(h + g) + i\,\mathrm{Im}(h - g)$. Thus,

$$f\left(\frac{w-1}{w+1}\right) = \frac{1}{6}\mathrm{Re}\left\{w^3 - 1\right\} + \frac{1}{4}i\,\mathrm{Im}\left\{w^2 - 1\right\}.$$

Using $w = u + iv$, and taking the real and imaginary parts, this becomes

$$\frac{1}{6}\left(u^3 - 3uv^2 - 1\right) + i\frac{1}{2}uv.$$

If we let $uv = 0$ (and so, $v = 0$), then the imaginary part vanishes and because $u > 0$, the real part varies from $-\frac{1}{6}$ to $+\infty$. Thus, for $uv = 0$, $f(\mathbb{D})$ contains the line segment on the real axis from $-\frac{1}{6}$ to $+\infty$. On the other hand, if we let $uv = c \neq 0$, then the imaginary part is constant and the real part is $\frac{u^3}{6} - \frac{c^2}{2u}$ which varies between $-\infty$ and $+\infty$. Thus, for any $c \neq 0$, $f(\mathbb{D})$ contains the entire line parallel to the real axis and through the point ic. Therefore, $f(\mathbb{D})$ is the entire complex plane except the slit on the negative real axis from $-\frac{1}{6}$ to $-\infty$.

EXERCISE 4.60. Let $f = h + \overline{g}$ with $h(z) - g(z) = \frac{z}{(1-z)^2}$ and $\omega(z) = z^2$. Compute h and g explicitly so that $f \in S_H^O$ and determine $f(\mathbb{D})$.
   Try it out!

EXPLORATION 4.61. If we shear $h(z) - g(z) = \frac{z}{(1-z)^2}$ with $\omega(z) = \frac{z^2 + az}{1 + az}$, where $-1 < a \leq 1$, then the image of $\mathbb{D}$ under $f = h + \overline{g}$ is a slit domain (like the image of $\mathbb{D}$ under the analytic and harmonic Koebe maps) with the tip of the slit varying as a varies. Use *ShearTool* to graph these domains while decreasing a from 1 to -1 [Note: You might find it beneficial to set the `Outer radius` value to less than 1.0. Try to place the cursor on the tip of the slit in the $f(z)$−domain box and look at the coordinates below to estimate the distance of the tip from the origin. How close to the origin can you get the tip of the slit? How far from the origin can you get the tip of the slit? Try finding other valid ω expressions that result in slit domains when shearing $h(z) - g(z) = \frac{z}{(1-z)^2}$ (This exploration is developed further in Small Project 4.71).
   Try it out!

Because of the importance and extremal nature of the analytic Koebe function (see Example 4.23), it is conjectured that this "harmonic" Koebe function in Example 4.59 is extremal in similar ways. Note that the coefficients of this harmonic function

(51)
$$f(z) = \sum_{n=0}^{\infty} a_n z^n + \sum_{n=1}^{\infty} b_n \overline{z}^n$$
$$= \mathrm{Re}\left(\frac{z + \frac{1}{3}z^3}{(1-z)^3}\right) + i\,\mathrm{Im}\left(\frac{z}{(1-z)^2}\right)$$

satisfy the properties $|a_n| = \frac{1}{6}(n+1)(2n+1)$, $|b_n| = \frac{1}{6}(n-1)(2n-1)$, and $\big||a_n| - |b_n|\big| = n$.

CONJECTURE 1 (Harmonic Bieberbach).

Let $f(z) = \sum_{n=0}^{\infty} a_n z^n + \sum_{n=1}^{\infty} b_n \overline{z}^n \in S_H^O$. Then

(52)
$$|a_n| \leq \frac{1}{6}(n+1)(2n+1),$$
$$|b_n| \leq \frac{1}{6}(n-1)(2n-1),$$
$$\big||a_n| - |b_n|\big| \leq n.$$

In particular,

(53)
$$|a_2| \leq \frac{5}{2}$$

EXERCISE 4.62. Verify that the "harmonic" Koebe function given in (51) satisfies equality in (52) and (53).
Try it out!

Currently, the best established bound is that for all functions $f \in S_H^O$, $|a_2| < 49$ (see [**9**]). There is room for improvement here, if one can find a right approach.

LARGE PROJECT 4.63. Read and understand the proof that $|a_2| \leq 2$ for analytic functions in S (for example, see [**1**], section 5.1). Read and understand two different proofs that give bounds on $|a_2|$ for harmonic functions in S_H^O (see [**5**], Theorem 4.1 and [**9**], p. 96). Investigate ways to modify any of these proofs or other proofs in order to establish that for $f \in S_H^O$, $|a_2| \leq K$ for some K, where $\frac{5}{2} \leq K < 49$.
Optional

OPEN PROBLEM 4.64. Prove the Harmonic Bieberbach Conjecture.

In addition, because the tip of the the "harmonic" Koebe function is at $-\frac{1}{6}$, we have the following conjecture that is the analogue of the analytic $\frac{1}{4}$ Koebe Theorem.

CONJECTURE 2. The range of every function in class S_H^O contains the disk $G = \left\{ w : |w| < \frac{1}{6} \right\}$.

OPEN PROBLEM 4.65. Prove Conjecture 2. The best result so far is that the range of every $f \in S_H^O$ contains the disk $\left\{ w : |w| < \frac{1}{16} \right\}$ (see [**5**]), so it would be interesting to increase the radius of this disk to some K, where $\frac{1}{16} < K \leq \frac{1}{6}$.

Recall Definition 4.50 of a domain convex in the general direction φ. The shearing theorem by Clunie and Sheil-Small can easily be generalized to apply to such domains.

COROLLARY 4.66. A harmonic function $f = h + \overline{g}$ locally univalent in $\mathbb{D}$ is a univalent mapping of $\mathbb{D}$ onto a domain convex in the direction $\varphi \iff h - e^{2i\varphi} g$ is an analytic univalent mapping of $\mathbb{D}$ onto a domain convex in the direction φ.
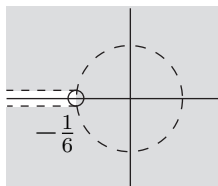
$$\text{disk} = \{w : |w| < \tfrac{1}{6}\}$$

FIGURE 4.24. The disk of radius $\frac{1}{6}$ is contained in the image of $\mathbb{D}$ under the "harmonic" Koebe function.

EXAMPLE 4.67. **Harmonic right half-plane map**

The analytic right half-plane function $\frac{z}{1-z}$ maps $\mathbb{D}$ onto a convex domain (that is, it is convex in all directions φ). So, in particular, it is convex in the direction of the imaginary axis ($\varphi = \frac{\pi}{2}$). Let's apply Corollary 4.66 with $\varphi = \frac{\pi}{2}$ to $\frac{z}{1-z}$ and use a dilatation that simplifies calculations. Note that with $\varphi = \frac{\pi}{2}$, we will set our analytic function $F(z) = \frac{z}{1-z}$ equal to $h(z) - e^{2i\varphi}g(z) = h(z) + g(z)$ instead of $h(z) - g(z)$. So we start with

(54)
$$h(z) + g(z) = \frac{z}{1-z}$$

and choose

$$\omega(z) = g'(z)/h'(z) = -z.$$

Computing h and g as in the previous two examples, yields

(55)
$$h(z) = \frac{z - \frac{1}{2}z^2}{(1-z)^2}$$
$$g(z) = -\frac{\frac{1}{2}z^2}{(1-z)^2}.$$

Hence, the harmonic function is

$$f(z) = h(z) + \overline{g(z)} = \operatorname{Re}\left(\frac{z}{1-z}\right) + i\operatorname{Im}\left(\frac{z}{(1-z)^2}\right) \in S_H^O.$$

EXERCISE 4.68. Verify that shearing $h(z) + g(z) = \frac{z}{1-z}$ with $\omega(z) = -z$, yields eq. (55).

Try it out!

This is the harmonic map we discussed in Example 4.40.

EXPLORATION 4.69. Shear $h + g = \frac{z}{1-z}$ using $\omega = -z^n$, where $n = 1, 2, 3, \ldots$ and sketch $f(\mathbb{D})$ using *ShearTool*. Describe what happens to $f(\mathbb{D})$ and n varies. Pay particular attention to the number of points the green lines go to as n increases.

Try it out!

275

One significance of $\mathbb{D}$ being mapped to the same domains under these two functions is that the uniqueness of the Riemann Mapping Theorem for analytic functions does not hold for harmonic functions. This leaves the open question:

OPEN PROBLEM 4.70. What is the analogue of the Riemann Mapping Theorem for harmonic functions?

SMALL PROJECT 4.71. Let $f = h + \bar{g}$ with $h(z) - g(z) = \dfrac{z}{(1-z)^2}$ and $\omega(z) = \dfrac{z^2 + az}{1 + az}$.

(1) Show that for $-1 \le a \le 1$, $|\omega(z)| < 1, \forall z \in \mathbb{D}$.
(2) Compute h and g explicitly so that $f \in S_H^O$.
(3) Show that for $a = -1$, $f(\mathbb{D})$ is a right half-plane.
(4) Show that for $-1 < a \le 1$, $f(\mathbb{D})$ is a slit domain like the Koebe domain. For each value of a, determine where the tip of the slit is located.

Optional

EXAMPLE 4.72. **Harmonic square map**
Here is one more shearing example that we will use in some later sections and in our discussion of minimal sections. Let

$$(56) \qquad h(z) - g(z) = \frac{1}{2} \log \left(\frac{1+z}{1-z} \right)$$

which is an analytic function that maps $\mathbb{D}$ onto a horizontal strip convex in the direction of the real axis and let

$$\omega(z) = g'(z)/h'(z) = -z^2.$$

Using the *ShearTool* we see that the shear of $h - g$ with $-z^2$ results in a univalent harmonic function that maps onto the interior of the region bounded by a square (see Figure 4.25).

Let's compute $h(z)$ and $g(z)$ explicitly and prove that the image is a square region. Applying the Shearing Method, we have

$$h'(z) - g'(z) = \frac{1}{1 - z^2} \Rightarrow h'(z) + z^2 h'(z) = \frac{1}{1 - z^2}$$

$$\Rightarrow h'(z) = \frac{1}{1 - z^4} = \frac{1}{4} \left[\frac{1}{1+z} + \frac{1}{1-z} + \frac{1}{i+z} + \frac{1}{i-z} \right].$$

Integrating $h'(z)$ and normalizing so that $h(0) = 0$, yields

$$(57) \qquad h(z) = \frac{1}{4} \log \left(\frac{1+z}{1-z} \right) + \frac{i}{4} \log \left(\frac{i+z}{i-z} \right).$$
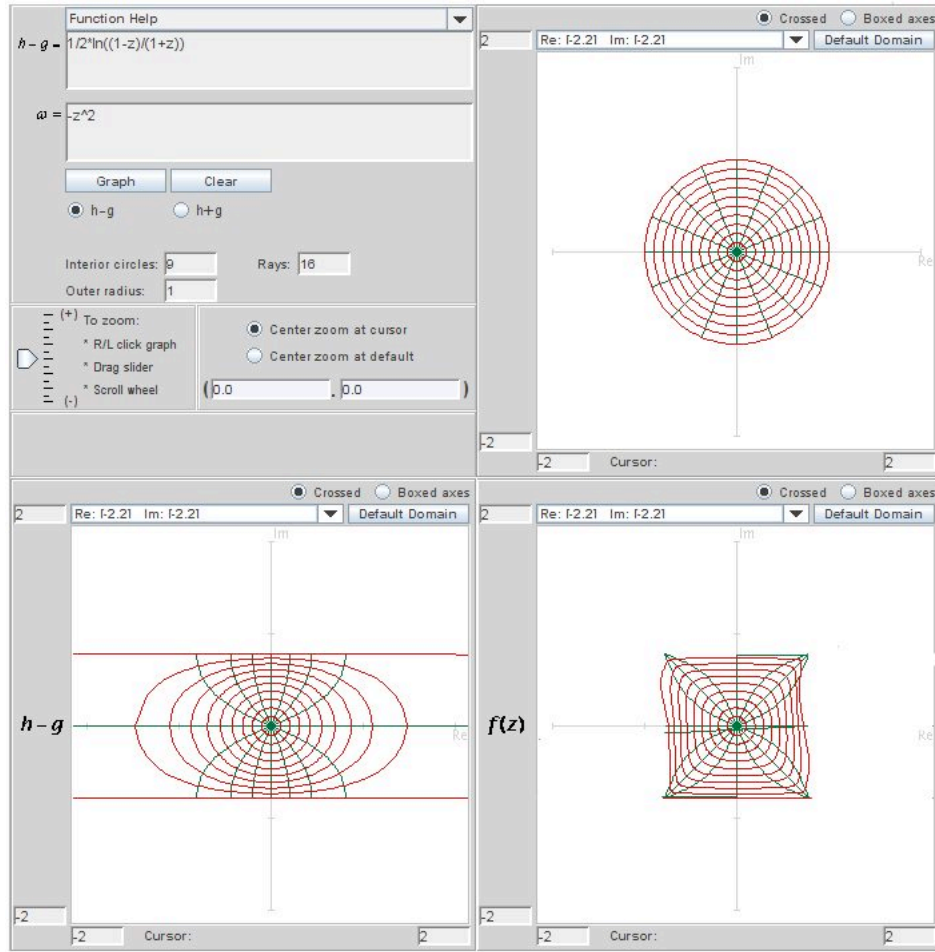
FIGURE 4.25. The image of $\mathbb{D}$ when shearing $h(z) - g(z) = \frac{1}{2}\log\left(\frac{1+z}{1-z}\right)$ with $\omega(z) = -z^2$.

We can use this same method to solve for normalized $g(z)$, where $g(0) = 0$. Note that we can also find $g(z)$ by using eqs. (56) and (57). Either way, we get

$$g(z) = -\frac{1}{4}\log\left(\frac{1+z}{1-z}\right) + \frac{i}{4}\log\left(\frac{i+z}{i-z}\right).$$

So

$$f(z) = h(z) + \overline{g(z)} = \text{Re}\left[\frac{i}{2}\log\left(\frac{i+z}{i-z}\right)\right] + i\,\text{Im}\left[\frac{1}{2}\log\left(\frac{1+z}{1-z}\right)\right] \in S_H^o.$$

277

What is $f(\mathbb{D})$? Notice that

$$f(z) = \left[-\frac{1}{2} \arg \left(\frac{i+z}{i-z} \right) \right] + i \left[\frac{1}{2} \arg \left(\frac{1+z}{1-z} \right) \right]$$
$$= u + iv.$$

Let $z = e^{i\theta} \in \partial \mathbb{D}$. Then

$$\frac{i+z}{i-z} = \frac{i + e^{i\theta}}{i - e^{i\theta}} \frac{-i - e^{-i\theta}}{-i - e^{-i\theta}} = \frac{1 - i(e^{i\theta} + e^{-i\theta}) - 1}{1 + i(e^{i\theta} - e^{-i\theta}) + 1} = -i \frac{\cos\theta}{1 - \sin\theta}.$$

Thus,

$$u = -\frac{1}{2} \arg \left(\frac{i+z}{i-z} \right) \bigg|_{z=e^{i\theta}} = \begin{cases} \frac{\pi}{4} & \text{if } \cos\theta > 0, \\ -\frac{\pi}{4} & \text{if } \cos\theta < 0. \end{cases}$$

Likewise, we can show that

$$v = \begin{cases} \frac{\pi}{4} & \text{if } \sin\theta > 0, \\ -\frac{\pi}{4} & \text{if } \sin\theta < 0. \end{cases}$$

In summary, we have that $z = e^{i\theta} \in \partial \mathbb{D}$ is mapped to

$$u + iv = \begin{cases} z_1 = \frac{\pi}{2\sqrt{2}} e^{i\frac{\pi}{4}} = \frac{\pi}{4} + i\frac{\pi}{4} & \text{if } \theta \in (0, \frac{\pi}{2}), \\ z_3 = \frac{\pi}{2\sqrt{2}} e^{i\frac{3\pi}{4}} = -\frac{\pi}{4} + i\frac{\pi}{4} & \text{if } \theta \in (\frac{\pi}{2}, \pi), \\ z_5 = \frac{\pi}{2\sqrt{2}} e^{i\frac{5\pi}{4}} = -\frac{\pi}{4} - i\frac{\pi}{4} & \text{if } \theta \in (\pi, \frac{3\pi}{2}), \\ z_7 = \frac{\pi}{2\sqrt{2}} e^{i\frac{7\pi}{4}} = \frac{\pi}{4} - i\frac{\pi}{4} & \text{if } \theta \in (\frac{3\pi}{2}, 2\pi). \end{cases}$$

Thus, this harmonic function maps $\mathbb{D}$ onto the interior of the region bounded by a square with vertices at z_1, z_3, z_5 and z_7.

Note that this same function can be obtained by using the Poisson integral formula. In fact, using the Poisson formula we can find a harmonic function that maps $\mathbb{D}$ onto a regular n-gon for any $n \geq 3$. In particular, let

$$f(z) = \frac{1}{2\pi} \int_0^{2\pi} Re \left(\frac{1 + ze^{-it}}{1 - ze^{-it}} \right) e^{i\phi(t)} dt,$$

where $\phi(t) = \frac{\pi(2k+1)}{n}$ $\quad (\frac{2\pi k}{n} \le t < \frac{2\pi(k+1)}{n}, k = 0, ..., n-1)$. In this case, we can derive that

$$h(z) = \sum_{m=0}^{\infty} \frac{1}{nm+1} z^{nm+1},$$

$$g(z) = \sum_{m=1}^{\infty} \frac{-1}{nm-1} z^{nm-1},$$

(58)

$$h'(z) = \frac{1}{1 - z^n},$$

$$g'(z) = \frac{-z^{n-2}}{1 - z^n}, \text{ and}$$

$$\omega(z) = -z^{n-2}.$$

EXERCISE 4.73. Let $f = h + \bar{g}$ with $h(z) - g(z) = \frac{1}{3} \log\left(\frac{1 + z + z^2}{1 - 2z + z^2}\right)$ and $\omega(z) = -z$.

(1) Show that $h'(z) = \frac{1}{1-z^3}$ and $g'(z) = \frac{-z}{1-z^3}$.
(2) According to the previous paragraph, what should be the image of $\mathbb{D}$ under $f = h + \bar{g}$?
(3) Use *ShearTool* to sketch the image of f.
(4) Compute h and g explicitly so $f \in S_H^O$.

Try it out!

EXERCISE 4.74. Let $f = h + \bar{g}$ with $h(z) + g(z) = \frac{1}{2} \log\left(\frac{1+z}{1-z}\right)$ and $\omega(z) = -z^2$ (note the difference between this exercise and Example 4.72). Compute h and g explicitly so that $f \in S_H^O$ and use *ComplexTool* to graph $f(\mathbb{D})$.

Try it out!

4.6. Properties of the dilatation

Because of the importance of the dilatation, we will examine some of its properties. A result from complex analysis is that an analytic function will map infinitesimal circles to infinitesimal circles at any point where its derivative is nonzero. For harmonic functions, this result does not hold. This can be seen in the following exploration.

EXPLORATION 4.75. In *ComplexTool*, enter values one at a time of 0.1, 0.4, 0.7, and 0.999 into the **Outer radius** box in the center panel to plot disks of various outer

radii under the analytic right half-plane map

$$F(z) = \frac{z}{1 - z}$$

and the harmonic right half-plane map

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

In all the cases look at the images of small circles under F and under f. What appears to be the image of circles under the harmonic function f?

Try it out!

Now, let's explore the ideas of the geometric dilatation, D_f, and the analytic dilatation, $\omega(z)$ which we discussed earlier.

EXERCISE 4.76.

(a.) Prove that the formulas

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

are equivalent to the formulas (46) and (47) given in section 4.4.

(b.) Prove that if $f = h + \overline{g}$ is harmonic, then

$$\frac{\partial f}{\partial z} = h'(z) \qquad \text{and} \qquad \frac{\partial f}{\partial \overline{z}} = \overline{g'(z)}.$$

Try it out!

Think of the differential of f as $df = f_z dz + f_{\overline{z}} d\overline{z}$. Thus we have

$$|df| = |f_z dz + f_{\overline{z}} d\overline{z}| = |h'(z)dz + \overline{g'(z)}d\overline{z}|.$$

Now bound this differential, and use the fact that if dz is very small, so also $d\overline{z}$ will be very small and approximately equal to dz. When we examine upper and lower bounds, we have

$$(|h'(z)| - |g'(z)|)|dz| \leq |df| \leq (|h'(z)| + |g'(z)|)|dz|.$$

Note that for sense-preserving harmonic functions, the left hand side is always positive. Now we take a ratio of the upper and lower bounds to get the *geometric dilatation D_f* defined by

$$D_f = \frac{|h'| + |g'|}{|h'| - |g'|}.$$

Since this is a ratio of the maximum to minimum $|df|$, if we evaluate D_f at a point z_0, that means that we will find a number that represents a ratio between the most and the least that an infinitesimal circle will be deformed by the function. Thus, if the function maps an infinitesimal circle to an infinitesimal ellipse, then the geometric dilatation gives the ratio of the major axis to the minor axis of the ellipse.

EXERCISE 4.77.

(a.) Prove that for analytic functions, D_f is always 1.

(b.) Prove that for sense-preserving harmonic functions, $D_f \geq 1$.

(c.) Find a formula for D_f for $z + \frac{\bar{z}^3}{3}$. What are the maximum and minimum of this function over the unit disk $\mathbb{D}$?

Try it out!

EXERCISE 4.78. Examine the geometric dilatation for $z + \frac{\bar{z}^3}{3}$ in greater detail. For the points $z = 0, 0.5, 0.9, 0.9e^{i\pi/4}$, and $0.9i$, find $D_f(z)$. Using *ComplexTool*, examine the images of circles of radius 0.05 centered at those points. Estimate the ratio of the major axis to the minor axis of the image ellipse. Does it match with your computation for D_f?

Try it out!

While the geometric dilatation provides some very useful information about the function, some information is lost when we take the modulus of $|h'(z)|$ and $|g'(z)|$. Instead, it is often useful to examine the analytic part versus the anti-analytic part of the function f. Thus we define what is sometimes called the *second complex dilatation* of f,

$$\omega_f(z) = \frac{\overline{f_{\bar{z}}(z)}}{f_z(z)} = \frac{g'(z)}{h'(z)},$$

where the representation in the last equality makes sense only for harmonic functions. When the function f is clear, we drop the subscript and refer only to the dilatation as ω. Because this function $\omega(z)$ is analytic if and only if $f(z)$ is harmonic, the second complex dilatation is also called the *analytic dilatation* of f.

EXERCISE 4.79.

(a.) Prove that $\omega(z) = \frac{\overline{f_{\bar{z}}(z)}}{f_z(z)}$ is analytic if and only if $f(z)$ is harmonic.

(b.) Prove that $\omega(z)$ is identically 0 if and only if f is analytic.

(c.) Prove that for sense-preserving non-analytic harmonic functions f, $0 < |\omega(z)| < 1$.

Try it out!

Now one can naturally ask what the relationship is between the geometric dilatation and analytic dilatation.

EXERCISE 4.80. Prove that $D_f(z) \leq K$ if and only if $|\omega_f(z)| \leq \frac{K-1}{K+1}$.

Try it out!

EXPLORATION 4.81. Re-examine the function $z + \frac{1}{3}\bar{z}^3$, now evaluating $\omega(z)$ at the points $z = 0, 0.5, 0.9, 0.9e^{i\pi/4}$, and $0.9i$. What do you observe about the relationship between ω at these points and images of a small circle centered at these points?

Try it out!

In Section 4.4, we remarked that if $f = h + \bar{g}$ is a sense-preserving harmonic map that has $|\omega(z)| = 1$ for all $z \in$ arc of $\partial\mathbb{D}$, then the image of the arc is either:

- a concave arc; or
- stationary.

To further explore this result, we will use *ShearTool* to graph the image of $\mathbb{D}$ under $f = h + \bar{g}$, when $h - g = z$ and ω has various values in order to see the effect of changing ω.

EXPLORATION 4.82.

(1) Shear $h(z) - g(z) = z$ using $\omega(z) = e^{i\pi n/6}z$, where $n = 0, \ldots, 6$ and sketch $f(\mathbb{D})$ using *ShearTool*. Describe what happens to $f(\mathbb{D})$ as n varies.
(2) Shear $h(z) - g(z) = z$ using $\omega(z) = z^n$, where $n = 1, 2, 3, 4$ and sketch $f(\mathbb{D})$ using *ShearTool*.
  (a) What patterns do you notice relating $f(\mathbb{D})$ and n?
  (b) Make a sketch on paper of what $f(\mathbb{D})$ looks like for $n = 5$. Then graph that shear using *ShearTool*.
  (c) Make a sketch on paper of what $f(\mathbb{D})$ looks like for $n = 6$. Then graph that shear using *ShearTool*.
(3) Shear $h(z) - g(z) = z$ using $\omega(z) = \frac{z+a}{1+\bar{a}z}$, for various values of $a \in \mathbb{D}$ and sketch $f(\mathbb{D})$ using *ShearTool*. Describe what happens to $f(\mathbb{D})$ as a varies.

Try it out!

SMALL PROJECT 4.83. Investigate the shearing of $h(z) - g(z) = z - \frac{1}{n^2}z^n$ ($n = 2, 3, 4, \ldots$) with ω for various values of ω (note that the image of $\mathbb{D}$ under the analytic function $z - \frac{1}{n}z^n$ is not CHD for $n = 4, 5, 6, \ldots$; however, it is if we use $z - \frac{1}{n^2}z^n$). Use the approach of Exploration 4.82 as a starting point and then explore new approaches.

Optional

Up to this point, we have only used dilatations that are finite Blaschke products. A *finite Blaschke product $B(z)$* can be expressed in the form

$$B(z) = e^{i\theta} \prod_{j=1}^{n} \left(\frac{z - a_j}{1 - \overline{a_j}z} \right)^{m_j},$$

where $\theta \in \mathbb{R}$, $|a_j| < 1$, and m_j is the multiplicity of the zero a_j. The dilatations given in Exploration 4.82 and finite products of them are examples of finite Blaschke products. Harmonic univalent mappings whose dilatation is a finite Blaschke product have been studied (see [**13**]). However, little is known about mappings whose dilatation is not a finite Blaschke product. And so an interesting problem is to investigate the properties of harmonic univalent mappings whose dilatation is not a finite Blaschke product. One important type of mappings that are not a finite Blaschke product is a singular inner function. Hence, we will now investigate harmonic univalent mappings whose dilatation is a singular inner function.

First, we need to know what singular inner functions are. Let $f : \mathbb{D} \to \mathbb{C}$ be an analytic function and denote its radial limit by

$$f^*(e^{i\theta}) := \lim_{r \to 1,\, r<1} f(re^{i\theta}).$$

DEFINITION 4.84. A bounded analytic function f is called an *inner function* if $|f^*(e^{i\theta})| = 1$ almost everywhere (with respect to Lebesgue measure on $\partial\mathbb{D}$). If f has no zeros on $\mathbb{D}$, then f is called a *singular inner function*.

Every inner function can be written in the form

$$f(z) = e^{i\alpha} B(z) e^{\left(-\int \frac{e^{i\theta} + z}{e^{i\theta} - z} \, d\mu(e^{i\theta}) \right)},$$

where $\alpha, \theta \in R$, μ is a positive measure on $\partial\mathbb{D}$, and $B(z)$ is a Blaschke product. The function $f(z) = e^{\frac{z+1}{z-1}}$ is an example of a singular inner function.

EXERCISE 4.85. Show that if $\omega(z) = e^{\frac{z+1}{z-1}}$, then $|\omega(z)| < 1, \forall z \in \mathbb{D}$.
Try it out!

It has been difficult to construct examples of harmonic mappings whose dilatation is a singular inner function. For awhile there were no known examples [**17**] until Weitsman [**26**] provided two examples. We present Weitsman's examples (see Example 4.87 and Example 4.92) giving a much shorter proof of his second example with this proof providing a method to find more examples.

One way to find an example of a harmonic map with a singular inner function as its dilatation is to use the shearing technique. However, being able to find a closed form for $f = h + \overline{g}$ is not often possible. For example, let $h(z) - g(z) = z$ and the dilatation be $\omega(z) = e^{\frac{z+1}{z-1}}$. Then by the shearing technique

$$h(z) = \int \frac{1}{1 - e^{\frac{z+1}{z-1}}} \, dz.$$

This integral does not have a closed form and so we cannot find an explicit representation for $f = h + \overline{g}$ in this case.

EXERCISE 4.86. Using the shearing technique with $h(z) - g(z) = z - \frac{1}{n}z^n$ and $\omega(z) = e^{\frac{z+1}{z-1}}$, express h as an integral. It is not possible to integrate h to get a closed-form solution.
Try it out!

The following is an example in which the shearing technique does allow us to find specific values for h and g.

EXAMPLE 4.87. Consider shearing the analytic function

$$h(z) - g(z) = \frac{z}{1 - z} + \frac{1}{2} e^{\frac{z+1}{z-1}}$$

with

$$\omega(z) = e^{\frac{z+1}{z-1}}.$$

To see that $h - g$ is convex in the direction of the real axis, we will use a remark by Pommerenke [**18**].

THEOREM 4.88. Let f be an analytic function in $\mathbb{D}$ with $f(0) = 0$ and $f'(0) \neq 0$, and let

$$\varphi(z) = \frac{z}{(1 + ze^{i\theta})(1 + ze^{-i\theta})},$$

where $\theta \in \mathbb{R}$. If

$$\mathrm{Re}\left\{\frac{zf'(z)}{\varphi(z)}\right\} > 0, \text{ for all } z \in \mathbb{D},$$

then f is convex in the direction of the real axis.

Note that in this example

$$h'(z) - g'(z) = \frac{1}{(1-z)^2}\left[1 - e^{\frac{z+1}{z-1}}\right].$$

Using $\theta = \pi$ in Theorem 4.88, we have

$$\mathrm{Re}\left[1 - e^{\frac{z+1}{z-1}}\right] > 0$$

because $\left|e^{\frac{z+1}{z-1}}\right| < 1$. Hence $h - g$ is convex in the direction of the real axis.

Shearing $h - g$ with $\omega(z) = e^{\frac{z+1}{z-1}}$ and normalizing yields

$$h(z) = \int \frac{1}{(1-z)^2}\,dz = \frac{z}{1-z},$$

and solving for g we get

$$g(z) = -\frac{1}{2}e^{\frac{z+1}{z-1}}.$$

The image given by this map is similar to the image given by the right half-plane map $\frac{z}{1-z}$ except in this case there are an infinite number of cusps (see Figure 4.26).

EXERCISE 4.89. Let $f = h + \bar{g}$ with $h(z) - g(z) = \frac{z}{1-z}$ and $\omega(z) = e^{\frac{z+1}{z-1}}$. Use the shearing method to compute h and g explicitly so $f \in S_H^O$ and use *ComplexTool* to sketch $f(\mathbb{D})$ [Hint: In finding the specific function h, use a u-substitution to evaluate the integral].
Try it out!

Another technique to find harmonic mappings whose dilatations are singular inner functions involves using the following theorem by Clunie and Sheil-Small [**5**].

THEOREM 4.90. Let $f = h + \bar{g}$ be locally univalent in $\mathbb{D}$ and suppose that $h + \epsilon g$ is convex for some $|\epsilon| \leq 1$. Then f is univalent.
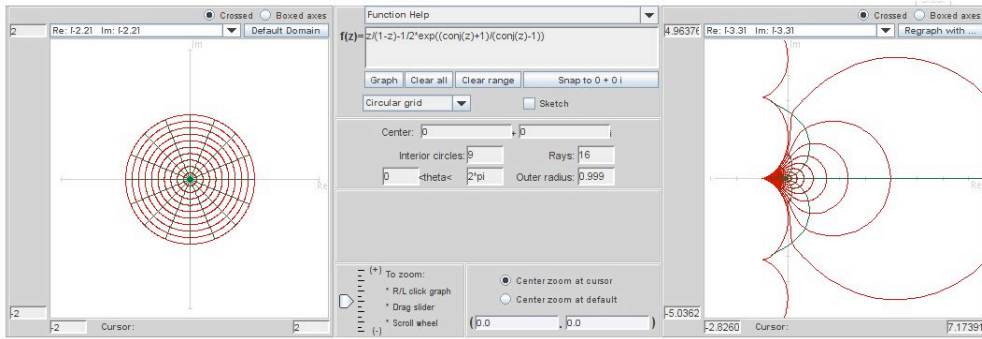
FIGURE 4.26. Image of $\mathbb{D}$ under $f(z) = \dfrac{z}{1-z} - \dfrac{1}{2}e^{\frac{\bar{z}+1}{\bar{z}-1}}$.

Theorem 4.90 gives us a nice way to show that a harmonic function is univalent. To develop our present technique, we will let $\epsilon = 0$ in the theorem. This means that if h is analytic convex and if ω is analytic with $|\omega(z)| < 1$, then $f = h + \bar{g}$ is a harmonic univalent mapping.

Actually, the original theorem by Clunie and Sheil-Small gives us a bit more information about f. It states that f is close-to-convex. Geometrically, a close-to-convex function f is one whose image $f(re^{i\theta})$ has no "large hairpin" turns; that is, the tangent vector at $f(re^{i\theta})$ does not turn backward through an angle greater or equal to π anyway along the image of the curve $|z| = r$. However, at this point we are just interested in showing that f is univalent, so you do not need to be concerned about the close-to-convex property.

Also, to establish that a function f is convex, the following theorem is useful.

THEOREM 4.91. Let f be analytic and univalent in $\mathbb{D}$. Then f maps onto a convex domain if and only if

$$\mathrm{Re}\left[1 + \frac{zf''(z)}{f'(z)}\right] \geq 0, \text{ for all } z \in \mathbb{D}.$$

In the following example we will show how these ideas can be used to construct a harmonic univalent function whose dilatation is a singular inner function.

EXAMPLE 4.92. Let

$$h(z) = z - \frac{1}{4}z^2 \quad \text{with} \quad \omega(z) = g'(z)/h'(z) = e^{\frac{z+1}{z-1}}.$$

We will use Theorem 4.91 to show that h is convex. Let

$$T(z) = 1 + \frac{zh''(z)}{h'(z)} = 1 + \frac{-\frac{1}{2}z}{1 - \frac{1}{2}z} = \frac{1-z}{1 - \frac{1}{2}z}.$$

285

Notice that $T(z)$ is a Möbius transformation. By the mapping properties of Möbius transformations we can show that T maps $\mathbb{D}$ onto $|z - \frac{2}{3}| = \frac{2}{3}$ which is the circle centered at $\frac{2}{3}$ of radius $\frac{2}{3}$. Hence, $\mathrm{Re}\{T(z)\} > 0$ and h is convex.

Now solve for g.

$$g(z) = \int h'(z)\omega(z)\, dz = \int (1 - \frac{1}{2}z)e^{\frac{z+1}{z-1}}\, dz = -\frac{1}{4}(z-1)^2 e^{\frac{z+1}{z-1}}.$$

Hence,

$$f(z) = h(z) + \overline{g(z)} = z - \frac{1}{4}z^2 - \frac{1}{4}(\overline{z}-1)^2 e^{\frac{\overline{z}+1}{\overline{z}-1}}.$$

Thus, by Theorem 4.90, $f = h + \overline{g}$ is univalent. The image of $\mathbb{D}$ under $f_1(z) = z - \frac{1}{4}z^2 - \frac{1}{4}(\overline{z}-1)^2 e^{\frac{\overline{z}+1}{\overline{z}-1}}$ is similar to the map of a harmonic polynomial but with an infinite number of cusps in the middle section on the right side (see Figure 4.27).
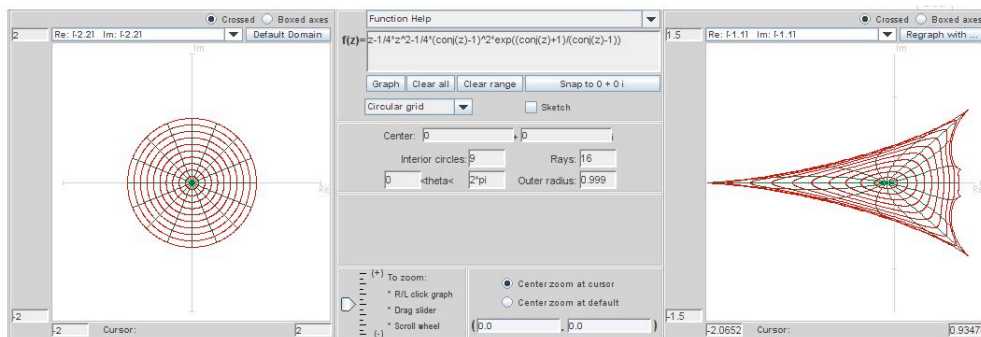
FIGURE 4.27. Image of $\mathbb{D}$ under $f_1(z) = z - \frac{1}{4}z^2 - \frac{1}{4}(\overline{z}-1)^2 e^{\frac{\overline{z}+1}{\overline{z}-1}}$.

EXERCISE 4.93. Let $h(z) = z + \frac{1}{11}z^3$ and $g(z) = -\frac{1}{11}(z-3)(z+1)^2 e^{\frac{z-1}{z+1}}$. Show that $f = h + \overline{g}$ is univalent and use *ComplexTool* to graph $f(\mathbb{D})$.

Try it out!

EXERCISE 4.94. Use the approach above to show that compute $f = h + \overline{g}$ is harmonic univalent, where $h(z) = z + 2\log(z+1)$ and $\omega(z) = e^{\frac{z-1}{z+1}}$. The graph of $\mathbb{D}$ under f is shown in Figure 4.28.

Try it out!

It was mentioned earlier that it is not often possible to find a closed form for $f = h + \overline{g}$ when the dilatation is a singular inner function. However, one can use *ShearTool* to explore images of $f(\mathbb{D})$ when the dilatation is a singular inner product.

EXPLORATION 4.95.

(1) If we shear $h(z) - g(z) = \frac{z}{(1-z)^2}$ with $\omega(z) = e^{\frac{z+1}{z-1}}$, then $f = h + \overline{g}$ will be univalent and convex in the direction of the real axis. Use *ShearTool* to sketch $f(\mathbb{D})$.
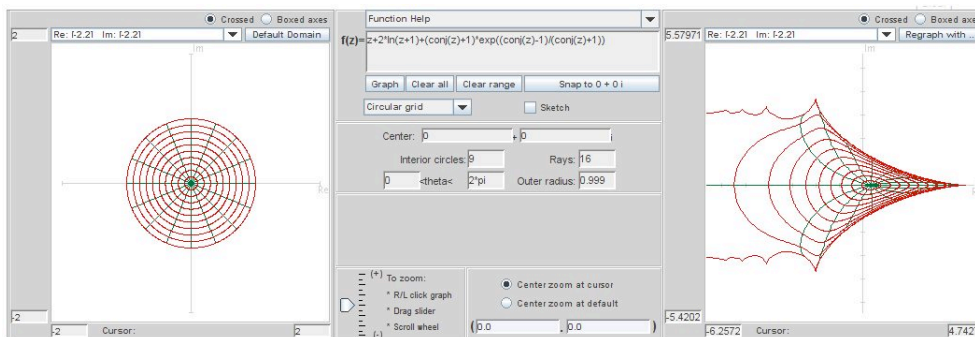
FIGURE 4.28. Image of $\mathbb{D}$ under $f(z) = h + \bar{g}$ in Exercise 4.94.

(2) Use *ShearTool* to sketch $f(\mathbb{D})$, where $h(z) - g(z) = z - \frac{1}{2}z^2$ with $\omega(z) = e^{\frac{z^2+1}{z^2-1}}$.

(3) Use *ShearTool* to sketch $f(\mathbb{D})$, where $h(z) - g(z) = \frac{1}{2}\log\left(\frac{1+z}{1-z}\right)$ with $\omega(z) = e^{\frac{z+1}{z-1}}$.

Try it out!

OPEN PROBLEM 4.96. What are the properties of harmonic univalent mappings whose dilatation is a singular inner product?

4.7. Harmonic Linear Combinations

A common way to try to construct new functions with a given property is to take the linear combination of two functions with that property. This is done with derivatives and integrals in beginning calculus. And in Exploration 4.25, this done with the analytic Koebe mapping, f_k, and the right half-plane mapping, f_r, where

$$f_k(z) = \frac{z}{(1-z)^2} \qquad \text{and} \qquad f_r(z) = \frac{z}{1-z},$$

to derive the univalent analytic map

$$f_3(z) = tf_k(z) + (1-t)f_r(z) = \frac{z - tz^2}{(1-z)^2},$$

where $0 \leq t \leq 1$.

Is it true that the linear combination of two $1-1$ functions is also a $1-1$ function? Let's look at the case for real-valued functions. Suppose $f_1 : \mathbb{R} \to \mathbb{R}$ and $f_2 : \mathbb{R} \to \mathbb{R}$ are $1-1$ functions. Will $f_3(x) = tf_1(x) + (1-t)f_2(x)$ also be $1-1$ when $0 \leq t \leq 1$? Not necessarily. Consider the example of $f_1(x) = x^3$, $f_2(x) = -x^3$, and $t = \frac{1}{2}$. Both f_1 and f_2 are $1-1$; they satisfy the horizontal line test. But $f_3(x) = tf_1(x) + (1-t)f_2(x) = 0$ which is not $1-1$.
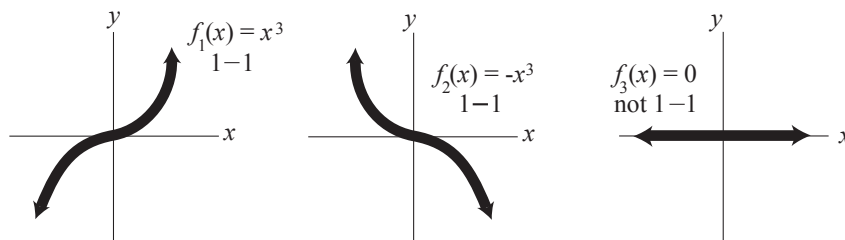
FIGURE 4.29. $f_1(x) = x^3$ and $f_2(x) = -x^3$ are $1 - 1$, but $f_3(x) = tf_1(x) + (1 - t)f_2(x) = 0$ s not $1 - 1$ when $t = \frac{1}{2}$.

In this case, the difficulty is that f_1 is an increasing $1-1$ function, f_2 is a decreasing $1 - 1$ function, and when $t = \frac{1}{2}$, the increase of f_1 cancels out the decrease of f_2. We can alleviate this difficulty by requiring that f_1, f_2 are either both increasing or both decreasing. This idea can be applied to complex-valued functions.

CONDITION A. Suppose f is complex-valued harmonic and non-constant in $\mathbb{D}$. There exists sequences z'_n, z''_n converging to $z = 1$, $z = -1$, respectively, such that

(59)
$$\lim_{n \to \infty} \mathrm{Re}\{f(z'_n)\} = \sup_{|z|<1} \mathrm{Re}\{f(z)\}$$
$$\lim_{n \to \infty} \mathrm{Re}\{f(z''_n)\} = \inf_{|z|<1} \mathrm{Re}\{f(z)\}.$$

Note that the normalization in (59) can be thought of in some sense as if $f(1)$ and $f(-1)$ are the right and left extremes in the image domain in the extended complex plane.

EXAMPLE 4.97. We will show that Condition A is satisfied by $f(z) = z + \frac{1}{3}\overline{z}^3$ (see Figure 4.30).
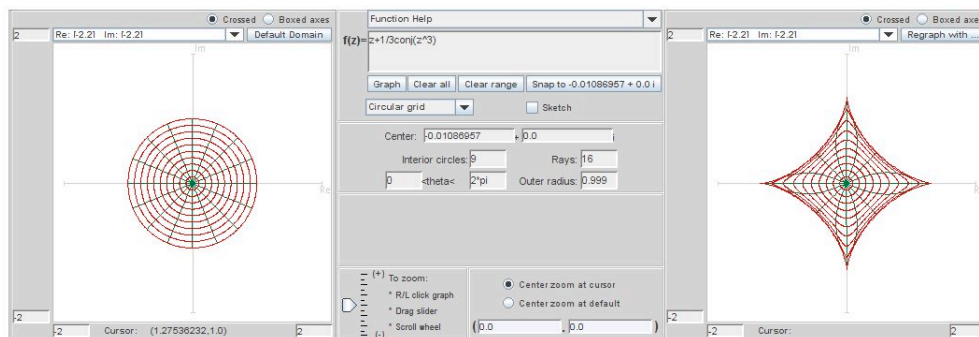
FIGURE 4.30. Image of $\mathbb{D}$ under $f(z) = z + \frac{1}{3}\overline{z}^3$

One can get a feel that f satisfies Condition A by using the Sketch option in *ComplexTool* to draw several paths $\{z'_n\} \in \mathbb{D}$ that approach 1 and see that the corresponding

288

images of these paths approach the right-side cusp of the image of $\mathbb{D}$, a hypocycloid of 4 cusps (see Figure 4.30). Likewise, various paths $\{z_n''\} \in \mathbb{D}$ that approach -1 result in image paths that approach the left-side cusp of the hypocycloid.

To prove that f satisfies Condition A, note that

$$f(e^{i\theta}) = e^{i\theta} + \frac{1}{3}e^{-3i\theta} = (\cos\theta + \frac{1}{3}\cos 3\theta) + i(\sin\theta - \frac{1}{3}\sin 3\theta).$$

So, $\text{Re}\{f(e^{i\theta})\} = \cos\theta + \frac{1}{3}\cos 3\theta$. Hence, $-\frac{4}{3} \leq \text{Re}\{f(e^{i\theta})\} \leq \frac{4}{3}$ which means $\sup_{|z|<1}\text{Re}\{f(z)\} = \frac{4}{3}$. Letting $z_n' = 1 - \frac{1}{n} \to 1$, we have that

$$\lim_{n\to\infty}\text{Re}\{f(z_n')\} = \frac{4}{3} = \sup_{|z|<1}\text{Re}\{f(z)\}.$$

Similarly, $\lim_{n\to\infty}\text{Re}\{f(z_n'')\} = -\frac{4}{3} = \inf_{|z|<1}\text{Re}\{f(z)\}$.

EXERCISE 4.98. Use the `Sketch` option in *ComplexTool* to determine which of the following harmonic functions satisfy Condition A:

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

(b) $\text{Re}\left[\frac{1}{2}\log\left(\frac{1+z}{1-z}\right)\right] + i\,\text{Im}\left[\frac{i}{2}\log\left(\frac{i+z}{i-z}\right)\right]$

(c) $\text{Re}\left(\frac{z}{1-z}\right) + i\,\text{Im}\left(\frac{z}{(1-z)^2}\right)$

(d) $\text{Re}\left[\frac{i}{2}\log\left(\frac{i+z}{i-z}\right)\right] + i\,\text{Im}\left[\frac{1}{2}\log\left(\frac{1+z}{1-z}\right)\right]$

(e) $\text{Re}\left(\frac{z+\frac{1}{3}z^3}{(1-z)^3}\right) + i\,\text{Im}\left(\frac{z}{(1-z)^2}\right)$

Try it out!

To prove a result about the linear combinations of harmonic functions, we will need the following result by Hengartner and Schober [**15**] that employs condition A. However, we won't use Theorem 4.99 afterwards.

THEOREM 4.99 (Hengartner and Schober). Suppose f is holomorphic (i.e., analytic) and non-constant in $\mathbb{D}$. Then

$$\text{Re}\{(1-z^2)f'(z)\} \geq 0, z \in \mathbb{D}$$

if and only if

(1) f is univalent in $\mathbb{D}$,
(2) f is convex in the imaginary direction, and
(3) condition A holds.

We now seek to study conditions under which f_3 is globally univalent.

289

THEOREM 4.100. Let $f_1 = h_1 + \overline{g_1}$, $f_2 = h_2 + \overline{g_2}$ be univalent harmonic mappings convex in the imaginary direction and $\omega_1 = \omega_2$. If f_1, f_2 satisfy condition A, then $f_3 = tf_1 + (1-t)f_2$ is convex in the imaginary direction (and univalent) $(0 \le t \le 1)$.

PROOF. To see that f_3 is locally univalent, use $g'_1 = \omega_1 h'_1$ and $g'_2 = \omega_2 h'_2 = \omega_1 h'_2$. Then

$$\omega_3 = \frac{tg'_1 + (1-t)g'_2}{th'_1 + (1-t)h'_2} = \frac{t\omega_1 h'_1 + (1-t)\omega_1 h'_2}{th'_1 + (1-t)h'_2} = \omega_1.$$

Next, by Clunie and Sheil-Small's shearing theorem (see Theorem 4.52), we know that each $h_j + g_j$ $(j = 1, 2)$ is univalent and convex in the imaginary direction. Also, $h_j + g_j$ satisfies Condition A since $\operatorname{Re}\{f_j\} = \operatorname{Re}\{h_j + g_j\}$. Applying Theorem 4.99 we have

$$\operatorname{Re}\{(1 - z^2)(h'_j(z) + g'_j(z))\} \ge 0, (j = 1, 2).$$

Consider

$$\begin{aligned}
\operatorname{Re}\{(1-z^2)&(h'_3(z) + g'_3(z))\} \\
&= \operatorname{Re}\{(1 - z^2)[t(h'_1(z) + g'_1(z)) + (1 - t)(h'_2(z) + g'_2(z))]\} \\
&= t \operatorname{Re}\{(1 - z^2)(h'_1(z) + g'_1(z))\} + (1 - t) \operatorname{Re}\{(1 - z^2)(h'_2(z) + g'_2(z))\} \ge 0.
\end{aligned}$$

By applying Theorem 4.99 in the other direction, we have that $h_3 + g_3$ is convex in the imaginary direction, and so by the shearing theorem, f_3 is convex in the imaginary direction. □

EXAMPLE 4.101. Consider the functions

$$f_1(z) = \operatorname{Re}\left[\frac{i}{2} \log\left(\frac{1+z}{1-z}\right)\right] + i \operatorname{Im}\left[-\frac{1}{2} \log\left(\frac{i+z}{i-z}\right)\right],$$

$$f_2(z) = \operatorname{Re}\left[\frac{1}{2} \log\left(\frac{1+z}{1-z}\right)\right] + i \operatorname{Im}\left[\frac{i}{2} \log\left(\frac{i+z}{i-z}\right)\right].$$

Now, f_1 maps $\mathbb{D}$ onto a square region (see Figure 4.31); the image is the same as for the harmonic square map in Example 4.72, but the function is different. In particular, f_1 has different arcs of the unit circle being mapped to the vertices, and the dilatation for f_1 is $\omega(z) = z^2$ which is different than the dilatation for the harmonic square map in Example 4.72. Condition A is satisfied for f_1 (see Example 4.104 for more details).

f_2 maps $\mathbb{D}$ onto a region similar to a hypocycloid with 4 cusps except instead of cusps the domain has ends that extend out to infinity (see Figure 4.32). The dilatation of f_2 is also $\omega(z) = z^2$ and Condition A is satisfied.

By Theorem 4.100, $f_3 = tf_1 + (1 - t)f_2$ is univalent. The image of $\mathbb{D}$ when $t = \frac{1}{2}$ is shown in Figure 4.33.
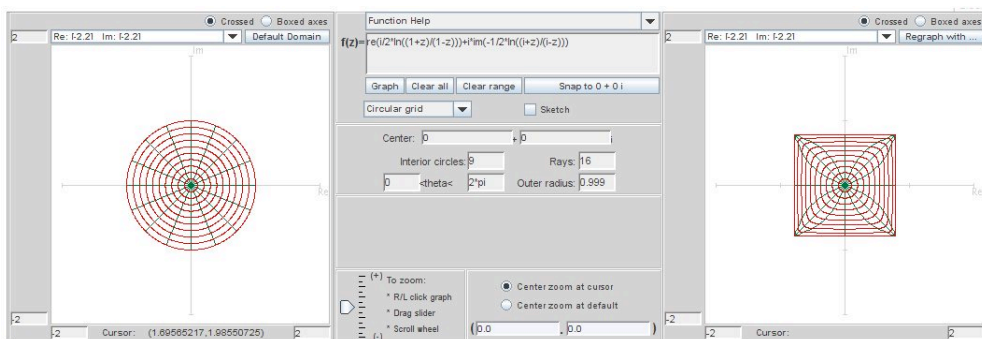
FIGURE 4.31. Image of $\mathbb{D}$ under $f_1(z) = \mathrm{Re}\left[\dfrac{i}{2}\log\left(\dfrac{1+z}{1-z}\right)\right] + i\,\mathrm{Im}\left[-\dfrac{1}{2}\log\left(\dfrac{i+z}{i-z}\right)\right]$.
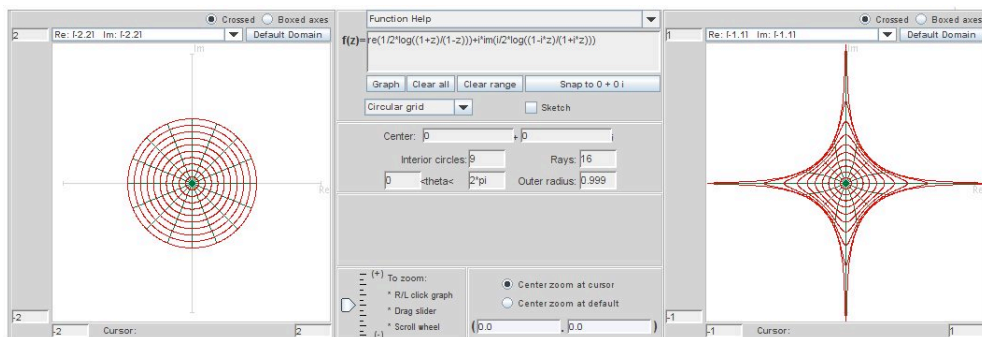
FIGURE 4.32. Image of $\mathbb{D}$ under $f_2(z) = \mathrm{Re}\left[\dfrac{1}{2}\log\left(\dfrac{1+z}{1-z}\right)\right] + i\,\mathrm{Im}\left[\dfrac{i}{2}\log\left(\dfrac{i+z}{i-z}\right)\right]$.
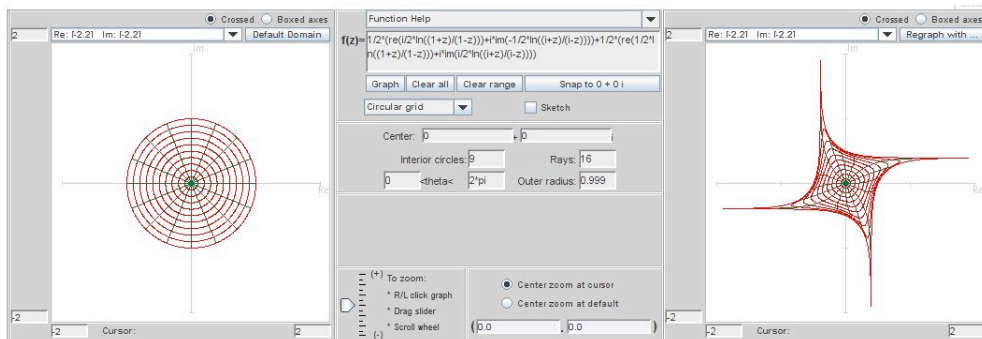
FIGURE 4.33. Image of $\mathbb{D}$ under $f_3(z) = \frac{1}{2}f_1(z) + \frac{1}{2}f_2(z)$

EXPLORATION 4.102. Let
$$f_1(z) = \mathrm{Re}\left[\frac{i}{\sqrt{3}}\ln\left(\frac{1+e^{-i\frac{\pi}{3}}z}{1+e^{i\frac{\pi}{3}}z}\right)\right] + i\,\mathrm{Im}\left[\frac{1}{3}\ln\left(\frac{1+z+z^2}{1-2z+z^2}\right)\right],$$
$$f_2(z) = \mathrm{Re}\left(\frac{z}{1-z}\right) + i\,\mathrm{Im}\left(\frac{z}{(1-z)^2}\right).$$

Show that f_1 and f_2 satisfies the conditions of Theorem 4.100 and then use *Com-plexTool* to plot images of $f_3 = tf_1 + (1 - t)f_2$ for various values of t.

Try it out!

LARGE PROJECT 4.103. Theorem 4.100 gives necessary but not sufficient conditions on f_1 and f_2 for the linear combination $f_3 = tf_1 + (1 - t)f_2$ to be univalent. That f_3 can be univalent when f_1 does not satisfy Condition A is demonstrated by the functions

$$f_1(z) = \text{Re}\left(\frac{z + 1/3z^3}{(1 - z)^3}\right) + i\,\text{Im}\left(\frac{z}{(1 - z)^2}\right),$$

$$f_2(z) = \text{Re}\left(\frac{z}{1 - z}\right) + i\,\text{Im}\left(\frac{z}{(1 - z)^2}\right).$$

Construct f_3 for various values of t and use *ComplexTool* to see the images of $\mathbb{D}$ under f_3.

In fact, the following functions suggest that several of the hypotheses of Theorem 4.100 can fail and still f_3 be univalent:

$$f_1(z) = z - \frac{1}{m}\overline{z}^m \qquad \text{and} \qquad f_2(z) = z - \frac{1}{n}\overline{z}^n,$$

where $m, n \geq 2$. For various values of m, n, and t construct f_3 and use *ComplexTool* to see the images of $\mathbb{D}$ under f_3.

Investigate the examples above, and then construct other examples in which f_3 is univalent but the hypotheses of Theorem 4.100 do not hold. Using these examples make a conjecture for hypotheses of a new theorem that guarantees f_3 will be univalent. Prove this new theorem.

Optional

Now, let us look at an example that initially is surprising and is related to the nonconvex polygons described by Duren, McDougall, and Schaubroeck [**11**].

EXAMPLE 4.104. Let $f_1 = h_1 + \overline{g_1}$ be the harmonic square map in Example 4.101, where

$$h_1(z) = \frac{i}{4}\log\left(\frac{1 + z}{1 - z}\right) - \frac{1}{4}\log\left(\frac{i + z}{i - z}\right)$$

$$g_1(z) = \frac{i}{4}\log\left(\frac{1 + z}{1 - z}\right) + \frac{1}{4}\log\left(\frac{i + z}{i - z}\right).$$

We can write this as

$$f_1(z) = \text{Re}\left[\frac{i}{2}\log\left(\frac{1 + z}{1 - z}\right)\right] + i\,\text{Im}\left[-\frac{1}{2}\log\left(\frac{i + z}{i - z}\right)\right].$$

Using the same approach as in Example 4.72, we see that $z = e^{i\theta} \in \partial\mathbb{D}$ is mapped to

$$
u_1 + iv_1 = \begin{cases}
z_1 = \frac{\pi}{2\sqrt{2}}\, e^{i\frac{\pi}{4}} & \text{if } \theta \in (\frac{-\pi}{2}, 0), \\
z_3 = \frac{\pi}{2\sqrt{2}}\, e^{i\frac{3\pi}{4}} & \text{if } \theta \in (0, \frac{\pi}{2}), \\
z_5 = \frac{\pi}{2\sqrt{2}}\, e^{i\frac{5\pi}{4}} & \text{if } \theta \in (\frac{\pi}{2}, \pi), \\
z_7 = \frac{\pi}{2\sqrt{2}}\, e^{i\frac{7\pi}{4}} & \text{if } \theta \in (\pi, \frac{3\pi}{2}).
\end{cases}
$$

So, f_1 maps $\mathbb{D}$ onto a square region with vertices at z_1, z_3, z_5 and z_7 (see Figure 4.34). The dilatation for f_1 is $\omega = z^2$ and Condition A is satisfied. For example, for any sequence of points, z'_n, in the fourth quadrant approaching 1,

$$
\lim_{n\to\infty} \mathrm{Re}\{f_1(z'_n)\} = \frac{\pi}{2\sqrt{2}} = \sup_{|z|<1} \mathrm{Re}\{f_1(z)\}
$$

and for any sequence of points, z''_n, in the second quadrant approaching -1,

$$
\lim_{n\to\infty} \mathrm{Re}\{f_1(z''_n)\} = -\frac{\pi}{2\sqrt{2}} = \inf_{|z|<1} \mathrm{Re}\{f_1(z)\}.
$$
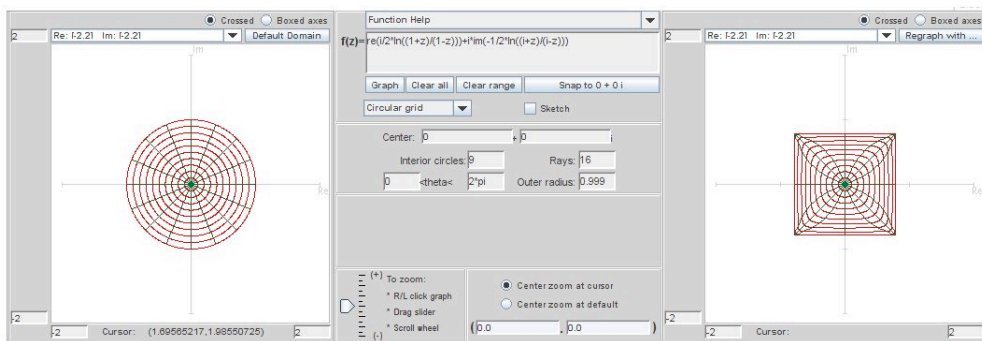
FIGURE 4.34. Image of $\mathbb{D}$ under f_1

Next, let $f_2 = h_2 + \overline{g_2}$, where

$$
h_2(z) = -\frac{1}{4}e^{-i\frac{3\pi}{4}} \log\left(\frac{e^{i\frac{\pi}{4}} + z}{e^{i\frac{\pi}{4}} - z}\right) - \frac{1}{4}e^{-i\frac{\pi}{4}} \log\left(\frac{e^{i\frac{3\pi}{4}} + z}{e^{i\frac{3\pi}{4}} - z}\right)
$$

$$
g_2(z) = \frac{1}{4}e^{i\frac{3\pi}{4}} \log\left(\frac{e^{i\frac{\pi}{4}} + z}{e^{i\frac{\pi}{4}} - z}\right) + \frac{1}{4}e^{i\frac{\pi}{4}} \log\left(\frac{e^{i\frac{3\pi}{4}} + z}{e^{i\frac{3\pi}{4}} - z}\right).
$$

Similar to above

$$f_2(z) = \text{Re}\left\{ \frac{i}{2\sqrt{2}} \left[\log\left(\frac{e^{i\frac{\pi}{4}} + z}{e^{i\frac{\pi}{4}} - z} \right) + \log\left(\frac{e^{i\frac{3\pi}{4}} + z}{e^{i\frac{3\pi}{4}} - z} \right) \right] \right\}$$

$$+ i\,\text{Im}\left\{ \frac{1}{2\sqrt{2}} \left[\log\left(\frac{e^{i\frac{\pi}{4}} + z}{e^{i\frac{\pi}{4}} - z} \right) - \log\left(\frac{e^{i\frac{3\pi}{4}} + z}{e^{i\frac{3\pi}{4}} - z} \right) \right] \right\},$$

$z = e^{i\theta} \in \partial\mathbb{D}$ is mapped to

$$u_2 + iv_2 = \begin{cases} z_0 = \frac{\pi}{2\sqrt{2}} & \text{if } \theta \in \left(\frac{-\pi}{4}, \frac{\pi}{4} \right), \\ z_2 = \frac{i\pi}{2\sqrt{2}} & \text{if } \theta \in \left(\frac{\pi}{4}, \frac{3\pi}{4} \right), \\ z_4 = -\frac{\pi}{2\sqrt{2}} & \text{if } \theta \in \left(\frac{3\pi}{4}, \frac{5\pi}{4} \right), \\ z_6 = -\frac{i\pi}{2\sqrt{2}} & \text{if } \theta \in \left(\frac{5\pi}{4}, \frac{7\pi}{4} \right). \end{cases}$$

That is, f_2 maps $\mathbb{D}$ onto a rotated square region with vertices at z_0, z_2, z_4 and z_6 (see Figure 4.35) with $\omega = z^2$, and it also satisfies Condition A.
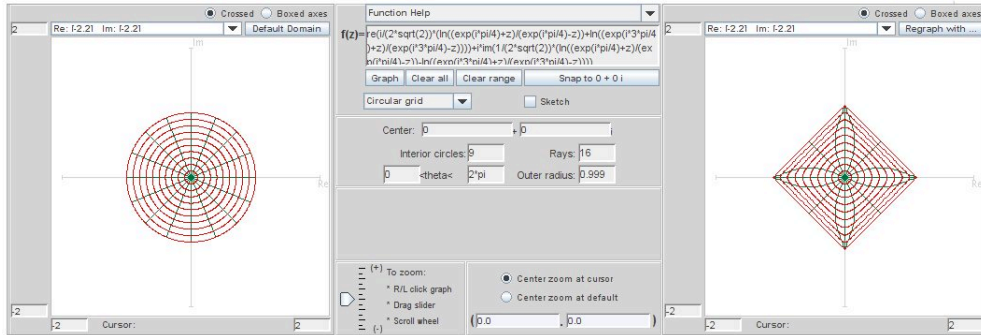
FIGURE 4.35. Image of $\mathbb{D}$ under f_2

By Theorem 4.100, $f_3 = tf_1 + (1 - t)f_2$ is univalent. What is the image of $\mathbb{D}$ under f_3? Let's look at the specific case when $t = \frac{1}{2}$. You might think that $f_3(\mathbb{D})$ would be just an overlaying of $f_1(\mathbb{D})$ and $f_2(\mathbb{D})$ (see Figure 4.36(a)). However, it is not. Instead, it is the nonconvex star shown in Figure 4.36(b).

Why is the correct image the nonconvex star in Figure 4.36(b)? Let's look where arcs of the unit circle are mapped under f_3. Notice that $f_1(e^{i\theta})$ and $f_2(e^{i\theta})$ depend upon which of eight arcs θ is in. For example, if $\theta \in (-\frac{\pi}{4}, 0)$, then $f_1(e^{i\theta}) = z_1$ and $f_2(e^{i\theta}) = z_0$, and so in this interval $f_3(e^{i\theta}) = \frac{z_1 + z_0}{2}$ (that is, it is the midpoint between z_1 and z_0). However, if $\theta \in (0, \frac{\pi}{4})$, then $f_1(e^{i\theta}) = z_3$ and $f_2(e^{i\theta}) = z_0$, and $f_3(e^{i\theta}) = \frac{z_3 + z_0}{2}$.

(a)   (b)
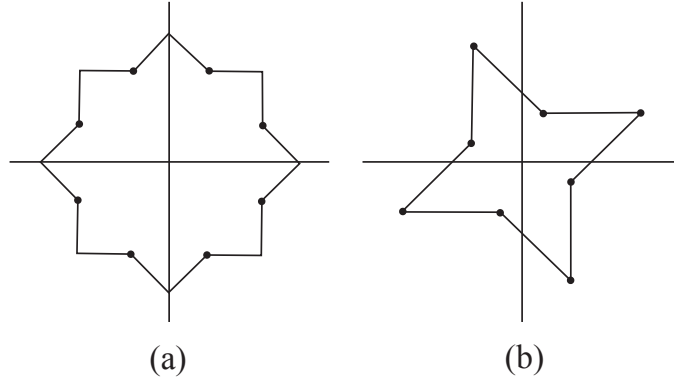
FIGURE 4.36. Which is the image of $f_3(\mathbb{D})$?

Specifically,

$$f_3(e^{i\theta}) = \begin{cases} w_1 = \frac{z_1+z_0}{2} = \frac{\pi}{2\sqrt{2}}\cos\frac{\pi}{8}e^{i\frac{\pi}{8}} & \text{if } \theta \in \left(\frac{-\pi}{4}, 0\right), \\ w_2 = \frac{z_3+z_0}{2} = \frac{\pi}{2\sqrt{2}}\cos\frac{3\pi}{8}e^{i\frac{3\pi}{8}} & \text{if } \theta \in \left(0, \frac{\pi}{4}\right), \\ w_3 = \frac{z_3+z_2}{2} = \frac{\pi}{2\sqrt{2}}\cos\frac{\pi}{8}e^{i\frac{5\pi}{8}} & \text{if } \theta \in \left(\frac{\pi}{4}, \frac{\pi}{2}\right), \\ w_4 = \frac{z_5+z_2}{2} = \frac{\pi}{2\sqrt{2}}\cos\frac{3\pi}{8}e^{i\frac{7\pi}{8}} & \text{if } \theta \in \left(\frac{\pi}{2}, \frac{3\pi}{4}\right), \\ w_5 = \frac{z_5+z_4}{2} = \frac{\pi}{2\sqrt{2}}\cos\frac{\pi}{8}e^{i\frac{9\pi}{8}} & \text{if } \theta \in \left(\frac{3\pi}{4}, \pi\right), \\ w_6 = \frac{z_7+z_4}{2} = \frac{\pi}{2\sqrt{2}}\cos\frac{3\pi}{8}e^{i\frac{11\pi}{8}} & \text{if } \theta \in \left(\pi, \frac{5\pi}{4}\right), \\ w_7 = \frac{z_7+z_6}{2} = \frac{\pi}{2\sqrt{2}}\cos\frac{\pi}{8}e^{i\frac{13\pi}{8}} & \text{if } \theta \in \left(\frac{5\pi}{4}, \frac{3\pi}{2}\right), \\ w_8 = \frac{z_1+z_6}{2} = \frac{\pi}{2\sqrt{2}}\cos\frac{3\pi}{8}e^{i\frac{15\pi}{8}} & \text{if } \theta \in \left(\frac{3\pi}{2}, \frac{7\pi}{4}\right). \end{cases}$$

Note that the vertices w_1, w_3, w_5 and w_7 lie equally spaced on a circle of radius $r_{outer} = \frac{\pi}{2\sqrt{2}}\cos\frac{\pi}{8} \approx 1.026$, while the vertices w_2, w_4, w_6 and w_8 lie equally spaced on a circle of radius $r_{inner} = \frac{\pi}{2\sqrt{2}}\cos\frac{3\pi}{8} \approx 0.425$.

We can visualize the boundary of $f_3(\mathbb{D})$ by plotting the eight vertices $z_0, z_1, \ldots z_7$ and drawing the midpoints $w_1, \ldots, w_8$ (see Figure 4.37).
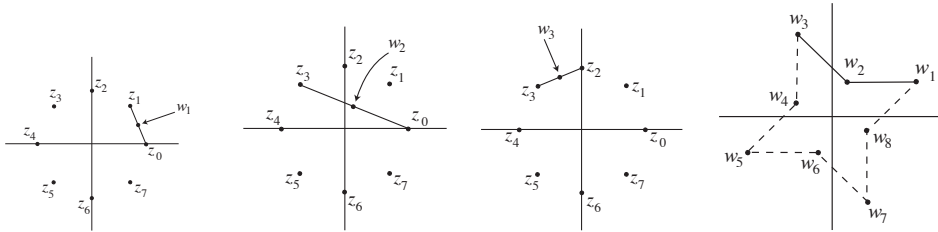
FIGURE 4.37. Visualizing the image of the boundary of $f_3(\mathbb{D})$

295

We can also explore the linear combination of these two functions that mapped onto rotated square regions by using the applet, *LinComboTool* (see Figure 4.38).
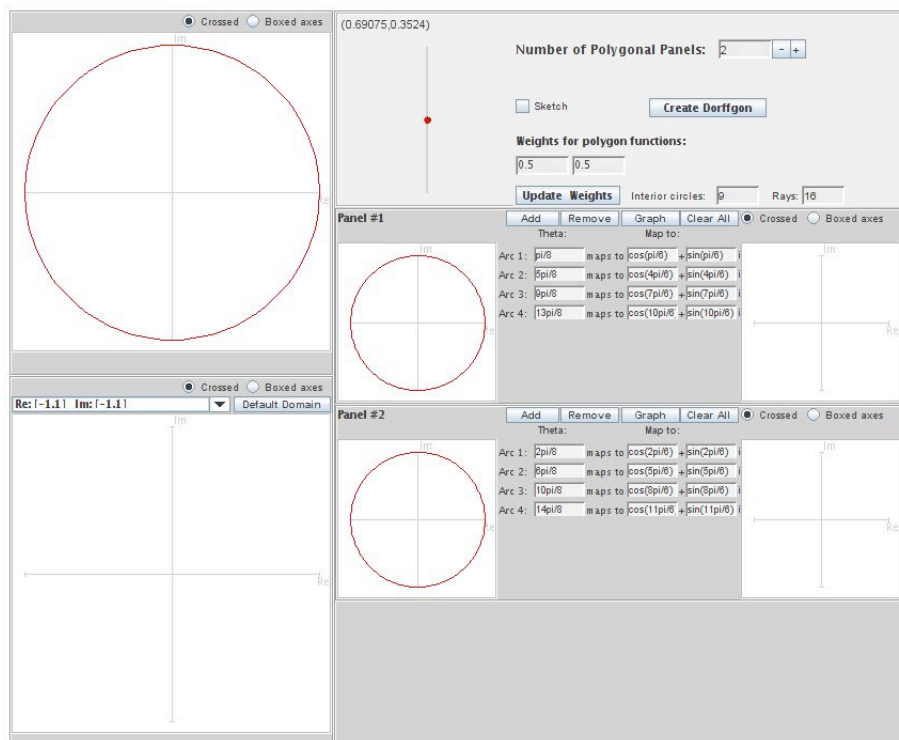
FIGURE 4.38. The applet *LinComboTool*

Open up *LinComboTool*. Make sure that at the top of the page, the **Number of Polygonal Pa...** is 2. In **Panel #1** enter the left end points of the intervals for the arcs of the unit circle used in function f_1 (these endpoints need to be positive numbers). Then enter the real and imaginary values of the image of this arc under the function f_1. For example, if we take the interval $(0, \frac{\pi}{2})$ for Arc 1 (note that we are starting with this interval because we need to use nonnegative values), then enter 0 for Arc 1, `pi/(2*sqrt(2))*cos(pi/4)` for the real value of its image, and `pi/(2*sqrt(2))*sin(pi/4)` for the imaginary value of its image. Remember that for Arc 4, we will use `3pi/2`. If there are not enough boxes for the arcs, click on the **Add** button to add an arc. Similarly, click on **Remove**, if there are too many boxes for the arcs. When you are done entering the points, click on **Graph** to produce the image $f_1(\mathbb{D})$. Then go to **Panel #2** and enter the points for f_2 and graph $f_2(\mathbb{D})$. After these are both graphed, click on **Create LinCombogon** and the corresponding image will appear in the lower lefthand box (see Figure 4.39).

EXERCISE 4.105. Using *LinComboTool*, start with the same arc values and corresponding point values as in Example 4.104. Note that you can change the value
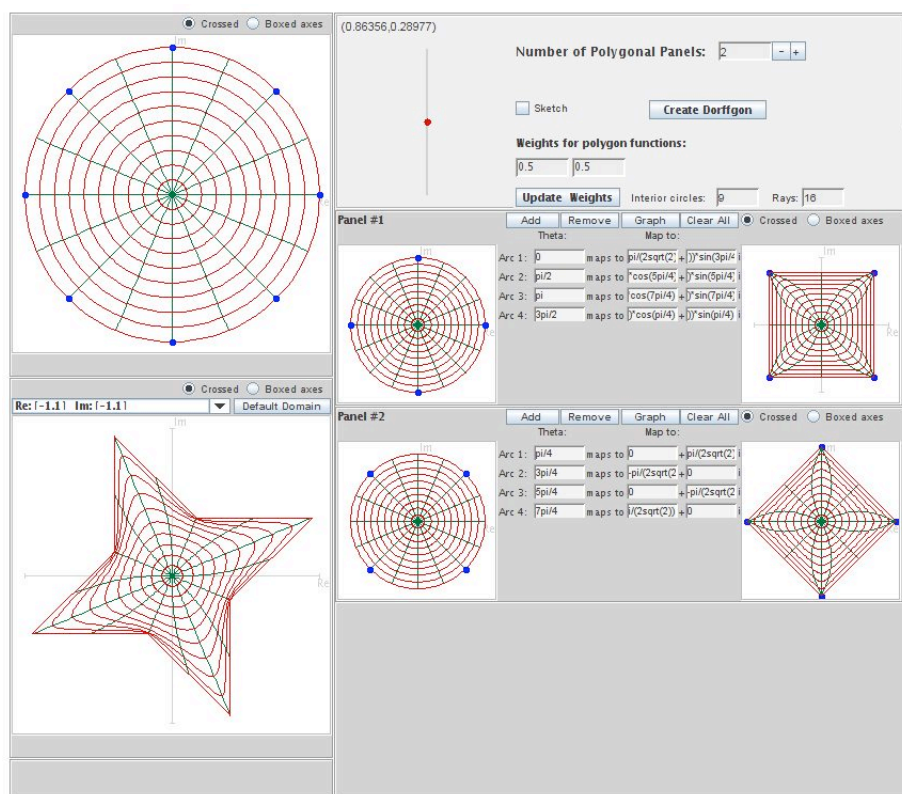
FIGURE 4.39. Image of $\mathbb{D}$ under f_3

of t by sliding up and down the red dot near the top of the page. Describe what happens as t varies from 0 to 1. In the example above we showed that when $t = \frac{1}{2}$, $r_{outer} = \frac{\pi}{2\sqrt{2}} \cos \frac{\pi}{8} \approx 1.026$ and $r_{inner} = \frac{\pi}{2\sqrt{2}} \cos \frac{3\pi}{8} \approx 0.425$. Compute r_{outer} and r_{inner} for any t $(0 \leq t \leq 1)$.

Try it out!

REMARK 4.106. In Theorem 4.100, we do not need that $\omega_1 = \omega_2$. Looking over the proof of this theorem, what is really needed is just that f_3 is locally univalent. This can be achieved if we have that

(60)
$$|\omega_3| = \left| \frac{t g_1' + (1-t) g_2'}{t h_1' + (1-t) h_2'} \right| < 1.$$

EXERCISE 4.107. We can have one pair of functions f_1, f_2 mapping onto image domains G_1, G_2, respectively, and another pair of functions $\tilde{f}_1$, $\tilde{f}_2$ that also map onto these same image domains G_1, G_2, but the linear combinations f_3 and $\tilde{f}_3$ map onto a different image domains.

297

Repeat the steps in Example 4.104 using the same function for f_2 but replacing f_1 with the harmonic square map in Example 4.72, where

$$h_1(z) = \frac{1}{4} \log \left(\frac{1+z}{1-z} \right) + \frac{i}{4} \log \left(\frac{i+z}{i-z} \right)$$

$$g_1(z) = -\frac{1}{4} \log \left(\frac{1+z}{1-z} \right) + \frac{i}{4} \log \left(\frac{i+z}{i-z} \right).$$

Note that

$$h_1'(z) = \frac{1}{1-z^4}, \qquad g_1'(z) = \frac{-z^2}{1-z^4}$$

$$h_2'(z) = \frac{1}{1+z^4}, \qquad g_2'(z) = \frac{z^2}{1+z^4}.$$

(a.) In this case, $\omega_1(z) = -z^2$ while $\omega_2(z) = z^2$. Using eq. (60) in Remark 4.106 above, show that f_3 is locally univalent.
(b.) Use *LinComboTool* find the image of $f_3(\mathbb{D})$ using this f_1 and f_2.
(c.) Explain why this happens by using the approach in Example 4.104 to compute the new values of $w_1, \ldots, w_8$ and then use the visualization technique in the example to plot the eight vertices $z_0, \ldots, z_7$ and draw the midpoints $w_1, \ldots, w_8$.

Try it out!

EXERCISE 4.108. Repeat the steps in Exercise 4.107 using the same function for f_1 but replacing f_2 with the harmonic hexagon map that can be derived from eq (58) for h' and g' at the end of Example 4.72 , where

$$h_2'(z) = \frac{1}{1-z^6} \Rightarrow$$

$$h_2(z) = \frac{1}{6} \log \left(\frac{1+z}{1-z} \right) + \frac{e^{\frac{-i\pi}{3}}}{6} \log \left(\frac{1 + e^{\frac{i\pi}{3}} z}{1 - e^{\frac{i\pi}{3}} z} \right) + \frac{e^{\frac{-i2\pi}{3}}}{6} \log \left(\frac{1 + e^{\frac{i2\pi}{3}} z}{1 - e^{\frac{i2\pi}{3}} z} \right)$$

$$g_2'(z) = \frac{-z^4}{1-z^6} \Rightarrow$$

$$g_2(z) = -\frac{1}{6} \log \left(\frac{1+z}{1-z} \right) - \frac{e^{\frac{i\pi}{3}}}{6} \log \left(\frac{1 + e^{\frac{i\pi}{3}} z}{1 - e^{\frac{i\pi}{3}} z} \right) - \frac{e^{\frac{i2\pi}{3}}}{6} \log \left(\frac{1 + e^{\frac{i2\pi}{3}} z}{1 - e^{\frac{i2\pi}{3}} z} \right).$$

(a.) In this case, $\omega_1(z) = -z^2$ while $\omega_2(z) = -z^4$. Using eq. (60) in the remark above, show that f_3 is locally univalent.
(b.) Use *LinComboTool* find the image of $f_3(\mathbb{D})$ using this f_1 and f_2.
(c.) Explain why this happens by using the approach in Example 4.104 to compute the new values of the vertices of $f_3(\mathbb{D})$.

Try it out!
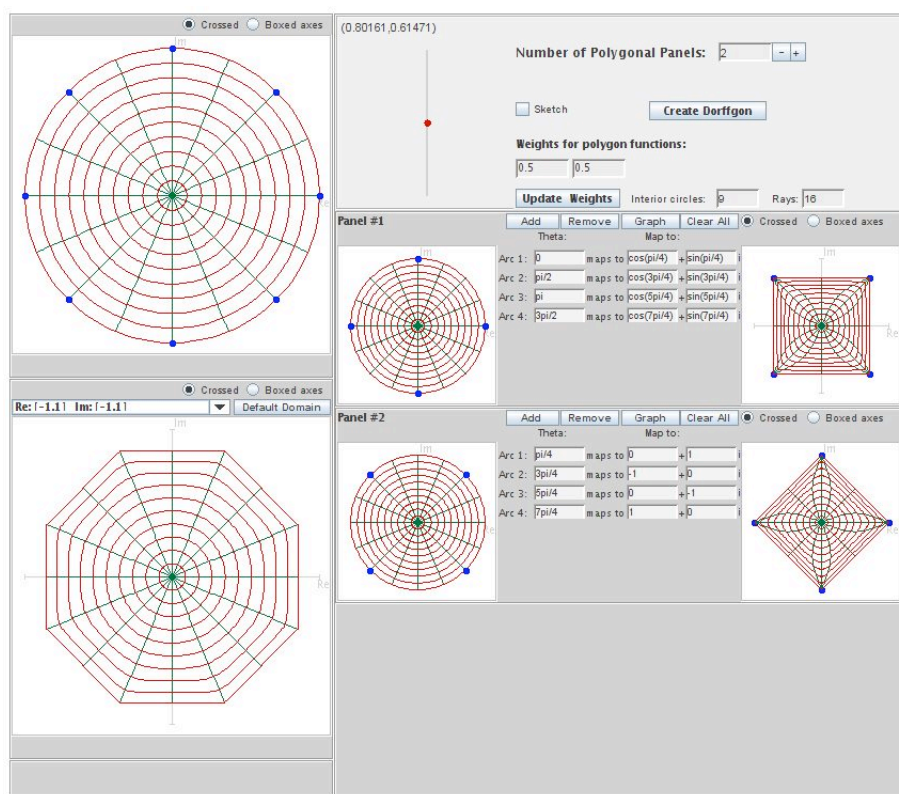
FIGURE 4.40. Image of $\mathbb{D}$ under f_3 in Exercise 4.107

We can generalize Theorem 4.100 to include the linear combination of n functions $f_1, \ldots, f_n$.

THEOREM 4.109. Let $f_1 = h_1 + \overline{g_1}$, ..., $f_n = h_n + \overline{g_n}$ be n univalent harmonic mappings convex in the imaginary direction and $\omega_1 = \cdots = \omega_n$. If f_1, ..., f_n satisfy condition A, then $F = t_1 f_1 + \cdots + t_n f_n$ is convex in the imaginary direction, where $0 \leq t_n \leq 1$ and $t_1 + \cdots + t_n = 1$.

EXERCISE 4.110. Prove Theorem 4.109.
Try it out!

EXPLORATION 4.111. Using the Theorem 4.109, create three maps in three different panels of *LinComboTool*, where each map takes 4 arcs on the unit circle to 4 vertices of a square. Make sure that these maps satisfy the conditions of the theorem. Then click on the **Create LinCombogon** button to see the resulting image domain. Explore this idea by using different maps in the panels. For an example, see Figure 4.41
Try it out!
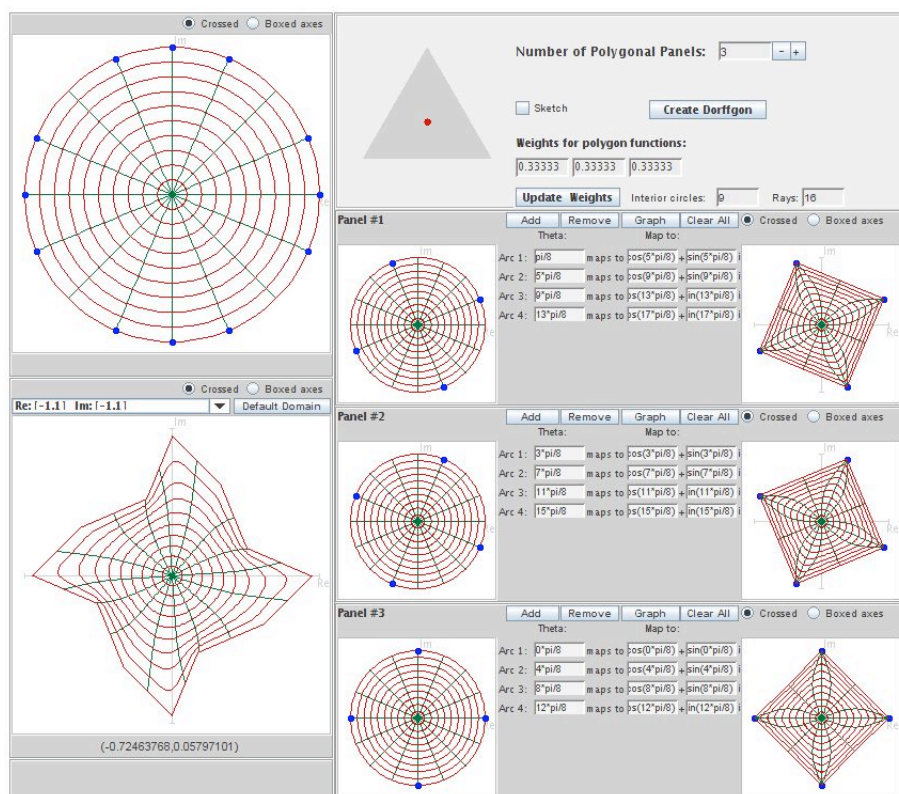
FIGURE 4.41. Example of image of $\mathbb{D}$ under the linear combination of three squares

LARGE PROJECT 4.112. In Example 4.104 and in Exercise 4.107 we took the linear combinations of two harmonic square mappings and ended up with fundamentally different images. Explore this with other n-gons. In particular, use *LinComboTool* to determine what and how many fundamentally different (i.e., not rotations or not scalings) images can be constructed when taking the linear combination with $t = \frac{1}{2}$ of two harmonic 5-gon maps? 6-gon maps? n-gon maps. Make sure that Condition A holds in every case and that $|\omega_3| < 1$.
Optional

LARGE PROJECT 4.113. In Exercise 4.108 we took the linear combinations of a harmonic 4-gon mapping and a harmonic 6-gon mapping with dilatations $-z^2$ and $-z^4$, respectively. Use *LinComboTool* to determine what combinations are possible and what images can be constructed when taking the linear combination with $t = \frac{1}{2}$ of a harmonic m-gon and n-gon, where $m < n$. Make sure that Condition A holds in every case and that $|\omega_3(z)| < 1$.
Optional

4.8. Convolutions

Another way of combining two univalent functions is the Hadamard product or convolution. For analytic functions

$$f(z) = \sum_{n=0}^{\infty} a_n z^n \quad \text{and} \quad F(z) = \sum_{n=0}^{\infty} A_n z^n,$$

their convolution is defined as

$$f(z) * F(z) = \sum_{n=0}^{\infty} a_n A_n z^n.$$

EXAMPLE 4.114. Consider the convolution of the right half-plane function (see Example 4.22)

$$f(z) = \frac{z}{1-z} = \sum_{n=1}^{\infty} z^n$$

and the Koebe function (see Example 4.23)

$$F(z) = \frac{z}{(1-z)^2} = \sum_{n=1}^{\infty} n z^n.$$

Then

$$
\begin{aligned}
f(z) * F(z) &= \frac{z}{1-z} * \frac{z}{(1-z)^2} \\
&= \sum_{n=1}^{\infty} z^n * \sum_{n=1}^{\infty} n z^n \\
&= (z + z^2 + z^3 + z^4 + \cdots) * (z + 2z^2 + 3z^3 + 4z^4 + \cdots) \\
&= (z + 2z^2 + 3z^3 + 4z^4 + \cdots) \\
&= \frac{z}{(1-z)^2}.
\end{aligned}
$$
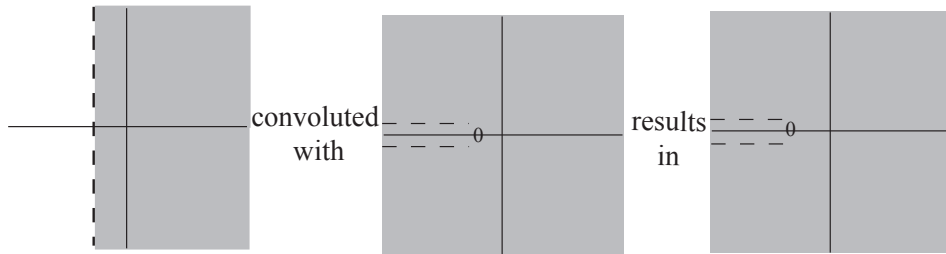
FIGURE 4.42. Right half-plane map convoluted with the Koebe function yields the Koebe function.

301

EXAMPLE 4.115. Now, consider the convolution of the Koebe function, $f(z) = \dfrac{z}{(1-z)^2}$, and the horizontal strip map, $F(z) = \dfrac{1}{2}\log\left(\dfrac{1+z}{1-z}\right)$. What is the Hadamard product, $f(z) * F(z)$? We need to compute the Taylor series for F. To do so, notice that

$$\log(1-z) = \int \frac{-1}{1-z}\, dz = -\int \sum_{n=0}^{\infty} z^n\, dz = \sum_{n=0}^{\infty} \frac{-1}{n+1} z^{n+1}.$$

Likewise,

$$\log(1+z) = \sum_{n=0}^{\infty} (-1)^{n+1} \frac{1}{n+1} z^{n+1}.$$

Hence,

$$\frac{1}{2}\log\left(\frac{1+z}{1-z}\right) = \sum_{n=0}^{\infty} (-1)^{n+1} \frac{1}{n+1} z^{n+1} - \sum_{n=0}^{\infty} \frac{-1}{n+1} z^{n+1}$$

$$= \sum_{n=0}^{\infty} \frac{1}{2n+1} z^{2n+1}$$

$$= z + \frac{1}{3} z^3 + \frac{1}{5} z^5 + \cdots$$

Thus,

$$f(z) * F(z) = \frac{z}{(1-z)^2} * \frac{1}{2}\log\left(\frac{1+z}{1-z}\right)$$

$$= \sum_{n=1}^{\infty} n z^n * \sum_{n=0}^{\infty} \frac{1}{2n+1} z^{2n+1}$$

$$= (z + 2z^2 + 3z^3 + 4z^4 + 5z^5 + \cdots) * (z + \frac{1}{3} z^3 + \frac{1}{5} z^5 + \cdots)$$

$$= z + z^3 + z^5 + \cdots.$$

Since $\dfrac{1}{1-z} = 1 + z + z^2 + z^3 + \cdots$, we have that $\dfrac{1}{1-z^2} = 1 + z^2 + z^4 + z^6 + \cdots$ and $\dfrac{z}{1-z^2} = z + z^3 + z^5 + \cdots$. That is,

$$f(z) * F(z) = \frac{z}{(1-z)^2} * \frac{1}{2}\log\left(\frac{1+z}{1-z}\right) = \frac{z}{1-z^2}.$$

EXERCISE 4.116. Let $f(z) = -\log(1-z)$ and $F(z) = \dfrac{z}{(1-z)^2}$. Determine $f(z) * F(z)$.
Try it out!

PROPOSITION 4.117.
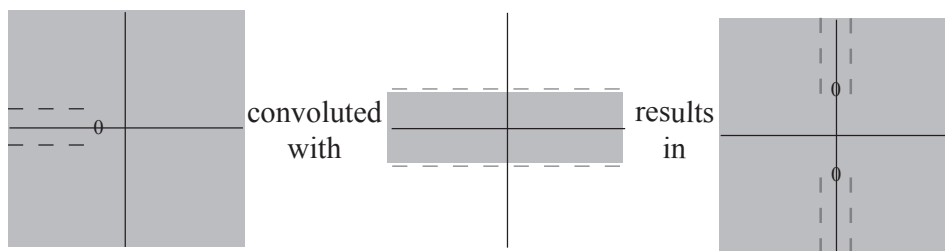
FIGURE 4.43. The Koebe function convoluted with a horizontal strip map yields a double-slit map.

(a) The right half-plane mapping, $f(z) = \dfrac{z}{1-z}$, acts as the convolution identity; that is, if F is an analytic function, then $\dfrac{z}{1-z} * F(z) = F(z)$.

(b) The Koebe function, $f(z) = \dfrac{z}{(1-z)^2}$, acts as a differential operator; that is, if $F(z)$ is an analytic function, then $\dfrac{z}{(1-z)^2} * F(z) = zF'(z)$.

(c) Convolution is commutative; that is, if f_1, f_2 are analytic functions, then $f_1 * f_2 = f_2 * f_1$.

(d) If f_1, f_2 are analytic functions, then $(f_1(z) * f_2(z))' = zf_1'(z) * f_2(z)$.

EXERCISE 4.118. Prove Proposition 4.117 (a)-(d).
Try it out!

Note that if $f_1, f_2 \in S$, then $f_1 * f_2$ may not be in S. For example,

$$\frac{z}{(1-z)^2} * \frac{z}{(1-z)^2} = \sum_{n=1}^{\infty} nz^n * \sum_{n=1}^{\infty} nz^n$$

$$= \sum_{n=1}^{\infty} n^2 z^n \notin S.$$

Why do we know that $\sum_{n=1}^{\infty} n^2 z^n \notin S$?

However, we do have the following results by Ruscheweyh and Sheil-Small. Note that if the analytic function, $f \in S$, maps onto a domain that is convex, then we will denote that by writing $f \in K$. Similarly, if the harmonic function, $f \in S_H$, maps onto a domain that is convex, then we will write $f \in K_H$.

THEOREM 4.119. Let $f, f_1 \in K$. Then $f * f_1 \in K$. In addition, if $f_2, f_3 \in S$ map onto a close-to-convex, and a starlike domain, respectively. Then $f * f_2, f * f_3$ are in S and map onto a close-to-convex, and a starlike domain, respectively.

Now let's consider the case of harmonic convolutions.

DEFINITION 4.120. For harmonic univalent functions

$$f(z) = h(z) + \overline{g}(z) = z + \sum_{n=2}^{\infty} a_n z^n + \sum_{n=1}^{\infty} \overline{b_n}\, \overline{z}^n \text{ and}$$

$$F(z) = H(z) + \overline{G}(z) = z + \sum_{n=2}^{\infty} A_n z^n + \sum_{n=1}^{\infty} \overline{B_n}\, \overline{z}^n,$$

define the *harmonic convolution* as

$$(61) \quad f(z) * F(z) = h(z) * H(z) + \overline{g(z) * G(z)} = z + \sum_{n=2}^{\infty} a_n A_n z^n + \sum_{n=1}^{\infty} \overline{b_n B_n}\, \overline{z}^n.$$

As mentioned above, for the convolution of analytic functions it is known that if $f_1, f_2 \in K$, then $f_1 * f_2 \in K$. Is such a similar result true for harmonic univalent convex mappings?

There are a few known results about harmonic convolutions of functions on $\mathbb{D}$.

THEOREM 4.121 (Clunie and Sheil-Small). If $f \in K_H$ and $\varphi \in S$, then the functions

$$f * (\alpha \overline{\varphi} + \varphi) \in S_H$$

map $\mathbb{D}$ onto a close-to-convex domain, where $(|\alpha| \leq 1)$.

Clunie and Sheil-Small posed the following open problem (see [**5**]).

OPEN PROBLEM 4.122. Let $f \in K_H$, then what are the collection of harmonic functions F such $f * F \in K_H$?

As partial answers to this open question, there are the following results.

THEOREM 4.123 (Ruscheweyh and Salinas). Let g be analytic in $\mathbb{D}$. Then

$$f \,\widetilde{*}\, g = \mathrm{Re}\{f\} * g + \overline{\mathrm{Im}\{f\} * g} \in K_H$$

for all $f \in K_H \iff$ for each $\gamma \in \mathbb{R}, g + i\gamma z g'$ is convex in the imaginary direction.

THEOREM 4.124 (Goodloe). Let $f_m, f_n \in K_H$ be the canonical harmonic functions that map $\mathbb{D}$ onto the regular m-gon and n-gon, respectively. Then $f_m * f_n \in K_H$ and the image of $\mathbb{D}$ is a p-gon, where $p = lcm(m,n)$.

EXERCISE 4.125. Compute $f_k = f_4 * f_6$, where $f_4 = h_4 + \overline{g_4}$ is the canonical square map (see Example 4.101) given by

$$h_4(z) = \frac{1}{4} \log\left(\frac{1+z}{1-z}\right) + \frac{i}{4} \log\left(\frac{i+z}{i-z}\right) = \int \frac{1}{1-z^4}\, dz$$

$$g_4(z) = -\frac{1}{4} \log\left(\frac{1+z}{1-z}\right) + \frac{i}{4} \log\left(\frac{i+z}{i-z}\right) = \int \frac{-z^2}{1-z^4}\, dz.$$

and $f_6 = h_6 + \overline{g_6}$ is the canonical regular hexagon map (see Exercise 4.108) given by

$$h_6(z) = \frac{1}{6}\log\left(\frac{1+z}{1-z}\right) + \frac{e^{\frac{-i\pi}{3}}}{6}\log\left(\frac{1+e^{\frac{i\pi}{3}}z}{1-e^{\frac{i\pi}{3}}z}\right) + \frac{e^{\frac{-i2\pi}{3}}}{6}\log\left(\frac{1+e^{\frac{i2\pi}{3}}z}{1-e^{\frac{i2\pi}{3}}z}\right) = \int \frac{1}{1-z^6}\,dz$$

$$g_6(z) = -\frac{1}{6}\log\left(\frac{1+z}{1-z}\right) - \frac{e^{\frac{i\pi}{3}}}{6}\log\left(\frac{1+e^{\frac{i\pi}{3}}z}{1-e^{\frac{i\pi}{3}}z}\right) - \frac{e^{\frac{i2\pi}{3}}}{6}\log\left(\frac{1+e^{\frac{i2\pi}{3}}z}{1-e^{\frac{i2\pi}{3}}z}\right) = \int \frac{-z^4}{1-z^6}\,dz.$$

Sketch $f_k(\mathbb{D})$ using *ComplexTool*.
    Try it out!

    In considering Open Problem 4.122, let's look at a simple problem: if $f_1, f_2 \in K_H$, then when is $f_1 * f_2 \in S_H$?
    Recall Lewy's Theorem that $f = h + \overline{g}$ with $h'(z) \neq 0$ in $\mathbb{D}$ is locally univalent and sense-preserving if and only if $\omega(z) = g'(z)/h'(z)| < 1, \forall z \in \mathbb{D}$.

    THEOREM 4.126. Let $f_1 = h_1 + \overline{g}_1, f_2 = h_2 + \overline{g}_2 \in K_H$ with $h_k(z) + g_k(z) = \frac{z}{1-z}$ for $k = 1, 2$. If $f_1 * f_2$ is locally univalent and sense-preserving, then $f_1 * f_2 \in S_H$ and is convex in the direction of the real axis.

    PROOF. Since $h(z) + g(z) = \dfrac{z}{1-z}$ and $F(z) * \dfrac{z}{1-z} = F(z)$ for any analytic function F, we have that

$$\begin{aligned}
h_2 - g_2 &= (h_1 + g_1) * (h_2 - g_2) \\
&= h_1 * h_2 - h_1 * g_2 + h_2 * g_1 - g_1 * g_2 \\
h_1 - g_1 &= (h_1 - g_1) * (h_2 + g_2) = \\
&\quad h_1 * h_2 + h_1 * g_2 - h_2 * g_1 - g_1 * g_2.
\end{aligned}$$

Thus,

(62) $$h_1 * h_2 - g_1 * g_2 = \tfrac{1}{2}[(h_1 - g_1) + (h_2 - g_2)].$$

    We will now show that $(h_1 - g_1) + (h_2 - g_2)$ is convex in the direction of the real axis. Note that

$$h'(z) - g'(z) = \big(h'(z) + g'(z)\big)\left(\frac{h'(z) - g'(z)}{h'(z) + g'(z)}\right) = \big(h'(z) + g'(z)\big)\left(\frac{1 - \omega(z)}{1 + \omega(z)}\right) = \frac{p(z)}{(1-z)^2},$$

where $\operatorname{Re}\{p(z)\} > 0, \forall z \in \mathbb{D}$.
So, letting $\varphi(z) = z/(1-z)^2$, we have

$$\begin{aligned}
\operatorname{Re}\left\{\frac{z[(h_1'(z) - g_1'(z)) + (h_2'(z) - g_2'(z))]}{\varphi(Z)}\right\} &= \operatorname{Re}\left\{\frac{\frac{z}{(1-z)^2}[p_1(z) + p_2(z)]}{\frac{z}{(1-z)^2}}\right\} \\
&= \operatorname{Re}\left\{p_1(z) + p_2(z)\right\} > 0.
\end{aligned}$$

Therefore, by Theorem 4.88 in Section 4.6 and eq.(62), $h_1 * h_2 - g_1 * g_2$ is convex in the direction of the real axis.

Finally, since we assumed that $f_1 * f_2$ is locally univalent, we apply Clunie and Sheil-Small's Shearing Theorem (see Theorem 4.52) to get that $f_1 * f_2 = h_1 * h_2 - g_1 * g_2$ is convex in the direction of the real axis. $\qquad\square$

It is known (see [**7**]), that for any right half-plane mapping $f = h + \bar{g} \in K_H$,

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

Hence, Theorem 4.126 applies to harmonic right half-plane mappings.

EXAMPLE 4.127. Let $f_0 = h_0 + \bar{g}_0$ be the canonical right half-plane mapping given in Example 4.40 with $h_0(z) + g_0(z) = \frac{z}{1-z}$ with $\omega(z) = -z$. Then

$$h_0(z) = \frac{z - \frac{1}{2}z^2}{(1 - z)^2}$$

$$g_0(z) = -\frac{\frac{1}{2}z^2}{(1 - z)^2}.$$

Next, let $f_1 = h_1 + \bar{g}_1$, where $h_1(z) + g_1(z) = \frac{z}{1-z}$ with $\omega(z) = z$. Then

$$h_1(z) = \frac{1}{4} \log \left(\frac{1 + z}{1 - z} \right) + \frac{1}{2} \frac{z}{1 - z}$$

$$g_1(z) = -\frac{1}{4} \log \left(\frac{1 + z}{1 - z} \right) + \frac{1}{2} \frac{z}{1 - z}.$$

Note that f_1 is a right half-strip mapping (see Figure 4.44).
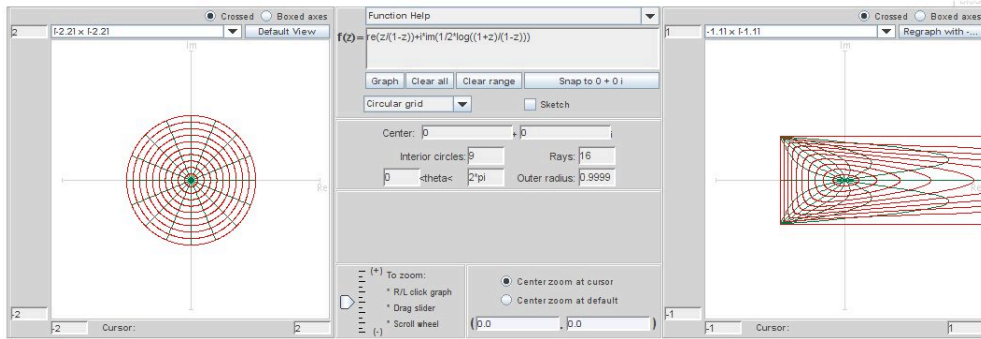
FIGURE 4.44. Image of $\mathbb{D}$ under f_1.

306

Consider $F_1 = f_0 * f_1 = H_1 + \overline{G_1}$. Note that

$$H_1(Z) = h_0(z) * h_1(z) = \frac{1}{2}\left[h_1(z) + zh_1'(z)\right] = \frac{1}{8}\log\left(\frac{1+z}{1-z}\right) + \frac{\frac{3}{4}z - \frac{1}{4}z^3}{(1-z)^2(1+z)}$$

$$G_1(z) = g_0(z) * g_1(z) = \frac{1}{2}\left[g_1(z) - zg_1'(z)\right] = -\frac{1}{8}\log\left(\frac{1+z}{1-z}\right) + \frac{\frac{1}{4}z - \frac{1}{2}z^2 - \frac{1}{4}z^3}{(1-z)^2(1+z)},$$

with

$$\widetilde{\omega}(z) = -z\left(\frac{2z^2 + z + 1}{z^2 + z + 2}\right).$$

The image of $\mathbb{D}$ under $F_1 = f_0 * f_1$ is shown in Figure 4.45.
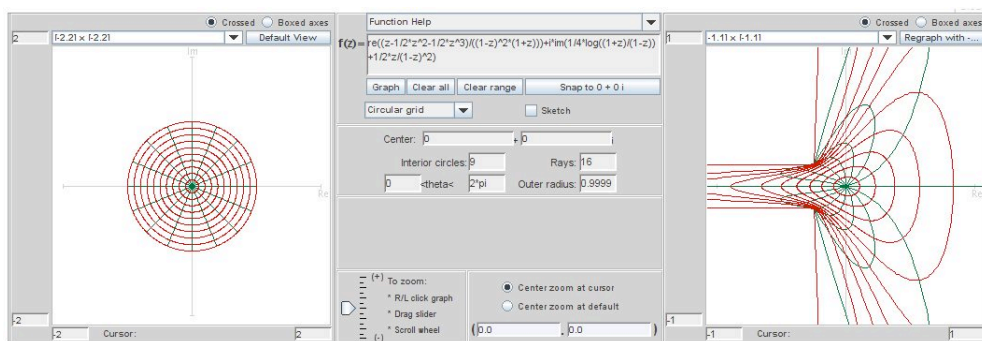
FIGURE 4.45. Image of $\mathbb{D}$ under $F_1 = f_0 * f_1$.

EXERCISE 4.128. Compute $F = H + \overline{G}$, where $F = f_0 * f_0$. Sketch $F(\mathbb{D})$ using *ComplexTool*.
Try it out!

Throughout the rest of this section we will consider the question "For which dilatation functions, $\omega = g'/h'$, is the function $f = h + \overline{g}$ locally univalent?" In doing so, let

$$f_0(z) = h_0(z) + \overline{g_0(z)} = \frac{z - \frac{1}{2}z^2}{(1-z)^2} - \frac{\overline{\frac{1}{2}z^2}}{(1-z)^2}$$

be the canonical right half-plane mapping used in Example 4.127.

Also, as mentioned in the proof above, the collection of functions $f = h + \overline{g} \in S_H^O$ that map $\mathbb{D}$ onto the right half-plane, $R = \{w : \operatorname{Re}(w) > -1/2\}$, have the form

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

We will use the following method to prove that local univalency holds:

METHOD 1. (Cohn's Rule, see [**19**], p 375) Given a polynomial

$$f(z) = a_0 + a_1 z + \cdots + a_n z^n$$

of degree n, let

$$f^*(z) = z^n \, \overline{f(1/\,\overline{z})} = \overline{a_n} + \overline{a_{n-1}} z + \cdots + \overline{a_0} z^n.$$

Denote by p and s the number of zeros of f inside the unit circle and on it, respectively. If $|a_0| < |a_n|$, then

$$f_1(z) = \frac{\overline{a_n} f(z) - a_0 f^*(z)}{z}$$

is of degree $n - 1$ with $p_1 = p - 1$ and $s_1 = s$ the number of zeros of f_1 inside the unit circle and on it, respectively.

THEOREM 4.129. Let $f = h + \overline{g} \in K_H^O$ with $h(z) + g(z) = \frac{z}{1-z}$ and $\omega(z) = e^{i\theta} z^n$ ($n \in \mathbb{N}$ and $\theta \in \mathbb{R}$). If $n = 1, 2$, then $f_0 * f \in S_H^O$ and is convex in the direction of the real axis.

PROOF. Let the dilatation of $f_0 * f$ be given by $\widetilde{\omega} = (g_0 * g)'/(h_0 * h)'$. By Theorem 4.126 and by Lewy's Theorem, we just need to show that $|\widetilde{\omega}(z)| < 1, \forall z \in \mathbb{D}$.

First, note that if F is analytic in $\mathbb{D}$ and $F(0) = 0$, then

$$h_0(z) * F(z) = \frac{1}{2} \big[F(z) + z F'(z) \big]$$

$$g_0(z) * F(z) = \frac{1}{2} \big[F(z) - z F'(z) \big].$$

Also, since $g'(z) = \omega(z) h'(z)$, we know $g''(z) = \omega(z) h''(z) + \omega'(z) h'(z)$.

Hence

(63) $$\widetilde{\omega}(z) = -\frac{z g''(z)}{2 h'(z) + z h''(z)} = \frac{-z \omega'(z) h'(z) - z \omega(z) h''(z)}{2 h'(z) + z h''(z)}.$$

Using $h(z) + g(z) = \frac{z}{1-z}$ and $g'(z) = \omega(z) h'(z)$, we can solve for $h'(z)$ and $h''(z)$ in terms of z and $\omega(z)$:

$$h'(z) = \frac{1}{(1 + \omega(z))(1 - z)^2}$$

$$h''(z) = \frac{2(1 + \omega(z)) - \omega'(z)(1 - z)}{(1 + \omega(z))^2 (1 - z)^3}.$$

Substituting these formulas for h' and h'' into the equation for $\widetilde{\omega}$, we derive:

(64)
$$\widetilde{\omega}(z) = \frac{-z\omega'(z)h'(z) - z\omega(z)h''(z)}{2h'(z) + zh''(z)}$$
$$= -z\frac{\omega^2(z) + [\omega(z) - \frac{1}{2}\omega'(z)z] + \frac{1}{2}\omega'(z)}{1 + [\omega(z) - \frac{1}{2}\omega'(z)z] + \frac{1}{2}\omega'(z)z^2}.$$

Now, consider the case in which $\omega(z) = e^{i\theta}z$. Then eq (64) yields

$$\widetilde{\omega}(z) = -ze^{2i\theta}\frac{\left(z^2 + \frac{1}{2}e^{-i\theta}z + \frac{1}{2}e^{-i\theta}\right)}{\left(1 + \frac{1}{2}e^{i\theta}z + \frac{1}{2}e^{i\theta}z^2\right)} = -ze^{2i\theta}\frac{p(z)}{q(z)}.$$

Note that $q(z) = z^2\ \overline{p(1/\overline{z})}$. In such a situation, if z_0 is a zero of p, then $\frac{1}{\overline{z_0}}$ is a zero of q. Hence,

$$\widetilde{\omega}(z) = -ze^{2i\theta}\frac{(z + A)(z + B)}{(1 + \overline{A}z)(1 + \overline{B}z)}.$$

Using Method 1, we have

$$p_1(z) = \frac{\overline{a_2}p(z) - a_0p^*(z)}{z} = \frac{3}{4}z + \left(\frac{1}{2}e^{-i\theta} - \frac{1}{4}\right).$$

So, p_1 has one zero at $z_0 = \frac{1}{3} - \frac{2}{3}e^{-i\theta} \in \mathbb{D}$. By Cohn's Rule, p has two zeros, namely A and B, with $|A|, |B| < 1$.

Next, consider the case in which $\omega(z) = e^{i\theta}z^2$. In this case,

$$|\widetilde{\omega}(z)| = |z^2|\left|\frac{z^3 + e^{-i\theta}}{1 + e^{i\theta}z^3}\right| = |z|^2 < 1.$$

$\square$

EXAMPLE 4.130. Let $f_2 = h_2 + \overline{g_2}$ be the harmonic mapping in $\mathbb{D}$ such that $h_1(z) + g_2(z) = \frac{z}{1-z}$ and $\omega_2(z) = -z^2$. Then we can compute

$$h_2(z) = \frac{1}{8}\ln\left(\frac{1+z}{1-z}\right) + \frac{1}{2}\frac{z}{1-z} + \frac{1}{4}\frac{z}{(1-z)^2}$$
$$g_2(z) = -\frac{1}{8}\ln\left(\frac{1+z}{1-z}\right) + \frac{1}{2}\frac{z}{1-z} - \frac{1}{4}\frac{z}{(1-z)^2}$$

and the image of $\mathbb{D}$ under f_2 is the right half-plane, $R = \{w \in \mathbb{C}\ |\ \text{Re}\{w\} \geq -\frac{1}{2}\}$. Note that $f_2(e^{it}) = \frac{1}{2} + i\frac{\pi}{16}$, if $0 < t < \pi$, and $f_2(e^{it}) = \frac{1}{2} - i\frac{\pi}{16}$, if $\pi < t < 2\pi$ (see Figure 4.46).

Let

$$F_2 = h_0 * h_2 + \overline{g_0 * g_2} = H_2 + \overline{G_2}.$$
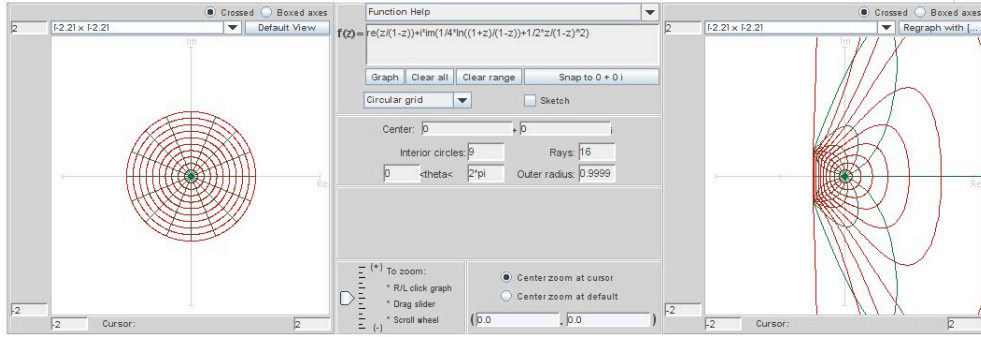
FIGURE 4.46. Image of $\mathbb{D}$ under f_2.

Then we can compute that

$$H_2(z) = \frac{1}{16} \ln\left(\frac{1+z}{1-z}\right) + \frac{1}{4}\frac{z}{1-z} + \frac{1}{8}\frac{z}{(1-z)^2} + \frac{1}{2}\frac{z}{(1-z)^3(1+z)}$$

(65)

$$G_2(z) = -\frac{1}{16} \ln\left(\frac{1+z}{1-z}\right) + \frac{1}{4}\frac{z}{1-z} - \frac{1}{8}\frac{z}{(1-z)^2} + \frac{1}{2}\frac{z^3}{(1-z)^3(1+z)}.$$

It can be shown analytically that $F_2(\mathbb{D})$ is the entire complex plane minus two half-lines given by $x \pm \frac{\pi}{16}i$, $x \leq -\frac{1}{4}$. This is not clear if we use *ComplexTool* with the standard settings to view this image (see Figure 4.47). However, using both this image and the image of just the unit circle (see Figure 4.48), this result seems reasonable [Note: to graph the image of $\partial\mathbb{D}$ in *ComplexTool*, change the settings in the middle box of *ComplexTool* to **Interior circles: 0** and **Rays: 0**].
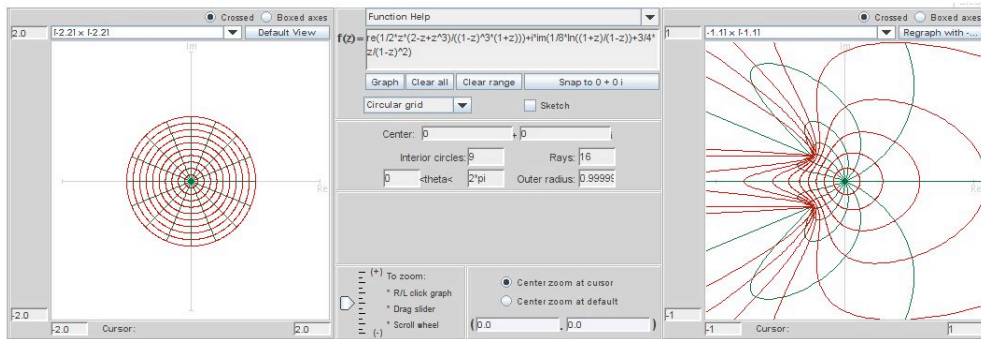
FIGURE 4.47. Image of $\mathbb{D}$ under $F_2 = f_0 * f_2$.

REMARK 4.131. If we assume the hypotheses of the previous theorem with the exception of making $n \geq 3$, then for each n we can find a specific $\omega(z) = e^{i\theta}z^n$ such

310
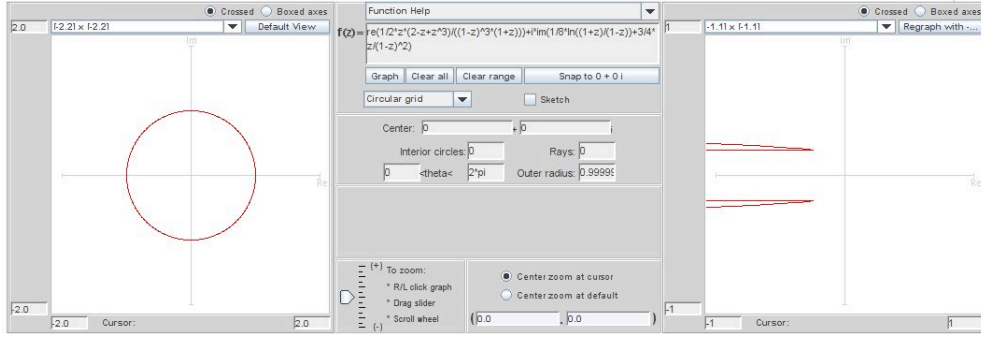
FIGURE 4.48. Image of $\partial \mathbb{D}$ under $F_2 = f_0 * f_2$.

that $f_0 * f \notin S_H^O$. For example, if n is odd, let $\omega(z) = -z^n$ and then eq (64) yields

$$\widetilde{\omega}(z) = -z^n \frac{z^{n+1} + \left(\frac{n}{2} - 1\right)z - \frac{n}{2}}{1 + \left(\frac{n}{2} - 1\right)z^n - \frac{n}{2}z^{n+1}}.$$

It suffices to show that for some point $z_0 \in \mathbb{D}$, $|\widetilde{\omega}(z_0)| > 1$. Let $z_0 = -\frac{n}{n+1} \in \mathbb{D}$. Then

$$\widetilde{\omega}(z_0) = \left(\frac{n}{n+1}\right)^n \frac{\left(\frac{n}{n+1}\right)^{n+1} - \left(\frac{n}{2} - 1\right)\left(\frac{n}{n+1}\right) - \frac{n}{2}}{1 - \left(\frac{n}{2} - 1\right)\left(\frac{n}{n+1}\right)^n - \left(\frac{n}{2}\right)\left(\frac{n}{n+1}\right)^{n+1}}$$

(66)

$$= 1 + \frac{\left[\left(\frac{n+1}{n}\right)^n - \left(\frac{n}{n+1}\right)^{n+1}\right] + \left[1 - \frac{n}{n+1}\right]}{\left(\frac{n}{2} - 1\right) + \left(\frac{n}{2}\right)\left(\frac{n}{n+1}\right) + \left(\frac{n+1}{n}\right)^n}.$$

Note that $\left[\left(\frac{n+1}{n}\right)^n - \left(\frac{n}{n+1}\right)^{n+1}\right] + \left[1 - \frac{n}{n+1}\right] > 0$. Also, $\left(\frac{n}{2} - 1\right) + \left(\frac{n}{2}\right)\left(\frac{n}{n+1}\right) + \left(\frac{n+1}{n}\right)^n > 0$ since $\left(\frac{n}{2} - 1\right) + \left(\frac{n}{2}\right)\left(\frac{n}{n+1}\right) > n - \frac{3}{2} > e$ and $\left(\frac{n+1}{n}\right)^n$ is an increasing series converging to e. Thus, if $n \geq 5$ is odd, $|\widetilde{\omega}(z_0)| > 1$. If $n = 3$, it is easy to compute that $|\widetilde{\omega}(z_0)| = \left(\frac{3}{4}\right)^3 \left|\frac{3^4 - \frac{1}{2} \cdot 3 \cdot 4^3 - \frac{3}{2} \cdot 4^4}{4^4 - \frac{1}{2} \cdot 3^3 \cdot 4 - \frac{3}{2} \cdot 3^4}\right| > 2$. Now, if n is even, let $\omega(z) = z^n$ and $z_0 = -\frac{n}{n+1}$. This simplifies to the same $\widetilde{\omega}(z_0)$ given eq (66) and the argument above also holds for $n \geq 6$. If $n = 4$, $|\widetilde{\omega}(z_0)| = \left(\frac{4}{5}\right)^4 \left|\frac{4^5 - 4 \cdot 5^4 - 2 \cdot 5^5}{5^5 - 4^4 \cdot 5 - 2 \cdot 4^5}\right| > 15$.

EXPLORATION 4.132. Using *ComplexPlot*, graph $\widetilde{\omega}(\mathbb{D})$ given in eq (64) for $\omega(z) = -z^n$, where $n = 1, 2, 3, 4$. Explain how these images support Theorem 4.129 and Remark 4.131.
   Try it out!

THEOREM 4.133. Let $f = h + \overline{g} \in K_H^O$ with $h(z) + g(z) = \frac{z}{1-z}$ and $\omega(z) = \frac{z+a}{1+az}$ with $a \in (-1, 1)$. Then $f_0 * f \in S_H^O$ and is convex in the direction of the real axis.

311

PROOF. Using $\omega(z) = \frac{z+a}{1+az}$, where $-1 < a < 1$, we have

$$\widetilde{\omega}(z) = -z\frac{\left(z^2 + \frac{1+3a}{2}z + \frac{1+a}{2}\right)}{\left(1 + \frac{1+3a}{2}z + \frac{1+a}{2}z^2\right)}$$

$$= -z\frac{f(z)}{f^*(z)}$$

$$= -z\frac{(z+A)(z+B)}{(1+\ \overline{A}z)(1+\ \overline{B}z)}$$

$$= -z\frac{p(z)}{q(z)}.$$

Again using Method 1,

$$p_1(z) = \frac{\overline{a_2}p(z) - a_0 p^*(z)}{z} = \frac{(a+3)(1-a)}{4}z + \frac{(1+3a)(1-a)}{4}.$$

So p_1 has one zero at $z_0 = -\frac{1+3a}{a+3}$ which is in the unit circle since $-1 < a < 1$. Thus, $|A|, |B| < 1$. $\qquad\square$

LARGE PROJECT 4.134. In Theorem 4.126, we require that the resulting convolution function satisfy the dilatation condition

$$|\omega(z)| = \left|\frac{g'(z)}{h'(z)}\right| < 1, \forall z \in \mathbb{D}.$$

Determine various ω functions for which the dilatation condition holds and ones for which it does not hold. See Theorem 4.129, Theorem 4.133, and Remark 4.131 for examples.

Optional

LARGE PROJECT 4.135. Similar to the right half-plane map, $f = h + \ \overline{g}$ is an asymmetric vertical strip map if $h(z) + g(z) = \frac{1}{2i\sin\alpha}\log\left(\frac{1+ze^{i\alpha}}{1+ze^{-i\alpha}}\right)$, where $0 < \alpha < \pi$. Theorem 4.126 can be stated in terms of asymmetric vertical strip mappings instead of right half-plane mappings.

> THEOREM: Let $f = h + \overline{g} \in K_H^O$ with $\omega = g'/h'$ be such that $h + g = \frac{1}{2i\sin\alpha}\log\left(\frac{1+ze^{i\alpha}}{1+ze^{-i\alpha}}\right)$, where $0 < \alpha < \pi$.
> Then $f_0 * f \in S_H^O$ and is convex in the direction of the real axis.

Determine various ω functions for which the dilatation condition holds for this theorem and ones for which it does not hold.

Optional

4.9. Conclusion

We have presented an introduction to harmonic univalent mappings and described a few topics to entice a beginner. Our emphasis is on the geometric aspects of harmonic univalent mappings that students can explore using the exercises, the exploratory problems, and the projects along with the applets. There are more interesting and deeper topics in harmonic univalent mappings. Here is a short list along with some resources: (a) coefficient estimates and conjectures ([**5**], [**9**], [**23**]); (b) a generalized Riemann Mapping Theorem ([**9**], [**14**]); (c) properties of special classes of functions such as convex, close-to-convex, starlike, and typically real ([**5**], [**9**], [**23**]); (d) harmonic polynomials ([**4**], [**24**], [**27**]); (e) extremal problems ([**9**]); (f) harmonic meromorphic functions ([**16**], [**25**]); (g) inner mapping radius ([**2**], [**8**]); and (h) multiply connected domains ([**9**], [**10**]). Another topic is the connection between harmonic mappings and minimal surfaces. This topic is discussed in the chapter on minimal surfaces in this book. In addition, there are several nice general resources to learn more about harmonic univalent functions. These include Peter Duren's book [**9**], Clunie and Sheil-Small's original article [**5**], Bshouty and Hengartner's article [**3**], and Schober's article [**22**]. Finally, Bshouty and Hengartner complied a list of open problems and conjectures [**2**].

4.10. Additional Exercises

Anamorphosis and Möbius Maps

EXERCISE 4.136. Determine the explicit image of horizontal lines $\operatorname{Im} z = k$ under the inversion map $f(z) = \frac{1}{z}$ and analytically prove your result.

EXERCISE 4.137. Analytically determine the image of the left half-plane $\{z \mid \operatorname{Re}\{z\} < 0\}$ under the map

$$M(z) = \frac{z - i}{z}.$$

Use *ComplexTool* to check your answer.

EXERCISE 4.138. The strip $D = \{z \mid 0 < \operatorname{Im}\{z\} < 1\}$ can be thought of as the intersection of the two domains $D_1 = \{z \mid 0 < \operatorname{Im}\{z\}\}$ and $D_2 = \{z \mid \operatorname{Im}\{z\} < 1\}$. Analytically determine the image of the strip D under the Möbius transformation $M(z) = \frac{z-i}{z+i}$.

EXERCISE 4.139. Determine a Möbius transformation the maps the left half-plane $\{z \mid \operatorname{Re}\{z\} < 0\}$ onto the disk $\{z \mid |z - 1| < 1\}$. Use *ComplexTool* to check your answer.

Try it out!

EXERCISE 4.140. Prove that if z_1, z_2, z_3 are distinct points and w_1, w_2, w_3 are distinct points, then the Möbius transformation T satisfying $T(z_1) = w_1$, $T(z_2) = w_2$, $T(z_3) = w_3$ is unique.

EXERCISE 4.141. Let $M(z) = \frac{az+b}{cz+d}$ with $ad \neq bc$ be a Möbius transformation. Prove that M is a rotation if and only if $M(0) = 0$ and M preserves distances (Note: preserving distances means $|z_1 - z_2| = |M(z_1) - M(z_2)|$).

The Family S of Analytic, Normalized, Univalent Functions

EXPLORATION 4.142.
  (a) Using *ComplexTool* guess the smallest $k > 0$ such that $(z + k)^2$ is univalent in $\mathbb{D}$.
  (b) Prove your guess from (a).
  (c) Using *ComplexTool* guess the smallest $k > 0$ such that $(z + k)^3$ is univalent in $\mathbb{D}$.
  (d) Prove your guess from (c).

EXERCISE 4.143. Show that $f(z) = z + a_3 z^3$ is univalent in $\mathbb{D} \iff |a_3| \leq \frac{1}{3}$. Determine $f(\mathbb{D})$ analytically when $a_3 = -\frac{1}{3}$.

EXERCISE 4.144. Work out the details to show that $\frac{z}{(1-z)^2} = \sum_{n=1}^{\infty} n z^n = z + 2z^2 + 3z^3 + \cdots$

EXERCISE 4.145. Determine $f(\mathbb{D})$ analytically $f(z) = \dfrac{z - cz^2}{(1 - z)^2}$, where $0 < c < 1$.

EXERCISE 4.146. Prove that $f(z) = \dfrac{1}{2} \log \left(\dfrac{1+z}{1-z} \right)$ is univalent. Determine $f(\mathbb{D})$ analytically.

EXPLORATION 4.147. Consider the function
$$f_c(z) = \frac{1}{2c} \left[\left(\frac{1+z}{1-z} \right)^c - 1 \right].$$

(a) Show that if $c = 2$, then $f_c(z)$ is the Koebe function.
(b) Show that if $c = 1$, then $f_c(z)$ is the right half-plane mapping.
(c) Use *ComplexTool* to view the image of $\mathbb{D}$ under f_c for various values $0 < c < 2$. For what values of c does f_c appear to be univalent.

EXERCISE 4.148. Find the image of $\mathbb{D}$ analytically under the univalent function $f(z) = \dfrac{z}{1 - z^2}$.

The Family S_H of Normalized, Harmonic, Univalent Functions

EXERCISE 4.149. Determine if $f(x, y) = u(x, y) + iv(x, y) = (x^3 + xy^2) + i(x^2 y + y^3)$ is complex-valued harmonic.

EXERCISE 4.150. Prove that $f(x, y) = u(x, y) + iv(x, y)$ is harmonic $\Longleftrightarrow \dfrac{\partial^2 f}{\partial z \partial \bar{z}} = 0$.

EXERCISE 4.151. Rewrite $f(x, y) = u(x, y) + iv(x, y) = (x - \frac{1}{2}x^2 + \frac{1}{2}y^2) + i(y - xy)$ in terms of z and $\bar{z}$ and then determine if f is analytic.

EXERCISE 4.152. Prove that for all functions $f \in S_H^O$, the sharp inequality $|b_2| \leq \frac{1}{2}$ holds.

EXERCISE 4.153. Verify that the image of $\mathbb{D}$ under the harmonic function $f(z) = z + \frac{1}{2}z^2$ is a hypocycloid with 3 cusps.

EXERCISE 4.154. If a domain is convex in the direction $e^{i\varphi}$ for every value of $\varphi \in [0, \pi)$, then the domain is called a *convex* domain. For example, a disk is a convex domain. For which values of $n = 1, 2, 3, \ldots$ are the following functions that map $\mathbb{D}$ onto a convex domain?

(a) $f(z) = z^n$,
(b) $f(z) = z - \frac{1}{n}z^n$ (see Example 4.19 and Definition 4.20),
(c) $f(z) = \dfrac{z}{(1-z)^n}$ (see Examples 4.22 and 4.23 to get you started).

The Shearing Technique

EXPLORATION 4.155. Let $f = h + \bar{g}$ with $h(z) - g(z) = z - \frac{1}{n}z^n$ and $\omega(z) = z^{n-1}$. Use *ShearTool* to sketch the graph of $f(\mathbb{D})$ for different values of n and then compute h and g explicitly so that $f \in S_H^O$.

EXERCISE 4.156. Let $f = h + \bar{g}$ with $h(z) - g(z) = \frac{z}{(1-z)^2}$ and $\omega(z) = -z$. Compute h and g explicitly so that $f \in S_H^O$ and determine $f(\mathbb{D})$.

EXERCISE 4.157. Let $f = h + \bar{g}$ with $h(z) - g(z) = \dfrac{z}{(1-z)^2}$ and $\omega(z) = z \dfrac{z + \frac{1}{2}z}{1 + \frac{1}{2}z}$.

(a) Show that $|\omega(z)| < 1, \forall z \in \mathbb{D}$.
(b) Compute h and g explicitly so that $f \in S_H^O$.
(c) Show that $f(\mathbb{D})$ is a slit domain like the Koebe domain. Determine where the tip of the slit is located.
(d) What is the significance of this example in relationship to the Riemann Mapping Theorem?

EXERCISE 4.158. Let $f = h + \bar{g}$ with $h(z) + g(z) = \dfrac{z}{1 - z}$ and $\omega(z) = e^{i\theta}z$, where $\theta \in [0, 2\pi)$. Use *ShearTool* to sketch the graph of $f(\mathbb{D})$ for different values of n and then compute h and g explicitly so that $f \in S_H^O$.

EXPLORATION 4.159. We can find harmonic functions, $f_n = h_n + \bar{g}_n$, that map onto regular n-gons by generalizing the ideas from Example 4.72. Use *ShearTool* to explore the images of $\mathbb{D}$ under $f = h + \bar{g}$, where f comes from shearing

$$h_n(z) - g_n(z) = \sum_{k=0}^{n-1} \frac{-2\cos\left(\frac{2\pi k}{n}\right)}{n} \log\left(1 - ze^{i\frac{2\pi k}{n}}\right)$$

with $\omega(z) = -z^{n-2}$.

EXPLORATION 4.160. Let $f = h + \bar{g}$ with $h(z) - g(z) = \dfrac{1}{2}\log\left(\dfrac{1+z}{1-z}\right)$ and $\omega(z) = m^2 z^2$, where $m = e^{i\theta}(0 \le \theta \le \frac{\pi}{2})$. Use *ShearTool* to sketch the graph of $f(\mathbb{D})$ for different values of n and then compute h and g explicitly so that $f \in S_H^O$ [Preview: This Exploration fits nicely with minimal surfaces, because when $m = 1$ f lifts to a canonical minimal surface, Scherk's doubly-periodic, and when $m = i$ f lifts to a different canonical minimal surface, the helicoid].

Dilatations

EXPLORATION 4.161. Shear $h(z) - g(z) = \frac{z}{1-z}$ using $\omega(z) = az$, where $-1 \le a \le 1$ and sketch $f(\mathbb{D})$ using *ShearTool*. Describe what happens to $f(\mathbb{D})$ as a varies.

EXPLORATION 4.162. Shear $h(z) - g(z) = \frac{z}{1-z}$ using $\omega(z) = z^n$, where $n = 1, 2, 3, 4, 5$ and sketch $f(\mathbb{D})$ using *ShearTool*. Describe what happens to $f(\mathbb{D})$ as a varies.

EXPLORATION 4.163. Shear $h(z) - g(z) = \log\left(\frac{1-z}{1+z}\right)$ using $\omega(z) = e^{i\pi n/6}z$, where $n = 0, \ldots, 6$ and sketch $f(\mathbb{D})$ using *ShearTool*. Describe what happens to $f(\mathbb{D})$ as n varies.

EXPLORATION 4.164. Shear $h(z) - g(z) = \frac{z}{1-z} + ae^{\frac{z+1}{z-1}}$ using $\omega(z) = e^{\frac{z+1}{z-1}}$, where $-0.5 \le a \le 0.5$ and sketch $f(\mathbb{D})$ using *ShearTool*. Describe what happens to $f(\mathbb{D})$ as a varies.

EXERCISE 4.165. Let $h_\alpha(z) = \frac{z}{1+ze^{-i\alpha}}$, where $0 < \alpha < \pi$, and $\omega_\alpha(z) = e^{-i\left(\frac{z+e^{-i\alpha}}{1+ze^{-i\alpha}}\right)}$. Compute $f_\alpha = h_\alpha + \overline{g}_\alpha$ and show that $f_\alpha \in S_H^O$. Use *ComplexTool* to sketch $f_\alpha(\mathbb{D})$ for various values of α [Note: as α approaches 0, you should get the image shown in Figure 4.26].

EXERCISE 4.166. Let $h_\gamma(z) = \frac{1}{2i\sin\gamma} \log\left(\frac{1+ze^{i\gamma}}{1+ze^{-i\gamma}}\right)$, where $\frac{\pi}{2} \le \gamma < \pi$, and $\omega_\gamma(z) = e^{-\left(\frac{2\sin(\pi-\gamma)}{(\pi-\gamma)}h(z)-1\right)}$. Compute $f_\gamma = h_\gamma + \overline{g}_\gamma$ and show that $f_\gamma \in S_H^O$. Use *ComplexTool* to sketch $f_\gamma(\mathbb{D})$ for various values of γ [Note: as γ approaches π, you should get the image shown in Figure 4.26, but each $f_\gamma(\mathbb{D})$ is different than any $f_\alpha(\mathbb{D})$ in the Exercise 4.165].

Harmonic Linear Combinations

EXERCISE 4.167. Let

$$f_1(z) = \operatorname{Re}\left\{ -z - 2\log(1-z)\right\} + i\operatorname{Im}\left\{z\right\},$$
$$f_2(z) = \operatorname{Re}\left\{\frac{z + 1/3z^3}{(1-z)^3}\right\} + i\operatorname{Im}\left\{\frac{z}{(1-z)^2}\right\}.$$

(a) Show that f_1 can be derived by shearing $h(z) - g(z) = z$ with $g'(z) = zh'(z)$.
(b) Use *ComplexTool* to plot the image of $\mathbb{D}$ under f_1. Recall that f_2 is the "harmonic Koebe" function. What is the image of $\mathbb{D}$ under f_2?
(c) Use *ComplexTool* to see that $f_3 = tf_1 + (1-t)f_2$ is not univalent for at least one value of $t, (0 \le t \le 1)$. Why does this not contradict Theorem 4.100?

EXPLORATION 4.168. Let

$$f_1(z) = \operatorname{Re}\left\{\frac{z}{(1-z)^2}\right\} + i\operatorname{Im}\left\{\frac{1}{2}\log\left(\frac{1+z}{1-z}\right)\right\},$$
$$f_2(z) = \operatorname{Re}\left\{\frac{i}{2}\log\left(\frac{1-iz}{1+iz}\right)\right\} + i\operatorname{Im}\left\{\frac{1}{2}\log\left(\frac{1+z}{1-z}\right)\right\}.$$

Show that f_1 and f_2 satisfies the conditions of Theorem 4.100 and then use *ComplexTool* to plot images of $f_3 = tf_1 + (1-t)f_2$ for various values of t.

EXPLORATION 4.169. Let

$$f_1(z) = \operatorname{Re}\left\{\frac{z}{1-z} - \frac{1}{2}e^{\frac{z+1}{z-1}}\right\} + i\operatorname{Im}\left\{\frac{z}{1-z} + \frac{1}{2}e^{\frac{z+1}{z-1}}\right\},$$
$$f_2(z) = \operatorname{Re}\left\{z - \frac{1}{4}z^2 - \frac{1}{4}(z-1)^2 e^{\frac{z+1}{z-1}}\right\} + i\operatorname{Im}\left\{z - \frac{1}{4}z^2 + \frac{1}{4}(z-1)^2 e^{\frac{z+1}{z-1}}\right\}.$$

Show that f_1 and f_2 satisfies the conditions of Theorem 4.100 and then use *ComplexTool* to plot images of $f_3 = tf_1 + (1-t)f_2$ for various values of t.

EXERCISE 4.170. Repeat the steps in Example 4.104 using the same function for f_1 but replacing f_2 with the harmonic square map in Example 4.72, where

$$h_2(z) = \frac{1}{4}\log\left(\frac{1+z}{1-z}\right) + \frac{i}{4}\log\left(\frac{i+z}{i-z}\right)$$

$$g_2(z) = -\frac{1}{4}\log\left(\frac{1+z}{1-z}\right) + \frac{i}{4}\log\left(\frac{i+z}{i-z}\right).$$

(a.) In this case, $\omega_1(z) = z^2$ while $\omega_2(z) = -z^2$. Using eq. (60) in the remark above, show that f_3 is locally univalent.

(b.) Use *LinComboTool* find the image of $f_3(\mathbb{D})$ using this f_1 and f_2.

(c.) Explain why this happens by using the approach in Example 4.104 to compute the new values of $w_1, \ldots, w_8$ and then use the visualization technique in the example to plot the eight vertices $z_0, \ldots, z_7$ and draw the midpoints $w_1, \ldots, w_8$.

EXPLORATION 4.171. Using *LinComboTool*, start with the same arc values and corresponding point values as in Example 4.104. In **Panel #1** increase the arc values by increments of $\frac{\pi}{16}$ while not changing the point values, and decrease the arc values in **Panel #2** by the same amount. Note that you can do this either by changing the specific value in the **Arc** n box or by just using the cursor to move the four blue dots the same amount in the same direction in **Panel #1** and the same amount in the opposite direction in **Panel #2** on the unit circle of the domain in each panel. Describe how the image domain changes as the arc values in **Panel #1** increase by a total of $\frac{\pi}{4}$ and decrease in **Panel #2** by the same amount.

EXPLORATION 4.172. Using *LinComboTool*, create a map in **Panel #1** that maps 3 arcs on the unit circle to 3 vertices of an equilateral triangle. Then create a second map in **Panel #2** that maps 3 different arcs on the unit circle to 3 vertices of a rotated equilateral triangle. Make sure that these maps satisfy the conditions of Theorem 4.100. Click on the **Create LinCombogon** button to see the resulting image domain. Explore this idea by using different maps in the panels to get at least three different resulting image domains.

EXPLORATION 4.173. Using *LinComboTool*, create a map in **Panel #1** that maps 6 arcs on the unit circle to 6 vertices of a regular hexagon. Then create a second map in **Panel #2** that maps 6 different arcs on the unit circle to 6 vertices of a rotated regular hexagon. Make sure that these maps satisfy the conditions of Theorem 4.100. Click on the **Create LinCombogon** button to see the resulting image domain. Explore this idea by using different maps in the panels to get at least three different resulting image domains.

Convolutions

EXERCISE 4.174. Let

$$f(z) = \int \frac{1}{1 - z^2} \, dz = \frac{1}{2} \log \left(\frac{1+z}{1-z} \right)$$

and

$$F(z) = \int \frac{1}{1 - z^3} \, dz = \frac{1}{3} e^{\frac{i5\pi}{3}} \log \left(1 + e^{\frac{i\pi}{3}} z \right) + \frac{1}{3} e^{\frac{i5\pi}{3}} \log \left(1 + e^{\frac{i5\pi}{3}} z \right) - \frac{1}{3} \log \left(1 - z \right).$$

Using eq (58) at the end of Example 4.72, determine $f * F$ and the image of $\mathbb{D}$ under this convolution. In general, what is $f * F$ when $f'(z) = \frac{1}{1-z^m}$ and $F(z) = \frac{1}{1-z^n}$?

EXERCISE 4.175. In Theorem 4.121, let f be the canonical right half-plane mapping $f_0 \in K_H$ and let $\varphi(z) = \frac{z}{(1-z)^2} \in S$. Compute $F = f_0 * (\overline{\varphi} + \varphi)$ and use *ComplexTool* to sketch $F(\mathbb{D})$.

EXERCISE 4.176. Derive the expressions for H_2 and G_2 given in eq (65).

SMALL PROJECT 4.177. Compute $f_a = h_a + \overline{g}_a$, where $h_a(z) + g_a(z) = \frac{z}{1-z}$ and $\omega(z) = \frac{z+a}{1+az}$. From Theorem 4.133, we know that $F_a = f_0 * f_a \in S_H$ for $-1 < a < 1$. Compute F_a and use *ComplexTool* to sketch $F_a(\mathbb{D})$ for various values of $a(-1 < a < 1)$. Describe what happens as a varies between 1 and -1.

Bibliography

[1] L. Ahlfors, *Conformal Invariants: Topics in Geometric Function Theory*, McGraw-Hill, Inc., New York, 1973.

[2] D. Bshouty and W. Hengartner (editors), *Problems and conjectures for harmonic mappings*, from a workshop held at the Technion, Haifa, 1995.

[3] D. Bshouty and W. Hengartner, Univalent harmonic mappings in the plane, *Handbook of complex analysis: geometric function theory*, Vol. 2, 479-506, Elsevier, Amsterdam, 2005.

[4] D. Bshouty and A. Lyzzaik, On Crofoot-Sarason's conjecture for harmonic polynomials, *Comput. Methods Funct. Theory* **4** (2004), no. 1, 35-41.

[5] J. Clunie and T. Sheil-Small, Harmonic univalent functions, *Ann. Acad. Sci. Fenn. Ser. A.I Math.* **9** (1984), 3-25.

[6] M. Dorff, Convolutions of planar harmonic convex mappings, *Complex Variables Theory Appl.*, **45** (2001), no. 3, 263-271.

[7] M. Dorff, Harmonic mappings onto asymmetric vertical strips, in *Computational Methods and Function Theory 1997*, (N. Papamichael, St. Ruscheweyh and E. B. Saff, eds.), 171-175, World Sci. Publishing, River Edge, NJ, 1999.

[8] M. Dorff and T. Suffridge, The inner mapping radius of harmonic mappings of the unit disk, *Complex Variables Theory Appl.* **33** (1997), no. 1-4, 97-103.

[9] P. Duren, *Harmonic mappings in the plane*, Cambridge Tracts in Mathematics, 156, Cambridge University Press, Cambridge, 2004.

[10] P. Duren and W. Hengartner, Harmonic mappings of multiply connected domains, *Pacific J. Math.* **180** (1997), no. 2, 201-220.

[11] P. Duren, J. McDougall, and L. Schaubroeck, Harmonic mappings onto stars, *J. Math. Anal. Appl.* **307** (2005), no. 1, 312-320.

[12] M. Goodloe, Hadamard products of convex harmonic mappings, *Complex Var. Theory Appl.* **47** (2002), no. 2, 81-92.

[13] P. Greiner, Geometric properties of harmonic shears, *Comput. Methods Funct. Theory*, **4** (2004), no. 1, 77-96.

[14] W. Hengartner and G. Schober, Harmonic mappings with given dilatation, *J. London Math. Soc.* (2) **33** (1986), no. 3, 473483.

[15] W. Hengartner and G. Schober, On schlicht mappings to domains convex in one direction, *Comment. Math. Helv.* **45** (1970), 303-314.

[16] W. Hengartner and G. Schober, Univalent harmonic functions, *Trans. Amer. Math. Soc.* **299** (1987), no. 1, 131.

[17] R. Laugesen, Planar harmonic maps with inner and Blaschke dilatations, *J. London Math. Soc.* (2) **56** (1997), 37-48.

[18] C. Pommenrenke, On starlike and close-to-convex functions, *Proc. London Math. Soc. (3)* **13** (1963), 290-304.

[19] Q. I. Rahman and G. Schmeisser, *Analytic Theory of Polynomials*, London Mathematical Society Monographs New Series, 26, Oxford University Press, Oxford, 2002.

[20] St. Ruscheweyh and L. Salinas, On the preservation of direction-convexity and the Goodman-Saff conjecture, *Ann. Acad. Sci. Fenn., Ser. A. I. Math.* **14** (1989), 63-73.

[21] St. Ruscheweyh and T. Sheil-Small, Hadamard products of schlicht functions and the Polya-Schoenberg conjecture, *Comment. Math. Helv.*, **48** (1973), 119-135.

[22] G. Schober, Planar harmonic mappings, *Computational methods and function theory* (Valparaso, 1989), 171–176, Lecture Notes in Math., **1435**, Springer, Berlin, 1990.

[23] T. Sheil-Small, Constants for planar harmonic mappings, *J. London Math. Soc.* (2) **42** (1990), no. 2, 237-248.

[24] T. Suffridge, Harmonic univalent polynomials, *Complex Variables Theory Appl.*, **35** (1998), no. 2, 93-107.

[25] J. Thompson, A family of meromorphic harmonic mappings, *Complex Var. Theory Appl.* **48** (2003), no. 8, 627-648.

[26] A. Weitsman, Harmonic mappings whose dilatations are singular inner functions, unpublished manuscript.

[27] A.S. Wilmshurst, The valence of harmonic polynomials, *Proc. Amer. Math. Soc.* **126** (1998), no. 7, 2077-2081.

CHAPTER 5

Mappings to Polygonal Domains

Jane McDougall and Lisbeth Schaubroeck (text), Jim Rolf (applets)

5.1. Introduction

A rich source of problems in analysis is determining when, and how, one can create a one-to-one function of a particular type from one region onto another. In this chapter, we consider the problem of mapping the unit disk $\mathbb{D}$ onto a polygonal domain by two different classes of functions. First for analytic functions we give an overview and examples of the well known Schwarz-Christoffel transformation. We then diverge from analytic function theory and consider the Poisson integral formula to find harmonic functions that will serve as mapping functions onto polygonal domains. Proving that these harmonic functions are univalent requires us to explore some less known theory of harmonic functions and some relatively new techniques.

Because of the Riemann Mapping Theorem, we can simplify our mapping problem for *either* class of function to asking when we can map the unit disk $\mathbb{D} = \{z : |z| < 1\}$ univalently onto a target region. This is because if we want to map one domain (other than the entire set of complex numbers) onto another, we can first map it to $\mathbb{D}$ by an analytic function, and subsequently apply an analytic or harmonic mapping from $\mathbb{D}$ to the other domain (recall that the composition of a harmonic function with an analytic function is harmonic).

We begin in Section 1.2 with the Schwarz-Christoffel formula to find univalent analytic maps onto polygonal domains, and so set the stage for the corresponding problem for harmonic functions with the Poisson integral formula in Section 1.3. Perhaps because of their importance in applications, many first books on complex analysis introduce Schwarz-Christoffel mappings through examples, without emphasis on subtleties of the deeper theory. Our approach here will be the same, and the examples we include are chosen to bring together ideas found elsewhere in this book, such as the shearing technique from Chapter 4 and the construction of minimal surfaces from Chapter 5.

We also include an example of a Schwarz-Christoffel map onto a regular star, a polygonal domain that is highly symmetric but also non-convex. The problem of using the Poisson integral formula to construct a univalent harmonic function onto a non-convex domain is not at all well understood. In Section 1.3, after developing the theory

for convex domains, we explore the example of finding harmonic maps onto regular star domains in detail, and lead the student to further investigation.

Terminology and technology: We use the term "univalent" for one-to-one, and take domain to mean open connected set in the complex plane. The applets used in this chapter are:

(1) *ComplexTool* - used to plot the image of domains in $\mathbb{C}$ under complex-valued functions.
(2) *PolyTool* - used to visualize the harmonic function that is the extension of a particular kind of boundary correspondence. The user of this applet can dynamically change the boundary correspondence and watch the harmonic function change.
(3) *StarTool* - used to examine the functions that map the unit disk $\mathbb{D}$ onto an n-pointed star. The user can modify the shape of the star (by changing n and r) and the boundary correspondence (by changing p).

5.2. Schwarz-Christoffel Maps

In this section we consider conformal maps from the unit disk and the upper half-plane onto various simply connected polygonal domains. By the Riemann Mapping Theorem, we can map the unit disk conformally onto any simply connected domain that is a proper subset of the complex numbers, with a mapping function that is essentially unique.

While the Riemann Mapping Theorem tells us that we **can** find a univalent analytic function to map $\mathbb{D}$ onto our domain, finding an actual Riemann mapping function is no easy task. Even for a simple domain such as a square, the mapping function from the disk cannot be expressed in terms of elementary functions. One situation however in which this problem *is* relatively simple is in mapping a region bounded by a circle or line in the complex plane to another such region, using fractional linear transformations. For this problem, we only need to pick three points on the bounding line or circle in the domain, and map them (in order) to three arbitrarily chosen points on the boundary of the target region (see for example [**10**], [**15**], [**19**], or [**20**]). This selection of three pairs of points determines the fractional linear transformation completely, and works for example, in finding a conformal map from $\mathbb{D}$ onto any planar region bounded by a line or circle.

EXERCISE 5.1. Show that the fractional linear transformation $z' = \phi(z) = i\frac{1+z}{1-z}$ maps the unit disk to the upper half-plane by finding the images of three boundary points. Then show that its inverse function $\phi^{-1}(z') = z = \frac{z'-i}{z'+i}$ maps the upper half-plane to the unit disk. ***Try it out!***

How can a mapping function be found in the case when the target region is more complicated? This question is relevant to solving the heat equation or the study of fluid flows, as discussed in Chapter 3.

The Schwarz-Christoffel transformation frequently enables us to find a function mapping onto a polygonal domain. In most texts the formula is presented as a mapping from the upper half-plane $\mathbb{H}$ onto the target polygon. We now develop this formula - for a more thorough treatment we refer the reader to [**13**], [**3**] and [**16**]. Suppose our target polygon has interior angles $\alpha_k\pi$ and exterior angles $\beta_k\pi$, where $\alpha_k + \beta_k = 1$ and $\alpha_k > 0$. The exterior angle measures the angle through which a bug, traversing the polygon in the counterclockwise direction, would turn at each vertex. This angle could be positive or negative, following the usual convention in mathematics that a counter-clockwise rotation is positive while a clockwise rotation is negative. For example, in Figure 5.1 the angle marked by $\alpha_2\pi$ is $\frac{3\pi}{2}$, so $\alpha_2 = 3/2$. We can see from $\alpha_2 + \beta_2 = 1$ that $\beta_2 = -1/2$, which coincides with the description of $\beta_2\pi$ as a clockwise turn on the boundary. We can also obtain the exterior angle by extending one side of the polygon, and then seeing through what angle you would rotate that side to get to the next side of the polygon, as shown in the figure. For a simple closed polygon (that is, one with no self-intersections), it is always possible to describe the interior and exterior angles using coefficients α_k and β_k as described. As a final note about terminology, we describe a vertex such as the one described with β_1 in Figure 5.1 as a **convex corner** and the one described with β_2 as a **non-convex** corner.
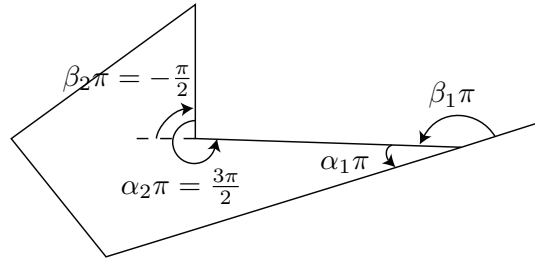
FIGURE 5.1. A sample polygon with both a convex and a non-convex corner

The Schwarz-Christoffel formula for the half-plane $\mathbb{H}$ to the polygon with exterior angles described by coefficients β_k as above is

$$(67) \qquad f(z) = A_1 \int_0^z \frac{1}{(w - x_1)^{\beta_1}(w - x_2)^{\beta_2}\cdots(w - x_n)^{\beta_n}}\,dw + A_2, \quad z \in \mathbb{H}.$$

The real values x_i are preimages of the n vertices of the polygon, which we will refer to from now on as **prevertices**. Different choices of the constants A_1 and A_2 rotate, scale and/or translate the target n-gon.

In Equation 67, we use w as the variable of integration, and the limits of integration are chosen to make the definite integral into a function of z. The (abitrary) choice of 0 as a fixed point might have to be altered if it corresponds to a point of discontinuity of the integrand.

EXERCISE 5.2. You may be familiar with the sine and arcsine functions on the complex plane. Verify that the Schwarz-Christoffel mapping of $\mathbb{H}$ onto the infinite half strip described by $|\operatorname{Re}(z)| < \frac{\pi}{2}$ and $\operatorname{Im}(z) > 0$ is given by the arcsine function. Use the prevertices $x_1 = -1$, $x_2 = 1$ in formula 67. ***Try it out!***

We can observe that the angles at the vertices are represented in the formula, but nowhere do we see an accomodation for the side-lengths of the target polygon. In fact the side-lengths are influenced by the choice of prevertices x_i, but in a nonlinear and non-obvious way.

In the following example, we will apply the Schwarz-Christoffel formula to map the upper half-plane onto a rectangle. We will make a somewhat arbitrary choice of prevertices, and then evaluate the resulting integral to determine the target rectangle. In computing the integral, we come across a first example of a special function, and find that we need to learn some of the basics of elliptic integrals. We will find that just as the prevertices were arbitrarily chosen, so are the sidelengths of our target rectangle.

EXAMPLE 5.3. In this example, we map the upper half-plane $\mathbb{H}$ onto a rectangle. We will choose the prevertices $x_1 = -3, x_2 = -1, x_3 = 1$, and $x_4 = 3$. Since our target image is a rectangle, all of the exterior angles are $\pi/2$, so each $\beta_i = 1/2$. Using equation 67, we find that

$$f(z) = A_1 \int_0^z \frac{1}{(w-1)^{1/2}(w-3)^{1/2}(w+1)^{1/2}(w+3)^{1/2}}dw + A_2, \quad z \in \mathbb{H}.$$

The constant A_1 allows us to scale and rotate the image of $\mathbb{H}$, and A_2 allows for a translation. By chosing $A_1 = 1$ and $A_2 = 0$ we simplify to

$$(68) \qquad\qquad f(z) = \int_0^z \frac{1}{\sqrt{(w^2-1)(w^2-9)}}dw.$$

This choice of constants does not affect the **aspect ratio** (ratio of adjacent sides) of the resulting rectangle. However the integral cannot be evaluated using techniques in standard calculus texts. Instead it is a special function known as an elliptic integral (of the first kind, with parameter $k = \frac{1}{3}$).

DEFINITION 5.4. An elliptic integral of the first kind is an integral of the form

$$F(\phi, k) = \int_0^{\sin\phi} \frac{1}{\sqrt{(1-w^2)(1-k^2w^2)}}dw.$$

An alternate form is $F(\phi, k) = \int_0^\phi \frac{1}{\sqrt{1-k^2\sin^2\theta}}d\theta$.

The two integrals in Definition 5.4 are identical after the change of variables $w = \sin\theta$, $dw = \cos\theta d\theta = \sqrt{1-w^2}d\theta$ which connects them. (Technology note: The computer algebra system *Mathematica* uses the alternate form, representing the integral by EllipticF$[\phi, m]$, where $m = k^2$.)

EXERCISE 5.5. Carry out the change of variables $w = \sin\theta$, $dw = \cos\theta d\theta = \sqrt{1-w^2}d\theta$ to show that

$$F(\phi, k) = \int_0^{\sin\phi} \frac{1}{\sqrt{(1-w^2)(1-k^2w^2)}}dw = \int_0^\phi \frac{1}{\sqrt{1-k^2\sin^2\theta}}d\theta.$$

Try it out!

Returning to our integral from Equation 68, we now work on rewriting it so that we may recognize it as an elliptic integral of the first kind.

$$
\begin{aligned}
f(z) &= \int_0^z \frac{1}{\sqrt{(w^2-1)(w^2-9)}}dw \\
&= \int_0^z \frac{1}{\sqrt{(w^2-1)\left(w^2 - \frac{1}{1/9}\right)}}dw \\
&= \int_0^z \frac{1}{\sqrt{\frac{1}{1/9}}}\frac{1}{\sqrt{(w^2-1)\left(\frac{1}{9}w^2-1\right)}}dw \\
&= \frac{1}{3}\int_0^z \frac{1}{\sqrt{(w^2-1)\left(\frac{1}{9}w^2-1\right)}}dw \\
&= \frac{1}{3}\int_0^z \frac{1}{\sqrt{(1-w^2)\left(1-\frac{1}{9}w^2\right)}}dw \\
&= \frac{1}{3}F(\arcsin z, \frac{1}{3}).
\end{aligned}
$$

As we will see, our initial choice for the prevertices $(-3, -1, 1, 3)$ directly impacts the aspect ratio of the rectangle.

EXERCISE 5.6. Follow these steps to determine the aspect ratio of the rectangle that is the image of $\mathbb{H}$ under the function

$$f(z) = \frac{1}{3}F(\arcsin z, \frac{1}{3}).$$

(1) Explain why the integral

$$K_1 = \int_0^1 \frac{1}{\sqrt{(1-w^2)\left(1-\frac{1}{9}w^2\right)}}dw$$

must be a real number (hint: use geometry of the integrand). Conclude that

$$f(1) = \frac{1}{3}F\left(\arcsin 1, \frac{1}{3}\right) = \frac{1}{3}F\left(\pi/2, \frac{1}{3}\right) = K_1/3.$$

By symmetry, show that $f(-1) = -K_1/3$. Thus the length of one side of the rectangle is $2|K_1|/3$.

(2) Determine that since $f(3)$ is the next vertex of the target rectangle (moving counterclockwise), and $f(3) = f(1) + iK_2$ where K_2 is some real constant. Combine this with the fact that

$$f(3) = \frac{1}{3}F\left(\arcsin 3, \frac{1}{3}\right)$$

to show that $iK_2 = \frac{1}{3}\left(F(\arcsin 3, \frac{1}{3}) - F(\pi/2, \frac{1}{3})\right)$. The choice of sign for K_2 could be either positive or negative, depending on our choice of $\sqrt{-1}$. For consistency with our choice of angles, K_2 should be positive (note that *Mathematica* uses the "wrong" branch of the square root for this function on the real axis).

(3) Combine the findings above to determine that the aspect ratio of the rectangle is

$$\frac{2F\left(\pi/2, \frac{1}{3}\right)}{F\left(\arcsin 3, \frac{1}{3}\right) - F\left(\pi/2, \frac{1}{3}\right)} \approx 1.279...$$
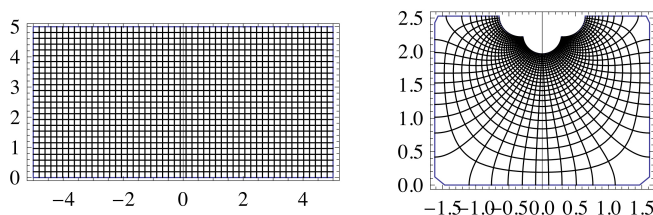
FIGURE 5.2. Portion of upper half-plane (left) and portion of target rectangle (right) to which it maps under the function of Exercise 5.3.

The mapping for Example 5.3 is illustrated in Figure 5.2. Only a portion of the upper half-plane and its image are shown. This explains why the target rectangle is incompletely filled in the upper central area. However we can see that the aspect ratio is at least approximately the same as the one we calculated.

SMALL PROJECT 5.7. Rework Example 5.3 for a more general situation. Use the prevertices $x_1 = -\lambda$, $x_2 = -1$, $x_3 = 1$, and $x_4 = \lambda$, where $\lambda > 1$. You can find an equation involving λ as a variable that, chosen correctly, would force the target rectangle to be a square. To make the target rectangle a square, you find that λ is part of an equation that cannot be explicitly solved, and must be approximated numerically. This is a standard problem in some introductory complex analysis books (see for instance example 22 of section 14, [**19**]).

One observation we can make based on this example is that while it is straightforward to write down a mapping function that has the correct angles, there is no

simple way to prescribe the side-lengths. Also, there are only a few rare cases when our integral can be expressed in terms of elementary functions, and in general it is not easy to evaluate. In order to find and evaluate specific Schwarz-Christoffel mappings, it is usually helpful to use symmetry of the target polygon (and of the prevertices) to simplify the computations.

A further simplification to the Schwarz-Christoffel formula that is frequently employed is to set one of the prevertices to be ∞, which effectively removes one factor from the denominator of the Schwarz-Christoffel formula. An example where this simplification is helpful is in mapping onto a triangle.

We note here that there is no guarantee that the Schwarz-Christoffel formula will result in a univalent function (see [**9**]). The only thing we can say for sure is that a map from the upper half plane to a simply connected polygonal domain that is conformal MUST take the form of equation 67 for some choice of constants and prevertices (for more detail, see [**3**]).

In Exercise 5.6, we made use of symmetry by choosing the prevertices to be ± 1 and ± 3. This symmetry simplified our calculations of the rectangle's vertices. However we unable to easily find a mapping onto a square. In mapping onto a square, or onto any regular polygon, it makes sense to adapt the Schwarz-Christoffel formula to map the unit disk to the polygon, using symmetrically placed points on the unit circle in place of the x_i in our existing formula. This can be accomplished by precomposing our mapping function found with Equation 67 with a Möbius transformation from the unit disk to the upper half-plane discussed in Exercise 5.1.

In addition to the problem of prescibing the lengths of the sides of the target polygon, a further problem arises with this approach for target polygons more complicated than a rectangle. Typically we will produce an integral that cannot be evaluated, even with special functions. These two issues are nicely resolved if we instead obtain a Schwarz-Christoffel formula that maps the unit disk directly to the target polygon and with the preimages of the vertices falling on the unit circle. These points can be chosen, for example, to be roots of unity to give some symmetry in the integral. To obtain this formula we carry out a change of variables that maps the disk to the upper half-plane, using the map defined in Exercise 5.1. Interestingly, the transformed integral formula is of exactly the same form.

EXERCISE 5.8. Set $w = \phi(z) = i\frac{1+z}{1-z}$ which maps the disk in the z plane to the upper half w-plane (see Exercise 5.1). Show that the Schwarz-Christoffel formula retains the same form as equation 67. ***Try it out!***

The Schwarz-Christoffel map that we will use on the unit disk is then

$$f(z) = C_1 \int_0^z \frac{1}{(w - \zeta_1)^{\beta_1}(w - \zeta_2)^{\beta_2} \cdots (w - \zeta_n)^{\beta_n}} dw + C_2, \ z \in \mathbb{D},$$

where $\beta_i \pi$ is the exterior angle of the ith vertex of the target polygon, and the preimages ζ_i of the vertices are on the unit circle. Here we use ζ_i instead of x_i to emphasize

that the prevertices are not on the real axis. As with Equation 67, the complex constants C_1 and C_2 with $C_1 \neq 0$ rotate, resize and translate the polygon.

With our original Schwarz-Christoffel formula from the upper half plane, it is not at all obvious how we could obtain a map onto a regular polygon. However, we can exploit the symmetries of the roots of unity by choosing ζ_i to be the symmetrically placed nth roots of unity corresponding to symmetrically placed vertices in the target polygon. Consequently, all side lengths of the target polygon will be equal.

EXAMPLE 5.9. We obtain the Schwarz-Christoffel map onto a regular n-gon. The exterior angles of a regular n-gon are $2\pi/n$, so $\beta_i = 2/n$.

$$\int_0^z \frac{1}{(w - \zeta_1)^{\beta_1}(w - \zeta_2)^{\beta_2} \cdots (w - \zeta_n)^{\beta_n}} dw = \int_0^z \frac{1}{[(w - \zeta_1)(w - \zeta_2) \cdots (w - \zeta_n)]^{2/n}} dw$$

Suppose that the ζ_i are the nth roots of unity. Now we can use the fact that

$$\prod_{i=1}^{n}(w - \zeta_i) = w^n - 1$$

to simplify to $\int_0^z \frac{1}{(w^n - 1)^{2/n}} dw$. By factoring out $(-1)^{2/n}$ we can adjust the multiplicative constant and chose our mapping function

(69) $$f(z) = (-1)^{-2/n} \int_0^z \frac{1}{(w^n - 1)^{2/n}} dw = \int_0^z \frac{1}{(1 - w^n)^{2/n}} dw.$$

Here f has been defined from the Schwarz-Christoffel formula with choices of constant $C_1 = (-1)^{-2/n}$ (which rotates the figure by $4\pi/n$ radians) and $C_2 = 0$. This last formula cannot be evaluated using the usual methods from calculus, but can be readily evaluated using hypergeometric functions.

5.2.1. Basic Facts about Hypergeometric Functions. The integral in the last example cannot be expressed in terms of elementary functions, but can be easily evaluated and plotted using a computer algebra system by using some basic facts about some special power series known as hypergeometric functions. Hypergeometric functions, besides their many other applications, can be used to evaluate the integrals obtained above. A geometric series is a power series in which ratios of successive terms are constant. Generalizing this, for a hypergeometric series ratios of successive terms are rational functions of the index rather than just constants. Here we will make use of the most widely utilized hypergeometric functions–the so-called "two F ones," where the rational function has numerator and denominator of the second order.

DEFINITION 5.10. The hypergeometric function $_2F_1(a, b; c; z)$ is the power series

$$_2F_1(a, b; c; z) = \sum_{n=0}^{\infty} \frac{(a)_n (b)_n}{(c)_n} \frac{z^n}{n!},$$

where a, b and $c \in \mathbb{C}$ and

$$(x)_n = x(x+1) \cdots (x+n-1)$$

is the **shifted factorial**, or **Pochhammer symbol**.

EXERCISE 5.11. Use simple algebra to check that $(x)_{n+1} / (x)_n = x + n$. ***Try it out!***

If we compute the ratio of two successive terms in the geometric series $\sum_{n=0}^{\infty} r^n z^n$ we obtain simply the ratio r times z. In the next exercise we carry out the same computation for a hypergeometric series.

EXERCISE 5.12. Show that the ratio of two successive terms in the series $_2F_1 (a, b; c; z)$ is

$$\frac{(a+n)(b+n)z}{(c+n)(n+1)}.$$

Try it out!

The formula obtained in this last exercise motivates the use of the term "hypergeometric." Whereas for a geometric series the ratio of successive terms is a single constant times z, for a hypergeometric function this ratio is a rational function of n, multiplied by z.

EXERCISE 5.13. Apply the ratio test to show that we get convergence of the hypergeometric function $_2F_1 (a, b; c; z)$ on compact subsets of the unit disk. ***Try it out!***

Any function which is useful or widely applicable typically earns the status of "special function." A number of well known special functions can be written as hypergeometric series. For example

$$\log \frac{1+z}{1-z} = 2z \, _2F_1 \left(1/2, 1; 3/2; z^2\right),$$
$$(1-z)^{-a} = \, _2F_1 \left(a, b; b; z\right), \text{ and}$$
$$\arcsin z = z \, _2F_1 \left(1/2, 1/2; 1; z^2\right).$$

The functions $\sin(z)$ and $\cos(z)$ themselves can each be obtained as limiting cases of a "two F one" series.

Now we consider another example that involves a $_2F_1$ hypergeometric series, $z \, _2F_1(\frac{1}{2}, \frac{1}{4}; \frac{5}{4}; z^4)$. We will see shortly that this function is a Schwarz-Christoffel transformation that maps the unit disk onto a square.

EXERCISE 5.14. Use Definition 5.10 to find the first several terms in the series of $z \, _2F_1(\frac{1}{2}, \frac{1}{4}; \frac{5}{4}; z^4)$. The following table gives the first several values of the necessary Pochhammer symbols. If you graph your result using *ComplexTool*, you should get a picture similar to Figure 5.3.

n	$(1/2)_n$	$(1/4)_n$	$(5/4)_n$	Coefficient of z^{4n+1}
0	1	1	1	1
1	1/2	1/4	5/4	1/10
2	3/4	5/16	45/16	1/24
3	15/8	45/64	585/64	5/208
4	105/16	585/256	9945/256	35/2176
5	945/32	9945/1024	208845/1024	3/256

Try it out!
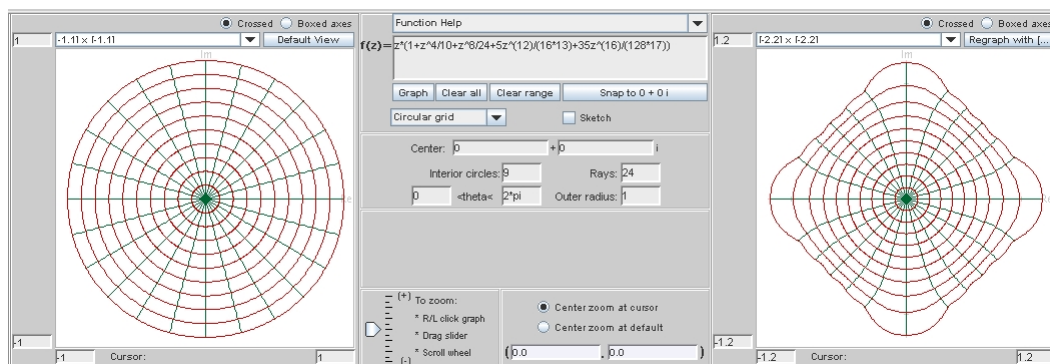
FIGURE 5.3. *ComplexTool* image of an approximation of the conformal map (using the first 5 terms)

The rate of convergences in this example is such that even with just a few non-zero terms of the series, we obtain a map whose image is approximately a square. We now see how to derive the formula for a Schwarz-Christoffel map onto the square.

DEFINITION 5.15 (Euler representation). The hypergeometric function $_2F_1\left(a, b; c; z\right)$ can be written in integral form as

$$_2F_1\left(a, b; c; z\right) = \frac{\Gamma(c)}{\Gamma(b)\Gamma(c-b)} \int_0^1 \frac{t^{b-1}(1-t)^{c-b-1}}{(1-tz)^a} dt,$$

which is known as the *Euler representation* of the $_2F_1$ function.

The symbol Γ stands for the Gamma function, defined for z in the right half plane by

$$\Gamma(z) = \int_0^\infty t^{z-1}e^{-t}dt.$$

The function can be extended analytically to the whole plane except for the negative integers $-1, -2, -3, \ldots$

EXERCISE 5.16. For integer values of n, the Gamma function is related to the factorial by $\Gamma(n) = (n-1)!$. Prove this by directly evaluating $\Gamma(1) = 1$ and then showing that $\Gamma(z+1) = z\Gamma(z)$ (hint: use integration by parts). ***Try it out!***

EXERCISE 5.17. In this exercise, you will show that Definitions 5.10 and 5.15 are equivalent.

(1) Expand the factor $(1 - tz)^{-a}$ with the binomial theorem as a power series to obtain

$$(1 - tz)^{-a} = \sum_{n=0}^{\infty} \frac{(a)_n}{n!} t^n z^n.$$

(2) Use the formula (see [**14**] Theorem 7, p. 19 for a proof)

$$\Gamma(p) \Gamma(q) = \Gamma(p + q) \cdot \int_0^1 t^{p-1} (1 - t)^{q-1} \, dt$$

(where p and q have positive real parts) to show

$$\int_0^1 t^{n+b-1} (1 - t)^{c-b-1} dt = \frac{\Gamma(b + n) \Gamma(c - b)}{\Gamma(c + n)}.$$

The integral on the right is also an important special function known as the **beta function** in the variables p and q.

(3) Show

$$\frac{(b)_n}{(c)_n} = \frac{\Gamma(c)}{\Gamma(b) \Gamma(c - b)} \frac{\Gamma(b + n) \Gamma(c - b)}{\Gamma(c + n)}.$$

(4) Put the previous facts together to obtain the required formula. Substitute the power series for the denominator term, and then interchange summation and integral, to show that

$$\int_0^1 \frac{t^{b-1}(1 - t)^{c-b-1}}{(1 - tz)^a} dt = \frac{\Gamma(b) \Gamma(c - b)}{\Gamma(c)} \, _2F_1(a, b; c; z).$$

Try it out!

It takes a little work to establish the relationship of the beta function with the Gamma function used in part (2). For an excellent exposition of this fact and an introduction to special functions in general, see [**14**].

EXAMPLE 5.18. For a square (a regular 4-gon), the Schwarz Christoffel map from the unit disk is given by $z \, _2F_1(\frac{1}{2}, \frac{1}{4}; \frac{5}{4}; z^4)$. These numbers probably seem to have been essentially "pulled from a hat," but when we apply the Euler integral representation we will see that these are the numbers we need (after a transformation) to evaluate the given integral. To see this, with $n = 4$ in the integral representation of equation 69 we obtain

$$\int_0^z \frac{1}{\sqrt{1 - w^4}} dw.$$

Let $a = 1/4$, $b = 1/2$, and $c = 5/4$. We use the Euler integral representation for $_2F_1(a, b; c; z)$ to evaluate

$$\begin{aligned}
_2F_1\left(1/2, 1/4; 5/4; z^4\right) &= \frac{\Gamma(5/4)}{\Gamma(1/4)\Gamma(1)} \int_0^1 \frac{t^{-3/4}(1-t)^0}{(1 - tz^4)^{1/2}} dt \\
&= 1/4 \int_0^1 \frac{1}{t^{3/4}} \frac{1}{\sqrt{1 - tz^4}} dt.
\end{aligned}$$

Now change variables by letting $w^4 = tz^4$ (so $t = (w^4/z^4)$) and $4w^3 dw = z^4 dt$. Then

$$\begin{aligned}
_2F_1\left(1/2, 1/4; 5/4; z^4\right) &= 1/4 \int_0^z \frac{z^3}{w^3} \frac{1}{\sqrt{1 - w^4}} \frac{4w^3 dw}{z^4} \\
&= \frac{1}{z} \int_0^z \frac{1}{\sqrt{1 - w^4}} dw.
\end{aligned}$$

Thus

$$f(z) = \int_0^z \frac{1}{\sqrt{1 - w^4}} dw = z\, _2F_1\left(1/2, 1/4; 5/4; z^4\right).$$
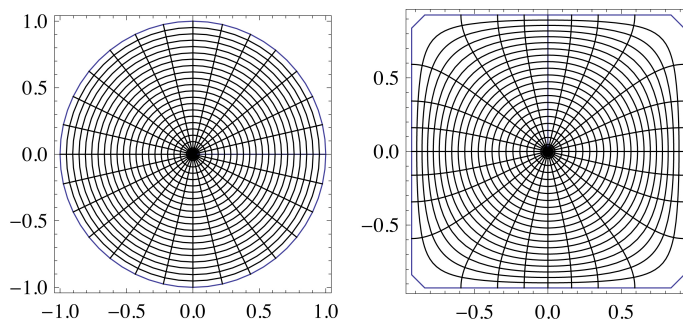
FIGURE 5.4. Unit disk (left) and target square (right) to which it maps under the function of Exercise 5.18.

Contrast the map in Figure 5.4, where the domain is the (bounded) unit disk, with the earlier map onto a rectangle in Figure 5.2. One advantage with the disk map is that we can see the entire mapping domain and the entire target square is filled. Note also the rotational and reflectional symmetry obtained by using nth roots of unity on the unit circle as prevertices.

EXERCISE 5.19. Show that the conformal map from the disk onto the regular n-gon is (up to rotations, translations and scalings) given by $z\, _2F_1\left(2/n, 1/n; (n+1)/n; z^n\right)$. **Try it out!**

We describe another situation where this technique of integration can be useful, for readers who have worked through Chapters 2 or 4. In Chapter 4, Section 4.5 discusses

the shear construction, and in Chapter 2, Section 2.6 the shear construction and its relationship to minimal surfaces is discussed.

SMALL PROJECT 5.20. Define a non-convex 6-sided polygon P with $\beta_1 = \beta_4 = -1/3$ and $\beta_2 = \beta_3 = \beta_5 = \beta_6 = 2/3$. (Draw this polygon!) Find a representation for the Schwarz-Christoffel transformation that maps the unit disk $\mathbb{D}$ onto P, with the prevertices ζ_i being the 6th roots of unity. It works well if $\zeta_1 = 1$ and $\zeta_4 = -1$, with the other 6th roots of unity going in order counterclockwise around the circle. Verify that the analytic function $f(z) : P \to \mathbb{D}$ is given by

$$f(z) = z \, _2F_1\left(\frac{2}{3}, \frac{1}{6}; \frac{7}{6}; z^6\right) - \frac{z^3}{3} \, _2F_1\left(\frac{2}{3}, \frac{1}{2}; \frac{3}{2}; z^6\right).$$

Now, in the language of Chapter 4, let $h(z) - g(z)$ be the function given above as $f(z)$ and let the dilatation $\omega(z) = z^2$. Find the harmonic function $h(z) + \overline{g(z)}$. Verify that by using

$$h(z) = z \, _2F_1\left(\frac{2}{3}, \frac{1}{6}; \frac{7}{6}; z^6\right)$$

and

$$g(z) = \frac{z^3}{3} \, _2F_1\left(\frac{2}{3}, \frac{1}{2}; \frac{3}{2}; z^6\right),$$

it is indeed true that $\omega(z) = \frac{h'}{g'} = z^2$. Use a computer algebra system to create a picture of the image of $\mathbb{D}$ under the function $h(z) + \overline{g(z)}$.

Furthermore, if you have studied Chapter 2, you can find the minimal surface that lifts from this harmonic function. You should find that it is defined by

$$
\begin{aligned}
x_1 &= \mathrm{Re}\left(z \, _2F_1\left(\frac{2}{3}, \frac{1}{6}; \frac{7}{6}; z^6\right) + \frac{z^3}{3} \, _2F_1\left(\frac{2}{3}, \frac{1}{2}; \frac{3}{2}; z^6\right) \right) \\
x_2 &= \mathrm{Im}\left(z \, _2F_1\left(\frac{2}{3}, \frac{1}{6}; \frac{7}{6}; z^6\right) - \frac{z^3}{3} \, _2F_1\left(\frac{2}{3}, \frac{1}{2}; \frac{3}{2}; z^6\right) \right) \\
x_3 &= \mathrm{Im}\left(z^2 \, _2F_1\left(\frac{2}{3}, \frac{1}{3}; \frac{3}{2}; z^6\right) \right).
\end{aligned}
$$

We now examine the conformal map onto a symmetric non-convex polygon in the shape of a star. We intend in the next section to find harmonic maps onto the same figure, and will find that while mappings onto polygons with convex corners are relatively easy to construct, there is little or no supporting theory when non-convex corners are involved.

EXAMPLE 5.21. Suppose we want to map onto a (non-convex) m-pointed star, so there are $n = 2m$ vertices. The interior angles alternate between $\pi\alpha_1$ and $\pi\alpha_2$ where $\alpha_1 < 1 < \alpha_2$ (so a "sharp point" of the star occurs at α_1). Corresponding exterior angles then alternate between a positive value β_1 and a negative value β_2 (assuming we have a non-convex star).
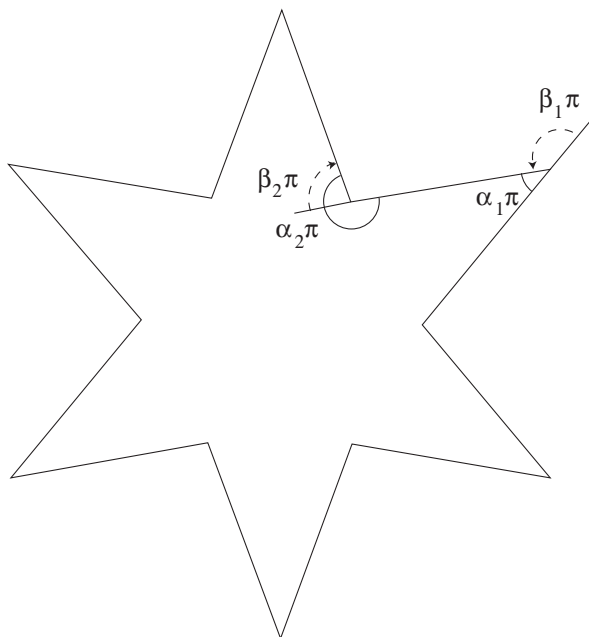
FIGURE 5.5. Interior and exterior angles of a symmetric star

Also, $\beta_1 + \beta_2$ must satisfy $m(\beta_1 + \beta_2) = 2$ so $\beta_1 + \beta_2 = 2/m = 4/n$. We use $\beta_{odd} = \beta_1 > 0$ and $\beta_{even} = -\beta_2 > 0$. Thus we have

$$\int_0^z \frac{\displaystyle\prod_{i \text{ even}} (w - \zeta_1)^{\beta_i}}{\displaystyle\prod_{i \text{ odd}} (w - \zeta_i)^{\beta_i}} \, dw.$$

Letting ζ_i be nth roots of unity,

$$\prod_{i \text{ even}}^n (z - \zeta_i) = z^m - 1 \text{ and } \prod_{i \text{ odd}}^n (z - \zeta_i) = z^m + 1,$$

so

$$\int_0^z \frac{(w^m - 1)^{-\beta_2}}{(w^m + 1)^{\beta_1}} \, dw,$$

where $\beta_1 + \beta_2 = 4/n$. Apart from constants chosen to expand or rotate the figure as necessary, we obtain the mapping function

$$f(z) = \int_0^z \frac{(1 - w^m)^{-\beta_2}}{(1 + w^m)^{\beta_1}} \, dw$$

from the disk onto the star shape in Figure 5.5.

335

EXAMPLE 5.22. The following appears as an exercise in [**13**] (Chapter V). Prove that the integral that maps the unit disk exactly onto a 5-pointed star with interior angles alternating at $\pi/5$ and $7\pi/5$ is given by

$$f(z) = \int_0^z \frac{(1-w^5)^{2/5}}{(1+w^5)^{4/5}} dw.$$

The corresponding exterior angles are $4\pi/5$ and $-2\pi/5$, so $\beta_1 = 4/5$ and $\beta_2 = -2/5$. Thus we have $n = 10$ and $m = 5$ and

$$\int_0^z \frac{(z^m - 1)^{-\beta_2}}{(z^m + 1)^{\beta_1}} dz = \int_0^z \frac{(z^5 - 1)^{2/5}}{(z^5 + 1)^{4/5}} dz$$

To compute this integral we must use the Appell F_1 function of two variables defined below.

DEFINITION 5.23. The Appell F_1 function is defined by

$$F_1(a; b_1, b_2; c; x, y) = \sum_{n=0}^\infty \sum_{m=0}^\infty \frac{(a)_{n+m}(b_1)_m(b_2)_n}{m! n! (c)_{n+m}} x^m y^n,$$

In *Mathematica* one can use the command $\texttt{AppelF1(a,b_1,b_2,c,x,y)}$. This special function also has an integral form, just as the hypergeometric functions have the Euler representation. We do not include a derivation here but refer the interested reader to [**2**] (Chapter 9) for the integral formula

$$F_1(a; b_1, b_2; c; x, y) = \frac{\Gamma(c)}{\Gamma(a)\Gamma(c-a)} \int_0^1 u^{a-1}(1-u)^{c-a-1}(1-ux)^{-b_1}(1-uy)^{-b_2} du.$$

Working in reverse we find that

$$F_1\left(1/5; 4/5, -2/5; 6/5; z^5, -z^5\right)$$
$$= \frac{\Gamma(6/5)}{\Gamma(1/5)\Gamma(1)} \int_0^1 u^{-4/5}(1-u)^0\left(1-uz^5\right)^{-4/5}\left(1+uz^5\right)^{2/5} du$$
$$= 1/5 \int_0^1 \frac{(1-uz^5)^{2/5}}{u^{4/5}(1+uz^5)^{4/5}} du.$$

To obtain our Schwarz-Christoffel formula we must now change variables, letting $w^5 = uz^5$ so $5w^4 dw = z^5 du$. Then

$$F_1\left(1/5; 4/5, -2/5; 6/5; z^5, -z^5\right) = 1/5 \int_0^z \frac{z^4}{w^4} \frac{(1-w^5)^{2/5}}{(1+w^5)^{4/5}} \frac{5w^4 dw}{z^5}$$
$$= \frac{1}{z} \int_0^z \frac{(1-w^5)^{2/5}}{(1+w^5)^{4/5}} dw$$

Thus

$$f(z) = \int_0^z \frac{(1 - w^5)^{2/5}}{(1 + w^5)^{4/5}} dw = z\, F_1\left(1/5; 4/5, -2/5; 6/5; z^5, -z^5\right)$$

and the mapping function is shown in Figure 5.6.
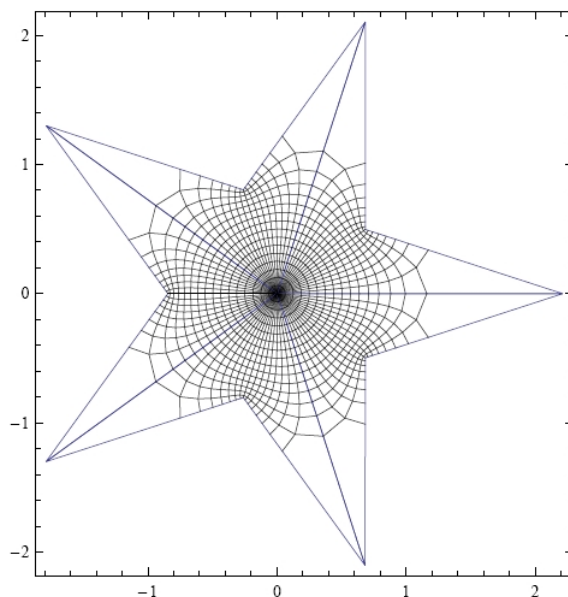
FIGURE 5.6. Image of the conformal map of the unit disk onto the 5-pointed star

EXERCISE 5.24. Show that the conformal map from the disk onto the m pointed star with exterior angle $\beta_1 > 0$, and $\beta_2 = 2/m - \beta_1$ (up to rotations, translations and scalings) is given by $z\, F_1\left(1/n; \beta_1, \beta_2, (n+1)/n; z^n, -z^n\right)$ where F_1 is the Appell F_1 function. **Try it out!**

5.3. The Poisson Integral Formula

While the Schwarz-Christoffel formula gives analytic and thus angle-preserving (conformal) functions from $\mathbb{D}$ to any polygon, we can see that it often starts with an integral that requires advanced mathematics to evaluate. If our goal is not necessarily an analytic function, we could work with the Poisson integral formula. This does not give us an analytic function, but instead, a harmonic function from the unit disk to the target domain. We first recall the definition of a harmonic function.

DEFINITION 5.25. A real-valued function $u(x, y)$ is harmonic provided that

$$u_{xx} + u_{yy} = 0.$$

337

A complex-valued function $f(z) = f(x + iy) = u(x, y) + iv(x, y)$ is harmonic if both u and v are harmonic.

The definition of a complex-valued harmonic function does not require that the functions u and v be harmonic conjugates, so while all analytic functions are harmonic, a complex-valued harmonic function is not necessarily analytic. In fact, the functions we work with for the rest of the chapter will not be analytic, and thus not conformal.

You may be familiar with the Poisson integral formula as a way of constructing a real-valued harmonic function that satisfies certain boundary conditions. For example, if the boundary conditions give the temperature of the boundary of a perfectly insulated plate, then the harmonic function gives the steady-state temperature of the interior of the plate. Another application is to find electrostatic potential given boundary conditions. A brief summary of that procedure is given here. For more detailed discussion, consult [**15**] or [**20**].

THEOREM 5.26 (Poisson Integral Formula). Let the complex valued function $\hat{f}(e^{i\theta})$ be piecewise continuous and bounded for θ in $[0, 2\pi]$. Then the function $f(z)$ defined by

(70)
$$f(z) = \frac{1}{2\pi} \int_0^{2\pi} \frac{1 - |z|^2}{|e^{it} - z|^2} \hat{f}(e^{it}) dt$$

is the unique harmonic function in the unit disk that satisfies the boundary condition

$$\lim_{r \to 1} f(re^{i\theta}) = \hat{f}(e^{i\theta})$$

for all θ where $\hat{f}$ is continuous.

Here, we present the proof in the special case where the boundary function $\hat{f}(e^{i\theta})$ is the real part of a function that is analytic on a disk with radius larger than 1. This proof can be found in any standard complex analysis textbook, for example, [**10**] or [**15**]. The interested reader may find the full result in Chapter 6 of [**?**] and Chapter 8 of [**10**].

PROOF. (Special Case) First observe that Cauchy's integral formula tells us that if we have a function $f(z)$ that is analytic inside and on the circle $|z| = R$, then, for $|z| < R$,

(71)
$$f(z) = \frac{1}{2\pi i} \int_{|\zeta|=R} \frac{f(\zeta)}{\zeta - z} d\zeta.$$

Here we are using the Greek letter zeta (ζ) as the variable of integration in the integral. In the discussion that follows here, we will be thinking about evaluating the function $f(z)$ at some fixed value of z, so the variable under consideration is now ζ. We also observe (for reasons that will become obvious in a few sentences) that for fixed z, with $|z| < 1$, the function $\dfrac{f(\zeta)\,\overline{z}}{1 - \zeta\,\overline{z}}$ is analytic in the variable ζ on and inside

338

$|\zeta| < 1$, since the denominator is non-zero. (Exercise for the reader: Think about why the denominator is non-zero.) Thus, by the Cauchy Integral Theorem, we know that

(72)
$$\frac{1}{2\pi i} \int_{|\zeta|=1} \frac{f(\zeta)\,\overline{z}}{1 - \zeta\,\overline{z}} = 0.$$

Combining Equations 71 and 72, we see that

$$
\begin{aligned}
f(z) &= \frac{1}{2\pi i} \int_{|\zeta|=1} \frac{f(\zeta)}{\zeta - z} d\zeta + 0 \\
&= \frac{1}{2\pi i} \int_{|\zeta|=1} \left(\frac{f(\zeta)}{\zeta - z} + \frac{f(\zeta)\,\overline{z}}{1 - \zeta\,\overline{z}} \right) d\zeta \\
&= \frac{1}{2\pi i} \int_{|\zeta|=1} \frac{1 - \zeta\,\overline{z} + \overline{z}(\zeta - z)}{(\zeta - z)(1 - \zeta\,\overline{z})} f(\zeta) d\zeta \\
&= \frac{1}{2\pi i} \int_{|\zeta|=1} \frac{1 - |z|^2}{(\zeta - z)(1 - \zeta\,\overline{z})} f(\zeta) d\zeta.
\end{aligned}
$$

Now we parameterize the circle $|\zeta| = 1$ by $\zeta(t) = e^{it}$, giving $d\zeta = ie^{it}dt$ and

$$
\begin{aligned}
f(z) &= \frac{1}{2\pi i} \int_0^{2\pi} \frac{1 - |z|^2}{(e^{it} - z)(1 - e^{it}\,\overline{z})} f(e^{it}) ie^{it} dt \\
&= \frac{1 - |z|^2}{2\pi} \int_0^{2\pi} \frac{1}{(e^{it} - z)e^{it}(e^{-it} - \overline{z})} f(e^{it}) e^{it} dt \\
&= \frac{1 - |z|^2}{2\pi} \int_0^{2\pi} \frac{f(e^{it})}{(e^{it} - z)(e^{-it} - \overline{z})} dt.
\end{aligned}
$$

Taking the real part of both sides of the equation gives us a harmonic function (since the real part of an analytic function is harmonic), and also yields Equation 70.

□

EXERCISE 5.27. Verify that the "Poisson kernel," $\dfrac{1 - |z|^2}{|e^{it} - z|^2}$, can be rewritten as

$\mathrm{Re}\left(\dfrac{e^{it} + z}{e^{it} - z} \right) = \mathrm{Re}\left(\dfrac{1 + ze^{-it}}{1 - ze^{-it}} \right)$. ***Try it out!***

The integral in Equation 70 is, in general, very difficult to integrate. However, if there is an arc on which the function $\hat{f}(e^{it})$ is constant, then the integration is easy to do.

EXERCISE 5.28. Verify that
(73)
$$\frac{1}{2\pi} \int_a^b K \, \mathrm{Re}\left(\frac{e^{it} + z}{e^{it} - z} \right) dt = K \frac{b - a}{2\pi} + \frac{K}{\pi} \arg\left(\frac{1 - ze^{-ib}}{1 - ze^{-ia}} \right) = \frac{K}{\pi} \left[\arg\left(\frac{e^{ib} - z}{e^{ia} - z} \right) - \frac{b - a}{2} \right].$$

Helpful hints: Swap the order of integration and taking the real part of a function. The algebraic identity $\frac{1+w}{1-w} = 1 + \frac{2w}{1-w}$ will be helpful.

Try it out!

The beauty of the last formulation of equation (73) is that it can be visualized geometrically. Consider the picture with e^{ia} and e^{ib} on the unit circle, and z somewhere inside the unit circle. Then the vector from z to e^{ia} is $e^{ia} - z$, and the vector from z to e^{ib} is $e^{ib} - z$, so that the angle between those two vectors is given by $\arg\left(\frac{e^{ib} - z}{e^{ia} - z}\right)$, as is shown if Figure 5.7.
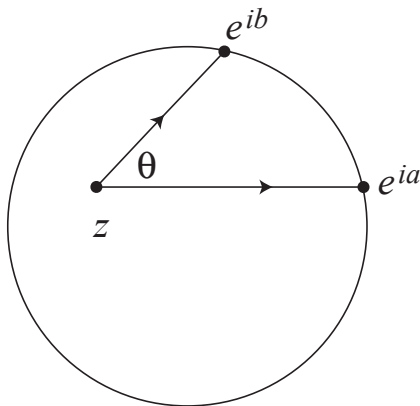
FIGURE 5.7. Geometric interpretation of $\theta = \arg\left(\dfrac{e^{ib} - z}{e^{ia} - z}\right)$.

EXAMPLE 5.29. Assume that the unit disk is a thin insulated plate, with a temperature along the boundary of 50 degrees for the top semicircle and 20 degrees along the bottom semicircle. From physics, we know that the function which describes the temperature within the unit disk must be a harmonic function. Use the results above to find that harmonic function.

Solution: Apply the formula given above to the situation where $a_1 = 0, b_1 = \pi, K_1 = 50$ and then add it to the result where $a_2 = \pi, b_2 = 2\pi, K_2 = 20$. The result is the function $f(z) = \frac{1}{2\pi}(70\pi + 60\arg\left(\frac{1+z}{1-z}\right))$. (Notice that the 70 is $50 + 20$, and that $60 = 2(50 - 20)$.) When z ranges across the unit disk, the function $\frac{1+z}{1-z}$ covers the right half-plane (you can check this experimentally by graphing the function $\frac{1+z}{1-z}$ using *ComplexTool*), so the argument of it is between $-\pi/2$ and $\pi/2$. This gives function values for $f(z)$ between 20 and 50, which makes good sense. Another way of thinking of the solution is that it gives the average temperature $\pm$ half of the difference between the maximum and minimum temperatures.

EXERCISE 5.30. Referring to $f(z) = \frac{1}{2\pi}(70\pi + 60 \arg\left(\frac{1+z}{1-z}\right))$ in the solution to Example 5.29, find $f(0)$, $f(i/2)$ and $f(-i/2)$. Do your answers make sense?

Using the result of Exercise 5.28, we can see that computing the Poisson integral formula for a *piecewise constant* boundary is particularly simple. Many applications of the Poisson integral formula come from having the boundary correspondence remain constant on arcs of the unit circle.

Most introductory analysis books give the Poisson integral formula for real-valued $\hat{f}(e^{i\theta})$. It can also be applied to create a harmonic function for complex-valued $\hat{f}(e^{i\theta})$, but the univalence of the harmonic function is not at all apparent. Let's first explore what could happen if we try to use the Poisson integral formula with complex boundary values.

EXAMPLE 5.31. The simplest example of this is obtained by letting the first third of the unit circle (that is, the arc from 0 to $e^{i2\pi/3}$) map to 1, the next third to $e^{i2\pi/3}$ and the last third to $e^{i4\pi/3}$. Let's work through the details of this integration, working from equation (73). We compute

$$
\begin{aligned}
f(z) &= \frac{1}{2\pi}\left(\left(\frac{2\pi}{3} - 0\right) + 2\arg\left(\frac{1 - ze^{-i2\pi/3}}{1 - ze^0}\right)\right. \\
&\quad + e^{i2\pi/3}\left(\frac{4\pi}{3} - \frac{2\pi}{3}\right) + 2e^{i2\pi/3}\arg\left(\frac{1 - ze^{-i4\pi/3}}{1 - ze^{-i2\pi/3}}\right) \\
&\quad \left. + e^{i4\pi/3}\left(2\pi - \frac{4\pi}{3}\right) + 2e^{i4\pi/3}\arg\left(\frac{1 - ze^{-2\pi i}}{1 - ze^{-i4\pi/3}}\right)\right) \\
&= \frac{2\pi}{3(2\pi)}\left(1 + e^{i2\pi/3} + e^{i4\pi/3}\right) \\
&\quad + \frac{1}{\pi}\left(\arg\left(\frac{1 - ze^{-i2\pi/3}}{1 - ze^0}\right) + e^{i2\pi/3}\arg\left(\frac{1 - ze^{-i4\pi/3}}{1 - ze^{-i2\pi/3}}\right) + e^{i4\pi/3}\arg\left(\frac{1 - ze^{-2\pi i}}{1 - ze^{-i4\pi/3}}\right)\right) \\
&= 0 + \frac{1}{\pi}\left(\arg\left(\frac{1 - ze^{-i2\pi/3}}{1 - z}\right) + e^{i2\pi/3}\arg\left(\frac{1 - ze^{-i4\pi/3}}{1 - ze^{-i2\pi/3}}\right) + e^{i4\pi/3}\arg\left(\frac{1 - z}{1 - ze^{-i4\pi/3}}\right)\right).
\end{aligned}
$$

Figure 5.8 shows the image of the unit disk as graphed in *ComplexTool*. Notice that it appears to be one-to-one on the interior of the unit disk. It certainly is not one-to-one on the boundary! (Entering this formula into *ComplexTool* is a bit unwieldy, so this function is one of the Pre-defined functions, the one called Harmonic Triangle. We will soon use the *PolyTool* applet, as described on the next page, to graph other similar functions.)

EXERCISE 5.32. Find a general formula that maps the unit disk harmonically to the interior of a convex regular n-gon. ***Try it out!***
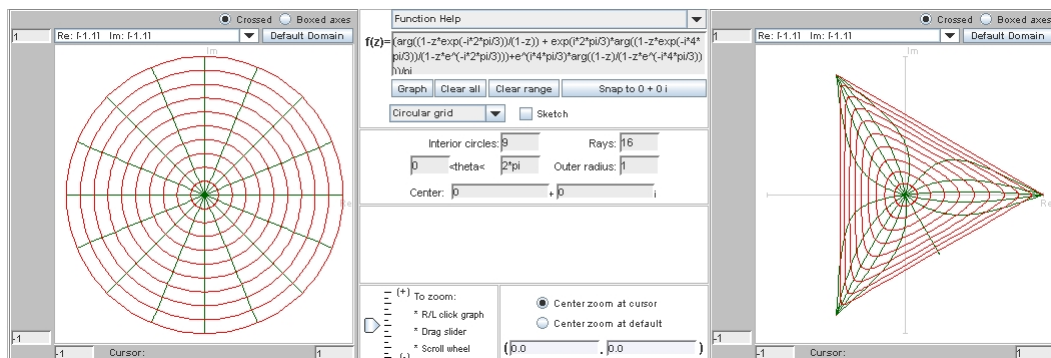
FIGURE 5.8. *ComplexTool* image of the harmonic function mapping to the triangle

SMALL PROJECT 5.33. Refer to Chapter 4 and its discussion of the shear construction. Find the pre-shears of the polygonal mappings in exercise 5.32. In other words, what analytic function do you shear to get that polygonal function? A good first step is to determine the dilatation of this function. See [**4**] for more details.

EXERCISE 5.34. For a non-convex example, consider the function that maps quarters of the unit circle to the four vertices $\{1, i, -1, \frac{i}{2}\}$. Verify that this function is

$$\frac{3i}{8} + \frac{1}{\pi}\left(\arg\left(\frac{1+iz}{1-z}\right) + i\arg\left(\frac{1+z}{1+iz}\right) - \arg\left(\frac{1-iz}{1+z}\right) + \frac{i}{2}\arg\left(\frac{1-z}{1-iz}\right)\right).$$

When we graph this function in *ComplexTool*, we notice that it appears to NOT be one-to-one. Furthermore, $f(0) = \frac{3i}{8}$, so that the image of $\mathbb{D}$ is not the interior of the polygon. This function is another one of the Pre-defined functions in *ComplexTool*. ***Try it out!***
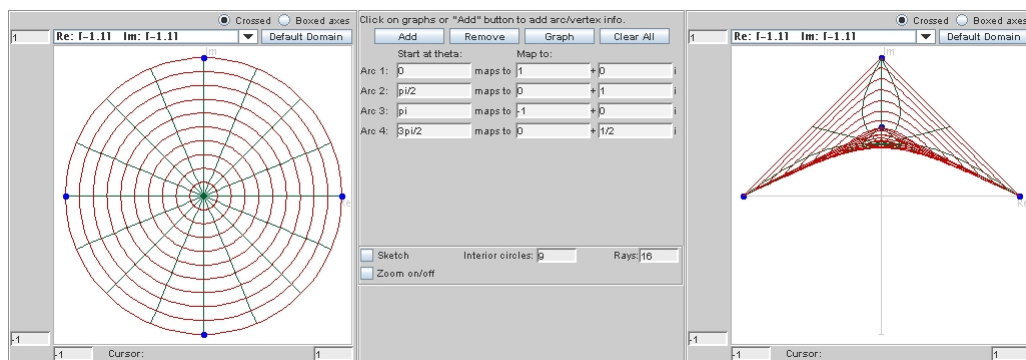
FIGURE 5.9. The *PolyTool* Applet

At this point, you should start using the *PolyTool* applet. In this applet, you can specify which arcs of the unit circle will map to which points in the range, and

342

the applet will compute and graph the harmonic function defined by extending that boundary correspondence to a function on $\mathbb{D}$. When you first open this applet, you see a circle on the left and a blank screen on the right. You can create a harmonic function that maps portions of the boundary of the circle to vertices of a polygon in one of two ways. First, you can click on the unit circle in the left panel to denote an arc endpoint, and continue choosing arc endpoints there, and then choose the target vertices by clicking in the right graph. (Note that as you click, text boxes in between the panels fill with information about where you clicked.) Once you have the boundary correspondence you want, click the `Graph` button. Alternatively, you can click the button that says `Add` to get text boxes for input. For example, to create the function in Exercise 5.34, click `Add`, then fill in the first row of boxes for `Arc 1:` with `0 maps to 1+0i`. When you want another set of arcs, click `Add` again. Note that the `Arc` boxes denote the starting point of the arc (i.e. for the arc from 0 to $\pi/3$, use 0 in the `Arc` box). Continue filling, and when you are ready to compute the Poisson integral to get the harmonic function, click the `Graph` button. Once you have a function graphed, you can "drag" around either the arc endpoints (in the domain on the left) or the target points (in the range on the right) and watch the function dynamically change.

EXPLORATION 5.35. Are there ways of rearranging the boundary conditions to make the function created in Exercise 5.34 univalent? For example, what if the bottom half of the unit circle gets mapped to $i/2$, and the top half of the unit circle is divided into thirds for the other three vertices? This isn't univalent, but in some sense is closer to univalent than the mapping defined in Exercise 5.34. Is there a modification to be made so that it is univalent? **_Try it out!_**

EXERCISE 5.36. This is an extension of Exploration 5.35. Sheil-Small [**17**] proved (by techniques other than those discussed so far) that the harmonic extension of the boundary correspondence below maps the unit disk univalently onto the desired shape:

arc from	to	maps to
$-i$	i	i
i	$-3/5 + 4i/5$	-1
$-3/5 + 4i/5$	$-3/5 - 4i/5$	$i/2$
$-3/5 - 4i/5$	$-i$	1

For this function, first convince yourself that it appears to be univalent, and then find the function $f(z)$. **_Try it out!_**

We will be working a lot with harmonic functions that are extensions of a piecewise constant boundary correspondence, as in the example above. To have a framework for future discussions, we make the following formal definition.

DEFINITION 5.37. Let $\{e^{it_k}\}$ be a partition of $\partial\mathbb{D}$, where $t_0 < t_1 < \ldots < t_n = t_0 + 2\pi$. Let $\hat{f}(e^{it}) = v_k$ for $t_{k-1} < t < t_k$. We call the harmonic extension of this step function (as defined by the Poisson integral formula) $f(z)$.

EXAMPLE 5.38. To better understand the definition, we demonstrate how the notation in Definition 5.37 is used for the function in Exercise 5.36. Since the arc from $-i$ to i can be thought of as the arc along the unit circle from $e^{-i\pi/2}$ to $e^{-i\pi/2}$, we say that $t_0 = -\pi/2$ and $t_1 = \pi/2$. These points map to the vertex at i, so $v_1 = i$. The next arc set is a little more difficult, because we need to find the angle t_2 that goes with the point in the plane $z = \frac{-3}{5} + \frac{4}{5}i = e^{it_2}$. Unfortunately, we can only get a numerical estimate of the angle, found by $\pi + \arctan\left(\frac{4/5}{-3/5}\right) = \pi + \arctan\left(-4/3\right) \approx 2.2143$. (We add π because the output of arctangent is always in the first or fourth quadrant, while the angle in question is in the second quadrant.) Thus we have $t_2 \approx 2.2143$ and $v_2 = -1$. Continuing this process, we have $t_3 \approx 4.0689$ and $v_3 = i/2$. Then we finish with $t_4 = 3\pi/2$ and $v_4 = 1$. Notice that as in the definition, $t_4 = t_0 + 2\pi$. Then the function $f(z)$ from Definition 5.37 is the harmonic function that appears to be univalent when graphed in *PolyTool*.

EXERCISE 5.39. Combine the result of Exercise 73 (on page 339) with Definition 5.37 to show that the function $f(z)$ in Definition 5.37 can be written

$$
\begin{aligned}
f(z) \;=\; & v_1 \left(\frac{t_1 - t_0}{2\pi}\right) + \frac{v_1}{\pi} \arg\left(\frac{1 - ze^{-it_1}}{1 - ze^{-it_0}}\right) \\
& + v_2 \left(\frac{t_2 - t_1}{2\pi}\right) + \frac{v_1}{\pi} \arg\left(\frac{1 - ze^{-it_2}}{1 - ze^{-it_1}}\right) \\
& + \ldots + v_n \left(\frac{t_n - t_{n-1}}{2\pi}\right) + \frac{v_n}{\pi} \arg\left(\frac{1 - ze^{-it_n}}{1 - ze^{-it_{n-1}}}\right).
\end{aligned}
$$

Try it out!

Since the Poisson integral formula gives rise to a harmonic function, we must learn some of the basics of the theory of harmonic functions before proceeding too far.

5.4. Harmonic Function Theory

Chapter 4 gives a detailed explanation of harmonic functions, as does [**5**]. Much of that material will be helpful for our future investigations, so we repeat it here.

5.4.1. The Basics. Any harmonic function f can be written as $f = h + \overline{g}$, where h and g are analytic functions. The analytic dilatation $\omega(z) = \frac{g'(z)}{h'(z)}$ is, in some sense, a measure of how much the harmonic function does not preserve angles. A dilatation of $\omega(z) \equiv 0$ means that the function is analytic, so must be conformal. A dilatation with modulus near 1 indicates that the function distorts angles greatly. (For more intuition about the dilatation, read Section 4.6 of Chapter 4.) A result of Lewy states that a harmonic function has nonzero Jacobian (denoted $J_f(z) = |h'|^2 - |g'|^2$) if it is locally univalent. This result is in line with our understanding of the relationship between local univalence and a nonvanishing derivative for analytic functions.

THEOREM 5.40 (Lewy's Theorem). For a harmonic function f defined on a domain Ω, if f locally univalent in Ω, then $J_f(z) \neq 0$ for all $z \in \Omega$.

Note that this is equivalent to Lewy's Theorem in Chapter 4.

A nice consequence of Lewy's Theorem is that if a function is locally univalent in Ω, then its analytic dilatation either satisfies $|\omega(z)| < 1$ for all $z \in \Omega$, or $|\omega(z)| > 1$ for all $z \in \Omega$. In our work, we will only study functions that are locally univalent in some domain Ω and satisfy $|\omega(z)| < 1$ for all $z \in \Omega$. These functions are called *sense-preserving* because they preserve the orientation of curves in Ω.

EXERCISE 5.41. Verify that the condition that $J_f(z) \neq 0$ is equivalent to $|\omega(z)| \neq 1$, as long as $h'(z) \neq 0$. Conclude that a function that is locally univalent and sense-preserving must have $J_f(z) > 0$ and $|\omega(z)| < 1$. ***Try it out!***

Of particular interest in this setting is determining how to split up the argument function (which is harmonic and sense-preserving) into h and $\overline{g}$.

EXERCISE 5.42. Show that the function $f(z) = K \arg(z)$ has canonical decomposition $h(z) = \dfrac{1}{2i} K \log(z)$ and $g(z) = \dfrac{1}{2i} \overline{K} \log(z)$. ***Try it out!***

Another consequence of the canonical decomposition of a harmonic function is that we can write the analytic functions h and g defined in some domain Ω in terms of their power series expansions, centered at some $z_0 \in \Omega$, as

$$(74) \qquad f(z) = a_0 + \sum_{k=n}^{\infty} a_k(z - z_0)^k + \overline{b_0 + \sum_{k=m}^{\infty} b_k(z - z_0)^k}.$$

If f is sense-preserving, then we necessarily have that either $m > n$ or $m = n$ with $|b_n| < |a_n|$. In either case, when f is represented by Equation 74, we say that f has a zero of order n at z_0.

EXERCISE 5.43. In this exercise, we prove that the zeros of a sense-preserving harmonic function are isolated.

(a) Assume that $f(z)$ is a sense-preserving locally univalent function with series expansion as given in Equation 74. Show that if $f(z_0) = 0$, there exists a positive δ and a function ψ such that, for $0 < |z - z_0| < \delta$ we can write

$$(75) \qquad f(z) = h(z) + \overline{g(z)} = a_n(z - z_0)^n(1 + \psi(z))$$

where

$$\psi(z) = \frac{a_{n+1}}{a_n}(z - z_0) + \frac{a_{n+2}}{a_n}(z - z_0)^2 + \cdots + \overline{b_m(z - z_0)^m}\frac{1}{a_n(z - z_0)^n} + \cdots .$$

(b) Show that part (a) implies that $|\psi(z)| < 1$ for z sufficiently close to z_0, since $m \geq n$ and $|b_n/a_n| < 1$ if $m = n$.

(c) Show that part (b) implies that the zeros of a sense-preserving harmonic function are isolated, since $f(z) \neq 0$ near z_0 (except, of course, at z_0).

Try it out!

5.4.2. The Argument Principle.

5.4.2.1. *Analytic Argument Principle.* The Argument Principle for analytic functions gives a very nice way to describe the number of zeros and poles inside a contour. We take time to explore this topic, even though it is in many introductory complex analysis courses, to emphasize the geometric nature of the result. We first provide a formal definition of a common phrase, the winding number of the image of a contour about the origin.

DEFINITION 5.44. The winding number of the image under $f(z)$ of a simple closed contour Γ about the origin is the net change in argument of $f(z)$ as z traverses Γ in the positive (counterclockwise) direction, divided by 2π. It can be denoted by $\frac{1}{2\pi}\Delta_\Gamma \arg f(z)$.

To explore the relationship between the winding number of the image of a contour about the origin and the number of zeros and poles contained within that contour, do the following exploration.

EXPLORATION 5.45. Open *ComplexTool*. Change the `Interior circles` to 1 and the `Rays` to 0. This will graph just the boundary of the circle of interest. As you graph the following functions, examine the image of the circle "winds around," or encloses, the origin. Count how many times the image of the circle winds around the origin, making sure that you count the counterclockwise direction as positive and the clockwise direction as negative. If the image of the circle winds around the origin once, you know that there must be a zero of $f(z)$ inside that circle. (Think about this last sentence and make sure you understand it.)

- Graph $f(z) = z^2$, using a circle of radius 1. Use the "Sketch" button and trace around the circle in the domain to get a good feeling for how many times its image winds around the origin. You should already know the answer. (Any other radius will work too. Why is that?)
- Graph $f(z) = z(z - 0.3)$. Use circles of radius $0.2, 0.3$ and 0.5. (You may have to zoom in on the image of the circle of radius 0.2 to really understand what it is doing.) If you want, you may check the `Vary radius` checkbox and use the slider to change the radius of the circle.
- Graph $f(z) = z^4 - 6z + 3$. Use circles of radius $0.9, 1.5, 1.7$, and 2.
- Graph $f(z) = \frac{z^4 - 6z + 3}{z - 1}$, using circles of radius 0.9 and 1.5.
- Graph $f(z) = \frac{z^4 - 6z + 3}{(z-1)^2}$, using circles of radius $0.9, 1.5$ and 2.
- Go back through the previous 3 items, now changing the function while keeping the radius fixed.
- For all of the previous functions, find the locations of all of the zeros and poles, paying particular attention to how far they are from the origin.
- Make up your own function and do some more experiments.

Based on the explorations above, what is your connection between the winding number of the image of a circle about the origin and the number of zeros and poles inside the circle?

Try it out!

THEOREM 5.46 (Argument Principle for Analytic Functions). Let C be a simple closed contour lying entirely within a domain D. Suppose f is analytic in D except at a finite number of poles inside C and that $f(z) \neq 0$ on C. Then

$$\frac{1}{2\pi i} \oint_C \frac{f'(z)}{f(z)} dz = N_0 - N_p,$$

where N_0 is the total number of zeros of f inside C and N_p is the total number of poles of f inside C. In determining N_0 and N_p, zeros and poles are counted according to their order or multiplicity.
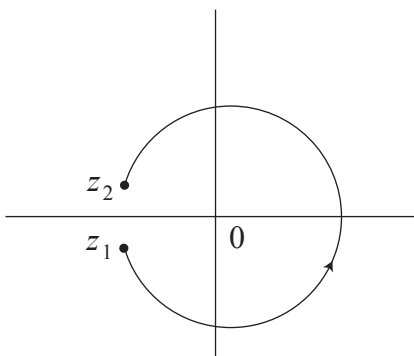
FIGURE 5.10. A path around a branch cut

Before proving Theorem 5.46, we explore the connection between the winding number and the integral $\dfrac{1}{2\pi i} \oint_C \dfrac{f'(z)}{f(z)} dz$. To see this connection, we start with a related integral, $\int_{z_1}^{z_2} \frac{f'(z)}{f(z)} dz$, where z_1 and z_2 are points very close to each other, but lying on opposite sides of a branch cut of $\log f(z)$, and we take a "counterclockwise" path along C from z_1 to z_2. (See Figure 5.10.) Now we do the following computation:

$$
\begin{aligned}
\int_{z_1}^{z_2} \frac{f'(z)}{f(z)} dz &= \log f(z)|_{z_1}^{z_2} \\
&= \ln|f(z_2)| - \ln|f(z_1)| + i(\arg(f(z_2)) - \arg(f(z_1))).
\end{aligned}
$$

When we take the limit as $z_1 \to z_2$, we get that $\ln|f(z_2)| - \ln|f(z_1)| \to 0$ and $i(\arg(f(z_2)) - \arg(f(z_1))) \to 2\pi i \cdot (\text{winding number})$. (Think carefully about this last statement and make sure you understand it.)

PROOF. The proof of the Argument Principle relies on the Cauchy integral formula and deformation of contours. Take a moment to review these important concepts. We begin by deforming the contour C to a series of smaller contours around the isolated zeros and poles of f. If there are no zeros or poles, then $\frac{f'(z)}{f(z)}$ is analytic, so the integral is zero, as desired. We then analyze the zeros and poles individually, and add the results together to get the desired conclusion. More formally, when f has zeros or poles inside C, they must be isolated, and because f is analytic on C, there are only a finite number of distinct zeros or poles inside C. Denote the zeros and poles by z_j, for $j = 1, 2, \ldots, n$. Let γ_j be a circle of radius $\delta > 0$ centered at z_j, where δ is chosen small enough that the circles γ_j all lie in D and do not meet each other. Join each circle γ_j to C by a Jordan arc λ_j in D. Consider the closed path Γ formed by moving around C in the positive (counterclockwise) direction while making a detour along each λ_j to γ_j, running once around this circle in the clockwise (negative) direction, then returning along λ_j to C. See Figure 5.11.
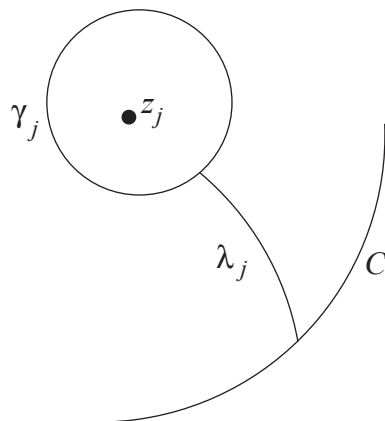
FIGURE 5.11. The contour Γ

This curve Γ contains no zeros or poles of f, so $\Delta_\Gamma \arg f(z) = \dfrac{1}{2\pi i} \oint_\Gamma \dfrac{f'(z)}{f(z)} dz = 0$ by the argument above. When considering the total change in argument along Γ of $f(z)$, the contributions of the arcs λ_j along Γ cancel out, so that

$$\Delta_C \arg f(z) = \sum_{j=1}^{n} \Delta_{\gamma_j} \arg f(z),$$

where each of the circles γ_j is now traversed in the positive (counterclockwise) direction. Thus now we may consider each individual γ_j and sum the results.

Suppose that f has a zero of order n at $z = z_j$. Then $f(z) = (z - z_j)^n f_n(z)$, where $f_n(z)$ is an analytic function satisfying $f_n(z_j) \neq 0$. (If you can't remember why this is

348

true, look in any standard introductory complex analysis book.) Then

$$f'(z) = n(z - z_j)^{n-1} f_n(z) + (z - z_j)^n f'_n(z)$$

and

$$
\begin{aligned}
\frac{f'(z)}{f(z)} &= \frac{n(z - z_j)^{n-1} f_n(z) + (z - z_j)^n f'_n(z)}{(z - z_j)^n f_n(z)} \\
&= \frac{n}{z - z_j} + \frac{f'_n(z)}{f_n(z)}.
\end{aligned}
$$

Now we note that when we integrate the above expression along γ_j, we get $n(2\pi i) + 0$, because $\frac{f'_n(z)}{f_n(z)}$ is analytic inside the contour.

Now suppose that f has a pole of order m at $z = z_k$. This means that f can be rewritten as $f(z) = (z - z_k)^{-m} f_m(z)$, where f_m is analytic an nonzero at $z = z_k$. Then, as previously, we have

$$
\begin{aligned}
\frac{f'(z)}{f(z)} &= \frac{-m(z - z_k)^{-m-1} f_m(z) + (z - z_k)^{-m} f'_m(z)}{(z - z_k)^{-m} f_m(z)} \\
&= \frac{-m}{z - z_k} + \frac{f'_m(z)}{f_m(z)}.
\end{aligned}
$$

Once again, when we integrate the above expression along γ_k, we get $-m(2\pi i) + 0$.

Summing over $j = 1, 2, \ldots n$ gives us the integral over Γ and the desired result.

$\square$

5.4.2.2. *Argument Principle for Harmonic Functions.* There are many versions of the argument principle for harmonic functions. We only need the simple proof presented in this section, developed by Duren, Hengartner, and Laugesen ([**6**]).

THEOREM 5.47 (Argument Principle for Harmonic Functions). Let D be a Jordan domain with boundary C. Suppose f be a sense-preserving harmonic function on D, continuous in $\overline{D}$ and $f(z) \neq 0$ on C. Then $\Delta_C \arg f(z) = 2\pi N$, where N is the total number of zeros of $f(z)$ in D, counted according to multiplicity.

PROOF. First, we suppose that f has no zeros in D. This means that $N = 0$ and the origin is not an element of $f(D \cup C)$. A fact from topology says that in this case, $\Delta_C \arg f(z) = 0$, and the theorem is proved. We will prove this fact. Let ϕ be a homeomorphism of the closed unit square S onto $D \cup C$ with the restriction of ϕ to the boundary, $\hat{\phi} : \partial S \to C$, also a homeomorphism. See Figure 5.12.

The composition $F = f \circ \phi$ is a continuous mapping of S onto the plane with no zeros, and we want to prove that $\Delta_{\partial S} \arg F(z) = 0$. Begin by subdividing S into finitely many small squares S_j so that on each S_j, the argument of F varies by at most $\pi/2$. Then $\Delta_{\partial S_j} \arg F(z) = 0$ (since $F(S_j)$ cannot enclose the origin). Now when we consider $\Delta_{\partial S} \arg F(z)$, it is the sum $\sum_j \Delta_{\partial S_j} \arg F(z)$ because the contributions to
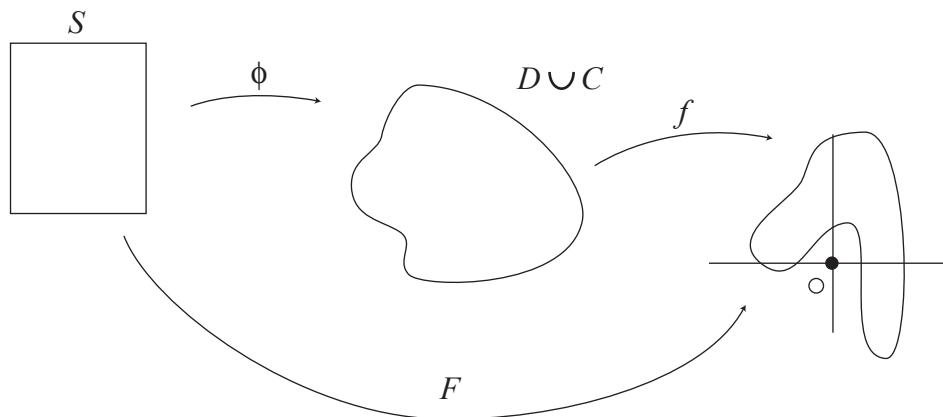
FIGURE 5.12. The composition of f and ϕ.

the sum from the boundaries of each S_j cancel out, except where the boundary of S_j agrees with the boundary of S. Thus $\Delta_{\partial S} \arg F(z) = 0$, as desired.

Now consider the case where f does have zeros in D. Because the zeros are isolated (as proven in Exercise 5.43), and because f is not zero on C, there are only a finite number of distinct zeros in D. We proceed in a manner similar to the proof of the analytic argument principle, and, denote the zeros by z_j, for $j = 1, 2, \ldots, n$. Let γ_j be a circle of radius $\delta > 0$ centered at z_j, where δ is chosen so small that the circles γ_j all lie in D and do not meet each other. Join each circle γ_j to C by a Jordan arc λ_j in D. Consider the closed path Γ formed by moving around C in the positive direction while making a detour along each λ_j to γ_j, running once around this circle in the clockwise (negative) direction, then returning along λ_j to C. (See Figure 5.11 on page 348.) This curve Γ contains no zeros of f, so $\Delta_\Gamma \arg F(z) = 0$ by the first case in this proof. When considering the total change in argument along Γ of $f(z)$, the contributions of the arcs λ_j along Γ cancel out, so that

$$\Delta_C \arg f(z) = \sum_{j=1}^{n} \Delta_{\gamma_j} \arg f(z),$$

where each of the circles γ_j is now traversed in the positive (counterclockwise) direction. Thus now we may consider each individual γ_j and sum the results.

Now suppose that f has a zero of order n at a point z_0. Then, as observed in Exercise 5.43 on page 345, f can be locally written as

$$f(z) = a_n(z - z_0)^n(1 + \psi(z))$$

where $a_n \neq 0$ and $|\psi(z)| < 1$ on a sufficiently small circle γ defined by $|z - z_0| = \delta$. This shows that

(76) $\qquad \Delta_\gamma \arg f(z) = n\Delta_\gamma \arg(z - z_0) + \Delta_\gamma \arg(1 + \psi(z)) = 2\pi n.$

350

Therefore, if f has zeros of order n_j at the points z_j, the conclusion is that

$$\Delta_C \arg f(z) = \sum_{j=1}^{n} \Delta_{\gamma_j} \arg f(z) = 2\pi \sum_{j=1}^{n} n_j = 2\pi N,$$

and the theorem is proved.

$\square$

EXERCISE 5.48. Justify why $\Delta_\gamma \arg(1 + \psi(z)) = 0$ in Equation 76. **Try it out!**

For the next section, we shift our focus back to the polygonal maps defined in Section 5.3. We will be using the Argument Principle for Harmonic Functions in the proof of the Rado-Kneser-Choquet Theorem.

5.5. Rado-Kneser-Choquet Theorem

As you examine the image of the unit disk using the examples in Section 5.3, you may notice that some of the functions seem to be one-to-one on the interior of the domain, while others do not seem to be univalent. Look again at the examples, and compare functions which map to convex domains versus functions that map to non-convex domains.

EXPLORATION 5.49. Make a conjecture about when functions are one-to-one, using the exercises from Section 5.3 as a springboard. Do this before reading the Rado-Kneser-Choquet Theorem! **Try it out!**

In general, we completely understand the behavior of harmonic extensions (as defined in Definition 5.37) that map to convex regions:

THEOREM 5.50 (Rado-Kneser-Choquet Theorem). Let Ω be a subset of $\mathbb{C}$ that is a bounded convex domain whose boundary is a Jordan curve Γ. Let $\hat{f}$ map $\partial \mathbb{D}$ continuously onto Γ and suppose that $\hat{f}(e^{it})$ runs once around Γ monotonically as e^{it} runs around $\partial \mathbb{D}$. Then the harmonic extension given in the Poisson integral formula is univalent in $\mathbb{D}$ and defines a harmonic mapping of $\mathbb{D}$ onto Ω.

For the proof of this important theorem, we use the following lemma.

LEMMA 5.51. Let ψ be a real-valued function harmonic in $\mathbb{D}$ and continuous in $\mathbb{D}$. Suppose ψ has the property that, after a rotation of coordinates, $\psi(e^{it}) - \psi(e^{-it}) \geq 0$ on the interval $[0, \pi]$, with strict inequality $\psi(e^{it}) - \psi(e^{-it}) > 0$ on some subinterval $[a, b]$ with $0 \leq a < b \leq \pi$. Then ψ has no critical points in $\mathbb{D}$.

The condition on ψ seems a bit mysterious at first, and so we should discuss it. One kind of function for which this property holds is a ψ that is at most bivalent on $\partial \mathbb{D}$. What does "at most bivalent" mean? We know that univalent means that a function is one-to-one. Bivalent means that a function is two-to-one, or that there may be $z_1 \neq z_2$ such that $f(z_1) = f(z_2)$, but that if $f(z_1) = f(z_2) = f(z_3)$, then at least 2 of z_1, z_2, z_3

must be equal. Alternatively, another kind of function ψ described in Lemma 5.51 is one that is continuous on $\partial \mathbb{D}$ where $\psi(e^{it})$ rises from a minimum at $e^{-i\alpha}$ to a maximum at $e^{i\alpha}$, then decreases again to its minimum at $e^{-i\alpha}$ as e^{it} runs around the unit circle, without having any other local extrema, but allowing arcs of constancy. See Figure 5.13.
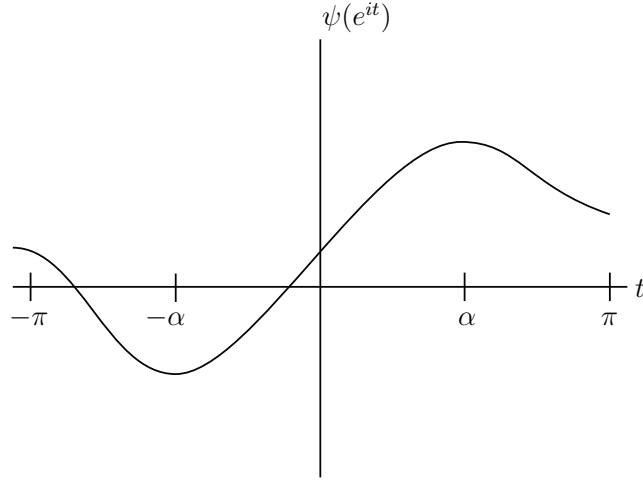
FIGURE 5.13. Boundary condition of one possible ψ satisfying the hypotheses of Lemma 5.51.

PROOF OF LEMMA 5.51. To show that ψ has no critical points in $\mathbb{D}$, we must show that $\frac{\partial \psi}{\partial z} \neq 0$ in $\mathbb{D}$. This is equivalent to saying that

$$\frac{1}{2}\left(\frac{\partial \psi}{\partial x} - i\frac{\partial \psi}{\partial y}\right) \neq 0.$$

At this point, we will simplify the proof by simply proving that $\psi_z(0) \neq 0$, and claim that will be sufficient. Indeed, if z_0 is some other point in $\mathbb{D}$, consider the function $\varphi(z) = \frac{z_0 - z}{1 - \overline{z_0}z}$ that is a conformal self-map of $\mathbb{D}$ with $\varphi(0) = z_0$, and consider the composition $F(\zeta) = \psi(\varphi(\zeta))$. Observe that F is harmonic in $\mathbb{D}$, continuous in $\overline{\mathbb{D}}$, and satisfies the same condition about $F(e^{it}) - F(e^{-it})$ as ψ does. Applying the chain rule to $F(\zeta)$ gives that $F_\zeta(\zeta) = \psi_z(\varphi(\zeta))\varphi'(\zeta)$, since φ is analytic and thus has $\varphi_{\overline{\zeta}} = 0$. (In general, the chain rule is more complicated for harmonic functions. Here, since φ is analytic, the chain rule takes its familiar form.) Plugging in 0 for ζ gives $F_\zeta(0) = \psi_z(z_0)\varphi'(0)$, implying that if $F_\zeta(0) = 0$ then also $\psi_\zeta(z_0) = 0$. Thus when we have proven that $\psi_z(0) \neq 0$, we will be able to generalize to $\psi_z(z_0) \neq 0$ for all z_0 in $\mathbb{D}$.

Now we use the Poisson integral formula to prove that $\psi_z(0) \neq 0$. Substituting in ψ (or $\hat{\psi}(e^{it}) = \lim_{r \to 1} \psi(re^{it})$ on $\partial \mathbb{D}$) gives

$$\psi(z) = \frac{1}{2\pi} \int_0^{2\pi} \frac{1 - |z|^2}{|e^{it} - z|^2} \hat{\psi}(e^{it}) dt = \frac{1}{2\pi} \int_0^{2\pi} \frac{1 - z\,\overline{z}}{(e^{it} - z)(e^{-it} - \overline{z})} \hat{\psi}(e^{it}) dt.$$

When we differentiate both sides with respect to z, the integral depends only on t, so we are just left differentiating the integrand. Doing this, we find

$$
\begin{aligned}
\frac{\partial}{\partial z} \left(\hat{\psi}(e^{it}) \frac{1 - z\,\overline{z}}{(e^{it} - z)(e^{-it} - \overline{z})} \right) &= \frac{\hat{\psi}(e^{it})}{e^{-it} - \overline{z}} \frac{\partial}{\partial z} \left(\frac{1 - z\,\overline{z}}{e^{it} - z} \right) \\
&= \left(\frac{\hat{\psi}(e^{it})}{e^{-it} - \overline{z}} \right) \cdot \left(\frac{e^{it}(e^{-it} - \overline{z})}{(e^{it} - z)^2} \right) \\
&= \hat{\psi}(e^{it}) \left(\frac{e^{it}}{(e^{it} - z)^2} \right),
\end{aligned}
$$

leading to the conclusion that

$$\psi_z(0) = \frac{1}{2\pi} \int_0^{2\pi} \hat{\psi}(e^{it}) e^{-it} dt.$$

From the hypotheses of the lemma, we know that there is some $t \in (0, \pi)$ such that $\psi(e^{it}) - \psi(e^{-it}) > 0$. Thus

$$
\begin{aligned}
\operatorname{Im} \psi_z(0) &= \operatorname{Im} \left(\frac{1}{2\pi} \int_0^{2\pi} \hat{\psi}(e^{it}) e^{-it} dt \right) \\
&= -\frac{1}{2\pi} \int_0^{2\pi} \hat{\psi}(e^{it}) \sin(t) dt \\
&= -\frac{1}{2\pi} \left(\int_0^{\pi} \hat{\psi}(e^{it}) \sin(t) dt + \int_{-\pi}^0 \hat{\psi}(e^{it}) \sin(t) dt \right) \text{ since } \hat{\psi} \text{ is periodic} \\
&= -\frac{1}{2\pi} \left(\int_0^{\pi} \hat{\psi}(e^{it}) \sin(t) dt - \int_0^{\pi} \hat{\psi}(e^{-it}) \sin(t) dt \right) \\
&= -\frac{1}{2\pi} \int_0^{\pi} (\hat{\psi}(e^{it}) - \hat{\psi}(e^{-it})) \sin(t) dt < 0.
\end{aligned}
$$

The last inequality relies on the fact that $\sin(t)$ is non-negative on the interval $[0, \pi]$. We have now shown that $\operatorname{Im} \psi_z(0) \neq 0$, thus proving the lemma. $\qquad \square$

PROOF OF THEOREM 5.50. Without loss of generality, assume that $\hat{f}(e^{it})$ runs around Γ in the counterclockwise direction as t increases. (Otherwise, take conjugates.) We will show that if the function f is not locally univalent in $\mathbb{D}$, then Lemma 5.51 will give a contradiction.

Suppose that $f = u + iv$ is not locally univalent, or that the Jacobian of f vanishes at some point z_0 in $\mathbb{D}$. This means that the matrix $\begin{pmatrix} u_x & v_x \\ u_y & v_y \end{pmatrix}$ has a determinant of 0 at z_0. From linear algebra, we know that this means that the system of equations

$$\begin{aligned} au_x + bv_x &= 0 \\ au_y + bv_y &= 0 \end{aligned}$$

has a nonzero solution (a, b). Thus the real-valued harmonic function $\psi = au + bv$ has a critical point at z_0 (since $(a, b) \neq (0, 0)$). However, the hypothesis of Theorem 5.50 implies that ψ satisfies the hypothesis of Lemma 5.51. Thus we have a contradiction, so f must be locally univalent.

Now that we see that f is locally univalent, we apply the argument principle to show that f is univalent in $\mathbb{D}$. Since f is sense-preserving on $\partial\mathbb{D}$ and locally univalent, f is sense-preserving throughout $\mathbb{D}$. Now, if f is not univalent, there are two points z_1 and z_2 in $\mathbb{D}$ such that $f(z_1) = f(z_2)$. However, that would imply that the function $f(z) - f(z_1)$ has two zeros in $\mathbb{D}$, so that the winding number of $f(z) - f(z_1)$ about the origin is 2, which contradicts the hypotheses about the boundary correspondence. This completes the proof.

$\square$

EXERCISE 5.52. Give a detailed proof of the statement, "However, the hypothesis of Theorem 5.50 implies that ψ satisfies the hypothesis of Lemma 5.51." ***Try it out!***

Notice that the description of $\hat{f}$ in Theorem 5.50 does not require that it be one-to-one on $\partial\mathbb{D}$, but permits arcs of constancy. Furthermore, the Rado-Kneser-Choquet Theorem is actually true in the case where $\hat{f}$ has jump discontinuities, as long as the image of $\partial\mathbb{D}$ is not contained in a straight line. This requires some additional justification, so we state it separately as a corollary.

COROLLARY 5.53. Let $f(z)$ be defined as in Definition 5.37 on 343. Suppose the vertices $v_1, v_2, \ldots v_n$, when traversed in order, define a convex polygon, with the interior of the polygon denoted by Ω. Then the function $f(z)$ is univalent in $\mathbb{D}$ and defines a harmonic mapping from $\mathbb{D}$ onto Ω.

Here is some intuition behind the proof of Corollary 5.53. Consider a sequence of functions $\hat{f}_m(e^{it})$ that are continuous and converge to the boundary correspondence $\hat{f}(e^{it})$ of Definition 5.37. (One possible such sequence of functions can be described by having $\hat{f}_m(e^{it}) = v_k$ for t-values in the interval $(t_{k-1} + \frac{t_k - t_{k-1}}{2m}, t_k - \frac{t_k - t_{k-1}}{2m})$ while for t-values in the interval $(t_k - \frac{t_k - t_{k-1}}{2m}, t_k + \frac{t_{k+1} - t_k}{2m})$, $\hat{f}_m(e^{it})$ maps the interval linearly to the segment between v_k and v_{k+1}.) Each of these functions $\hat{f}_m(e^{it})$ satisfies the conditions of the Rado-Kneser-Choquet Theorem, so extends to a univalent harmonic function, $f_m(z)$, in the unit disk. But the functions f_m converge uniformly on compact subsets of $\mathbb{D}$, so the entire sequence converges uniformly to f in $\mathbb{D}$. Therefore, $f(z)$ inherits

354

the univalence from the sequence. The fact that the limit function is still univalent is not immediately apparent–full details may be found in [**8**].

Interestingly enough, this theorem does not guarantee anything about univalence if the domain Ω is not convex. In fact, the expectation is that univalence will not be achieved. For example, look at Exercise 5.34 on page 342.

EXPLORATION 5.54. Extend the explorations begun in Exploration 5.35 on page 343. Now, instead of modifying the boundary correspondence, start with the correspondence in Exercise 5.36. Then, move the vertex that is at $i/2$, moving it closer to i. A very nice picture comes from having the vertex set be $\{1, i, -1, \frac{9i}{10}\}$. In this last we see the lack of univalence very clearly. ***Try it out!***

5.5.1. Boundary behavior. In this section, we explore what seems to be true with some of the above examples: There appears to be some very interesting boundary behavior of our harmonic extensions of step functions. Examine this behavior in the following exploration.

EXPLORATION 5.55. Using *ComplexTool* or *PolyTool*, graph the function from $\mathbb{D}$ to a triangle in Example 5.31. Now investigate the behavior of the boundary using the sketching tool of the applet. In particular, approach the break point between arcs (such as $z = 1$) along different paths. First approach radially, then approach along a line that is not a radius of the circle. Observe how these different paths that approach 1 cause the image of the path to approach different points along the line segment that makes up a portion of the boundary of the range. (As you get very close to an arc endpoint, the image of the sketch may jump to a vertex–here, examine where the image is immediately before that jump.) Technology hint: in *PolyTool* and *ComplexTool*, you can hit the `Graph` button to clear all previous sketching but keep the polygonal map. Repeat this exercise with some of the other examples of polygonal functions. Try to answer some of the following questions:

(1) Given a point ζ on the boundary of the polygon, is it possible to find a path γ approaching $\partial\mathbb{D}$ such that $\gamma(z)$ approaches ζ?
(2) As you approach an arc endpoint in $\partial\mathbb{D}$ radially, what point on the boundary of the polygon do you approach?

Try it out!

As you performed the exploration above, you probably discovered some of the known properties of the boundary behavior of harmonic extensions of step functions. These results were originally proven by Hengartner and Schober [**8**], who proved a more general form of the theorem below. We now restate their theorem as it applies to the step functions of Definition 5.37. In the theorem below, the *cluster set* of f at a point e^{it_k} is the set of all possible limits of sequences $\{z_n\}$, where z_n are inside Γ, and $\lim_{n\to\infty}(z_n) = e^{it_k}$.

THEOREM 5.56. Let f be the harmonic extension of a step function $\hat{f}(e^{it_k})$ in Definition 5.37. Denote by Γ the polygon defined by the vertices v_k. By definition, the radial limits $\lim_{r \to 1} f(re^{it})$ lie on Γ for all t except those in the set $\{t_k\}$. Then the unrestricted limit

$$\hat{f}(e^{it}) = \lim_{z \to e^{it}} f(z)$$

exists at every point $e^{it} \in \partial\mathbb{D}\setminus\{e^{it_1}, e^{it_2}, ..., e^{it_n}\}$ and lies on Γ. Furthermore,

(1) the one-sided limits as $t \to t_k$ are

$$\lim_{t \to t_k^-} \hat{f}(e^{it}) = v_k \qquad \text{and} \qquad \lim_{t \to t_k^+} \hat{f}(e^{it}) = v_{k+1}; \text{ and}$$

(2) the cluster set of f at each point $e^{it_k} \in \{e^{it_1}, e^{it_2}, ..., e^{it_n}\}$ is the linear segment joining v_k to v_{k+1}.

PROOF. Part (1) of the theorem follows directly from the definition of the function f and from properties of the Poisson integral formula. That is, since the boundary correspondence is defined in Definition 5.37, the limits follow. Now we need to show part (2). Now let us consider e^{it_k}. If z approaches e^{it_k} along the circular arc

(77) $$\arg\left(\frac{e^{it_k} + z}{e^{it_k} - z}\right) = \frac{\lambda\pi}{2}, \qquad -1 < \lambda < 1,$$

then $f(z)$ converges to the value

$$\frac{1}{2}(1 - \lambda)v_k + \frac{1}{2}(1 + \lambda)v_{k+1}.$$

Therefore the cluster set of f at t_k is the line segment joining v_k and v_{k+1}, and part (2) is proven.

$\square$

EXERCISE 5.57. Use basic ideas from analytic geometry to observe that equation 77 is a circular arc. (Hint: First consider the cases where z is either on the circle or is the center of the circle for some intuition.) **Try it out!**

EXERCISE 5.58. It is important to note that Theorem 5.56 holds for even non-univalent mappings. Go back to some of the previous examples and identify the line segment that connnects the vertices. In particular, regraph the example from Exercise 5.34. Using the `Sketch` utility in either *ComplexTool* or *Polytool*, check that the limit as you approach one of the t_k does appear to be that line segment. **Try it out!**

5.6. Star Mappings

From the Rado-Kneser-Choquet Theorem, we see that harmonic functions mapping the unit disk $\mathbb{D}$ to convex polygons are well-understood. That is, if we define a harmonic function mapping from the unit disk to a convex polygon as in Definition 5.37, the function is univalent in $\mathbb{D}$. Theorem 5.56 describes the boundary behavior

fully, showing that the limit of the function as we approach one of the break points between vertex pre-images, t_k, gives the line segment joining the vertices.

However, non-convex polygons are not nearly as well-understood. We first examine non-convex polygons in their simplest mathematical form: the ones of regular stars.

DEFINITION 5.59. By an n-pointed "star," or "r-star", we mean an equilateral $2n$-gon with the vertex set,

$$\left\{ r\alpha^{2k}, \alpha^{2k+1} : k = 1, 2, \ldots, n \text{ and } \alpha = e^{i\pi/n} \right\},$$

where r is some real constant.

Notice that when $r = 1$, the n-pointed star is a regular $2n$-gon, and when $r < \cos(\pi/n)$ or $r > \sec(\pi/n)$, the star is a strictly non-convex $2n$-gon. Our preimages of the vertices of the $2n$-gon will be arcs *centered* at the $2n$th roots of unity (this is different from our previous examples).
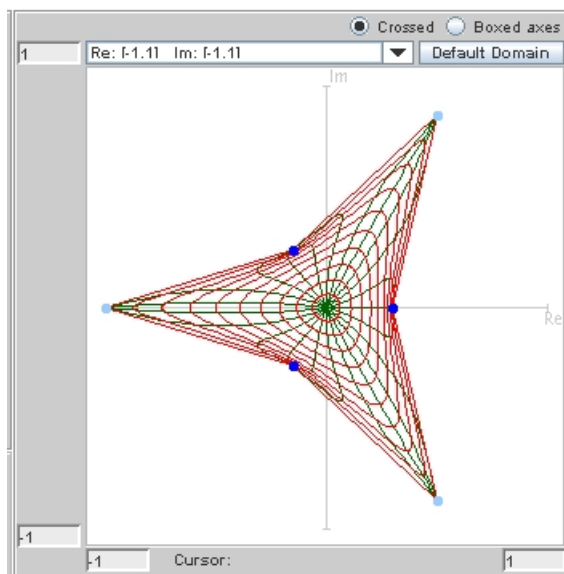
FIGURE 5.14. The 0.3-star for $n = 3$

EXAMPLE 5.60. We will find a harmonic mapping of the unit disk into the 0.3-star. More precisely, will find the harmonic extension of the following boundary correspondence.

for t from	to	$\hat{f}(e^{it})$
$-\pi/6$	$\pi/6$	0.3
$\pi/6$	$\pi/2$	$e^{i\pi/3}$
$\pi/2$	$5\pi/6$	$0.3e^{i2\pi/3}$
$5\pi/6$	$7\pi/6$	-1
$7\pi/6$	$3\pi/2$	$0.3e^{i4\pi/3}$
$3\pi/2$	$11\pi/6$	$e^{i5\pi/3}$

After going through details similarly to previous examples, we discover the harmonic extension is

$$
\begin{aligned}
f(z) \;=\; \frac{1}{\pi} \Bigg[& 0.3 \arg\left(\frac{1 - ze^{-i\pi/6}}{1 - ze^{i\pi/6}} \right) \\
& + e^{i\pi/3} \arg\left(\frac{1 + iz}{1 - ze^{-i\pi/6}} \right) + 0.3e^{i2\pi/3} \arg\left(\frac{1 - ze^{-i5\pi/6}}{1 + iz} \right) \\
& - \arg\left(\frac{1 - ze^{-i7\pi/6}}{1 - ze^{-i5\pi/6}} \right) + 0.3e^{i4\pi/3} \arg\left(\frac{1 - iz}{1 - ze^{-7i\pi/6}} \right) \\
& + e^{i5\pi/3} \arg\left(\frac{1 - ze^{-i11\pi/6}}{1 - ze^{-i3\pi/2}} \right) \Bigg].
\end{aligned}
$$

Graph this function using *ComplexTool* (it is one of the `Pre-defined functions`). Notice that it appears to be univalent. We certainly have not yet proved its univalence.

EXERCISE 5.61. Prove that if $f(z)$ is the harmonic extension to the r-star as defined in Definition 5.59, then $f(0) = 0$. Interpret this result geometrically. ***Try it out!***

EXERCISE 5.62. Modify the function in Exercise 5.60 to have $r = 0.15$ and see whether it appears univalent. To graph this new function in *ComplexTool*, choose the previous star as one of the `Pre-defined functions` and then modify the equation that shows in the function box. ***Try it out!***

To work with these stars, we may sometimes want to vary the boundary correspondence. That is, we may want to not split up $\partial\mathbb{D}$ completely evenly among the $2n$ vertices. It will become useful to us to have an unequal correspondence in the boundary arcs, but maintain some symmetry. To do this, we will still consider arcs centered at the $2n$-th roots of unity, but alternating between larger and smaller arcs. If we examine the geometry of this matter, we realize that an even split would make each arc have length $\frac{2\pi}{2n} = \pi/n$. Two consecutive arcs would together have length $2\pi/n$. To still maintain some symmetry, but let the arcs alternate in size, we want two consecutive arcs to still add to $2\pi/n$, but not split evenly. We introduce the parameter p, with $0 < p < 1$, as a tool to explain how the arcs are split. We will want two consecutive arcs split into $p2\pi/n$ and $(1-p)2\pi/n$. Note that the sum is still $2\pi/n$. This is formally described in the definition below.

DEFINITION 5.63. Let $n \geq 2$ be a fixed integer, r be a positive real number, and $\alpha = e^{i\pi/n}$. Define a boundary correspondence for all but a finite number of points on $\partial \mathbb{D}$ by mapping arcs with endpoints $\{\alpha e^{-ip\pi/n}, \alpha e^{ip\pi/n}, 0 \leq k \leq n-1\}$ as follows:

$$(78) \qquad \hat{f}(e^{it}) = \begin{cases} r\alpha^{2k}, & e^{it} \in (\alpha^{2k}e^{-ip\pi/n}, \alpha^{2k}e^{ip\pi/n}) \\ \alpha^{2k+1}, & e^{it} \in (\alpha^{2k+1}e^{-i(1-p)\pi/n}, \alpha^{2k+1}e^{i(1-p)\pi/n}) \end{cases} \quad .$$

Let f be the Poisson extension of $\hat{f}$.
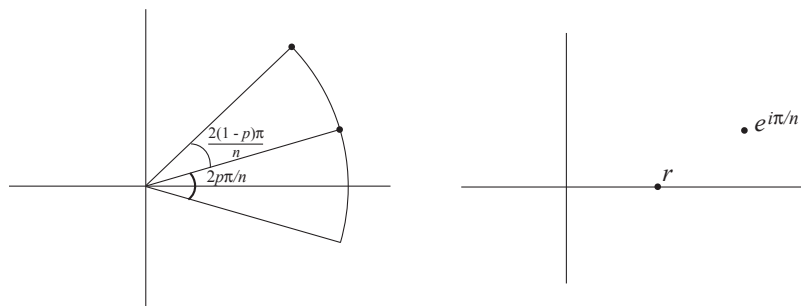
FIGURE 5.15. The first two arcs and their images according to Definition 5.63. The dots on the left-hand-side indicate points of discontinuity of the boundary correspondence.

Note that the arc $(\alpha^{2k}e^{-ip\pi/n}, \alpha^{2k}e^{ip\pi/n})$ centered at α^{2k} is mapped to the vertex $r\alpha^{2k}$ and the arc $(\alpha^{2k+1}e^{-i(1-p)\pi/n}, \alpha^{2k+1}e^{i(1-p)\pi/n})$ centered at α^{2k+1} is mapped to the vertex α^{2k+1}.

EXERCISE 5.64. Show that the interval in the second half of Equation 78,

$$(\alpha^{2k+1}e^{-i(1-p)\pi/n}, \alpha^{2k+1}e^{i(1-p)\pi/n}),$$

can be written more compactly as $(\alpha^{2k}e^{ip\pi/n}, \alpha^{2k+2}e^{-ip\pi/n})$. **_Try it out!_**

At this point, you should start using with the *StarTool* applet. The default for this applet is the 3-pointed star discussed in Example 5.60. Note that the arcs and their target vertices are color-coded (with a light blue arc mapping to a light blue vertex, for example). The default p-value is 0.5, which corresponds to evenly spaced arcs. You can use the slider bars (the plus/minus buttons for n) or type in the text boxes to change the values for n, p, and r. The maximum n-value allowed by the applet is $n = 18$, which is sufficient for the explorations below. As with *ComplexTool*, there is the option to `Sketch` on the graph to get a better feel for the mapping properties of these stars. There is also an option to `Show roots of` $\omega(z)$. The roots of $\omega(z)$ will be helpful in future discussion, but are not essential for the starting explorations. In general, try to first get a good feel for what happens for "small" values of n, such as $4, 5, 6$, or 7.
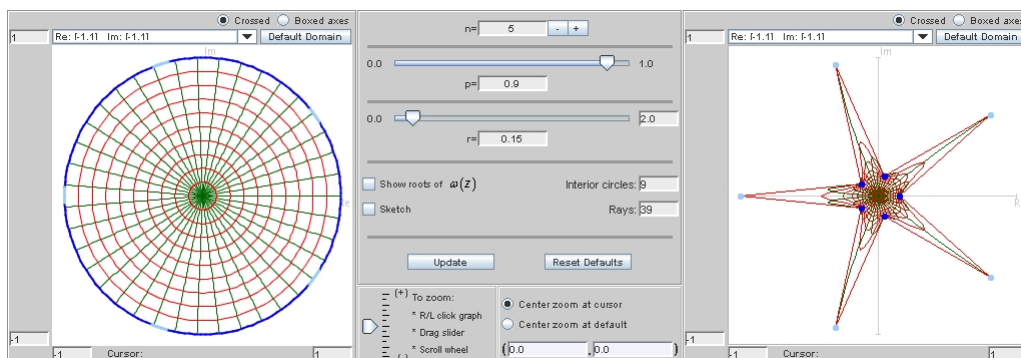
FIGURE 5.16. A star with $n = 5$, $r = 0.15$, and $p = 0.9$

EXPLORATION 5.65. Do some explorations with different values of n and r. See if you can find a pattern for univalence of the star function. Here are some avenues for exploration.

(1) What is the relationship between the r value you choose and the p-value necessary for univalence? Is there a range of p that works?
(2) What happens as p goes to 0 or 1?
(3) For a given p-value, can you determine the "minimal" r? That is, how small can you make r and maintain univalence?
(4) Is there a minimum r-value, one for which there is no p-value that will achieve univalence?
(5) For a fixed r, as you change n, what happens to the p that you need to achieve univalence?
(6) What is the full relationship between r, n, and p? (It is unlikely that you will answer this question now, but make some conjectures about it.)

Try it out!

5.7. Dilatations of Polygonal Maps are Blaschke Products

We are now ready to use the tools of harmonic functions to study the polygonal maps. To understand when the star maps are univalent, we must first examine their dilatation. As we discover in this section, the dilatation of a polygonal map is always in the form of a Blaschke product.

EXERCISE 5.66. Consider the function generated in Exercise 5.34 on page 342. Complete all of the following:

(a) Find the formulas for $h(z)$ and $g(z)$. Use the result of Exercise 5.42 for this computation.

(b) Find the derivatives $h'(z)$ and $g'(z)$, verifying that the derivatives simplify to

$$h'(z) = \frac{1}{(2\pi i)(1 - z^4)} \left((\frac{3}{2} - \frac{3}{2}i)z^2 - 3iz + (\frac{5}{2} + \frac{5}{2}i) \right)$$

and $g'(z) = \dfrac{1}{(2\pi i)(1 - z^4)} \left((\frac{5}{2} - \frac{5}{2}i)z^2 + 3iz + (\frac{3}{2} + \frac{3}{2}i) \right)$.

(c) Show that the zeros of $g'(z)$ are $z_1 \approx 0.9245 - 0.9245i$ and $z_2 \approx -0.3245 + 0.3245i$, and that the zeros of $h'(z)$ are $1/\ \overline{z_1}$ and $1/\ \overline{z_2}$. Thus we are able to write the dilatation as

$$\omega(z) = C \left(\frac{z_1 - z}{1 - \overline{z_1}z} \right) \left(\frac{z_2 - z}{1 - \overline{z_2}z} \right),$$

where C is some constant.

(d) What are $|z_1|$, $|z_2|$, and $|C|$? These will be helpful to recall as we look ahead to Theorem 5.81 in Section 5.8.

Try it out!

Motivated by the results of Exercise 5.66, we now examine functions that are of a particular form.

EXPLORATION 5.67. We examine the properties of functions of the form $B_{z_0}(z) = \dfrac{z_0 - z}{1 - \ \overline{z_0}z}$.

- Using *ComplexTool*, graph the image of the unit disk under the functions $B_{0.5}(z) = \frac{0.5-z}{1-0.5z}$, $B_{0.5\exp(i\pi/4)}(z) = \frac{0.5e^{i\pi/4}-z}{1-0.5e^{-i\pi/4}z}$, and $B_{0.5i}(z) = \frac{0.5i-z}{1+0.5iz}$. What is $B(0)$? What is the image of the unit disk under $B(z)$? Does $B(z)$ appear to be univalent?

- Now graph the image of the unit disk under the function $B_{2\exp(i\pi/4)}(z) = \frac{2e^{i\pi/4}-z}{1-2e^{-i\pi/4}z}$. What is $B(0)$? What is the image of the unit disk under $B(z)$? Does $B(z)$ appear to be univalent?

- If we consider $B_{z_0}(z)$, can we determine $B_{z_0}(0)$ in general? What effect does $\arg(z_0)$ have on the location of $B_{z_0}(0)$? What effect does $|z_0|$ have on the image of the unit disk?

- Now let's multiply the functions $B_{z_0}(z)$. First graph the image of the unit disk under $f_1(z) = (B_{0.5}(z))^2 = \left(\frac{0.5-z}{1-0.5z} \right)^2$. What is $f(0)$? What is the image of the unit disk under $f(z)$? Does $f(z)$ appear to be univalent? How does this compare to the function $f(z) = z^2$?

- Now graph the image of the unit disk under $f_2(z) = B_{0.5}(z)B_{0.5i}(z)$, $f_3(z) = B_{0.5}(z)B_{0.2i}(z)$, and $f_4(z) = B_{0.5\exp(i\pi/4)}(z)B_{0.2i}(z)$. What are $f_2(0)$, $f_3(0)$, and $f_4(0)$? How do $f_2(0)$, $f_3(0)$, and $f_4(0)$ relate to the values of z_0 in the functions $B_{z_0}(z)$?

Try it out!

DEFINITION 5.68. A Blaschke factor is $B_{z_0}(z) = \frac{z_0 - z}{1 - \overline{z_0} z}$, and a finite Blaschke product or order n is a product of n Blaschke factors, possibly multiplied by a constant ζ such that $|\zeta| = 1$:

$$\zeta \prod_{k=1}^{n} \frac{z_k - z}{1 - \overline{z_k} z}.$$

Note that the multiplication by ζ is simply a rotation.

REMARK 5.69. It is important to note that the Blashke product definition given above is a bit non-standard. The standard definition of Blaschke product, as given in Chapters 4 and 6, assumes that $|z_k| < 1$. Here we do not place that restriction on z_k for the purpose of simplifying our computations.

In this section, we use the result of Exercise 5.39 on page 344 to see that the dilatation of the harmonic polygonal functions to an n-gon is a Blaschke product of order at most $n - 2$. This result was proven by T. Sheil-Small in [**17**], and is also discussed in [**5**].

We use the notation $f_k(z)$ to denote the contribution to the function $f(z)$ that arises from applying the Poisson integral formula to the boundary correspondence for $t_{k-1} < t < t_k$. We then have $f(z) = \sum_{k=1}^{n} f_k(z)$. On the interval $t_{k-1} < t < t_k$, we observe that

$$f_k(z) = \frac{v_k}{2\pi}(t_k - t_{k-1}) + \frac{v_k}{\pi} \arg\left(\frac{1 - ze^{-it_k}}{1 - ze^{-it_{k-1}}} \right)$$

by Definition 5.37. We now consider the canonical decomposition of $f_k(z) = h_k(z) + \overline{g_k(z)}$. By Exercise 5.42, we have

$$
\begin{aligned}
h_k(z) &= \frac{v_k}{2\pi}(t_k - t_{k-1}) + \frac{v_k}{2\pi i} \log\left(\frac{1 - ze^{-it_k}}{1 - ze^{-it_{k-1}}} \right) \\
&= \frac{v_k}{2\pi}(t_k - t_{k-1}) + \frac{v_k}{2\pi i} \left(\log(1 - ze^{-it_k}) - \log(1 - ze^{-it_{k-1}}) \right)
\end{aligned}
$$

and

$$
\begin{aligned}
g_k(z) &= \frac{\overline{v_k}}{2\pi i} \log\left(\frac{1 - ze^{-it_k}}{1 - ze^{-it_{k-1}}} \right) \\
&= \frac{\overline{v_k}}{2\pi i} \left(\log(1 - ze^{-it_k}) - \log(1 - ze^{-it_{k-1}}) \right).
\end{aligned}
$$

The computations that follow will give some rigor to our intuition: since h and g are sums of logarithms, their derivatives are sums of terms that have $1 - ze^{-it_k}$ in the denominators. We will combine these factors, and hope that we can see each derivative as a Blaschke product. Now, when we look at $h(z) = \sum_{k=1}^{n} h_k(z)$ and take derivatives,

we see that:

$$
(79) \qquad h'(z) \;=\; \sum_{k=1}^{n} \frac{v_k}{2\pi i}\left(\frac{-e^{-it_k}}{1 - ze^{-it_k}} - \frac{-e^{-it_{k-1}}}{1 - ze^{-it_{k-1}}}\right)
$$

$$
(80) \qquad =\; \sum_{k=1}^{n} \frac{v_k}{2\pi i}\left(\frac{1}{z - e^{it_k}} - \frac{1}{z - e^{it_{k-1}}}\right).
$$

The function $g'(z)$ is identical except for having $\overline{v_k}$ instead of v_k:

$$
(81) \qquad g'(z) \;=\; \sum_{k=1}^{n} \frac{\overline{v_k}}{2\pi i}\left(\frac{-e^{-it_k}}{1 - ze^{-it_k}} - \frac{-e^{-it_{k-1}}}{1 - ze^{-it_{k-1}}}\right)
$$

$$
(82) \qquad =\; \sum_{k=1}^{n} \frac{\overline{v_k}}{2\pi i}\left(\frac{1}{z - e^{it_k}} - \frac{1}{z - e^{it_{k-1}}}\right).
$$

Combining like factors gives us a more compact form, with

$$
(83) \qquad h'(z) = \frac{1}{2\pi i}\sum_{k=1}^{n} \frac{v_k - v_{k+1}}{z - e^{it_k}} \qquad \text{and} \qquad g'(z) = \frac{1}{2\pi i}\sum_{k=1}^{n} \frac{\overline{v_k} - \overline{v_{k+1}}}{z - e^{it_k}}.
$$

It will be useful for the upcoming discussion to note that $\sum_{k=1}^{n}(v_k - v_{k+1}) = 0$, since $v_{n+1} = v_1$.

EXERCISE 5.70. Prove that $\sum_{k=1}^{n}(v_k - v_{k+1}) = 0$, since $v_{n+1} = v_1$. Interpret this result geometrically.
Try it out!

We rely heavily upon the observation that

$$
(84) \qquad \overline{h'(1/\,\overline{z})} = z^2 g'(z) \qquad \text{or} \quad \overline{g'(1/\,\overline{z})} = z^2 h'(z).
$$

363

Equation 84 arises from the following computation:

$$
\begin{aligned}
\overline{h'(1/\,\overline{z})} - z^2 g'(z) &= \frac{-1}{2\pi i} \sum_{k=1}^{n} \frac{\overline{v}_k - \overline{v}_{k+1}}{1/z - e^{-it_k}} - \frac{z^2}{2\pi i} \sum_{k=1}^{n} \frac{\overline{v}_k - \overline{v}_{k+1}}{z - e^{it_k}} \\
&= \frac{-z}{2\pi i} \sum_{k=1}^{n} \frac{e^{it_k}(\,\overline{v}_k - \overline{v}_{k+1})}{e^{it_k} - z} - \frac{z^2}{2\pi i} \sum_{k=1}^{n} \frac{\overline{v}_k - \overline{v}_{k+1}}{z - e^{it_k}} \\
&= \frac{z}{2\pi i} \sum_{k=1}^{n} \frac{e^{it_k}(\,\overline{v}_k - \overline{v}_{k+1})}{z - e^{it_k}} - \frac{z^2}{2\pi i} \sum_{k=1}^{n} \frac{\overline{v}_k - \overline{v}_{k+1}}{z - e^{it_k}} \\
&= \frac{1}{2\pi i} \sum_{k=1}^{n} \frac{(\,\overline{v}_k - \overline{v}_{k+1})(ze^{it_k} - z^2)}{z - e^{it_k}} \\
&= \frac{1}{2\pi i} \sum_{k=1}^{n} \frac{(\,\overline{v}_k - \overline{v}_{k+1})(z)(e^{it_k} - z)}{z - e^{it_k}} \\
&= \frac{-1}{2\pi i} \sum_{k=1}^{n} (\,\overline{v}_k - \overline{v}_{k+1})(z) \\
&= 0.
\end{aligned}
$$

EXERCISE 5.71. We have just proven the first half of Equation 84. Using that result, prove the second part of Equation 84 with minimal calculation. **Try it out!**

EXERCISE 5.72. Interpret this result geometrically. That is, note that if we have a value $z_0 \in \mathbb{D}$ such that $g'(z_0) = 0$, then what do we know about the zeros of h'? How are the locations of the zeros of h' related to the locations of the zeros of g'? Completion of this exercise will give some intuition about the proof that is to come. **Try it out!**

EXERCISE 5.73. Show that $h'(z)$ and $g'(z)$ of Exercise 5.66 satisfy Equation 84. Relate this to the previous two exercises–do the conclusions of those exercises also hold true for this example? **Try it out!**

For simplicity of notation, let us consider the functions $h'(z)$ and $g'(z)$. We can already tell that if we got a common denominator for $h'(z)$ or $g'(z)$, that the denominator would be $\prod_{k=1}^{n}(z - e^{it_k})$, and we would guess that the ratio of the two would give us a product of rational functions. At this point, that is all we can tell–it is not obvious that this product should be a Blaschke product, although we may expect it to be from the explorations we did in Exercise 5.66. The remainder of this section will be devoted to determining that this is, indeed, a Blaschke product, as well as finding the order of that product.

EXPLORATION 5.74. Based on the results of Exercise 5.66, and upon other examples in this chapter, make a conjecture about the number of Blaschke factors that should be in the dilatation of a harmonic function from $\mathbb{D}$ to an n-gon. ***Try it out!***

Obtaining a common denominator for both h' and g', we can look at them as

$$h'(z) = \frac{P(z)}{S(z)} \quad \text{and} \quad g'(z) = \frac{Q(z)}{S(z)},$$

where $S(z) = \prod_{k=1}^{n}(z - e^{it_k})$. Now we need to consider what P and Q look like. Consider that by brute force, each term of the $P(z)$ looks like $(v_k - v_{k+1}) \prod_{j=1; j \neq k}^{n}(z - e^{it_j})$, or, put more simply, a polynomial of degree at most $n-1$. Let us consider the z^{n-1} term of $P(z)$. It is simply $v_k - v_{k+1}$ for each piece of the sum, so it must be $\sum_{k=1}^{n}(v_k - v_{k+1})$, which we already observed to be 0. Thus we have shown that $P(z)$ has degree at most $n-2$. The same argument works for $Q(z)$, since it has the same structure as $P(z)$ but with conjugates over the v_k.

We now turn our attention to the denominator, which is $S(z) = \prod_{k=1}^{n}(z - e^{it_k})$.

EXERCISE 5.75. Show that the following equation holds:

$$S(1/\overline{z}) = \left(\frac{1}{\overline{z}}\right)^n (-1)^n \overline{S(z)} \prod_{k=1}^{n} e^{it_k}.$$

Try it out!

Put another way, we could write

(85) $$\overline{S(1/\overline{z})} = \left(\frac{1}{z}\right)^n (-1)^n S(z) \prod_{k=1}^{n} e^{-it_k}.$$

EXERCISE 5.76. Show that Equation 85 holds for the denominator of the derivatives in Exercise 5.66, $2\pi i(1 - z^4)$. ***Try it out!***

Now we can combine Equations 84 and 85 to get a relationship between $P(z)$ and $Q(z)$. Directly substituting into $\overline{h'(1/\overline{z})} = z^2 g'(z)$, we see that

$$\overline{h'(1/\bar{z})} = z^2 g'(z)$$

$$\overline{\frac{P(1/\bar{z})}{S(1/\bar{z})}} = z^2 \frac{Q(z)}{S(z)}$$

$$\frac{\overline{P(1/\bar{z})}}{\left(\frac{1}{z}\right)^n (-1)^n S(z) \prod_{k=1}^{n} e^{-it_k}} = z^2 \frac{Q(z)}{S(z)}.$$

This leads to the relationship

(86)
$$z^{n-2} \, \overline{P(1/\bar{z})} = (-1)^n Q(z) \prod_{k=1}^{n} e^{-it_k}.$$

EXERCISE 5.77. Using the result above, show that

(87)
$$z^{n-2} \, \overline{Q(1/\bar{z})} = (-1)^n P(z) \prod_{k=1}^{n} e^{-it_k}.$$

Try it out!

Since the function f is orientation-preserving, we know that $h'(z) \neq 0$ in $\mathbb{D}$. This implies that $P(z) \neq 0$ in $\mathbb{D}$. In particular, $P(0) \neq 0$. Substituting 0 into equation 87, we find that the left hand side must not be zero, which forces the degree of Q to be at least $n-2$. However, we had previously determined that the degree of Q must be at most $n-2$. Thus the degree of Q is $n-2$. Similarly, the degree of P is also $n-2$. Since the degree of Q is $n-2$, let us write

$$Q(z) = z^m \prod_{k=1}^{n-m-2} (z - z_k)$$

to show that Q may have m zeros at the origin and $n-m-2$ zeros elsewhere (note that the z_k need not be distinct). Now using Equation 87, we can write

$$z^{n-2} \, \overline{\left(\frac{1}{\bar{z}}\right)^m \prod_{k=1}^{n-m-2} \left(\frac{1}{\bar{z}} - z_k\right)} = (-1)^n P(z) \prod_{k=1}^{n} e^{-it_k}.$$

The left hand side of the above equation may be rewritten as

$$z^{n-m-2} \frac{1}{z^{n-m-2}} \prod_{k=1}^{n-m-2} \left(1 - z \, \overline{z_k}\right).$$

At this point we can see that since the zeros of Q are z_k, the zeros of P are the zeros of $\prod_{k=1}^{n-m-2} (1 - z \, \overline{z_k})$, which are precisely $1/\, \overline{z_k}$. Now we are able to see what the Blaschke product is.

EXERCISE 5.78. Find the relationship between the number of zeros of Q and the number of zeros of P. In particular, if Q has degree $n - 2$ with m zeros at the origin and $n - m - 2$ zeros away from the origin, then how many of the zeros of P are at the origin? How many of the zeros of P are away from the origin? **Try it out!**

We now summarize the results of our work in this section (as originally proved by T. Sheil-Small, [**17**] Theorem 1; see also [**5**]).

THEOREM 5.79. Let f be the harmonic extension of the step function $\hat{f}(e^{it})$ as given in Definition 5.37. Then

$$g'(z) = \frac{Q(z)}{S(z)} \qquad \text{and} \qquad h'(z) = \frac{P(z)}{S(z)},$$

where $Q(z), P(z)$, and $S(z)$ are defined as above, and P and Q are polynomials of degree at most $n - 2$. Furthermore, their ratio $\omega(z)$ satisfies $|\omega(z)| = 1$ when $|z| = 1$, so takes the form of a Blaschke product of degree at most $n - 2$.

5.8. An Important Univalence Theorem

In this section, we examine a theorem of Sheil-Small that tells when the harmonic function in Definition 5.37 is univalent. In particular, the location of the zeros of the analytic dilatation $\omega(z) = \frac{g'(z)}{h'(z)}$ are sufficient to tell when the harmonic function is univalent.

EXPLORATION 5.80. Open up the *StarTool* applet. Check the box in front of `Show roots of` $\omega(z)$. You will see extra dots appear in the right-hand pane (the range of the function), as well as a unit circle for reference. These dots denote the locations of the zeros of the dilatation $\omega(z)$. Now experiment with the values of p and r to see if there is a relationship between the roots of $\omega(z)$ and whether the resulting star is univalent. Do this for various values of n to see if your result seems to hold. Does your conjecture agree with the examination of a function that maps $\mathbb{D}$ to a different non-convex polygon, as in Exercise 5.66? **Try it out!**

The theorem below was proven by Sheil-Small, and is Theorem 11.6.6 of [**18**].

THEOREM 5.81. Let f be a harmonic function of the form in Definition 5.37. Here the function f is the harmonic extension of a piecewise constant boundary function with values on the m vertices of a polygonal region Ω, so that, by Theorem 5.79, the dilatation of f is a Blaschke product with at most $m - 2$ factors. Then f is univalent in $\mathbb{D}$ if and only if all zeros of ω lie in $\mathbb{D}$. In this case, f is a harmonic mapping of $\mathbb{D}$ **onto** Ω.

PROOF. First, suppose that f is univalent in $\mathbb{D}$. If a Blaschke factor is defined as $\varphi_{z_0}(z) = \frac{z_0 - z}{1 - \overline{z_0} z}$, with the constant z_0 not having modulus 1, then we notice that since the dilatation ω is a product of a finite number of Blaschke factors, $\omega(z) \neq 0$ on the unit circle. This is because the zero of the Blaschke factor is z_0, and if $|z_0| = 1$, we get

367

that $\varphi_{z_0}(z) = z_0$ for all z. If ω has a zero at some point z_0 outside of $\overline{\mathbb{D}}$, then it has a pole at $1/\overline{z_0} \in \mathbb{D}$. If it also has zeros in $\mathbb{D}$, then there are points in $\mathbb{D}$ where $|\omega(z)| < 1$ and other points where $|\omega(z)| > 1$. This implies that the Jacobian of f changes sign in $\mathbb{D}$, which would force the Jacobian to equal 0 at some point in $\mathbb{D}$, contradicting Lewy's theorem, which says that the Jacobian is non-zero since f is locally univalent. Thus there are only two possibilities for a univalent f: Either all of the zeros of $\omega(z)$ lie in $\mathbb{D}$, or all lie outside $\overline{\mathbb{D}}$. But if the zeros of ω lie outside of $\overline{\mathbb{D}}$, then $|\omega(z)| > 1$ in $\mathbb{D}$ and f has negative Jacobian, contradicting its construction as a sense-preserving boundary function. Therefore, all of the zeros of ω must lie in $\mathbb{D}$.

Conversely, assume all of the zeros of ω lie within $\mathbb{D}$. By the mapping properties of Blaschke products, $|\omega(z)| < 1$ in $\mathbb{D}$. We use the argument principle to show that f is univalent in $\mathbb{D}$ and maps $\mathbb{D}$ onto Ω. Choose an arbitrary point $w_0 \in \Omega$. Let C_ϵ be the path in $\overline{\mathbb{D}}$ consisting of arcs of the unit circle along with small circular arcs of radius ϵ about the points b_k (the points b_k are the arc endpoints in the domain disk), as shown in Figure 5.17.
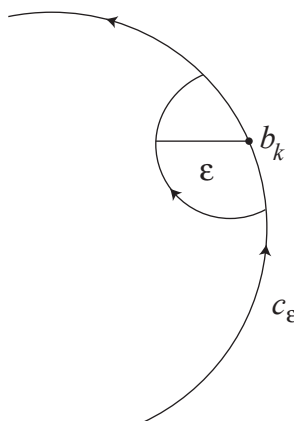
FIGURE 5.17. Tiny circles around the b_k

If ϵ is sufficiently small, the image of C_ϵ will not go through w_0, and will have winding number $+1$ around w_0. Since $|\omega(z)| < 1$ inside C_ϵ, it follows from the argument principle for harmonic functions that $f(z) - w_0$ has one simple zero inside C_ϵ (or, put another way, $f(z) = w_0$ has exactly one solution for $z \in \mathbb{D}$). Thus $\Omega \subset f(\mathbb{D})$. Now do a similar construction with $w_0 \notin \Omega$ to show that $w_0 \notin f(\mathbb{D})$. Thus f maps $\mathbb{D}$ univalently onto Ω. $\qquad\square$

To apply this theorem to the star mappings of Section 5.6, we study the dilatation of the star mappings in detail.

5.9. The Dilatation for Star Mappings

In this section, we will use Definition 5.63 of Section 5.6 as the starting point. We will build upon the basic formula for the functions $h'(z)$ and $g'(z)$, and then will simplify the dilatation $\omega(z)$ as a Blaschke product. By doing so, we completely determine which star functions are univalent.

Using Equations (79) and (81) of Section 5.7, along with Definition 5.63, we find the following equations for $h'(z)$ and $g'(z)$ for the star functions:

$$
\begin{aligned}
h'(z) \;=\;& \frac{r\alpha^0}{2\pi i}\left(\frac{1}{z-\alpha^0 e^{ip\pi/n}} - \frac{1}{z-\alpha^0 e^{-ip\pi/n}}\right) \\
&+ \frac{\alpha}{2\pi i}\left(\frac{1}{z-\alpha^2 e^{-ip\pi/n}} - \frac{1}{z-\alpha^0 e^{ip\pi/n}}\right) \\
&+ \frac{r\alpha^2}{2\pi i}\left(\frac{1}{z-\alpha^2 e^{ip\pi/n}} - \frac{1}{z-\alpha^2 e^{-ip\pi/n}}\right) + \ldots
\end{aligned}
$$

$$
\begin{aligned}
(88)\qquad =\;& \frac{r}{2\pi i}\sum_{k=0}^{n-1}\alpha^{2k}\left(\frac{1}{z-\alpha^{2k}e^{ip\pi/n}} - \frac{1}{z-\alpha^{2k}e^{-ip\pi/n}}\right) \\
&+ \frac{1}{2\pi i}\sum_{k=0}^{n-1}\alpha^{2k+1}\left(\frac{1}{z-\alpha^{2k+2}e^{-ip\pi/n}} - \frac{1}{z-\alpha^{2k}e^{ip\pi/n}}\right)
\end{aligned}
$$

and

$$
\begin{aligned}
(89)\qquad g'(z) \;=\;& \frac{r}{2\pi i}\sum_{k=0}^{n-1}\overline{\alpha^{2k}}\left(\frac{1}{z-\alpha^{2k}e^{ip\pi/n}} - \frac{1}{z-\alpha^{2k}e^{-ip\pi/n}}\right) \\
&+ \frac{1}{2\pi i}\sum_{k=0}^{n-1}\overline{\alpha^{2k+1}}\left(\frac{1}{z-\alpha^{2k+2}e^{-ip\pi/n}} - \frac{1}{z-\alpha^{2k}e^{ip\pi/n}}\right).
\end{aligned}
$$

Our goal is to express $\omega(z) = g'(z)/h'(z)$ as the ratio of Blaschke products guaranteed by Sheil-Small's work. To that end, we first establish a few general algebraic identities involving sums of the quantities of the type found in the expansions of $h'(z)$ and $g'(z)$ above.

It is a basic complex identity that if ζ is a primitive mth root of unity, then

$$
(90)\qquad \prod_{k=1}^{m}(z-\zeta^k) = z^m - 1.
$$

EXERCISE 5.82. Prove Equation (90). Interpret this result geometrically. **Try it out!**

EXERCISE 5.83. Again, let ζ be a primitive mth root of unity. Show that

$$(91) \qquad \prod_{k=0}^{m-1} (z - \zeta^k a) = z^m - a^m.$$

Hint: Replace z with z/a in Equation (90). ***Try it out!***

Now we work to answer the hard question: How do we add together all of the sums in equations (88) and (89), given that their numerators are not simply the constant 1? As an intermediate step toward achieving this, we establish the following identity.

LEMMA 5.84. If ζ is a primitive mth root of unity, then

$$(92) \qquad \sum_{k=0}^{m-1} \frac{\zeta^k}{z - \zeta^k a} = \frac{m a^{m-1}}{z^m - a^m}.$$

EXERCISE 5.85. Prove the lemma, using the following steps in the partial fraction decomposition.

(1) Recall Equation (91) and note how it fits in with this formula.
(2) Note that since we have n distinct linear factors in the denominator, we can expect to find that

$$\frac{m a^{m-1}}{z^m - a^m} = \sum_{k=0}^{m-1} \frac{a_k}{z - \zeta^k a}.$$

(3) We will find an arbitrary a_{k_0}. By setting $z = \zeta^{k_0} a$, establish that

$$a_{k_0} = \frac{m a^{m-1}}{\prod_{k \neq k_0} (\zeta^{k_0} a - \zeta^k a)} = \frac{m}{\prod_{k \neq k_0} (\zeta^{k_0} - \zeta^k)}.$$

(4) Show that $\prod_{k \neq k_0} (\zeta^{k_0} - \zeta^k) = m \zeta^{-k_0}$. It will be helpful to remember that ζ^{k_0} is an mth root of unity, so $\zeta^{k_0 m} = 1$.
(5) Conclude that $a_{k_0} = \zeta^{k_0}$.
(6) Equation (92) should follow.

Try it out!

We now recall the result of Exercise 5.64 on page 359 that

$$\alpha^{2k+2} e^{-ip\pi/n} = \alpha^{2k+1} e^{i(1-p)\pi/n} \qquad \text{and} \qquad \alpha^{2k} e^{ip\pi/n} = \alpha^{2k+1} e^{-i(1-p)\pi/n}.$$

Combining with the earlier work, we have

$$\frac{1}{2\pi i}\sum_{k=0}^{n-1}\frac{\alpha^{2k+1}}{z-\alpha^{2k+2}e^{-ip\pi/n}} = \frac{-1}{2\pi i}\left(\frac{ne^{i(1-p)(n-1)\pi/n}}{z^n+e^{i(1-p)\pi}}\right) \qquad \text{and}$$

$$-\frac{1}{2\pi i}\sum_{k=0}^{n-1}\frac{\alpha^{2k+1}}{z-\alpha^{2k}e^{ip\pi/n}} = \frac{1}{2\pi i}\left(\frac{ne^{-i(1-p)(n-1)\pi/n}}{z^n+e^{-i(1-p)\pi}}\right).$$

EXERCISE 5.86. Prove that

$$-\frac{1}{2\pi i}\sum_{k=0}^{n-1}\frac{\alpha^{2k+1}}{z-\alpha^{2k}e^{ip\pi/n}} = \frac{1}{2\pi i}\left(\frac{ne^{-i(1-p)(n-1)\pi/n}}{z^n+e^{-i(1-p)\pi}}\right).$$

Try it out!

Combining all of these together, we see that

$$h'(z) = \frac{n}{2\pi i}\left(\frac{re^{i(\frac{n-1}{n})p\pi}}{z^n-e^{ip\pi}} + \frac{-re^{-i(\frac{n-1}{n})p\pi}}{z^n-e^{-ip\pi}}\right.$$

$$\left. + \frac{-e^{i(\frac{n-1}{n})(1-p)\pi}}{z^n+e^{i(1-p)\pi}} + \frac{e^{-i(\frac{n-1}{n})(1-p)\pi}}{z^n+e^{-i(1-p)\pi}}\right). \tag{93}$$

We need to keep our goal in mind: We know from Theorem 5.79 that $\frac{g'(z)}{h'(z)}$ can be written as a Blaschke product. To do this, we will have to find a common denominator and combine the four terms of $h'(z)$ to see the quotient. As an initial step, we find that we can write the common denominator more simply than it first appears.

EXERCISE 5.87. Prove that $(z^n-e^{ip\pi})(z^n-e^{-ip\pi}) = (z^n+e^{i(1-p)\pi})(z^n+e^{-i(1-p)\pi})$. We will call this product $S_n(z)$. ***Try it out!***

EXERCISE 5.88. Using basic algebra (finding a common denominator, simplifying, and using properties of $z + \bar{z}$), prove that

$$h'(z) = \frac{n}{\pi S_n(z)}\left(z^n\left(r\sin\left(\frac{(n-1)p\pi}{n}\right) - \sin\left(\frac{(n-1)(1-p)\pi}{n}\right)\right)\right. \tag{94}$$

$$\left. + r\sin\left(\frac{p\pi}{n}\right) + \sin\left(\frac{(1-p)\pi}{n}\right)\right).$$

Try it out!

Through methods similar to those of the simplification of $h'(z)$, we can also prove that

$$g'(z) = \frac{nz^{n-2}}{\pi S_n(z)}\left(z^n\left(r\sin\left(\frac{p\pi}{n}\right) + \sin\left(\frac{(n-1)(1-p)\pi}{n}\right)\right)\right. \tag{95}$$

$$\left. + r\sin\left(\frac{(n-1)p\pi}{n}\right) - \sin\left(\frac{(n-1)(1-p)\pi}{n}\right)\right).$$

With this simplified form of $g'(z)$, we use

$$(96) \qquad c = \frac{\sin \frac{(n-1)(1-p)\pi}{n} - r \sin \frac{(n-1)p\pi}{n}}{r \sin \frac{p\pi}{n} + \sin \frac{(1-p)\pi}{n}}$$

to write

$$(97) \qquad g'(z) = \frac{nz^{n-2}}{\pi} \left(r \sin \left(\frac{p\pi}{n} \right) + \sin \left(\frac{(1-p)\pi}{n} \right) \right) \frac{z^n - c}{S_n(z)}.$$

Now we are ready to pull together the result of Theorem 5.81 and the dilatation that we just simplified. When the zeros of this dilatation are within the unit disk, then the harmonic function $f = h + \bar{g}$ that defines the star is univalent. By a straightforward computation, we find that the dilatation of f is

$$(98) \qquad \omega(z) = \frac{z^{n-2}(z^n - c)}{1 - z^n c}.$$

EXERCISE 5.89. Look again at Theorem 5.81 and verify that it does, indeed, hold for the star function. ***Try it out!***

EXPLORATION 5.90. Notice that f is univalent when $|c| < 1$. Using that observation, do the following:

- Use the *StarTool* applet to explore graphically what relationship there is between n, p, and c.
- For a fixed n, find the range of p-values that make $|c| < 1$.
- For a fixed p, find the range of n-values that make $|c| < 1$.

Try it out!

LARGE PROJECT 5.91. If you move just one vertex of the star, do the same results hold for the relationship between n and p? (For example, take the vertex at r, and move it to $r + \epsilon$ or $r - \epsilon$. Is the star still univalent?)

EXERCISE 5.92. For a given n, consider the formula for c in Equation 96 to be a function of p alone. Prove that any star configuration is possible; that is, prove that for any value of r, a value of p can be found to make $|c| < 1$. What ranges of p makes this happen? Conversely, prove that for all values of $r < \cos(\pi/n)$ or $r > \sec(\pi/n)$, a p can be chosen to make the function NOT univalent. Why is this not true for $\cos(\pi/n) \le r \le \sec(\pi/n)$? For more information, see Theorem 4 in [**7**]. ***Try it out!***

SMALL PROJECT 5.93. Refer to Chapter 2. For what values of c is the dilatation a perfect square? Find and describe the associated minimal surfaces. Can these surfaces be described as examples of other well-known surfaces? For more information on this project, see [**12**].

5.10. Open Questions

LARGE PROJECT 5.94. Can we map to any polygon univalently? The star setup takes full advantage of the symmetry. Once you lose that advantage, it is much harder to discover whether the zeros of the dilatation have modulus less than 1. This question is known as the Mapping Problem, proposed by T. Sheil-Small in [**17**].

SMALL PROJECT 5.95. Look at a function f that is not univalent. Now look at the set $S \subset \mathbb{D}$ of points on which the function f is univalent. First, how do you find that set? What is the shape of S? Is it starlike? Is it convex? Is it connected? Is it simply connected? Can $\mathbb{D}\backslash S$ be connected?

SMALL PROJECT 5.96. In this chapter, we discussed one way of proving that a harmonic function is univalent by looking at zeros of the analytic dilatation $\omega(z)$. In Chapter 4, there is another set of criteria for univalence, as demonstrated in Section 4.6. Connect these two avenues of investigation. For example, does one imply the other? How does the work with stars in this chapter generalize to the approach in Chapter 4? Are there results in this chapter that could not be found using the methods of Chapter 4?

Bibliography

1. Lars Ahlfors, *Complex analysis*, third ed., McGraw-Hill, New York, 1979.
2. W. N. Bailey, *Generalized hypergeometric series*, Cambridge Tracts in Mathematics and Mathematical Physics, No. 32, Stechert-Hafner, Inc., New York, 1964. MR MR0185155 (32 #2625)
3. Tobin A. Driscoll and Lloyd N. Trefethen, *Schwarz-christoffel mapping*, Cambridge University Press, Cambridge, 2002.
4. Kathy Driver and Peter Duren, *Harmonic shears of regular polygons by hypergeometric functions*, J. Math. Anal. Appl. **239** (1999), no. 1, 72–84. MR 2000k:30008
5. Peter Duren, *Harmonic mappings in the plane*, Cambridge Tracts in Mathematics, vol. 156, Cambridge University Press, Cambridge, 2004. MR 2 048 384
6. Peter Duren, Walter Hengartner, and Richard S. Laugesen, *The argument principle for harmonic functions*, Amer. Math. Monthly **103** (1996), no. 5, 411–415. MR 97f:30002
7. Peter Duren, Jane McDougall, and Lisbeth Schaubroeck, *Harmonic mappings onto stars*, J. Math. Anal. Appl. **307** (2005), no. 1, 312–320. MR MR2138992
8. W. Hengartner and G. Schober, *Harmonic mappings with given dilatation*, J. London Math. Soc. (2) **33** (1986), no. 3, 473–483. MR MR850963 (87j:30037)
9. Elgin Johnston, *A "counterexample" for the Schwarz-Christoffel transform*, Amer. Math. Monthly **90** (1983), no. 10, 701–703. MR MR723944 (85f:30009)
10. Serge Lang, *Complex analysis*, fourth ed., Graduate Texts in Mathematics, vol. 103, Springer-Verlag, New York, 1999. MR MR1659317 (99i:30001)
11. H. Lewy, *On the non-vanishing of the jacobian in certain one-to-one mappings*, Bull. Amer. Math. Soc. **42** (1936), no. 1, 689–692.
12. Jane McDougall and Lisbeth Schaubroeck, *Minimal surfaces over stars*, J. Math. Anal. Appl. **340** (2008), no. 1, 721–738. MR MR2376192
13. Zeev Nehari, *Conformal mapping*, Dover Publications Inc., New York, 1975, Reprinting of the 1952 edition. MR MR0377031 (51 #13206)
14. Earl D. Rainville, *Special functions*, The Macmillan Co., New York, 1960. MR MR0107725 (21 #6447)
15. E.B. Saff and A.D. Snider, *Fundamentals of complex analysis for mathematics, science, and engineering, 2nd ed.*, Prentice Hall, Upper Saddle River, NJ, 1993.
16. Roland Schinzinger and Patricio A. A. Laura, *Conformal mapping: Methods and applications*, Dover Publications Inc., New York, 2003, Reprinting of the 1991 edition.
17. T. Sheil-Small, *On the Fourier series of a step function*, Michigan Math. J. **36** (1989), no. 3, 459–475. MR 91b:30002
18. ———, *Complex polynomials*, Cambridge Studies in Advanced Mathematics, vol. 75, Cambridge University Press, Cambridge, 2002. MR MR1962935 (2004b:30001)
19. Richard A. Silverman, *Complex analysis with applications*, Dover Publications Inc., New York, 1984, Reprint of the 1974 original. MR MR781933 (86c:30004)
20. Dennis G. Zill and Patrick D. Shanahan, *A first course in complex analysis with applications*, Jones and Bartlett, Sudbury, MA, 2003.

CHAPTER 6

Circle Packing

KEN STEPHENSON (text and software)

Complex analysis is in many ways the ultimate in "continuous" mathematics. It presents you with a *smooth* world: continuous variables, infinitely differentiable functions, smooth surfaces, plus a complex arithmetic with its power series, line integrals, and all sorts of handy formulae. In this chapter, however, we will develop a quite different view of the topic. Here the geometry behind analytic functions moves strongly to the fore as we see how one might "discretize" complex analysis.

An analogy to keep in mind is that of a mountain stream. In normal experience we treat the stream as a continuous medium and use continuous variables and functions to understand it — pressure, velocity, vorticity — modeled, perhaps, by Navier-Stokes and other PDE's. Yet we know that the stream is in fact a collection of individual water molecules, a "discrete" medium, if you will. The macro behavior, the waves, eddies, and currents, emerge from the myriad local interactions among the discrete water molecules. So one might ask: *what purely local rules of interaction among these molecules could lead to the observed global behavior?*

In this chapter, we address this issue for continuous analytic functions. Circles will be our molecules, and "packing" conditions will provide the local rules for their interaction. Circle packings, configurations of myriad individual circles, each interacting only with its neighbors, will manifest macro behavior that we will recognize as a version of analyticity. Later in the chapter we will also see that our discrete objects converge under appropriate refinement to the familiar continuous counterparts. In effect, then, circle packing provides a *quantum complex analysis*, one that is classical in the limit.

This is a topic ideal for visualization and experimentation — in fact, it is accessible only through the computer. You will therefore find yourself tethered to the software more closely than in the earlier chapters, so don't bypass those *Try it out!* signs. Persevere and you may develop a new appreciation for complex analysis. We will jump right in with a quick experiment from the genesis of the topic in Ⓢ DiscreteRMT Ⓢ. *Try it out!* .

6.1. First Impressions

Circle packings were introduced to analysis by Bill Thurston in a talk in 1985. The context was the conformal mapping of bounded, simply connected, plane regions Ω

onto the unit disc $\mathbb{D}$. The central result is the famous Riemann Mapping Theorem stated in the Appendix, page 417. (Alternate statements are given in Theorem 3.33, page 210, and Theorem 4.16, page 250). For us, the preferred statement is this: *Given Ω and designate points w_0, w_1 in Ω, there exists a unique one-to-one analytic mapping F from $\mathbb{D}$ onto Ω so that $F(0) = w_0$ and so that w_1 lies on the image of the positive real axis.* Figure 6.1 therefore illustrates the **Discrete** Riemann Mapping Theorem.
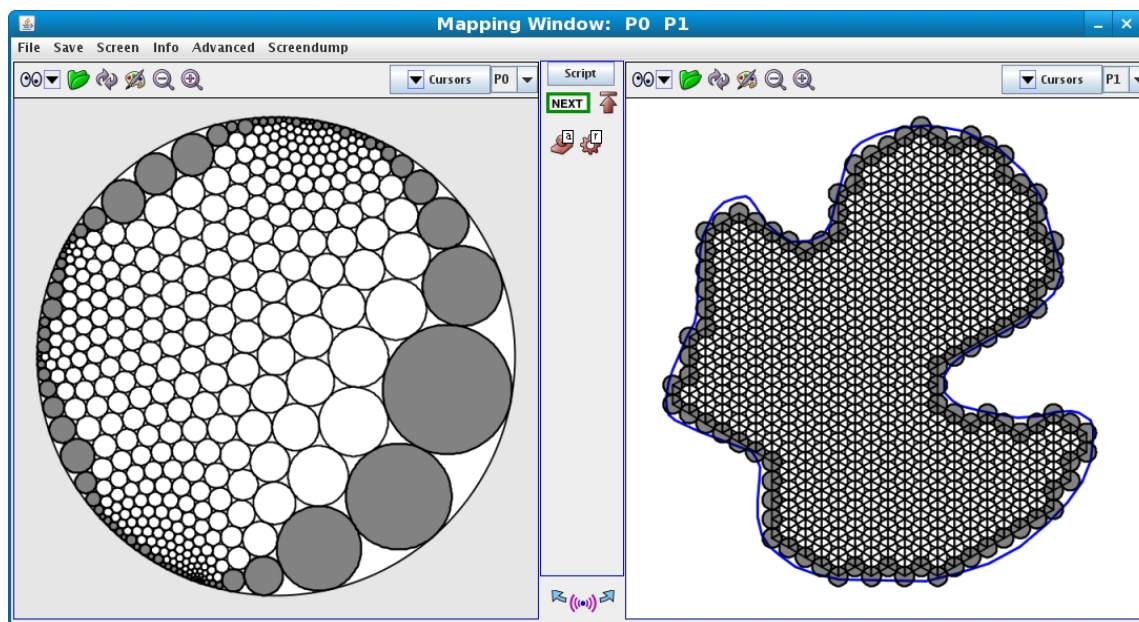
FIGURE 6.1. A conformal map display, domain (left) and range (right)

Figure 6.1 shows the mapping window as displayed by our experimental testbench, the software package *CirclePack*. Typical of our practice, the domain circle packing will be on the left in such side-by-side displays, and the range packing will be on the right. In this instance, the domain is $\mathbb{D}$ and the range is the given region Ω. The "function" itself can be interpreted in various ways. Each circle on the left has a corresponding circle on the right, so one can map circles to circles. Alternately, as we will see later, one can map the triangles formed by the circles on the left to those formed by the corresponding circles on the right. In any case, the two packings provide, in essence, a map from $\mathbb{D}$ onto Ω, a "discrete conformal map". We will see the details shortly, but assuming you have run ⓈDiscreteRMTⓈ and have this example before you on the computer, try clicking the left mouse button on some circle in one packing: both it and the corresponding circle of the other packing should be highlighted (likewise, for faces, using the middle mouse button). ***Try it out!***

About *CirclePack*: A few words are in order about *CirclePack* before we go on. Figure 6.2 is a screenshot showing the context for these mapping images, with various frames and panels for running the software. Unlike most of this book's software, *CirclePack* is a Java "application" rather than a Java "applet": it requires access to the file system for file manipulations, it is larger, more comprehensive (and reconfigurable), and it avails itself of shared C/C++ libraries for computationally intense experiments. It also has an important scripting feature: "script" files with prepared sequences of commands are provided with the book to guide your experiments. These files have file extension ".xmd" and are indicated in the text by Ⓢ Name Ⓢ. To run an experiment, then, you start *CirclePack*, from within it you load Name.xmd, then you click through the prepared commands.
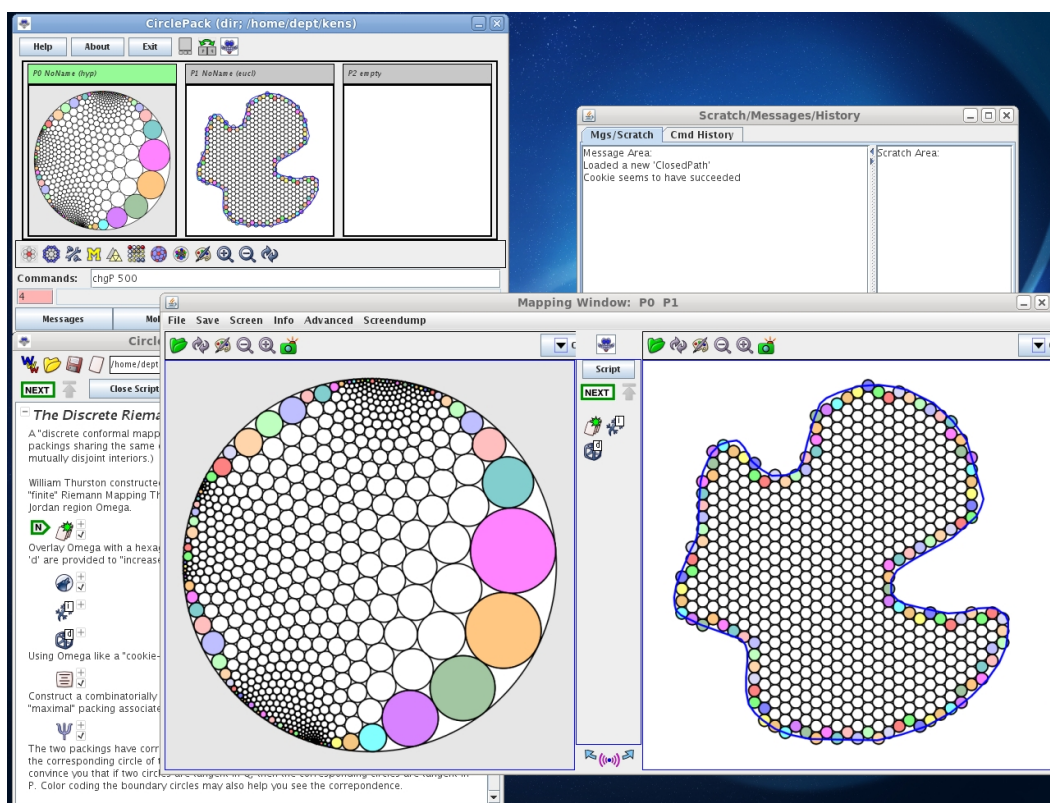
FIGURE 6.2. *CirclePack* in "domain/range" window mode

To recreate Figure 6.1, run the script Ⓢ DiscreteRMT Ⓢ. This is a very structured experiment, so move *CirclePack* through the commands by successively pressing the "NEXT" button (or press "enter" when the mouse is in the canvas) until the experiment is complete. ***Try it out!*** (If things become jumbled in *CirclePack*— and they sometimes will! — it may be that you have outrun the computations: just click the "up"

icon to restart at the beginning of the script again, or if things have really gone bad, exit and restart *CirclePack*.)

Even in the very passive experimental mode of Ⓢ`DiscreteRMT`Ⓢ you can interact with the images — moving, focusing, clicking circles and faces, and so forth, as we describe later. However, it is hoped that as your experience and curiosity grow, you move through the following additional stages to gain more experimental independence:

(1) The "`Open Script`" button will open the *CirclePack* script window. Here the commands are laid out linearly and typically are accompanied by text explanations. If you didn't understand what the experiment was purporting to show, try reading along as you execute the commands. Figure 6.3 illustrates the script window, with words, icons that encapsulate the commands, included files, etc.

FIGURE 6.3. A typical "script" for *CirclePack*

(2) In the Script you can open a command icon (by clicking the small "+" sign) to see the string of commands that it actually executes. (E.g., "`infile_path gamma.g;cookie;disp -w -c -cf b`" tells *CirclePack* to read the path "`gamma.g`" from the script's data section, cookie-cut the portion of the resident packing inside the path, then clear the canvas and display the circles of the resulting packing, with the boundary circles "filled".) Modify these commands and you can change the action in the experiment. In many of the scripts, hints are given about alterations you might try. In this first experiment, for example, you can change the radii of the circles in the initial hexagonal packing of the range. Try an adjustment and see what happens.

(3) If you want to dig deeper, click the "`Advanced`" button to bring up the *CirclePack* central control frame; this gives you access to full functionality. The program actually holds up to three packings, has panels with messages, history, and rudimentary error feedback, a "`Command:`" line where you can issue explicit strings of commands, and various "droppable" icons: you just drag one onto a packing window and a prepared string of commands will be applied to that packing.

(4) For the truly ambitious, the Java code for CirclePack is freely available under the GNU open source license. It includes a "`PackExtender`" class which allows anyone (with enough work) to create specialized data structures and commands having full access to the core functionality of *CirclePack*. Such PackExtenders will be used in later experiments.

Figure 6.4 provides an annotated screen shot of *CirclePack* with various of its screens and auxiliary panels. (Note that the screenshots are current as of publication, but *CirclePack* may change over time.)

A final note on the scripts: bare-bones scripts are quite easy to write for yourself, but following, reading, and modifying those of others can be a real challenge. The prepared scripts are built for ease of use (believe it or not), so the commands are often long and involved. For example, a simple "`disp -w -c`" will clear a screen and draw the circles. However, "`disp -w -cc20 i -ct5c180 b`" will clear the screen then draw the interior circles thicker and in blue, and the boundary circles filled with red. Don't let the complications bother you — go ahead and start your own scripts with simple commands and small steps. When things go wrong, adjust the commands. As you learn the *CirclePack* ropes, your scripts may become more involved, more sophisticated, but also, more effective.

We are going to move now to the basic definitions and notations we will use in our discrete analytic function theory. Although static images are helpful, there is nothing like running experiments, even in automatic mode, to bring the topic to life. The conformal mapping experiment we did above may help, and I certainly hope it
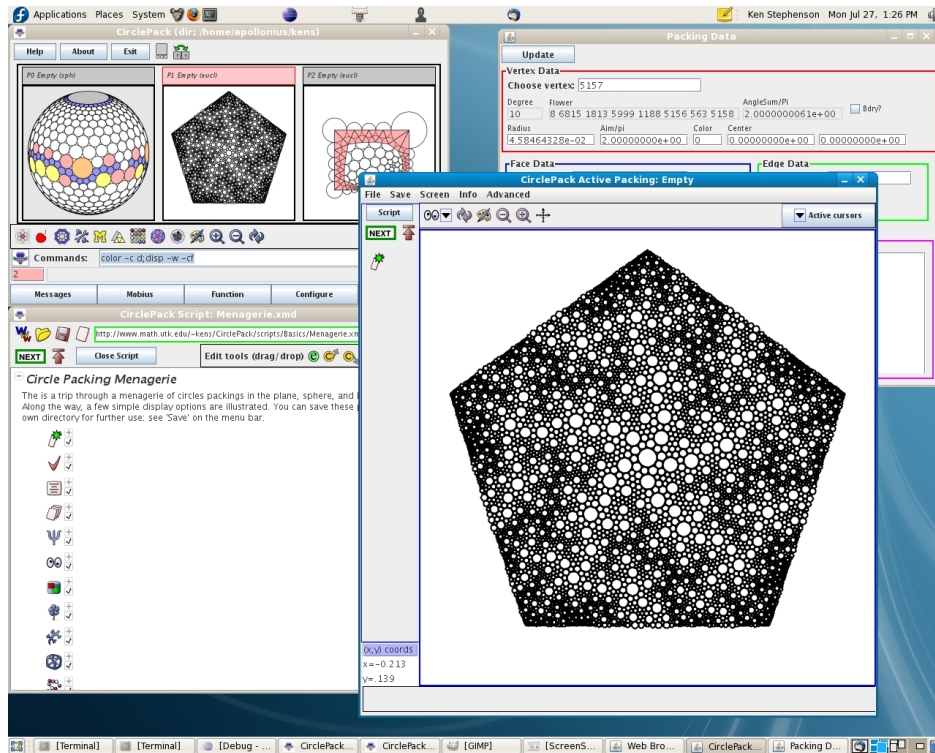
FIGURE 6.4. *CirclePack* in Mapping window mode

motivates the theory we will be investigating later. However, we have next a sequence of experiments tailored to display/explain the basics of the theory.

6.2. Basics of Circle Packing

DEFINITION 6.1. A **circle packing** is a configuration of circles with a prescribed pattern of tangencies.

Figure 6.5 illustrates a few circle packings from the "Menagerie" in Ⓢ Menagerie Ⓢ; run the script to see these and others. ***Try it out!*** Among the things you might observe as you tour our Menagerie: • The circles form mutually tangent **triples** and hence have triangular interstices. • The packings can live in the euclidean plane, $\mathbb{C}$, the hyperbolic plane, $\mathbb{D}$ (the unit disc in $\mathbb{C}$) or the Riemann sphere, $\mathbb{P}$, represented as the unit sphere at the origin in $\mathbb{R}^3$. • The "pattern" of tangencies shows up in the **carrier** of the packing, the geometric triangles formed by connecting the centers of tangent circles. • This pattern can be simply connected or multiply connected. • There are **interior** circles, those completely surrounded by their neighbors, and **boundary** circles, those on the edge of the configuration. • There are numerous manipulations, colors, and display options for displaying a circle packing and/or its carrier; these are invaluable in highlighting the structures you need to study.
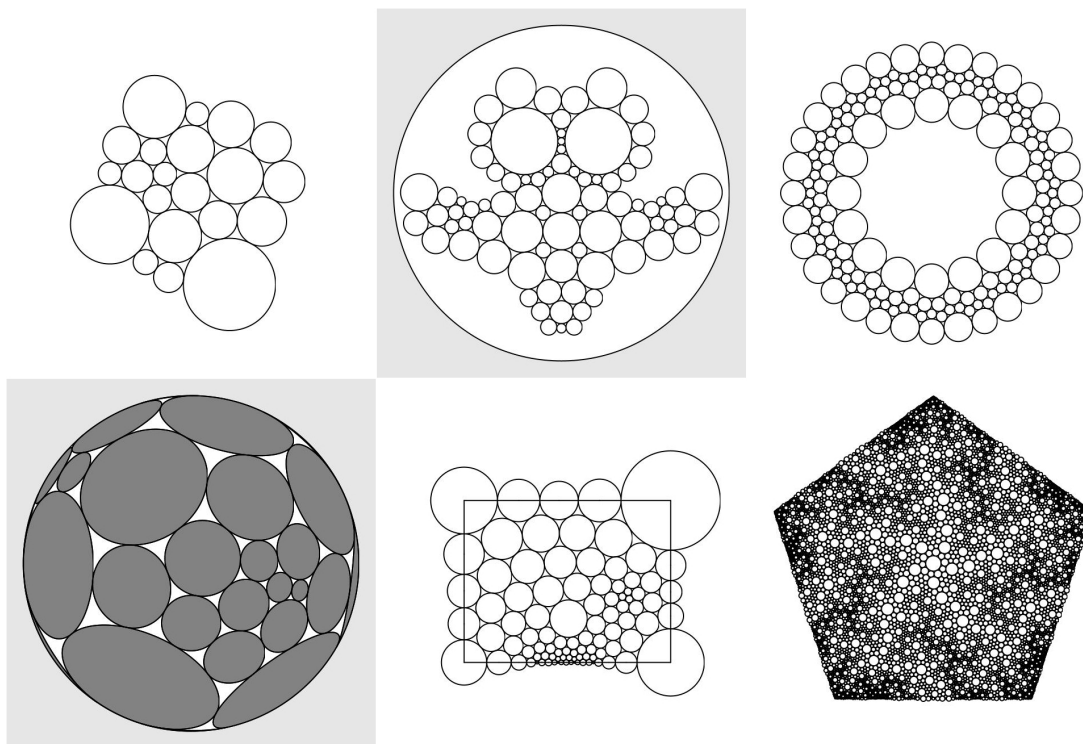
FIGURE 6.5. A small menagerie of circle packings.

We will have to make formal definitions shortly, but I would suggest that you first play around a little with the CirclePack Owl in Ⓢ OwlPlay Ⓢ. You might just run sequentially through the commands first, but then, by pressing the "Script" button, you can open the Script Window and follow the descriptions associated with the commands. There are suggestions along the way for free-form changes that you can make to see how the packings react. ***Try it out!***

Mathematics requires some formalities; annoying at times, they are essential if you want to understand what's going on and establish results that you and others can depend on. Let me lay out the formal objects involved in circle packing. Follow along in Ⓢ Objects Ⓢ, where these are illustrated explicitly in turn; read the descriptions in the script as you proceed. In other words, ***Try it out!***

• **Complex, K:** The "prescribed pattern" of tangencies for a circle packing is represented as an abstract triangulation K of a topological surface (technically, a simplicial 2-complex). This is a combinatorial object: K has **vertices** v, **edges** $\langle v, w \rangle$, and **faces** $\langle v, u, w \rangle$ formed by triples of vertices, but does not live in any geometric setting. It may be finite or infinite, may or may not have boundary, may be simply or multiply connected. We will always assume K is oriented. Namely,

in every face $\langle v, u, w \rangle$, the vertices are listed in positive (counterclockwise) order, and these orientations are globally consistent, meaning that if faces f, g share a directed edge $e = \langle v, w \rangle$, then e is positively oriented in one face and negatively oriented in the other.

Of course, I can't "show" you an abstract complex, but Ⓢ`Objects`Ⓢ begins with a random triangulation, one having no *a priori* connection with any circle packing — it is a simple graph you might have sketched by hand and then entered into some drawing program.

• **Circle Packing,** P: A circle packing for K is a configuration $P = \{c_v\}_{v \in K}$ of circles with these properties:

(1) There is a circle c_v associate with every vertex v of K.

(2) If vertices v and u are neighbors (this is denoted $v \sim u$ and means that $\langle v, u \rangle$ is an edge of K), then the circles c_v and c_u are (externally) tangent to one another.

(3) if $\langle v, u, w \rangle$ is an oriented face of K, then the circles c_v, c_u, c_w form a triple of mutually tangent circles which is positively (i.e., counterclockwise) ordered.

In *CirclePack*, vertices are indexed sequentially starting at 1; Ⓢ`Objects`Ⓢ highlights, successively: the circle c_5, a neighbor, c_9, a triple $\langle c_5, c_9, c_{27} \rangle$, and then the **flower** of c_5, namely, c_5 and the closed chain of neighboring **petal** circles.

The script goes on to highlight several other terms: **Interior** circles are those having closed flowers (their petals wrap completely around them), while the **boundary** circles are those on the edge, with flowers that don't close up. Likewise, the boundary/interior edges and faces are illustrated. P is said to be **univalent** if its circles have mutually disjoint interior.

• **Carrier, carr(**P**):** The carrier of P refers to the union of geometry triangles formed by connecting the centers of the triples of circles in P. In the script, the carrier is displayed as the "complex". P is **locally univalent** if for every interior vertex v, the faces in carr(P) corresponding to the flower of v have mutually disjoint interiors.

• **Label,** R: A label for K is a vector $R = \{r_v\}_{v \in K}$ of putative radii, one for each vertex of K. As you might imagine, it is the circle radii that are of central importance in circle packing. These are maintained in *CirclePack* and can be found using the "Info" button. (Note that labels depend on which of the three geometries you are working in.)

If you try to create a circle packing "by hand", say manipulating circle objects in some graphics program, you quickly learn how tightly choreographed their radii must be. The computational heart of circle packing involves the computation of

compatible radii; the neutral term "label" is used in place of "radii", because the algorithms typically start with values which do NOT fit together and then apply some iterative adjustments to approach a solution. (But see the §**??** for the spherical packing challenge.)

• **Angle Sums, θ_R:** Given a complex K and an associated label R, how might you determine if circles with those radii would actually fit together in the pattern prescribed by K? The key lies with angle sums. Consider an interior vertex v and let $\{v_1, v_2, \cdots, v_k\}$ be a list, in counterclockwise order, of its neighbors. Since the list is closed, it's convenient in computations to write $v_{k+1} = v_1$.

The faces containing v are $\langle v, v_j, v_{j+1} \rangle$, $j = 1, 2, \cdots, k$. Suppose r, r_j, r_{j+1} are the associated labels taken from R. For each such face we could place circles c_v, c_j, c_{j+1} in the appropriate geometry to form a triple. The triangle formed by their centers would have edge lengths which are the sums of radii, and we could appeal to the Law of Cosines to compute the angle in this triangle at v. In euclidean geometry, for instance, that angle would be

$$\alpha(r; r_j, r_{j+1}) = \arccos\Big[\frac{(r + r_j)^2 + (r + r_{j+1})^2 - (r_j + r_{j+1})^2}{2(r + r_j)(r + r_{j+1})}\Big].$$

The sum of these angles at v in all the faces containing v is known as the **angle sum**:

$$\theta_R(v) = \sum_{\langle v, v_j, v_{j+1} \rangle} \alpha(r; r_j, r_{j+1}z).$$

The neighbors of v will wrap precisely around v if and only if $\theta_R(v)$ is an integral multiple of 2π. This is illustrated concretely in ⓈObjectsⓈ, but note that one does not need to actually place any circles, as computations are done entirely with the labels and the combinatoric information from K.

• **Packing Label:** A label R for K is known as a packing label if $\theta_R(v)$ is an integral multiple of 2π for every interior vertex v. If P is a circle packing for K (in one of our geometries), then the list R of its radii will necessarily be a packing label. Conversely, one can show that if R is a packing label for K, then one can lay out circles, using the labels as radii, to form a circle packing P for K.

CirclePack is a platform for creating, manipulating, analyzing, and displaying circle packings. Its computational core lies in computing packing labels, and although the "packing" algorithms themselves are interesting and revealing mathematics, they are not our target here. We will instead treat *CirclePack* as a blackbox computational engine so that we can concentrate on the underlying geometry. (Recall that there are several levels of usage for scripts; the more skill you develop, the more you'll learn from experiments: (1) click "NEXT" and watch the action; (2) press "Open Script"

and follow the commands and associated descriptions in the script window; (3) modify current script commands; and (4) create new scripts from scratch.)

6.3. Circle Packing Manipulations

We base our introduction to circle packing on a sequence of scripts revealing both the objects of study and the principal manipulations of them available within *CirclePack*.

1. ⓈPatternsⓈ concentrates on combinatorics. Starting with an empty canvas, a "seed 7" command creates the simplest type of packing, namely a single **flower** — an interior circle surrounded by, in this case, seven petal circles. As you click through the script the first time, disregard the occasional repackings that are needed. The emphasis here is on building and adjusting the combinatorics. You will see how to add generations of new neighbors, how to add/delete individual boundary circles, how to cut open or zip up a string of boundary circles, how to "double" a packing across a boundary segment, how to adjoin two packings, how to cookie out one packing from a larger one, and how the process of "hex" refinement works. ***Try it out!***

2. ⓈGeometriesⓈ concerns geometries. *CirclePack* operates in any of the three standard ones: the euclidean plane, represented by the familiar complex plane $\mathbb{C}$; the hyperbolic plane, represented by the unit disc $\mathbb{D} \subset \mathbb{C}$ with the Poincaré metric (density $|ds| = |dz|/(1 - |z|^2)$); and spherical geometry represented by the Riemann sphere $\mathbb{P}$, the unit sphere $\{(x, y, z) : x^2 + y^2 + z^2 = 1\}$ in $\mathbb{R}^3$. These are nicely nested, $\mathbb{D} \subset \mathbb{C} \subset \mathbb{P}$ with this last inclusion via "stereographic" projection. Moreover, circles in one geometry are circles in the others (though with different centers and radii). The script illustrates changes in geometry, displays circles and faces in each, and shows the effects of automorphisms. Hyperbolic geometry is perhaps least familiar, but also in many ways the richest. The plane is the most comfortable for most of us. The sphere is also quite familiar — after all, we live on a sphere — but in circle packing it is by far the most challenging. ***Try it out!***

3. ⓈLayoutⓈ may help you understand the repacking/layout process; though we treat both these as blackbox operations in general, one should see at least a little about them. In the script, an abstract complex K is represented as a concrete graph (without any reference to circles). Arbitrary radii are assigned to the vertices and the corresponding circles are drawn. The "repack" command adjusts the radii iteratively based on their current values and the neighbor relationships encoded in K. Several intermediate stages of the repacking are shown: boundary radii are fixed, interior radii change, all centers are fixed for now. When the radii are "right" (i.e., comprise a "packing" label), then we can move the circles to the correct locations. Two tangent circles are laid out first, after which the rest can be placed unambiguously in succession.

The original combinatorics remain, but now the locations of its vertices are determined by the layout. We have a packing! ***Try it out!***

4. Ⓢ`BdryControl`Ⓢ illustrates manipulations critical to our later work: solutions of boundary value problems for radii and boundary angle sums for a small euclidean packing. The first section of the script emphasizes boundary angle sum manipulations. With the cursor and key presses you can prescribe boundary angles sums of π, $\pi/2$, or $3\pi/2$ at various boundary vertices; other boundary vertices have unrestrained anglesums which necessarily adjust during repacking. A key observation is that while manipulating circle packings, there are some things you control, but the packing pushes back — there is a mixture of flexibility and rigidity. Here are some things to think about: Can you prescribe boundary angle sums arbitrarily? Can you generate some problem situations? Suppose you prescribe all but one of the boundary angle sums; can you compute the remaining one? Can you prescribe angles to get the region to overlap itself? Some of these issues will come up in Exercise 6.5.

It is interesting to compare boundary controls in the other geometries. The script has you investigate quadrilaterals in the hyperbolic plane; you have a similar ability to control boundary angle sums, for instance, but the constraints of the geometry are different. ***Try it out!***

5. Ⓢ`Type67`Ⓢ investigates connections between combinatorics and geometry using constant-degree packings, those in which every circle has the same degree (the same number of neighbors). The takehome message involves these associations: degree 6 circles, flat (zero curvature); degree 5 or less, positive curvature; degree 7 or more, negative curvature.***Try it out!***

6. Ⓢ`MaxPack`Ⓢ investigates an especially important type of circle packing. A **maximal packing** for a complex K is a univalent circle packing whose carrier "fills" the underlying geometric space. In Ⓢ`Type67`Ⓢ above we saw examples for two infinite simply connected cases, the hexagonal packing, whose carrier is the euclidean plane, and the septagonal packing, whose carrier is the hyperbolic plane.

The foundational theorem in circle packing is the Koebe-Andreev-Thurston theorem below.

THEOREM 6.2 (KAT). Let K be any triangulation of a topological sphere. Then there is univalent circle packing P_K for K living on the Riemann sphere, $\mathbb{P}$. Moreover, P_K is unique up to Möbius transformations of the sphere (and inversions).

Note that we assume our complexes K and associated packings P have the same orientation, so we won't allow inversions. Also, we say a circle packing for K is **essentially unique** if it is unique up to conformal automorphisms of the underlying space. In our terminology, then, the KAT Theorem states: *Every complex K triangulating a sphere has a maximal packing P_K of the sphere, and P_K is essesntially unique.*

The final simply connnected situation we need (which actually lies behind all the others) involves complexes K which triangulate closed topological discs. Specifically, the conditions on K are that it be finite, simply connected, and have one connected component of boundary vertices. In this situation we appeal to the following theorem.

THEOREM 6.3. If K triangulates a closed topological disc, then there is a univalent packing P_K for K in the hyperbolic plane having the property that all its boundary circles are horocycles. Moreover, P_K is unique up to Möbius transformations of $\mathbb{D}$.

Ⓢ`MaxPack`Ⓢ illustrates various maximal packings. Some, of course, are just suggestive, since we cannot display infinitely many circles. All are, nonetheless, theoretically sound. The script also explores the close connection between the previous two theorems: the maximal packings in $\mathbb{D}$ of the last theorem are in fact, solutions of boundary value problems: just set boundary radii to infinity (in the hyperbolic plane this makes them horcycles). These maximal packings, in turn, provide the practical mechanism behind maximal packings of the sphere: see why in this script. **_Try it out!_**

7. Ⓢ`Branching`Ⓢ illustrates one further source of control beyond the boundary conditions we have studied so far: namely, branching. A circle c_v is **branched** if it is an interior circle whose neighbors wrap more than once around it. For "simple" branching, the angle sum at v is 4π and the neighbors wrap twice around it. In general, if the angle sum is $2\pi n$, the neighbors wrap n times around and the circle is said to have **branch order** $n-1$. Investigate the local geometry at branch circles in this script. **_Try it out!_**

8. Ⓢ`InputOutput`Ⓢ gives you a tour of _CirclePack_ input/output operations. After you work with _CirclePack_ for a while, you may want to read/write packings, save/load scripts, create PostScript images and *.jpg screendumps, etc. **_Try it out!_**

6.4. Discrete Function Theory

With experience from the previous scripts, the reader should be ready for the investigation of discrete analytic function theory. To see how _CirclePack_ displays functions, run through Ⓢ`DiscreteRMT`Ⓢ, which creates the discrete conformal mapping of Figure 6.1.

The script begins with a curve defining a Jordan region Ω. A regular hexagonal packing H_ϵ (i.e., all circles the same radius ϵ) overlays Ω in the convas on the right. Using Ω like a cookie cutter, a packing P of Ω is cut from H_ϵ. The associated complex K has a maximal packing $Q = P_K$ in $\mathbb{D}$, which _CirclePack_ computes and displays in the canvas on the left. We now have, more or less, the situation as pictured in Figure 6.1. This is the typical way we will represent our functions $f : Q \longrightarrow P$ in _CirclePack_: "domain" on the left, "range" on the right. Here's the official definition.

DEFINITION 6.4. A **discrete analytic function** is a map $f : Q \longrightarrow P$ between circle packings Q and P which preserves tangency and orientation.

As a practical matter, our discrete analytic functions will always have domain and range packings that share a common complex K (though, of course, K may vary from one example to another). If v is a vertex of K with circle C_v in Q, then $c_v = f(C_v)$ is the circle for v in P.

Note that because of Theorem 6.3, we actually have already generated the packings for many discrete analytic functions. Every packing, P, is the range of a discrete analytic function: just build its associated maximal packing P_K, and *viola*, $f : P_K \longrightarrow P$ is a discrete analytic function. You might well ask "What is the big deal about functions?" The function is not about the range (alone); it's about how the domain changes into the range — it's about the *mapping*. Which circles grow, which shrink? How does the boundary behave? Do the circles remain disjoint, or do some come to overlap? How does the geometry change under f? Which properties are preserved? which ones change?

Subsequent scripts will teach you how *CirclePack* can aid in generating mappings and investigating their properties. In the domain/range mode, as in Figure 6.1, you can click the left mouse button on a circle in one packing and both it and its counterpart in the other packing will be highlighted. Check a few pairs of tangent circles in Q to see that their counterparts are indeed tangent in P. Also observe that the circle centered at 0 in Q corresponds to the circle of P which is also at 0: this is what we will mean when we say $f(0) = 0$. Clicking the middle mouse button will highlight corresponding faces of Q and P. *CirclePack* provides various buttons for refreshing, zooming, etc.; I'll leave it to you to experiment with the interface. Your goal is to be able to probe the packings to discover properties of the mapping f.

6.4.1. Conformal Maps. The "cookie" method of Ⓢ`Cookie`Ⓢ is a very practical method for creating **discrete conformal** mappings, so called because they parallel the conformal (i.e., one-to-one analytic) mappings of the classical theory.

There is a second approach which is modeled on the Schwarz-Christoffel methods seen in Section 5.2. In the discrete setting, the machinery is remarkably close to the (largely hidden) numerical machinery behind the classical method. Given a target polygon Γ, one can read off the angles at its corners; in Ⓢ`DiscreteSC`Ⓢ we impose these as boundary angle sum conditions. The difficulty is in choosing **which** boundary vertices to place the restrictions on. These are the so-called "prevertices" of the classical Schwarz-Christoffel formula (the "x_j" in formula (67)).

EXERCISE 6.5. Before you jump into Ⓢ`DiscreteSC`Ⓢ, work out necessary conditions on boundary angle sums which can be realized by a euclidean packing. Consider the packing of Figure 6.6 and the polygon Γ formed by the boundary edges of its carrier. To start, what can you say about the interior angles of such a closed polygonal path? It is typical to consider the "turning angle" at the corners as you travel in the

positive direction (counterclockwise) about Γ. When you come to a corner, the turning angle is positive if you turn left onto the next side and negative if you turn right. This angle can be interpreted as $\pi - \theta$, where θ is the interior angle (on the left) at the corner. There is a necessary condition on the sum of turning angles. What is it? How can this be restated as a condition on the sum of interior angles at corners?
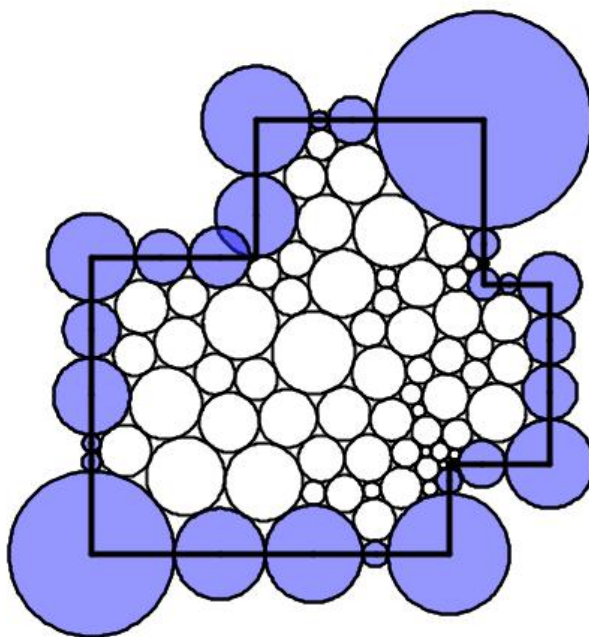
FIGURE 6.6. Packing of a Polygon

EXERCISE 6.6. The Riemann Mapping Theorem says that every simply connected proper subdomain Ω of the plane is conformally equivalent to a particular "model" simply connected domain, namely, the open unit disc. This has been extended to certain *multiply* connected regions Ω. For instance, if Ω is a connected domain in the plane bounded by some finite number n of disjoint Jordan curves, then it is called **n-connected**. (A "Jordan" curve is a closed curve without self-intersections.) In this caae, there is a conformal mapping of Ω onto a standard model multiply connected domain. There are various possible "models", but of particular recent interest are

are so-called "circle domains". A circle domain C in the plane is one bounded by a finite number of circles. Despite the several numerical methods for computing the maps for simply connected domains — such as Schwarz-Christoffel — the multiply connected cases are much tougher. The discrete version is quite elementary, however. See ⓈCircleDomainsⓈ. ***Try it out!***

6.4.2. Basic Function Theory. Let's look at scripts in which we construct and study additional types of functions. We start with maps of the unit disc into itself which are important in classical function theory and especially convenient for us since we can work in hyperbolic geometry. ⓈSchwarzⓈ starts with a somewhat random, rather small range packing P in $\mathbb{D}$. Incrementally increasing the boundary radii to infinity (in practice, 5 is sufficiently large!) and P morphs into its maximal packing P_K. The resulting map $f : P_K \longrightarrow P$ is a discrete analytic function of $\mathbb{D}$ into itself. It's convenient to introduce some notation here.

DEFINITION 6.7. Let $f : Q \longrightarrow P$ be a discrete analytic function, where Q and P are packings with common complex K. The associated **ratio** function $f^\# : K \longrightarrow \mathbb{R}$ is defined by

$$(99) \qquad f^\#(v) = \frac{\mathrm{radius}(f(C_v))}{\mathrm{radius}(C_v)}, v \in K,$$

where C_v is the circle for v in Q and $f(C_v)$ is the corresponding circle in P.

$f^\#(v)$ is the "stretch factor" at the v, the amount that the associated circle is stretched in going from Q to P. With this for notation, we can illustrates our first piece of **discrete function theory**, the Discrete Schwarz Lemma.

THEOREM 6.8 (Discrete Schwarz Lemma). Let $f : P_K \longrightarrow P$ be a discrete analytic self-map of the unit disc with $f(0) = 0$. Then $f^\#(0) \leq 1$. Moreover, equality holds if and only if f is a rotation; that is, $P = \lambda P_K$ for some complex λ with $|\lambda| = 1$.

EXERCISE 6.9. Run through ⓈSchwarzⓈ again, but more carefully. ***Try it out!*** Can you see the Discrete Schwarz Lemma in action? The ratio function plays the role in the discrete theory that the absolute value of the derivative, $|F'|$, plays in the classical theory, so the Discrete Schwarz Lemma precisely parallels its classical counterpart, even up to the equality statement.

EXERCISE 6.10. There is a classical extension of the Schwarz Lemma known as the Dieudonné-Schwarz Lemma. If $F : \mathbb{D} \longrightarrow \mathbb{D}$ is analytic and $F(0) = 0$, then by the Schwarz Lemma, $|F'(0)| \leq 1$. According to this extension, there is a certain universal constant $C, 0 < C < 1$ (i.e., independent of F) so that $|F'(z)| \leq 1$ whenever $|z| <= C$. ⓈDieudonneⓈ sets up experiments so you can try to estimate this constant: you manipulate a range packing P in $\mathbb{D}$, color coding those circles whose (euclidean) radii are smaller in P than in P_K. I'll give you a bit of help: C is larger than $1/2$ but quite a bit less than 1. ***Try it out!*** (The correct value can be found in the text at the end of the script.)

The Discrete Schwarz Lemma is an example of a "monotonicity" result — if you decrease the boundary radii of Q to those of P, then the radius of the circle at the origin also goes down. Details are given in [**6**], but the basic ideas can be seen in the following Exercise:

EXERCISE 6.11. Consider a mutually tangent triple of circles in euclidean geometry as illustrated in Figure 6.7. Using the euclidean Law of Cosines, show that if the radius of circle C_v is increased, then the angle α at C_v will decrease while the angles β, γ at the other two circles will increase.
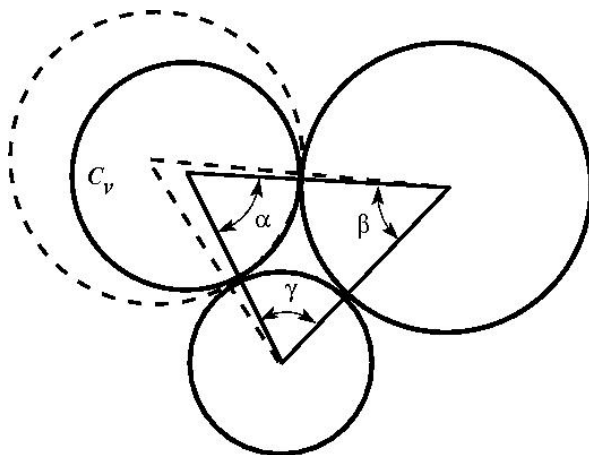
FIGURE 6.7. Monotonicity in euclidean triples

EXERCISE 6.12. Using the previous fact, prove the following:

THEOREM 6.13 (Discrete Maximum Principle). Let $f : Q \longrightarrow P$ be a discrete analytic function between euclidean packings Q and P, and suppose Q is locally univalent. Then the ratio function $f^{\#}(v)$ attains it maximum at some boundary vertex v.

Analogous results hold in hyperbolic geometry and are central to the computation of packing radii. The failure of this monotonicity in spherical geometry is tied in with the problems you address in §**??**.

EXERCISE 6.14. Show with a computation in spherical geometry (using the Spherical Law of Cosines) that the monotonicity of 6.11 can fail in spherical geometry. What appears to be the trigger for such failure? ***Try it out!***

The analogy between $f^{\#}$ and the classical $|F'|$ is not perfect. In particular, $|F'|$ is zero at branch points of F, whereas $f^{\#}$ will never be zero. The fact that discrete functions correctly realize the geometry underlying branching can be seen in the following Exploration.

EXPLORATION 6.15. An important class of classical analytic self-maps of the disc, the **finite Blaschke products**, was introduced in Chapter **??**. They are displayed there as products, each factor vanishing at one point of the disc, and thus their properties are subject to direct computation. For instance, every finite Blaschke product $B : \mathbb{D} \longrightarrow \mathbb{D}$ extends to a continous map on the unit circle, $\partial\mathbb{D}$, mapping $\partial\mathbb{D}$ n times around itself, where n is the number of factors in B.

However, there is a more geometric characterized of finite Blaschke products: they are precisely the **proper** analytic maps f of $\mathbb{D}$ onto itself. "Proper" means that whenever $\{z_j\}$ is a sequence of points in $\mathbb{D}$ converging to the unit circle, $|z_j| \to 1$, then the images, $\{f(z_j)\}$, also converge to the unit circle, $|f(z_j)| \to 1$. By orientation, B wraps the boundary around itself some number n times, and from this the argument principle implies that B will have $n - 1$ branch points inside the disc. It is these geometric properties which lead you to **discrete** finite Blaschke products, as you will see in Ⓢ `Blaschke` Ⓢ. ***Try it out!*** You also get a closeup look at branch points — a chance to demystify the local geometric behavior.

Analytic functions present this fascinating mixture of rigidity within flexibility — I'm convinced that this explains why they have attracted such attention for two centuries now. You get to manipulate some features, but then you have to live with the consequences that conformal rigidity brings. This mixture is rather difficult to get a handle on, but there are two particular notions mathematicians have invented that give some insight: **harmonic measure** and **extremal length**.

6.4.3. Harmonic Measure. Let Ω be a Jordan region Ω in the plane (that is, one bounded by a simple closed curve). Fix a subset A lying in the boundary of Ω. The **harmonic measure function**, $\omega(z) = \omega_{A,\Omega}(z)$, is the unique real harmonic function on Ω with the property that $\omega(w) = 1$ for $w \in A$ and $\omega(w) = 0$ for $w \in \partial\Omega \backslash A$. The **harmonic measure** of A with respect to Ω **at** $z \in \Omega$ is the value $\omega_{A,\Omega}(z)$. This is called a "measure" because if you fix Ω and z, then this is a probability measure on $\partial\Omega$. That is, it is a function of $A \subset \partial\Omega$: $\omega_{A,\Omega}(z) \in [0, 1]$, with $\omega_{\emptyset,\Omega}(z) = 0$, $\omega_{\partial\Omega,\Omega}(z) = 1$, while if A_1, A_2 are disjoint in $\partial\Omega$, then $\omega_{A_1 \cup A_2,\Omega}(z) = \omega_{A_1,\Omega}(z) + \omega_{A_2,\Omega}(z)$. (We necessarily suppress finer details about measurability and continuity here.)

There are two especially nice properties of harmonic measure in the classical setting:

(1) It is easy to compute in certain simple situations: In particular, if Ω is the unit disc $\mathbb{D}$ and z is taken as the origin, then $\omega_{A,\mathbb{D}}(0) = \mathrm{arclength}(A)/(2\pi)$, the proportion of the unit circle that A occupies as seen from the origin.

(2) It is a **conformal invariant**; that is, it is invariant under conformal maps. Put more explicitly, suppose $F : \Omega_1 \longrightarrow \Omega_2$ is a conformal map (i.e., a conformal homeomorphism) between two domains. Under quite weak conditions, it is known that F can be extended to map the boundary onto the boundary, $F : \partial\Omega_1 \to \partial\Omega_2$.

Given $A_1 \subset \partial\Omega_1$, define A_2 to be its image $A_2 = F(A_1)$ in $\partial\Omega_2$, and suppose $z_1 \in \Omega_1$ and $z_2 = F(z_1) \in \Omega_2$. Then $\omega_{A_1,\Omega_1}(z_1) = \omega_{A_2,\Omega_2}(z_2)$.

If you put these two properties together with the Riemann Mapping Theorem, one can, in theory, find the harmonic measure function for any simply connected region. How might this be useful? Here's but one of numerous applications. Suppose a homogeneous, thin metal plate is cutout in the shape of Ω. Suppose a segment A of the boundary is held at 100 degrees (centigrade), while the remainder of the boundary is held at 0 degrees. After the plate has had time to reach thermal equilibrium (so temperatures are no longer changing), what is the temperature at a point z in the interior of the plate? It is evidently somewhere between 0 and 100, it would approach 100 as z comes closer to A, and 0 as z gets closer to other parts of the boundary. But what is the temperature precisely at z? The answer is precisely $100\omega_{A,\Omega}(z)$.

In Ⓢ `HarmMeasure` Ⓢ we investigate the discrete analogy. I can share the definition now, since it is motivated directly by property (1). Suppose K triangulates a closed topological disc, A is a chain of edges in ∂K, and $v \in K$ is an interior vertex of K. Let P_K be the maximal packing for K in $\mathbb{D}$ which has the circle for v centered at the origin. Each edge in $e \in A$ corresponds to an arc $\alpha_e \subset \partial\mathbb{D}$ between the centers of the two horocycles forming e. (You have seen horocycles in *CirclePack*. A "horocycle" is a circle in $\mathbb{D}$ which is internally tangent to the unit circle and may be treated as hyperbolic circle with infinite hyperbolic radius and center at the point of tangency with $\partial\mathbb{D}$.)

DEFINITION 6.16 (Discrete Harmonic Measure). Given K, A, and v as described, the **discrete harmonic measure** of A at v with respect to K is

$$\omega_{A,K}(v) = \sum_{e \in A} \text{arclength}(\alpha_e)/(2\pi).$$

So, in fact, discrete harmonic measure is defined directly on K. The invariance in the discrete version of property (2) is now merely definition. In the script, however, you will see its geometric nature exhibited, much as you would see the temperature distributed about a plate in the application noted above. A couple of experiments are quite convincing that the same geometry is at work in both the discrete and the classical settings: we "graph" the harmonic measure for a classical and discrete representations of a region, and we illustrate the **extension of domain princple**.

6.4.4. Extremal Length. Extremal length is a slightly more nuanced measure of conformal "shape". Start with the Jordan region Ω again, but this time suppose we are given disjoint closed subarcs A and B of its boundary; the triple $\mathcal{Q} = \{\Omega; A, B\}$ is described as a "conformal quadrilateral".

A classical theorem states that there exists a conformal mapping $F : \Omega \longrightarrow R$, where R is a euclidean rectangle, so that F extends continuously to the boundary and so that the images $F(A), F(B)$ are opposite ends of R. F is unique up to euclidean

similarities. (You might sense a similarity to a Schwarz-Christoffel situation, as investigated earlier, but this is quite different. The four boundary points mapping to corners are specified, but the shape of the image rectangle is not set in advance.)

Figure 6.8 illustrates the analogous discrete result you will see in Ⓢ`ExtLength`Ⓢ. By a rigid motion, image rectangles can be positioned as shown here, with the images of A and B on the left and right, respectively. The **aspect ratio** of R will be defined as its width over its height, $\text{Aspect}(R) = W/H$.
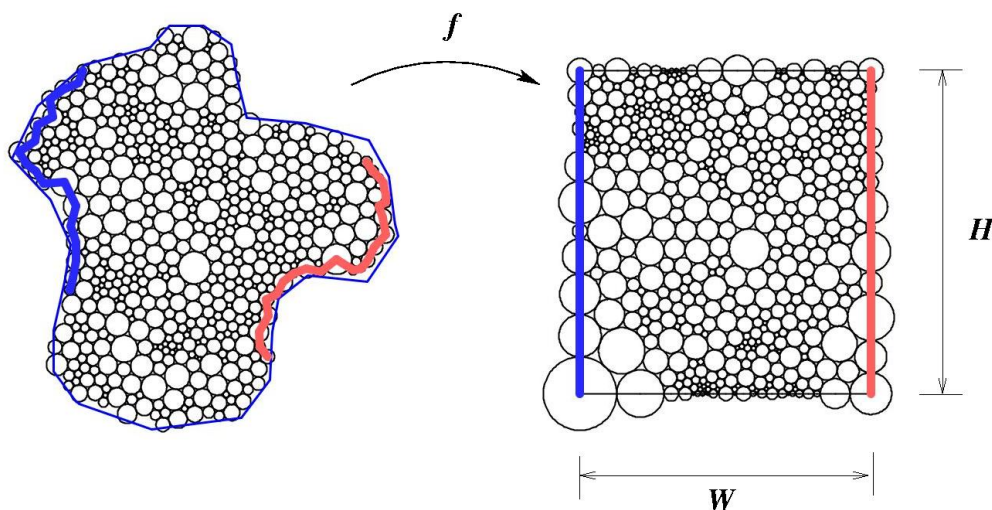
FIGURE 6.8. Packing for extremal length

Continuing our discussion of the classical setting, $\text{Aspect}(R)$ is known as the **extremal distance** from A to B, relative to Ω, or the **extremal length** of the quadrilateral $\mathcal{Q}$, denoted $\text{EL}(\mathcal{Q})$. As with harmonic measure, there are two key properties:

(i). Extremal length is easy to compute in certain simple situations: For instance, if Ω is a euclidean rectangle width W and height H, and A and B are its two ends, then clearly $\text{EL}(\{\Omega, A, B\}) = W/H$.

(ii). Extremal length is a **conformal invariant**; that is, if $\mathcal{Q}_1 = \{\Omega_1; A_1, B_1\}$ and $\mathcal{Q}_2 = \{\Omega_2; A_2, B_2\}$ are conformal quadrilaterals and $F : \mathcal{Q}_1 \longrightarrow \mathcal{Q}_2$ is a conformal mapping such that $F(A_1) = A_2$ and $F(B_1) = B_2$, then $\text{EL}(\mathcal{Q}_1) = \text{EL}(\mathcal{Q}_2)$

Also as with harmonic measure, there are physical interpretations. Suppose Ω is a uniform conducting metal plate in the plane, the boundary segment A is attached to a 1 volt power supply, the boundary segment B is grounded, and the remaining segements of the boundary are insulated. Then the current flow between A and B in amperes will be proportional to $\text{EL}(\{\Omega, A, B\})$, the constant depending on the properties of the metal.

Discrete analytic mappings to rectangles having designated boundary vertices as corners are quite easy to compute by methods you have already seen in Ⓢ RePack Ⓢ: you simply specify that all boundary vertices have angle sum π except for the four which are to be the "corners", and these you set to $\pi/2$. Ⓢ ExtLength Ⓢ illustrates the construction of Figure 6.8 and explains how *CirclePack* reports the extremal length when you have a rectangle. You are asked to modify the constructions to answer a question about current flow. You might then test your intuition with an extremal length analogue of the extension of domain principle we saw with harmonic functions.

One last script for this section, Ⓢ MiscFtnTheory Ⓢ, illustrates the range of miscellaneous behaviors you might encounter with discrete analytic functions. These can provide concrete insights you might carry back to the classical setting.

6.5. Function Construction

We have "hands-on" control of our discrete functions, but this is not like the formula-driven and algebraic control available in the classical setting. In the discrete world, we have limited ability to add or multiply functions, and essentially no way to consider compositions of the type studied in Chapter ??. To build functions we require something geometric to hook into. We have seen several examples already: to get maximal packings in $\mathbb{D}$, we send boundary radii to infinity; to build euclidean polygons, we control their boundary angle sums; to create Blaschke products, we designate branch circles. The next sequence of scripts give additional examples, some involving techniques, others relying on special circumstances or outright tricks. There's plenty of room for creative thinking in this.

6.5.1. Discrete Exponentials. Ⓢ Doyle Ⓢ challenges you to adjust parameters in a family of fascinating circle packings known as Doyle spirals. With patience you can build a range packing which clearly mimics the classical exponential function. ***Try it out!*** Read the script for guidance, but then experiment on your own. The answer is revealed at the end.

6.5.2. Discrete Rational Functions. A classical **rational** function is one of the form $F(z) = P(z)/Q(z)$, where $P(z), Q(z)$ are complex polynomials (assumed to have no common factors). Projecting stereographically to the sphere in both domain and range, F may be considered as a map from the Riemann sphere to itself, $F : \mathbb{P} \longrightarrow \mathbb{P}$, and that will be our view here. In particular, if $P(z) = 0$, then $F(z) = 0$, while if $Q(z) = 0$, then $F(z) = \infty$. The behavior at ∞ in the domain depends on relative degrees: if $\deg(P) > \deg(Q)$, then $F(\infty) = \infty$, if $\deg(P) < \deg(Q)$, then $F(\infty) = 0$, and if $\deg(P) = \deg(Q)$, $F(\infty)$ is a non-zero complex number.

The sphere is compact, and $F : \mathbb{P} \longrightarrow \mathbb{P}$ is an open continuous map, so key properties of rational maps are purely topological: F has a constant **valence**, $N > 0$; that is, for every point $p \in \mathbb{P}$, $F^{-1}(p)$ has N points (counting multiplicities at branch points). The valence N satisfies $N = max\{\deg(P), \deg(Q)\}$. Moreover, there are

precisely $2(N-1)$ branch points. (This counts multiplicities at branch points, which may include branching at infinity.) As you build examples you should watch for these properties directly.

Ⓢ`Rational1`Ⓢ (???? Poly) shows one technique for the construction of rational maps, since polynomials P are themselves rational maps (take $Q \equiv 1$): the valence N is just the degree of P. The discrete construction in Ⓢ`Rational1`Ⓢ illustrates these "polynomial-like" rational maps. To get a taste for what should be "more typical" rational maps Ⓢ`Rational2`Ⓢ gives a particular example that was constructed using the special geometry of "Schwarz" triangles on the sphere. This is a beautiful example: it is the discrete analogue of $F(z) = z^2(3z^5 - 1)/(z^5 + 3)$, with valence 7 and 12 simple branch points. Lovely as it is, it just highlights a major capability that is missing in our discrete theory: we have no algorithm for computing packing labels directly in spherical geometry.

6.5.3. Ratio Function Constructions. Recall that if $f : Q \longrightarrow P$ is a discrete analytic function, then the associated local scaling, the analogue of $|F'|$ for analytic functions, is given by its ratio function $f^{\#}$. (See Definition 6.7.) In particular, if we wish to build a discrete analogue f for some classical function F, we might begin from this differential level, mimicking $|F'|$.

Ⓢ Ⓢ builds a discrete analogue for the classical function $F(z) = (1 + z)^2$, which was illustrated with the applet *ComplexTool* in Figure 4.2.

Ratio function constructions also give an approach to entire functions. We illustrate in Ⓢ`Erf`Ⓢ for a function central to statistics, the **error function** $\mathrm{erf}(z)$ defined by

$$\mathrm{erf}(z) = \frac{2}{\sqrt{\pi}} \int_0^z e^{-\zeta^2} \, d\zeta.$$

When z is restricted to the real axis, $z = x$, then the derivative of $\mathrm{erf}(x)$ is the Gaussian distribution, the famous "bell-shaped" curve.

The construction of the discrete error function illustrates several construction issues. Let me highlight a few to prepare you.

(1) First, F, as an entire function, is defined on the whole complex plane. Discrete entire functions would likewise be defined on infinite complexes, as the discrete exponential we defined earlier on the regular hexagonal packing. Experiments and computations, however, are restricted to finite complexes. Our approach relies on "exhaustion": we mimic F using finite packings that take up successively larger, but finite, portions of $\mathbb{C}$.

(2) F has infinite valence: for every $w \in \mathbb{C}$, $F^{-1}\{w\}$ has infinitely many points. In fact, it's "mapping" behavior is quite fascinating to watch. Early in the construction, it may appear that we are omitting two points, w_1, w_2, much as the exponential omits the image value 0. That, however, would contradict the Little Picard Theorem, which tells us that a nonconstant entire function can omit at most one value in the complex plane.

Watch more stages in the construction and see how our growing function manages, in fact, to cover everything. It turns out that though not omitted, w_1 and w_2 are what is known as **asymptotic values**, that is, there are paths in the domain whose image paths converge to w_1 or w_2. Can you find such paths in the *CirclePack* experiments?

(3) Does the sequence of packings of this construction ultimately converge to an infinite circle packing? If so, it would answer in the negative one of the earliest questions in circle packing:

QUESTION 1. Are the regular hexagonal packings and the Doyle spirals the only locally univalent hexagonal circle packings in the plane?

But is there, in fact, an infinite packing behind this construction? Each stage of construction is finite, so this is a convergence issue, and it remains unresolved. The script suggests both positive and negative indicators.

6.5.4. Harmonic Mappings. Recall from Section **??** that harmonic mappings H on the disc can be written in the form $F(z) = H(z) + \overline{G}(z)$ where H, G are analytic in $\mathbb{D}$. While the "addition" here is not a geometric feature that can be carried directly by our circles, there is a discrete fact that captures the local behavior which we'll discuss in a moment. The upshot is that if one just jumps in with discrete functions, the fundamental global phenomena appear to persist. In this section we start with the mechanics, see some examples parallel to ones in Chapter **??**, then set up some lines of open experimentation for you.

Take a quick run through ⓈHGBasicsⓈ. ***Try it out!*** The way *CirclePack* operates, h and g will be discrete analytic functions on the disc, so, within each example they will share a complex K and have its maximal packing P_K as their common domain. In fact, there will be no need to view P_K, but in ⓈHGintroⓈ h is taken as the identity function, so P_K itself appears in p0 as the range of h. The image of g appears in p1. On execution of 'h_gbar', the sum $h + \overline{g}$ is displayed in p2. *CirclePack* requires that the range packings for h and g both be euclidean, and a key feature in ⓈHGintroⓈ is that their boundary radii are identical. The script starts with g also the identity, so the initial f is pretty boring. However, if you press 'b' on an interior circle of h, you will that to become a branch circle for g. The image of f is immediately more interesting. You can add additional branch points and see some of the behavior described in Chapter 4. In particular, since the boundary radii are identical for h and g, the complex dilatation function $w(z)$ is a discrete finite Blaschke product.

Also in ⓈHGintroⓈ we see, for purposes of comparison, the addition $h + g$ (note there is no conjugate on g now). The result $f = h + g$, again in p2, will approximate the classical sum. When you display the circles in p2, however, you see that they are not doing the work for us: these radii are simply the sums of radii from h and g, they do not form a packing label.

The next script, Ⓢ HGexamples Ⓢ, illustrates a few additional examples related to Chapter 4.

A last script, Ⓢ HGtrials Ⓢ, sets up various adjustments you can implement with mouse and keys for more open ended experiments. And, of course, you can also add actions of your own devising. Certain keys will increase boundary radii of g, and this introduces you to another feature: faces of f (in p2) are colored blue if f reverses their orientation. We know from the classical theory, that this can happen only if $|G'|/|H'| > 1$ on some part of $\partial \mathbb{D}$. In the discrete setting, absolute values of derivatives are replaced by sharp functions; if you return to the definition of the sharp function and recall that h, g share the same domain packing, you see that the discrete condition becomes $g^{\#}/h^{\#} > 1$ on some vertices of ∂K. Therefore, by manipulating the relative sizes of the boundary circles in p0 and p1, we affect the orientation of $f = h + \overline{g}$.

EXERCISE 6.17. Verify this bit of local geometry, which may account for the properties of these discrete harmonic functions. Consider euclidean triangles T and $\widetilde{T}$ with vertices $\{z_1, z_2, z_3\}$ and $\{\tilde{z}_1, \tilde{z}_2, \tilde{z}_3\}$, respectively. Assume that each of these triples is positively oriented. Now form a new triangle Δ with vertices $\{w_1, w_2, w_3\}$, where $w_j = z_j + \overline{\tilde{z}}_j, j = 1, 2, 3$. The issue is whether Δ is necessarily positively oriented.

In general, the answer is "no", but here is a pretty fact that you should try to prove: *If T and $\widetilde{T}$ happen to be formed by mutually tangent triples of circles $\{C_1, C_2, C_3\}$ and $\{\widetilde{C}_1, \widetilde{C}_2, \widetilde{C}_3\}$, and if $radius(\widetilde{C}_j/radius(C_j) <= 1, j = 1, 2, 3$, then Δ is positively oriented.* In other words, the classical boundary condition for preserving orientability of $H + \overline{G}$, namely $|G'|/|H'| < 1$, carries over to the discrete case at the local face level.

Use your favorite computer math package to build examples. Show first that Δ can indeed have reverse orientation in general. But even more is true: show that even if you arrange that the side lengths of $\widetilde{T}$ are individually less than the corresponding side lengths of T, Δ may still end up with reverse orientation. Mysteriously, however, when those side lengths are **sums of radii**, and the radii for $\widetilde{T}$ are individually less than those for T, then Δ **remains** positively oriented!

6.6. Convergence

We have introduced the basic local geometry of circle packings, moved on to defining mappings between packings, and investigated the emergent global geometric properties. The claim has been made — with convincing evidence, I hope — that these mappings deserve the name "discrete" analytic functions. Can we, in fact, move beyond mere parallels?

The answer is a resounding yes. Though granularity prevents discrete analytic functions from actually **being** analytic functions, they can be made to approximate analytic functions by finding a suitable *refinement* process to improve resolution. Circle packing got started with exactly this approximation issue with the foundational Rodin and Sullivan Theorem explained below. Let's see it in action first.

ⓈDiscreteRMTⓈ allows you to repeat the conformal mapping experiment of Section **??**, but now with a choice among several refinement levels. In particular, recall that you build a range packing P by cookie cutting a region Ω from a regular hexagonal packing. The maximal packing Q having the same combinatorics is the domain in $\mathbb{D}$. A standard *CirclePack* normalization puts the circle for the α vertex at the origin in both Q and P, and a γ vertex on the imaginary axis in each. The map $f : Q \longrightarrow P$ is then a discrete conformal map.

How well do these discrete mappings approximate the classical conformal mapping $F : \mathbb{D} \longrightarrow \Omega$ guaranteed by the Riemann Mapping Theorem? That is the content of the foundational theorem of the topic by Rodin and Sullivan, paraphrased here.

THEOREM 6.18 (Rodin and Sullivan, [**5**]). Let Ω be a bounded simply connected domain in the plane and for each $\epsilon > 0$ let f_ϵ be the discrete conformal mapping created as in the script, based on a regular hexagonal packing with circles of radius ϵ. Let $F : \mathbb{D} \to \Omega$ be the classical conformal mapping with the same normalization. Then the functions f_ϵ converge to F uniformly on compact subsets of $\mathbb{D}$ as $\epsilon \to 0$. Moreover, the ratio functions $f^\#$ converge uniformly on compact to the modulus classical derivative, $|F'|$.

The script illustrates the quality of the discrete mapping in two ways. One is rather self explanatory, imposing a spoke-and-wheel grid on the domain disc and having the discrete analytic function carry that forward to Ω. The result is compared in Figure **??** with the analogous result computed using Toby Driscoll's Schwarz-Christophel software (citeDSCT).

A second test of quality is illustrated in the script via color-coding. Let me briefly describe this. First, treat both Q and P as euclidean and concentrate on their carriers. We can reimagine $f : Q \longrightarrow P$ as a more familiar type of function; namely, as a function $f : \text{carr}(Q) \longrightarrow \text{carr}(P)$ mapping each face of $\text{carr}(Q)$ onto the corresponding face of $\text{carr}(P)$, with the vertices (circle centers) mapped to corresponding vertices. This can be accomplished nicely by defining f to be **piecewise affine**, that is affine on each face of $\text{carr}(Q)$. I will let the reader look into the defintion. For our purposes it is enough to note that if T_1 is a face of $\text{carr}(Q)$, then f maps T_1 to the corresponding face t_1 of $\text{carr}(P)$ and has the form

$$f(x,y) = (ax + by + x_0, cx + dy + y_0), \ (x,y) \in T_1,$$

for some real values a, b, c, d, x_0, y_0. Such a mapping is **quasiconformal** with **dilatation** given by $K = w + \sqrt{w^2 - 1}$, where $w = (a^2 + b^2 + c^2 + d^2)/(2(ad - bc))$. In particular, $K \geq 1$ and f is conformal (analytic) on T_1 iff $K = 1$, that is, iff T_1 and t_1 are similar triangles. In general, the size of K reflects the how close the mapping f is to being analytic. To do this, the experiments in ⓈDiscreteRMTⓈ color code the faces of the range packings. For each face, the further the dilatation of f on that face is from 1, the more red the color, up to some cutoff, (e.g. $K = 1.2$), after which the color is set to blue (???). In particular, the closer the color gets to white, the

closer the dilatation to 1 — i.e., the closer to being analtyic. An example was shown in Figure 6.1.

The convergence of discrete objects to their classical counterparts that we have seen in the Rodin/Sullivan theorem actually holds much more generally, indeed, in essentially all the settings we have discussed in this chapter. Here are some scripts that refine earlier constructions. A couple words of caution about increasing the complexity: • Complexity brings a computational penalty, so repacking and display times can increase dramatically. • Complexity can lead to roundoff errors that interfere with layout, display, and other operations. Though *CirclePack* can handle tens of thousands of circles, a few ill advised mouse clicks can push you into a computational neverland. (Of course, you can always kill *CirclePack* and start over!) • Complex packings are harder to manipulate — you can't as easily discern indices, boundary circles become harder to pick out with the mouse, etc.

Several ??? scripts investigate generalizations and refinements for the convergence involved in the Rodin/Sullivan Theorem.

EXERCISE 6.19. Ⓢ `ConformalRefine` Ⓢ follows up on the discrete conformal mapping in Ⓢ `Conformal` Ⓢ by permitting hex refinements. Figure **??** compares the images of a spoke/circle grid in $\mathbb{D}$ under two different approximations to the conformal map $F : \mathbb{D} \to \Omega$: the left applies a discrete conformal mapping as generated in by *CirclePack* in the script, while the right applies a classical numerical approximation generated by Tobin Driscoll's "SC Toolbox", a Schwarz-Christoffel mapping package for **matlab** (see Section **??**). In the script one should change the region Ω and then apply the commands to see the effects of hex refinement on the apparent accuracy.

EXERCISE 6.20. Ⓢ `RandomRefine` Ⓢ illustrates the use of random packings in place of hexagonal packing as the refinement method. Again, you should create you own simply connected region Ω, then apply the commands developed in the script to create and investigate the discrete conformal mappings at various refinement levels.

EXERCISE 6.21. Ⓢ `MultiRefine` Ⓢ illustrates maps between multiply connected regions, as in Ⓢ `????` Ⓢ. The best way to modify the script for other regions is to make a copy, then swap out the current "PATH" data files (at the end of the script) for ones the user creates, either in *CirclePack* or another package.

EXERCISE 6.22. Run Ⓢ `ErfRefine` Ⓢ to pursue hex refined packings for the "error" function we encountered in Ⓢ `?` Ⓢ (see [**6**, §14.4]). One method for judging how closely the results mimic the classical erf(z) is to compare the images of the unit circle under the discrete map and the classical map. Methods are given in the script to display the former. However, you should use your favorite math package to generate the latter; the script explains how to read and display a path given as a list of x, y coordinates.
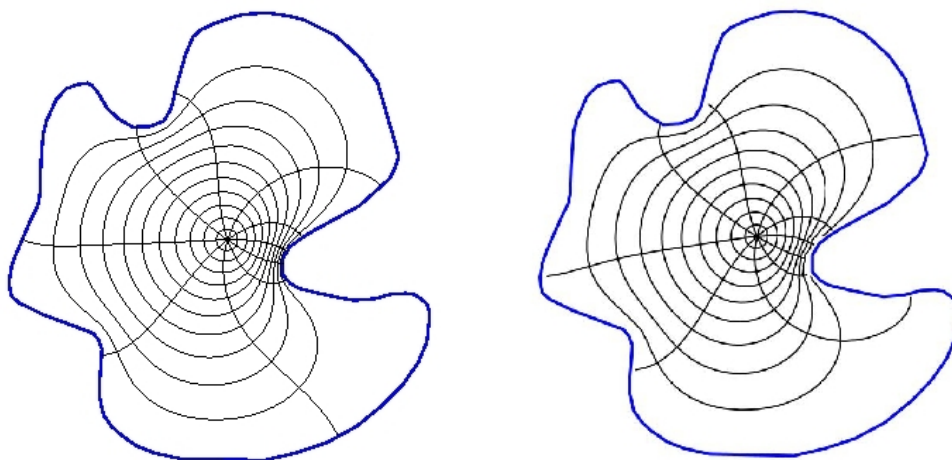
FIGURE 6.9. Two conformal approximations: (a) circle packing, (b) Schwarz-Christoffel

EXERCISE 6.23. This exercise sets a simple task: estimate the complex number $B(3i/4)$ (??? Maple to find this value) for a classical finite Blaschke product B. However, all you know about B is this: $B(0) = 0$, $B(1) = 1$, B has three zeros (counting multiplicies), and B has branching (i.e., B' vanishes) at $z_1 = (1+i)/2$ and $z_2 = -1/2$. It is known that appropriate discrete finite Blaschke products b_n will converge uniformly on compact subset of $\mathbb{D}$ to B (an extension of the Rodin/Sullivan theorem to branched packings). Ⓢ BlaschkeRefine Ⓢ allows you to create discrete functions b_n (at various refinement levels) based solely on this type of data. You will have to tailor the construction and normalizations to match this particular B. Then you are asked to apply two methods for estimating $B(3i/4)$: (1) Use the values $b_n(3i/4)$ for the b_n you constructed. (2) Locate the zeros of the b_n, use those to build a classical finite Blaschke product $\widehat{B}$ close to B, then compute $\widehat{B}(3i/4)$. ??? The point is that classical Blaschke products are best constructed from their zeros (see ???), while their discrete analogues are best constructed from their branch points.

Nearly every script from section **??** involves packings which are subject to the types of refinements we've just discussed, and invariably the discrete objects will converge under this refinement to their classical counterparts. The reader should pick a topic and create a script which implements appropriate refinement: discrete exponentials, discrete rational maps, discrete ratio function constructions, discrete harmonic functions — each project will have its distinct character.

6.7. Wrapup

I hope the reader has enjoyed this chapter. Perhaps it has given you a new way to look at and think about analytic functions. As discrete theories go, circle packing is arguably unique in the breadth, depth, and fidelity with which it captures its subject. The essential phenomena appear with even the simplest of packings, where you seem to glimpse the heart of analyticity. As complexity increases, that fidelity only grows, to the point that you can reliably investigate — perhaps even discover new phenomena — in the classical theory. So it is a full-featured "quantum" theory, classical in the limit, with experimental and visualization capabilities to boot.

This chapter may also have introduced you to a topic of interest in its own right. There is a lot of room for growth in circle packing, from packings on surfaces to randomized packings to applications. And wherever circle packings go, there, too, will be complex analysis.

6.8. Additional Exercises

Here are three additional exercises to stretch your understanding of circle packing. The first challenge in each is simply understanding the mechanisms being illustrated in the scripts. After that, however, you may see new directions for experiments, and you are encouraged to set your own path in studying them.

EXERCISE 6.24. Seeing the connection between triangulations and circle packings may have piqued your curiosity about more general triangulations. Whereas our complexes K have been simply connected up to this point, the situation in ⓢ Torus ⓢ is quite different: here K triangulates a torus. A torus is topologically equivalent to the surface of an inner tube, so you would rightly guess that one cannot realize a circle packing for K in the plane. Yet the script presents you with the "packing" P shown on the left in Figure **??**.

Interpretation of P rests on the color coded "side-pairings": The 32 shaded circles form the packing, but additional tangencies exist on the torus. Since we cannot realize these in the plane, we create ghost circles (like the unshaded ones on the left in the figure): the circles centered along one colored edge are identified with corresponding circles along the other edge sharing that color. You may be able to see the matchings in the image either from the combinatorics or the radii, but the script should help you explore. The image in the right of Figure **??** presents this in another fashion, copies of the packing tiling the plane.

Ultimately, however, it's up to you to come to terms with this represention for circle packings of tori. It may help to create your own triangulated torus and insert it into the script data section in place of the current packing. Encoding of complexes is described in the "Formats" tab of the Help Frame of *CirclePack*. Let me offer a specific challenge: create a combinatorial torus K having just 7 vertices — the smallest number possible for a proper triangulation (that is, one in which any two faces are disjoint, share a single vertex, or share an edge and its end vertices).
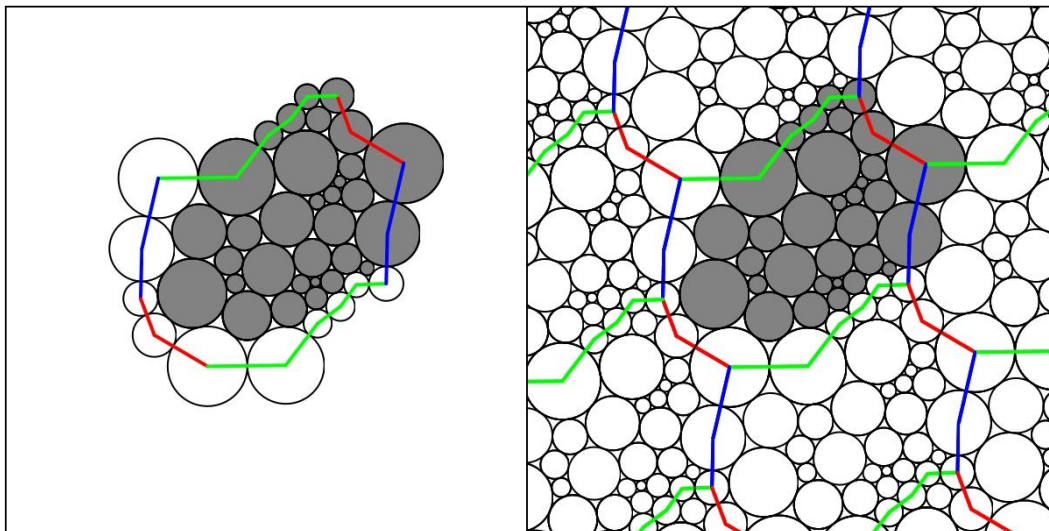
FIGURE 6.10. Circle Packing of a torus with color-coded side pairings

EXERCISE 6.25. An experimental capability of *CirclePack* that we touched on briefly in ???? involves the "adjoin" operation for triangulations. The key observation is that when two simply connected triangulations K_1 and K_2 have the same number of boundary vertices, they can be pasted to one another along their boundary edges. The result is a new triangulation K of the sphere. A packing P for K in $\mathbb{P}$ provides simultaneous packings for K_1 and K_2 while also giving an embedding for their common boundary curve, the so-called **welding curve**.

The basics are demonstrated with Owl in Ⓢ`OwlWelding`Ⓢ. Among other things, interest lies in how the shape of the welding curve is determined by our choices in attaching K_1 and K_2. There is a connection with conformality, but I can only say that it is related to a classical topic called **conformal welding** (see [**9**]). Learn about the "double" and "adjoin" operations, and let the experiments begin.

EXERCISE 6.26. The *CirclePack* algorithms for computing circle packings work in euclidean and hyperbolic geometry. Unfortunately, there is as yet no general algorithm known for computing circle packings directly in spherical geometry, and the spherical packings in previous scripts have been stereographic projections from the plane or disc. All is not lost, however. At least a smattering of examples can be created by hand using basic spherical geometry and symmetry. Figure **??** illustrates an example with the combinatorics of the soccer ball: (a) shows the maximal packing, while (b) is a branched packing having the same combinatorics. Investigate this further in Ⓢ`SoccerPack`Ⓢ. ***Try it out!*** .

Both packings in Figure **??** are constructed using the geometry of Schwarz triangles. The univalent packing Q shows the symmetry of the regular dodecahedron: twelve (12) uniformly spaced circles each with five neighbors, the remaining 20 circles each with 6
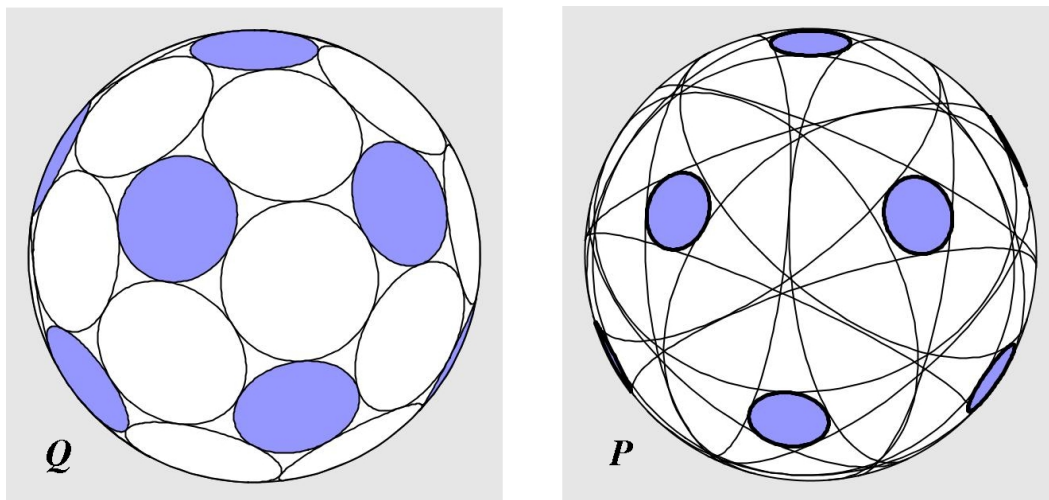
FIGURE 6.11. Circle packings with soccer ball combinatorics

neighbors. (This packing can, in fact, be created by standard methods in *CirclePack*; see Ⓢ Ⓢ). The packing P shares the same combinatorics, but the degree five circles are all branch circles, so P defines a 7-fold covering of the sphere. The map $f : Q \to P$ is a discrete rational function with valence 7 and 12 simple branch points.

Both Q and P can be created using the spherical triangles of Figure **??**, two of several triangles known as "Schwarz" triangles. The Schwarz $\langle 2, 3, 5 \rangle$ triangle t (i.e., having angles $\pi/2, \pi/3$, and $\pi/5$) is pictured in Figure **??**(a). A second triangle, the $\langle 2, 3, 5/2 \rangle$ Schwarz triangle T, is shown in (b) (at a different scale). Dashed lines on T show how it is built from 7 copies of t. A Pattern of two circular arcs is imprinted on t in (a), the analogous pattern is imprinted on T in (b).

A Schwarz triangle can be reflected across any of its three edges to form a copy of itself with reverse orientation. Repeated reflections will tile the sphere; in the case of t this tiling involves 120 copies. Carrying the imprinted arcs along during the reflection process yields a univalent circle packing with 42 circles, our maximal packing Q. The identical reflection process, but now starting with T, gives 120 copies which tile the sphere 7 layers deep and whose imprinted arcs give our packing P. In any copy of T, it takes 10 reflections around the vertex with angle $2\pi/5$ to close up, and this accounts for the 12 branch points of P. The script Ⓢ Soccer Ⓢ lets you check one of these visually to see the 5 (huge) neighboring circles wrap twice around it.

Schwarz triangles where first used to generate branched circle packings of the sphere in [**1**] (see Appendix H of [**6**]). However, the example of Figure **??** and other equally beautiful examples were created by Samantha Corveno, one of my REU (Research Experience for Undergraduates) students.

There are a handful of additional Schwarz triangles, and other options for circular markings, so the challenge to you is to create yet more examples. For each, you
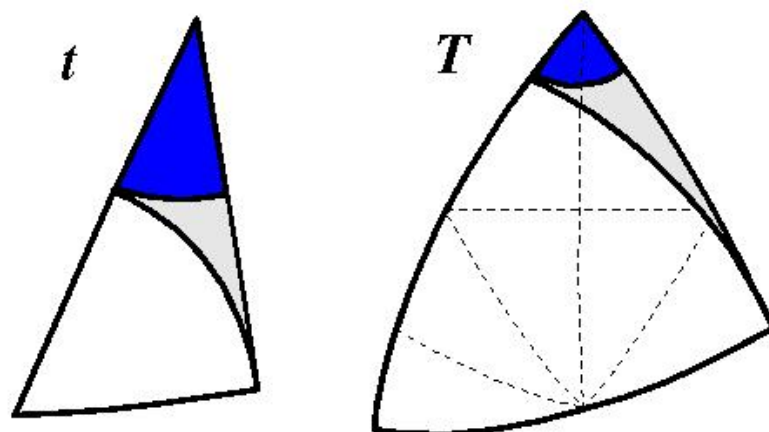
FIGURE 6.12. Schwarz triangles t and T
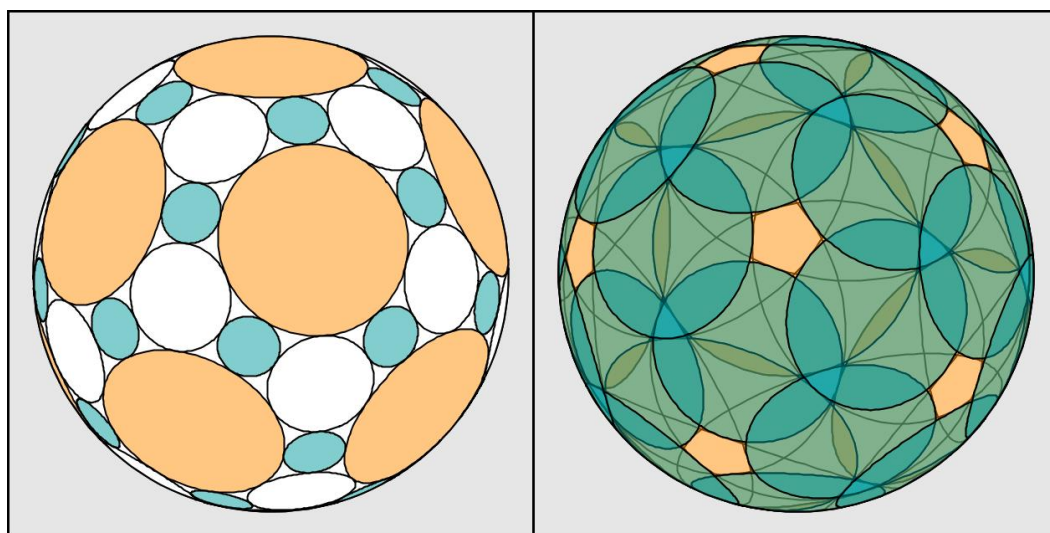
FIGURE 6.13. A branched packing due to Samantha Corvino

will need to encode the combinatorics K for *CirclePack*, which can then generate the associated maximal packing. *CirclePack* won't find the branched packing radii, but if you can compute those with spherical trigonometry, *CirclePack* will display your results. To get you started, I've put another of Samantha's examples with more details in ⓈTri432Ⓢ. ***Try it out!***

Bibliography

1. Philip L. Bowers and Kenneth Stephenson, *A branched Andreev-Thurston Theorem for circle packings of the sphere*, Proc. London Math. Soc., **73** (1996), 185–215.
2. Charles R. Collins and Kenneth Stephenson, *A circle packing algorithm*, Computational Geometry: Theory and Applications **25** (2003), 233–256.
3. Tobin A. Driscoll and Lloyd N. Trefethen, *Schwarz-Christoffel Mapping*, Camb. Univ. Press, New York, 2002.
4. Tobin Driscoll, *Schwarz-Christoffel Toolbox, for* **Matlab**, *http://www.math.udel.edu/∼driscoll/software/SC/*
5. Burt Rodin and Dennis Sullivan, *The convergence of circle packings to the Riemann mapping*, J. Differential Geometry **26** (1987), 349–360.
6. Kenneth Stephenson, *Introduction to circle packing: the theory of discrete analytic functions*, Camb. Univ. Press, New York, 2005.
7. Kenneth Stephenson, *Circle packing: A mathematical tale*, Notices Amer. Math. Soc. **50** (2003), no. 11, 1376–1388.
8. William Thurston, *The finite Riemann mapping theorem*, 1985, Invited talk, An International Symposium at Purdue University in celebrations of de Branges' proof of the Bieberbach conjecture, March 1985 (unpublished).
9. G. Brock Williams, *Discrete conformal welding*, Indiana Univ. Math. J., **53** (2004), 765–804.

Background

Notation Page

The following notation will be defined and employed throughout the text, and so we collect it here for easy reference.

$\mathbb{N}$ = the natural numbers.

$\mathbb{Q}$ = the rational numbers.

$\mathbb{R}$ = the real numbers.

$\overline{\mathbb{R}} = \mathbb{R} \cup \{\infty\}$ = the extended real numbers.

$\mathbb{R}^2 = \{(x, y) : x, y \in \mathbb{R}\}$ = the Euclidean plane.

$\mathbb{C} = \{x + iy : x, y \in \mathbb{R}; i = \sqrt{-1}\}$ = the complex numbers.

$\text{Re}(z)$ = the real part of z.

$\text{Im}(z)$ = the imaginary part of z.

$\overline{\mathbb{C}} = \mathbb{C} \cup \{\infty\}$ = the extended Complex plane or Riemann Sphere.

$C(a, r) = \{z \in \mathbb{C} : |z - a| = r\}$, Euclidean circle.

$\triangle(a, r) = \{z \in \mathbb{C} : |z - a| < r\}$, open Euclidean disk.

$\mathbb{D} = \triangle(0, 1)$ = the open unit disk centered at 0.

σ = the spherical metric on $\overline{\mathbb{C}}$.

$\triangle_\sigma(a, r) = \{z \in \overline{\mathbb{C}} : \sigma(z, a) < r\}$, open spherical disk.

$\text{Int}(E)$= interior of a set E in $\overline{\mathbb{C}}$.

$\overline{E} = closure(E)$ = closure of a set E in $\overline{\mathbb{C}}$.

∂E = boundary of a set E in $\overline{\mathbb{C}}$.

$domain(f)$ = domain set of function f.

$\text{Res}(f, z_0)$ = the residue of f at z_0.

This appendix is intended to be a summary of some of the major topics and theorems from a standard undergraduate complex analysis course. Since it is only a review of the material, we do not prove all the results here, but rather direct readers looking for more details to the main reference texts [**1**] and [**3**], and occasionally to the more advanced text [**2**]. Additionally, a few advanced topics are also included here. These include the Riemann Mapping Theorem, the Open Mapping Theorem, and the Schwarz Lemma. The extended notion of analyticity on the Riemann sphere $\overline{\mathbb{C}}$ is addressed separately in Appendix B.

A.1. Functions of a Complex Variable as Mappings

A complex valued function f defined on a set $R \subset \mathbb{C}$ is denoted $f : R \to \mathbb{C}$. We often denote the domain set of f by $domain(f)$, and without further mention use the convention that we may express $z = x + iy = (x, y)$ and $f(z) = f(x, y)$ with the implicit understanding that $x = \text{Re}(z)$ and $y = \text{Im}(z)$. Also, the set $f(R) = \{f(z) : z \in R\}$ is called the *range* of f. Lastly, if $A \subset \mathbb{C}$, we denote the *inverse image* (or preimage) of A under f by $f^{-1}(A) = \{z \in R : f(z) \in A\}$.

A.1.1. Linear Functions. We can examine the linear mappings $f(z) = az + b$, for constants $a, b \in \mathbb{C}$, by first considering the simpler functions $h(z) = az$ and $g(z) = z + b$.

Writing the variable $z = re^{i\theta}$ and the parameter $a = |a|e^{i\alpha}$ in polar form, we can express the image point as $h(z) = az = r|a|e^{i(\theta + \alpha)}$. Hence, the function h maps the point z to the point $h(z)$, which geometrically can be interpreted as a stretching or contraction of the modulus by $|a|$ and a rotation about the origin by the angle α.

We can understand the action of the function $g(z) = z + b$ geometrically as translating the point z by $|b|$ units in the direction of the *vector b*.

Putting this together we see that the function $f(z) = az + b$ is the composition $g(h(z))$ which moves a point z by stretching or contracting by $|a|$, rotating by the angle α, and then translating by the point (vector) b. See Figure A.1 for an example.
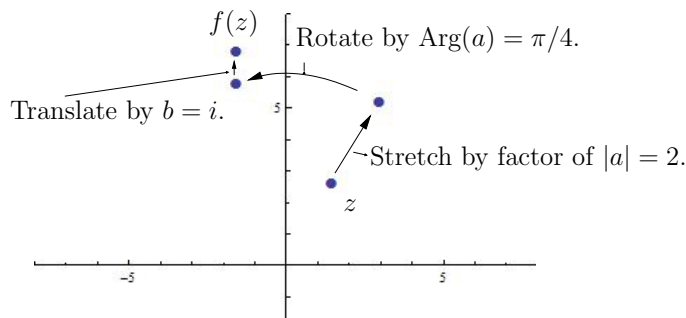

FIGURE A.1. An illustration of the map $f(z) = az + b$ for $a = 2e^{i\pi/4}$ and $b = i$.

A.1.2. Power Functions. Consider the map $f(z) = z^n$ for fixed $n \in \mathbb{N}$, where $\mathbb{N}$ denotes the natural numbers. Expressing the variable $z = re^{i\theta}$ in polar form, we see that $f(z) = (re^{i\theta})^n = r^n e^{in\theta}$; that is, under the action of f the point z has its modulus raised to the nth power and its argument multiplied by n. See Figure A.2 for an illustrative example.
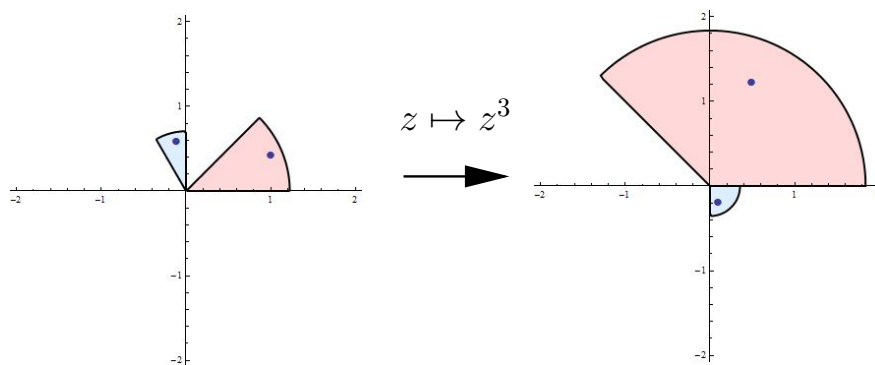

$$z \mapsto z^3$$

FIGURE A.2. An illustration of the map $z \mapsto z^3$.

A.1.2.1. *Roots of unity.* Fix a positive integer n. We call any of the n solutions to the equation $z^n = 1$ an *nth root of unity,* noting that each must be of the form $\omega_k = e^{2k\pi i/n}$ for $k = 0, \ldots, n-1$. We also note that these nth roots of unity can be expressed $1, \omega_1, \omega_1^2, \ldots, \omega_1^{n-1}$ and that these are equally spaced points on the unit circle $|z| = 1$.

A.1.3. The Exponential Function. The definition of the exponential function arises naturally out of Euler's formula $e^{i\theta} = \cos\theta + i\sin\theta$.

DEFINITION A.1. If $z = x + iy$, then

$$e^z = e^x(\cos y + i\sin y).$$

The preceding definition has two important consequences:
  (1) The exponential function maps $\mathbb{C}$ onto $\mathbb{C}\backslash\{0\}$.
  (2) The exponential function is periodic of period $2\pi i$.

A.1.4. The Logarithm Function. The essence of the logarithm function is in its role in providing an inverse of the exponential function. However, since the exponential function is not one-to-one, special consideration must be taken by first understanding $\arg(z)$.

DEFINITION A.2. For a complex number $z \neq 0$, the argument of z, denoted $\arg(z)$, is the set of all $\theta \in \mathbb{R}$ such that $z = |z|e^{i\theta}$. The *principal value* of the argument, denoted $\text{Arg}(z)$, is the unique such angle θ with $-\pi < \theta \leq \pi$.

DEFINITION A.3. The function $\log(z)$ is the multiple-valued function

$$\log(re^{i\theta}) = \ln(r) + i\theta,$$

where $\ln(r)$ denotes the real-valued logarithm. Alternatively, we may define

$$\log(z) = \ln|z| + i\arg(z).$$

Since $\log(z)$ is multiple-valued (because $\arg(z)$ is multiple-valued), it is not really a **function** in the traditional sense. To make it a function, we can use the *principal value* of the argument, $\mathrm{Arg}(z)$.

DEFINITION A.4. The *principal value* of the logarithm is the function defined by

$$w = \mathrm{Log}(z) = \ln|z| + i\mathrm{Arg(z)}.$$

It has domain set $\mathbb{C}\backslash\{0\}$ and range $-\pi < \mathrm{Im}(w) \leq \pi$, but is not continuous at any point of the negative real axis.

Important Technology Note: While mathematically we use $\mathrm{Log}(z)$ and $\mathrm{Arg}(z)$ to denote the principal values of the logarithm and argument functions, the applet *ComplexTool* uses $\log(z)$ and $\arg(z)$ to express these.

REMARK A.5. While the principal branch of the logarithm is the one most widely used, it is possible to define other branches of the logarithm function which are continuous, for example, along the negative real axis. Note that any branch of the logarithm must necessarily include 0 as a point of discontinuity. For more information, we refer the reader to [**1**] and [**3**].

A.1.5. Trigonometric Functions. The trigonometric functions of a complex number z are defined in terms of the exponential function.

DEFINITION A.6. Given any complex number z, we define

$$\sin z = \frac{e^{iz} - e^{-iz}}{2i} \qquad \text{and} \qquad \cos z = \frac{e^{iz} + e^{-iz}}{2}.$$

One can show that the image of the infinite vertical strip $-\pi/2 \leq \mathrm{Re}\, z \leq \pi/2$ under either of these maps is the entire plane. (Thus the trigonometric functions are unbounded in $\mathbb{C}$.)

A.2. Continuity and Analyticity in $\mathbb{C}$

DEFINITION A.7. Let $R \subset \mathbb{C}$ and consider a complex valued function $f : R \to \mathbb{C}$.
(1) We say that f is *continuous* at z_0 if $\lim_{z\to z_0} f(z) = f(z_0)$; i.e., for each $\epsilon > 0$ there exists $\delta > 0$ such that $|f(z) - f(z_0)| < \epsilon$ whenever $z \in R$ and $|z - z_0| < \delta$. We say f is continuous on a set $U \subset R$ if it is continuous at each point of U.
(2) Furthermore, f is called *differentiable* at z_0 when its derivative $f'(z_0) = \lim_{z\to z_0} \frac{f(z)-f(z_0)}{z-z_0}$ exists.

(3) When such a function is differentiable at all points of an *open* set R, the function f is said to be *analytic* on R. (Some authors refer to such a function as *holomorphic* in R.)

A.2.1. Cauchy-Riemann Equations. When we write $f(z) = f(x, y)$, we call $u(x, y) = \operatorname{Re} f(x, y)$ and $v(x, y) = \operatorname{Im} f(x, y)$. Then $f(x, y) = u(x, y) + iv(x, y)$ and thus we may define the following partial derivatives

$$(100) \qquad \frac{\partial f}{\partial x} = \frac{\partial u}{\partial x} + i\frac{\partial v}{\partial x} \qquad \text{and} \qquad \frac{\partial f}{\partial y} = \frac{\partial u}{\partial y} + i\frac{\partial v}{\partial y}.$$

We recall that when f is differentiable at z_0, we have $f'(z_0) = \frac{\partial f}{\partial x}(z_0) = -i\frac{\partial f}{\partial y}(z_0)$, which by equating real and imaginary parts yields the Cauchy-Riemann equations

$$(101) \qquad \frac{\partial u}{\partial x} = \frac{\partial v}{\partial y} \qquad \text{and} \qquad \frac{\partial u}{\partial y} = -\frac{\partial v}{\partial x}.$$

THEOREM A.8. A function $f(z) = u(x, y) + iv(x, y)$ is analytic in an open set R if and only if the first partial derivatives u_x, u_y, v_x, v_y exist, are continuous, and satisfy the Cauchy-Riemann equations in R. (See, e.g., [**1**], p. 63-66.)

Furthermore, by using the Cauchy-Riemann equations (101), one can show that if $f(z) = u(x, y) + iv(x, y)$ is analytic in an open set R, then each of the component functions u and v is *harmonic* in R, that is, $u_{xx} + u_{yy} = 0$ and $v_{xx} + v_{yy} = 0$.

A.2.2. Conformal Mappings. Recall that a function $f : R \to \mathbb{C}$ is called *univalent* if it is one-to-one; i.e., for any two points $z_1 \neq z_2$ in R, we then have $f(z_1) \neq f(z_2)$. An analytic function is **locally** univalent in a small neighborhood of a point z_0 if and only if its derivative is non-zero at z_0. A function that is both analytic and univalent on an open set is said to be *conformal* on that set. Geometrically, conformal means that the function is locally angle-preserving, preserving both the magnitude and sense of the angle. You can get a feel for this by graphing any univalent analytic function with the applet *ComplexTool*, using either a rectangular or circular grid, and zooming in on points in the range to see the angle preservation.

A.3. Complex Integration

A complex integral is an expression of the form $\int_C f(z)\, dz$. This integral can be evaluated for many kinds of functions (analytic or not) and any path (closed loop or not). In many cases, the value of the integral does depend on the actual geometry of the path, not just on the end points. It is important to note that the value of the integral does not depend on the parameterization used to describe the path as long as the different parameterizations trace out the path in the same direction. This will be evident in the formula given below.

A.3.1. Computing a Complex Integral. To compute a complex integral by brute force requires some parameterization of the path C. We describe the path directly in terms of z whenever possible. Here are some important examples of curves and how they could be parameterized:

(1) A circle of radius R centered at a: $z(t) = a + R\,e^{it}$, where $0 \leq t \leq 2\pi$.
(2) The line segment from z_1 to z_2: $z(t) = (1-t)z_1 + tz_2$, where $0 \leq t \leq 1$.

Once a parameterization is obtained, the formula for calculating the integral is

$$(102) \qquad \int_C f(z)\,dz = \int_{t=a}^b f(z(t))z'(t)\,dt.$$

A.3.2. Topology on the plane. [1]

DEFINITION A.9.

(1) A set is a *domain* if it is open and connected.
(2) A path C is a *closed curve* if its initial and terminal points coincide.
(3) A *simple closed curve* is a closed curve that does not cross itself.

A.3.3. Three Important Integral Theorems. The following results are central to any study of complex analysis.

THEOREM A.10 (Cauchy's Theorem). Let $f(z)$ be analytic *everywhere on and inside* a simple closed curve γ. Then

$$\int_\gamma f(z)\,dz = 0.$$

THEOREM A.11 (Cauchy's Integral Formula). Let $f(z)$ be analytic *everywhere on and inside* a simple closed positively-oriented curve γ. Let a be a point inside γ. Then

$$\int_\gamma \frac{f(z)}{z-a}\,dz = 2\pi\,i\,f(a).$$

THEOREM A.12 (Cauchy's Generalized Integral Formula). Let $f(z)$ be analytic *everywhere on and inside* a simple closed positively-oriented curve γ. Let a be a point inside γ. Then

$$\int_\gamma \frac{f(z)}{(z-a)^{n+1}}\,dz = \frac{2\pi\,i\,f^{(n)}(a)}{n!}.$$

---

[1]See Appendix B.3 for more detailed information regarding topology on both the plane and the Riemann sphere.

A.3.4. Path Integrals. If C is not a simple closed curve, then none of the above integral theorems can be used to calculate it. Fortunately it may not be necessary to use brute-force either. Suppose that C is a path that goes from a point z_1 to the point z_2. The main question is when does the value of the integral only depend on the endpoints and not on the particular path between them? The answer is exactly when $f(z)$ is analytic on an appropriate region. In this case the integral can be calculated by finding an anti-derivative and evaluating it at the endpoints. This is simply an application of the Fundamental Theorem of Calculus.

Here is a more precise statement: If C_1 and C_2 are paths between the same endpoints, then $\int_{C_1} f(z)\, dz = \int_{C_2} f(z)\, dz$ if $f(z)$ is *analytic everywhere on and in between both paths.*

A.4. Taylor Series and Laurent Series

Taylor series and Laurent series play crucial roles in our understanding of analytic functions.

A.4.1. Taylor Series. The theory of Taylor series carries over directly from the theory in real variables. However, it is even better as the following theorem shows:

THEOREM A.13. Let $f(z)$ be an *analytic function* in a domain D and let $z_0 \in D$. Then the following statements hold:

(1) $f(z)$ can be represented by a convergent power series in the disk $|z - z_0| < R$, where R is the distance from z_0 to the nearest singularity of $f(z)$. In particular, if $f(z)$ is entire, then $R = \infty$ and the series converges everywhere.
(2) The series representing $f(z)$ is the Taylor series; hence

$$f(z) = \sum_{n=0}^{\infty} \frac{f^{(n)}(z_0)}{n!} (z - z_0)^n.$$

(3) The Taylor series is unique, in other words, any power series representing an analytic function must in fact be the Taylor series of the function.

Note that power series can be added, subtracted, multiplied and divided. Caution: multiplication and division are not performed term by term; rather you must treat each series like a polynomial. Series can also be differentiated and integrated term-by-term to obtain new series.

A.4.2. Laurent Series. If a complex function has a singularity at $z = z_0$, then it certainly does not have a Taylor expansion at that point. It may, however, have a different sort of series expansion in a deleted neighborhood of the singularity called a *Laurent series.*

In general, a function f which is analytic throughout an annular domain $R_1 < |z - z_0| < R_2$ has a unique Laurent series representation

$$f(z) = \sum_{n=0}^{\infty} a_n (z - z_0)^n + \sum_{n=1}^{\infty} \frac{b_n}{(z - z_0)^n}$$

on that annular domain. It is allowed that we may have $R_1 = 0$ or $R_2 = +\infty$, or both.

We note that just as in the case of Taylor series, there are formulas that give the coefficients of the desired Laurent series. However, these coefficient formulas involve complicated complex integrals and are of more theoretical than practical value.

One of the main values of the Laurent expansion stems from its use in complex integration via residues. When f can be expressed by the Laurent series $f(z) = \sum_{n=0}^{\infty} a_n (z - z_0)^n + \sum_{n=1}^{\infty} \frac{b_n}{(z-z_0)^n}$ on a punctured disk with center z_0, the *residue* of f at z_0 is $\mathrm{Res}(f, z_0) = b_1$.

THEOREM A.14 (Residue Theorem). Let γ be a simple closed positively-oriented curve. If a function f is analytic inside and on γ except for a finite number of singular points z_k, for $k = 1, \ldots, n$ inside C, then $\int_\gamma f(z)\, dz = 2\pi i \sum_{k=1}^{n} \mathrm{Res}(f, z_k)$.

A.4.3. Isolated Singularities. Given a function f which is analytic on a deleted ϵ neighborhood of z_0, we can express f by its Laurent series

$$f(z) = \sum_{n=0}^{\infty} a_n (z - z_0)^n + \sum_{n=1}^{\infty} \frac{b_n}{(z - z_0)^n}$$

on $\Delta(z, \epsilon) \setminus \{z_0\}$. This situation gives us three different types of singularities classified as follows.

DEFINITION A.15 (Isolated singularities).
 (i) When $b_n = 0$ for every $n = 1, 2, \ldots$, then $f(z)$ is said to have a *removable singularity* at z_0. In such a case $f(z) = \sum_{n=0}^{\infty} a_n (z - z_0)^n$, and so defining $f(z_0) = a_0$ will make f analytic in the full neighborhood $\Delta(z, \epsilon)$ (and thus we have *removed* the singularity).
 (ii) When only finitely many $b_n \neq 0$, then there exists N such that $b_N \neq 0$ and $b_n = 0$ for all $n > N$. In this case, we say f has a *pole* of order N at z_0. We may then express f in a factored form $f(z) = (z - z_0)^{-N} h(z)$ where $h(z) = b_N + b_{N-1}(z - z_0) + \ldots$ is analytic on $\Delta(z, \epsilon)$ and $h(z_0) = b_N \neq 0$.
 (iii) When infinitely many $b_n \neq 0$, we say f has an *essential singularity* at z_0.

DEFINITION A.16. A function f is *meromorphic* in a domain D if it is analytic at every point of D except possibly at poles.

A.5. Key Theorems

This section records several key results that will be used in the text.

A.5.1. Maximum Modulus Theorem. We recall two key results relating to the maximum modulus of an analytic function. See [**1**], p. 176-178.

THEOREM A.17 (Maximum Modulus Theorem). If a function f is analytic and not constant in a given domain $D \subset \mathbb{C}$, then $|f(z)|$ has no maximum value in D. That is, there is no point $z_0 \in D$ such that $|f(z)| \leq |f(z_0)|$ for all $z \in D$.

COROLLARY A.18 (Corollary to Maximum Modulus Theorem). Suppose that a function f is continuous on a closed and bounded region $R \subset \mathbb{C}$ and that it is analytic and not constant in the interior of R. Then, the maximum value of $|f(z)|$ in R, which is always reached, occurs somewhere on the boundary of R and never in the interior.

A.5.2. Argument Principle. The Argument Principle for analytic functions gives a very nice way to count the number of zeros and poles inside a contour.

THEOREM A.19 (Argument Principle). Let γ be a simple closed curve lying entirely within a domain $D \subset \mathbb{C}$. Suppose f is analytic in D except at a finite number of poles inside γ and that $f(z) \neq 0$ on γ. Then

$$\frac{1}{2\pi i} \oint_\gamma \frac{f'(z)}{f(z)} dz = N_0 - N_p,$$

where N_0 is the total number of zeros of f inside γ and N_p is the total number of poles of f inside γ. In determining N_0 and N_p, zeros and poles are counted according to their order or multiplicities.

A.5.3. Schwarz Lemma.

THEOREM A.20 (Schwarz Lemma). Suppose $f : \mathbb{D} \to \overline{\mathbb{D}}$ is analytic and $f(0) = 0$. Then $|f'(0)| \leq 1$ and $|f(z)| \leq |z|$ for all $z \in \mathbb{D}$. Furthermore, unless f is of the form $f(z) = e^{i\theta} z$ for some $\theta \in \mathbb{R}$ (i.e., f is a rotation), we must have strict inequalities in both statements above.

A.6. More Advanced Results

A.6.1. Local Properties of Analytic Maps. The existence of power series representations of analytic maps is a powerful tool for understanding their local, and sometimes global, behavior. Consider an analytic function f defined in a domain $D \subset \mathbb{C}$. For any $z_0 \in D$, Theorem A.13 tells us that we may express

(103) $$f(z) = a_0 + a_1(z - z_0) + a_2(z - z_0)^2 + \dots$$

on any neighborhood $\Delta(z_0, r)$ which is contained in D.

Suppose for the moment that $f'(z_0) = a_1 \neq 0$. Then for z *very close* to z_0 we see that the linear contribution $a_0 + a_1(z - z_0)$ dominates the rest of the series in (103). Omitting the details, we can say that $f(z) \approx A_1(z) = a_0 + a_1(z - z_0)$ for $z \in \Delta(z_0, \epsilon)$ when $\epsilon > 0$ is very small. Since $A_1(z)$ is a linear map, with very well understood properties, we can reasonably expect that $f(z)$ will have the same properties as $A_1(z)$

when $z \in \Delta(z_0, \epsilon)$. In particular, for every w near $a_0 = f(z_0)$, there is exactly one value z_1 near z_0 which maps to w. This holds for A_1, and it also holds for f. Thus we say that f is *locally* one-to-one (also called locally univalent) at z_0 when $f'(z_0) \neq 0$. See Figure A.3.
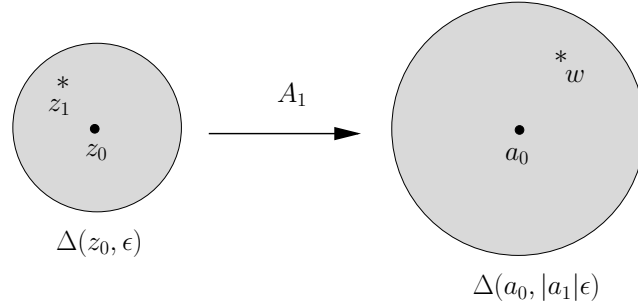

FIGURE A.3. An illustration of the map $A_1(z) = a_0 + a_1(z - z_0)$ mapping $\Delta(z_0, \epsilon)$ onto $\Delta(a_0, |a_1|\epsilon)$ in a one-to-one fashion.

Similarly, we can use approximations to understand the local behavior of $f(z)$ when $a_1 = 0$, i.e., $f'(z_0) = 0$. In such a case, we know that (unless f is constant) we may express f as

(104) $$f(z) = a_0 + a_k(z - z_0)^k + \dots$$

where a_k is the first non-zero coefficient (other than possibly a_0) in (103). Then for z *very close* to z_0, we see that the terms $a_0 + a_k(z - z_0)^k$ dominate the rest of the series in (104). Omitting the details, we can say that $f(z) \approx A_k(z) = a_0 + a_k(z - z_0)^k$ for $z \in \Delta(z_0, \epsilon)$ when $\epsilon > 0$ is very small. Since $A_k(z)$ has very well understood properties,[2] we can reasonably expect that $f(z)$ will have the same properties as $A_k(z)$ when $z \in \Delta(z_0, \epsilon)$. In particular, for every w near $a_0 = f(z_0)$, there are exactly k values $z_1, \dots, z_k$ symmetrically arranged near z_0 which maps to w. This holds for A_k, and it also holds for f, though since we only have an approximation the k values may not be exactly symmetrically arranged about z_0. See Figure A.4. Thus we say that f is *locally* k-to-one at z_0 when $f'(z_0) = f''(z_0) = \dots = f^{(k-1)}(z_0) = 0$, but $f^{(k)}(z_0) \neq 0$. We also describe this situation by saying that z_0 maps to $f(z_0)$ with **degree** (aka, multiplicity or valency) k.

When z_0 maps to $f(z_0)$ with multiplicity k, we can rewrite (104) as $f(z) = a_0 + (z - z_0)^k[a_k + a_{k+1}(z - z_0) + \dots]$. Noting that $a_k + a_{k+1}(z - z_0) + \dots$ determines an analytic map on D, we have the following.

LEMMA A.21. Let f be a function analytic at z_0 with multiplicity k. Then there is a map h which is analytic at z_0 such that $h(z_0) \neq 0$ and $f(z) = f(z_0) + (z - z_0)^k h(z)$.

---

[2]In particular, we note that $A_k(z)$ is a composition $h_3 \circ h_2 \circ h_1$ of the following simple maps $h_1(z) = z - z_0$ (translation), $h_2(z) = z^k$ (power function), and $h_3(z) = a_0 + a_k z$ (linear function).
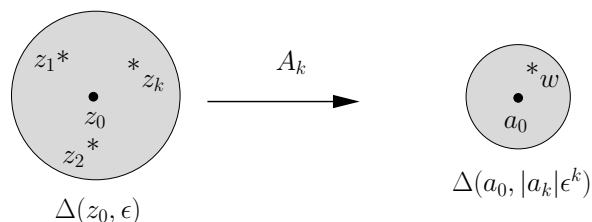
FIGURE A.4. An illustration of the map $A_k(z) = a_0 + a_k(z - z_0)^k$ for $k = 3$ mapping $\Delta(z_0, \epsilon)$ onto $\Delta(a_0, |a_k|\epsilon^k)$ in a k-to-one fashion.

In particular, if $f(z_0) = 0$, i.e., f has a zero of order k at z_0, then f has the form $f(z) = (z - z_0)^k h(z)$.

A.6.1.1. *Open Mapping Theorem.* A fact not usually introduced in a standard undergraduate complex variables course is that analytic maps are open maps, which we define more carefully below. Some chapters in this text reference this and other such results, and so we present them here. The interested reader can find formal proofs in more advanced texts such as [**2**], p. 344-348.

Consider a non-constant function f which is analytic at z_0. The discussion above shows that (whether or not $f'(z_0) = 0$) the image of a small neighborhood of z_0 must cover a small neighborhood of $f(z_0) = a_0$. This is enough to assert the following.

THEOREM A.22 (Open Mapping Theorem). If f is a function on a domain $D \subset \mathbb{C}$ which is analytic and non-constant, then the range $f(D)$ is an open set.

Since **open maps** are those for which the image of an open set is always an open set, non-constant analytic maps are open maps.

One application of this and the above theory is the following.

THEOREM A.23 (Inverse Function Theorem). Suppose that $D \subset \mathbb{C}$ is a domain and $f : D \to \mathbb{C}$ is a univalent analytic map. Then its inverse function $f^{-1} : f(D) \to D$ is also analytic.

A nice use for the Inverse Function Theorem is that it shows that *locally* one-to-one functions have local analytic inverse functions. More precisely, consider an analytic map f defined on a domain Ω, where f need not necessarily be one-to-one on all of Ω. If $f'(z_0) \neq 0$ for some $z_0 \in \Omega$, then we know from above that there is a small disk $D = \Delta(z_0, \epsilon)$ on which f is one-to-one. Hence by the Inverse Function Theorem, we see that there is a map $g : f(D) \to D$ (the *local inverse* of f) defined on the open set $f(D)$ which is analytic and satisfies $g \circ f(z) = z$ for all $z \in D$.

A.6.1.2. *Riemann Mapping Theorem.* We close with a very important theorem in complex analysis which states that all simply connected domains (other than the whole plane) are conformally equivalent to the unit disk. More precisely we have the following, whose proof can be found in [**2**], p. 420.

THEOREM A.24 (Riemann Mapping Theorem). Suppose that D is a simply connected domain in the complex plane, $D \neq \mathbb{C}$, and that $z_0 \in D$. Then, there exists a unique conformal mapping f of D onto the unit disk $\triangle(0,1)$ satisfying the conditions $f(z_0) = 0$ and $f'(z_0) > 0$.

Bibliography

1. Ruel V. Churchill and James Ward Brown. *Complex variables and applications.* McGraw-Hill Book Co., New York, eighth edition, 2009.
2. Bruce P. Palka. *An introduction to complex function theory: Undergraduate Texts in Mathematics.* Springer-Verlag, New York, 1991.
3. Edward B. Saff and Arthur David Snider. *Fundamentals of Complex Analysis with Applications to Engineering, Science, and Mathematics.* Prentice Hall, New Jersey, third edition, 2003.

APPENDIX B

The Riemann Sphere

The main purpose of this appendix is to refresh some of the key ideas about the concept of ∞ in complex analysis. In particular, we discuss the Riemann sphere via stereographic projection, the induced spherical metric σ and corresponding topology, as well as discuss how these notions tie in with the concepts of continuity and analyticity. We do not re-prove all the results here or give a complete exposition of these topics; our goal is to provide just enough background to allow the reader to fully grasp the material in those chapters of this book which require it. More details can be found in the reference texts [**1**], [**3**], and the more advanced text [**2**].

Complex arithmetic actually extends slightly beyond the complex plane. As points z in $\mathbb{C}$ move arbitrarily far away from the origin we informally say that they go to "infinity". We can make that quite precise, while also giving you one answer to the division-by-zero problem: we adjoin a point denoted ∞ to $\mathbb{C}$ to get the *Riemann sphere* (or *extended complex plane*), denoted here as $\overline{\mathbb{C}} = \mathbb{C} \cup \{\infty\}$.

B.1. Stereographic Projection and Spherical Geometry

We model the Riemann sphere $\overline{\mathbb{C}}$ by first identifying points in $\mathbb{C}$ with points in the unit sphere $S = \{(x_1, x_2, x_3) : x_1^2 + x_2^2 + x_3^2 = 1\} \subset \mathbb{R}^3$ through what is called *stereographic projection*. See Figure B.1. Label $N = (0, 0, 1)$ the "north pole" on the sphere S. Then for any given point $z = x + iy = (x, y, 0)$ in $\mathbb{C}$, consider the line in $\mathbb{R}^3$, between z and N.

Obviously, this line intersects S at precisely two points, namely N and a second point we call $Z = (x_1, x_2, x_3)$. We then associate the points z and Z. Formally, we define a map $\pi : \mathbb{C} \to S$ given by $\pi(z) = Z$. Omitting the details (which can be found in [**2**], p. 351), the precise formula for this map is $Z = \pi(z) = \pi(x + iy) = \left(\frac{2x}{|z|^2+1}, \frac{2y}{|z|^2+1}, \frac{|z|^2-1}{|z|^2+1} \right)$. However, we note that we rarely make use of this formula since it is the idea and the picture that provide the necessary understanding we require.

Note that the only point on S which is not associated with any complex number is the north pole N and that this association between $\mathbb{C}$ and $S \setminus \{N\}$ is bijective.[1] We would like to now decide what meaning there could be to associating a point to N.

Notice that if $|z|$ is very large (i.e., z is very far from the origin), then the corresponding point $Z \in S$ is very close to N. It is then natural to say that any point that

---

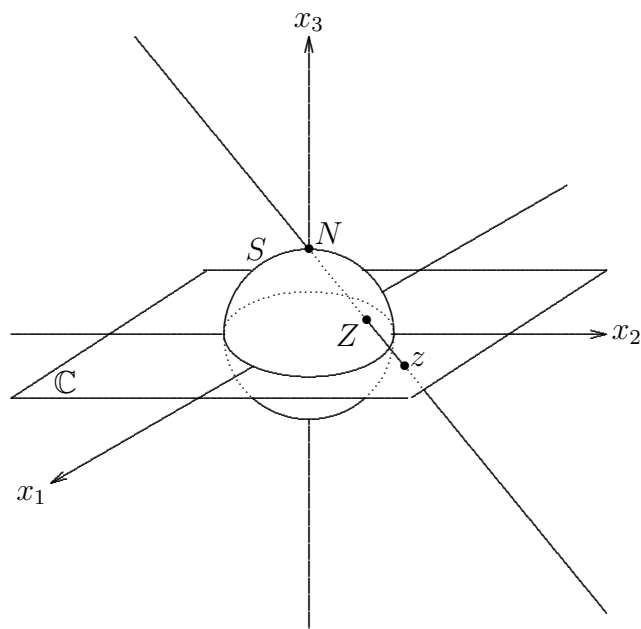[1]Recall that a map is bijective when it is both one-to-one and onto.

FIGURE B.1. Stereographic projection.

is associated with N must be "infinitely" far from the origin. It is for this reason that we adjoin ∞ to $\mathbb{C}$ and extend the definition of π by defining $\pi(\infty) = N$. We now have an identification, i.e., a bijection, between *all* of $\overline{\mathbb{C}} = \mathbb{C} \cup \{\infty\}$ and *all* of S. Thus when we speak of the Riemann sphere $\overline{\mathbb{C}}$ we can think of it as the usual complex plane together with the point ∞ being "infinitely" far from the origin, or we can think of it as being the associated sphere S where ∞ is understood to be the north pole N. It will be useful to think of $\overline{\mathbb{C}}$ in both these ways; let us do so with some examples.

The origin in $0 \in \mathbb{C}$ is identified with the "south pole" $(0, 0, -1)$, unit disk $\mathbb{D} = \{z : |z| < 1\}$ is identified with the "southern hemisphere", the unit circle $|z| = 1$ is the "equator", and the real axis is wrapped around the sphere S to form a circle running through the the south pole $(0, 0, -1)$, the north pole N, and the points $(1, 0, 0)$ and $(-1, 0, 0)$. Notice that as the real number x_1 goes to $-\infty$ and the real number x_2 goes to $+\infty$, the corresponding points $\pi(x_1)$ and $\pi(x_2)$ on the sphere S come together at the north pole N. So we see that in S there is only one ∞.

As an exercise you are asked to match up the curves in $\mathbb{C}$ with their projections onto $\overline{\mathbb{C}}$ given in Figure B.2.

B.2. The Spherical Metric σ

The standard way we measure the distance between two points z and w in $\mathbb{C}$ is by $|z - w|$, the Euclidean metric on $\mathbb{C}$. But can we come up with a natural way to define the distance between points on $\overline{\mathbb{C}}$ by using our understanding of the sphere model?
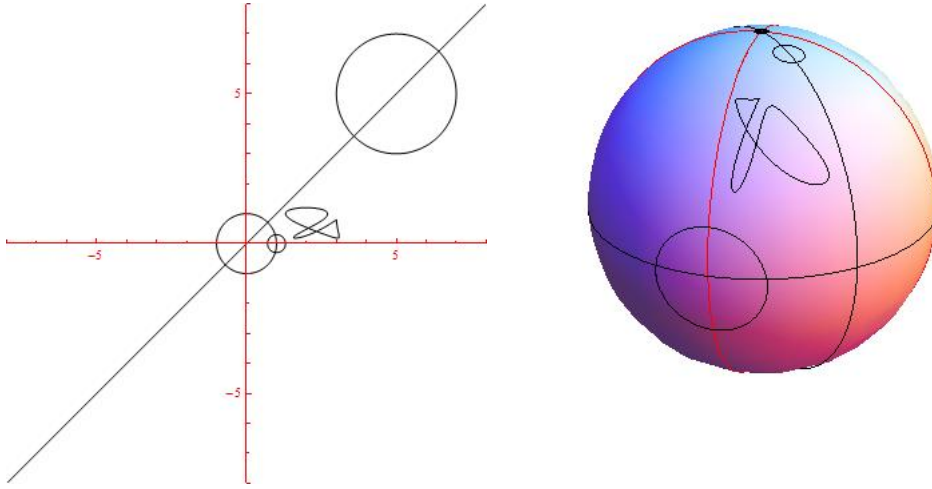
FIGURE B.2. Various curves drawn in the plane and projected onto the sphere, where ∞ is the marked point on top of the sphere $\overline{\mathbb{C}}$.

The answer is yes, we just need to use the natural spherical metric on S and transfer that back to $\overline{\mathbb{C}}$. Here's how.

The spherical distance between two points Z and W on S is defined to be the arclength of the shortest path on the sphere S which connects Z and W, which is, of course, the shorter arc of the great circle that runs through Z and W. We denote this distance by $d(Z, W)$. For example, the distance between $(1, 0, 0)$ and $N = (0, 0, 1)$ is the number $\pi/2$ since this is just $1/4$ of the circumference of a great circle with radius 1.

We can now transfer this metric from S to $\overline{\mathbb{C}}$ using the map $\pi : \overline{\mathbb{C}} \to S$ by defining the distance between two points $z, w \in \overline{\mathbb{C}}$ to be $\sigma(z, w) = d(Z, W) = d(\pi(z), \pi(w))$. This simply amounts to projecting z and w onto the sphere S and then taking the distance there. There does exist a formula for $\sigma(z, w)$, but we omit it since it is not important to our purposes. We will, however, note that this new *metric* (aka distance function) on $\overline{\mathbb{C}}$ does not treat ∞ as a special point. It simply plays the same role as any other point on the Riemann Sphere. This idea can take some getting used to, so let's look at some examples.

We first would like to understand what a *Spherical* disk $\triangle_\sigma(a, r) = \{z \in \mathbb{C} : \sigma(z, a) < r\}$, with center $a \in \overline{\mathbb{C}}$ and radius $r > 0$ looks like. Let's consider $\triangle_\sigma(\infty, \pi/2)$, i.e., the set of all points $z \in \overline{\mathbb{C}}$ such that $\sigma(z, \infty) < \pi/2$. By picturing these points on the sphere we see that the answer is the upper hemisphere, which we note can also be expressed as $\overline{\mathbb{C}} \setminus \overline{\mathbb{D}}$, i.e., the points in $\mathbb{C}$ outside of the closed unit disk together with ∞. In fact, denoting the *Euclidean* disk of radius $r > 0$ and center $a \in \mathbb{C}$ by $\triangle(a, r) = \{z \in \mathbb{C} : |z - a| < r\}$, we have the following more general result. For any

421

small $\epsilon > 0$, there is some large number $r > 0$ such that

$$(105) \qquad \Delta_\sigma(\infty, \epsilon) = \overline{\mathbb{C}} \setminus \overline{\Delta(0, r)}.$$

B.3. Topology in $\mathbb{C}$ and $\overline{\mathbb{C}}$

With the spherical metric in hand, we can now define the corresponding topological concepts on $\overline{\mathbb{C}}$. We begin by defining the interior, closure, and boundary of a set $E \subset \overline{\mathbb{C}}$ as follows: The **interior** of E, denoted $\text{Int}(E)$, is the set which contains all points $z \in E$ for which there exists $r > 0$ such that $\Delta_\sigma(z, r) \subset E$. The **closure** of E, denoted $\overline{E}$ or $closure(E)$, is the set which contains all points $z \in \overline{\mathbb{C}}$ such that for any $r > 0$ we have $\triangle_\sigma(z, r) \cap E \neq \emptyset$. The **boundary** of E, denoted ∂E, is the set which contains all points $z \in \overline{\mathbb{C}}$ such that for any $r > 0$ we have both $\Delta_\sigma(z, r) \cap E \neq \emptyset$ and $\Delta_\sigma(z, r) \cap (\overline{\mathbb{C}} \setminus E) \neq \emptyset$. Hence, $\partial E = \overline{E} \cap \overline{\overline{\mathbb{C}} \setminus E}$ and $\overline{E} = E \cup \partial E$.

EXAMPLE B.1. Without needing to provide a formal proof, the readers should convince themselves that $\text{Int}(\mathbb{D}) = \mathbb{D}, \overline{\mathbb{D}} = \{z \in \mathbb{C} : |z| \leq 1\}$, and $\partial \mathbb{D}$ is the unit circle $|z| = 1$. In general, $\partial\triangle(a, r) = C(a, r)$, where we define $C(a, r) = \{z \in \mathbb{C} : |z - a| = r\}$, i.e., the *Euclidean* circle of radius $r > 0$ and center $a \in \mathbb{C}$.

EXERCISE B.2. Identify the interior, closure, and boundary of each of the following subsets of $\overline{\mathbb{C}}$. No proof is required. ***Try it out!***

(a) $A = \emptyset$,
(b) $B = \{z \in \mathbb{C} : |z| \leq 1\}$,
(c) $C = \{z \in \mathbb{C} : |z| = 1\}$,
(d) $D = \mathbb{C} \setminus \mathbb{R}$,
(e) $E = \{x + iy \in \mathbb{C} : x, y \in \mathbb{Q}\}$, where $\mathbb{Q}$ is the set of rational numbers in $\mathbb{R}$,
(f) $F = \{1/n : n \in \mathbb{N}\}$,
(g) $G = \mathbb{C}$.

DEFINITION B.3 (Open sets).

(a) A set $A \subset \overline{\mathbb{C}}$ is called *open in $\overline{\mathbb{C}}$* if $\text{Int}(A) = A$, i.e., if for each point $z \in A$, there exist $r > 0$ such that $\triangle_\sigma(z, r) \subset A$.
(b) A set $A \subset \mathbb{C}$ is called *open in $\mathbb{C}$* if for each point $z \in A$, there exist $r > 0$ such that $\triangle(z, r) \subset A$.

DEFINITION B.4 (Closed sets).

(a) A set $A \subset \overline{\mathbb{C}}$ is called *closed in $\overline{\mathbb{C}}$* if its complement $\overline{\mathbb{C}} \setminus A$ is open in $\overline{\mathbb{C}}$.
(b) A set $A \subset \mathbb{C}$ is called *closed in $\mathbb{C}$* if its complement $\mathbb{C} \setminus A$ is open in $\mathbb{C}$.

REMARK B.5. Often, this text will refer to open or closed sets without explicitly referencing whether it is understood that these sets are open or closed in $\mathbb{C}$ or in $\overline{\mathbb{C}}$. However, context will make it clear which is meant and so no confusion should arise. It is also important to keep in mind the following two points that show that *often*, but not always, these two notions are equivalent anyway.

(1) A set $A \subset \mathbb{C}$ is open in $\overline{\mathbb{C}}$ if and only if it is open in $\mathbb{C}$.

(2) A bounded[2] subset E of $\mathbb{C}$ is closed in $\mathbb{C}$ if and only if it is closed in $\overline{\mathbb{C}}$. Note that the boundedness condition is crucial since, for example, the unbounded set $\mathbb{R}$ is closed in $\mathbb{C}$, but not closed in $\overline{\mathbb{C}}$.

The readers should take a moment to convince themselves that these two statements are indeed true.

An open set $A \subset \overline{\mathbb{C}}$ is called **connected** if given any two points $z, w \in A$ there exists a *polygonal line*[3] in A which connects z to w. If U is an open set and $z_0 \in U$, then the set $U(z_0)$ of all points $z \in U$ such that there is a polygonal line in U which connects z_0 to z is called the *component of U containing z_0*. Two facts regarding an open set U are as follows

(1) An open set U equals the union of components, i.e., $U = \cup_{z \in U} U(z)$.

(2) Components $U(z)$ and $U(w)$ equal each other exactly when there is a polygonal path in U which connects z to w.

The interested reader can read the details proving these facts in [**2**]; however, a less formal understanding of these concepts will suffice for this text.

A **domain** is a non-empty open connected set in $\overline{\mathbb{C}}$. Note that the *domain set* of a function need not be a domain in the sense just defined. A **Jordan domain** is any **simply connected domain** in $\mathbb{C}$, i.e., any domain D in $\mathbb{C}$ such that every simple closed curve in D encloses only points in D. Informally, this means that D has no "holes" in it. For example, the set $\mathbb{D}$ is a Jordan domain, but the set $\mathbb{D} \setminus \{0\}$ is not.

A **neighborhood** of a point $z \in \overline{\mathbb{C}}$ is any open set which contains z. A **deleted neighborhood** of a point $z \in \overline{\mathbb{C}}$ is any set $U \setminus \{z\}$ where U is a neighborhood of z. When a neighborhood U is of the form $\Delta_\sigma(z, \epsilon)$ or $\Delta(z, \epsilon)$ we often refer to this as an ϵ-**neighborhood** of z.

B.3.1. Convergence in the Riemann Sphere. With the spherical metric σ on $\overline{\mathbb{C}}$ and the corresponding notion of ϵ-neighborhood, we can now define convergence of sequences in $\overline{\mathbb{C}}$ as follows.

DEFINITION B.6. A sequence of points $z_n \in \overline{\mathbb{C}}$ is said to converge to $z \in \overline{\mathbb{C}}$ when $\sigma(z_n, z) \to 0$, in which case we write $z_n \to z$.

Note that the statement $z_n \to z$ when $z \in \mathbb{C}$ (i.e., $z \neq \infty$) as given above, conforms to our standard notion of convergence in $\mathbb{C}$. The following proposition records this fact and several more which relate the standard metric in $\mathbb{C}$ to the metric σ on $\overline{\mathbb{C}}$. The reader should become comfortable with these statements, though it is not crucial to be able to prove these with the rigor of an $\epsilon - \delta$ argument.

---

[2] A set E is called **bounded** in $\mathbb{C}$ if there exists some $R > 0$ such that $|z| \leq R$ for all $z \in E$.

[3] A *polygonal line* is a union of a finite number of line segments joined end to end. In the context of $\overline{\mathbb{C}}$, we allow a line segment to be any standard line segment in the plane $\mathbb{C}$ or any arc of a great circle that passes through ∞ (as viewed as the north pole N on the sphere S).

PROPOSITION B.7. For points z_n, z, and w in $\mathbb{C}$, we have the following:

(1) $\sigma(z_n, z) \to 0$ if and only if $|z_n - z| \to 0$,

(2) $|z| > |w|$ if and only if $\sigma(z, \infty) < \sigma(w, \infty)$,

(3) $z_n \to \infty$ if and only if $|z_n| \to +\infty$ if and only if $\sigma(z_n, \infty) \to 0$,

(4) $z_n \to 0$ if and only if $|z_n| \to 0$ if and only if $1/|z_n| \to +\infty$ if and only if $1/z_n \to \infty$.

We can use this proposition to solve the *division by zero* problem, but we must first define continuity in the context of $\overline{\mathbb{C}}$.

B.4. Continuity in the Riemann Sphere

Let $R \subset \overline{\mathbb{C}}$ and consider an *extended* complex valued function $f : R \to \overline{\mathbb{C}}$. We write $\lim_{z \to z_0} f(z) = L$ and we say the function f *approaches* L *as* z *approaches* z_0, if for each $\epsilon > 0$ there exists $\delta > 0$ such that $\sigma(f(z), L) < \epsilon$ whenever $z \in R$ and $0 < \sigma(z, z_0) < \delta$. We say that f is *continuous* at z_0 if $\lim_{z \to z_0} f(z) = f(z_0)$, i.e., for each $\epsilon > 0$ there exists $\delta > 0$ such that $\sigma(f(z), f(z_0)) < \epsilon$ whenever $z \in R$ and $\sigma(z, z_0) < \delta$. We say f is continuous on a set U if it is continuous at each point of U. Due to Proposition B.7, this notion of continuity conforms with our usual notion of continuity in $\mathbb{C}$ (see Section A.2).

EXAMPLE B.8 (The map $1/z$ and division by zero). Our definition of continuity on $\overline{\mathbb{C}}$ allows us to solve the *division by zero* problem by defining $1/0 = \infty$. This makes sense and is natural since it is this definition that makes the function $z \mapsto 1/z$ continuous on $\overline{\mathbb{C}}$. We leave it to the reader to use Proposition B.7 to show this, and in the process show that we also have $1/\infty = 0$.

We also point out an important observation considering a rotation of the sphere $S = \{(x_1, x_2, x_3) : x_1^2 + x_2^2 + x_3^2 = 1\} \subset \mathbb{R}^3$. Consider rotating the sphere S about the x_1 axis, which meets S at the points $(1, 0, 0)$ and $(-1, 0, 0)$, by $180° = \pi$ radians. This is a map which takes each point of S to another point of S and so let us call this map f. Identifying S with $\overline{\mathbb{C}}$ as above, we see that $f(\infty) = 0, f(0) = \infty, f(1) = 1, f(-1) = -1$ and $f(\overline{\mathbb{R}}) = \overline{\mathbb{R}}$, where $\overline{\mathbb{R}}$ denotes the extended real line $\mathbb{R} \cup \{\infty\}$. The reader should take some time to convince themself that f is exactly the map $f(z) = 1/z$. It is for this reason that we often call $f(z) = 1/z$ a rotation of the Riemann sphere $\overline{\mathbb{C}}$.

Some texts avoid the need to speak explicitly about the spherical metric while still speaking about continuity of functions which are defined at ∞ or which take on the value ∞ at finite points $z_0 \in \mathbb{C}$. Since this approach is also valuable, we present it here. The key is to use the simple rotation $z \mapsto 1/z$ of the Riemann sphere $\overline{\mathbb{C}}$ in judicious ways to "move" ∞ to 0 so that we can then use our standard notions of continuity (and, as we shall see, analyticity) in $\mathbb{C}$ without needing to make explicit reference to the spherical metric. This is how one may prove the following result.

THEOREM B.9. Let $z_0, w_0 \in \mathbb{C}$ and let $f(z)$ be a complex valued function defined in a deleted neighborhood of z_0. Then,

i) $\lim\limits_{z \to z_0} f(z) = \infty$ if and only if $\lim\limits_{z \to z_0} \dfrac{1}{f(z)} = 0$,

ii) $\lim\limits_{z \to \infty} f(z) = w_0$ if and only if $\lim\limits_{z \to 0} f\left(\dfrac{1}{z}\right) = w_0$,

iii) $\lim\limits_{z \to \infty} f(z) = \infty$ if and only if $\lim\limits_{z \to 0} \dfrac{1}{f(1/z)} = 0$.

By pre composing and/or post composing f with the map $z \mapsto 1/z$, the above theorem converts each statement on the left (involving ∞) to the corresponding statement on the right (which avoids a formal notion of ∞). By considering the topology of the Riemann sphere $\overline{\mathbb{C}}$ near ∞ and the continuity of the map $z \mapsto 1/z$, the details can easily be checked (and they can also be found on p. 51 of [**1**]).

The simplest and most useful class of maps which are continuous on all of $\overline{\mathbb{C}}$ is the class of rational maps.

DEFINITION B.10. *A* quotient of two polynomials is called a *rational function*.

Let $f(z) = \frac{P(z)}{Q(z)}$ be a rational function in reduced form (i.e., polynomials $P(z)$ and $Q(z)$ have no common factors). Although this *formula* for f is defined only for complex values where Q is not zero, we can regard f (as a mapping into $\overline{\mathbb{C}}$) as being both defined and continuous on all of $\overline{\mathbb{C}}$. In particular, if $Q(a) = 0$, then we set $f(a) = \infty$, and we also set $f(\infty) = \lim_{z \to \infty} f(z)$. We leave it to the reader to check that this gives f the desired continuity properties; however, we do illustrate this in the following examples.

EXAMPLE B.11. Let $f(z) = z^2 + 5i$ and $g(z) = \frac{3z^2-5}{z^2+2z}$. In line with the above discussion, it is then understood that $f(\infty) = \infty, g(\infty) = 3, g(-2) = \infty$, and $g(0) = \infty$. By defining these maps in this way, we see that both f and g are continuous at every point of $\overline{\mathbb{C}}$. Note that $f(\infty) = \lim_{z \to \infty} f(z) = \infty$ since (using Theorem B.9) $\lim_{z \to 0} \frac{1}{f(1/z)} = \lim_{z \to 0} \frac{1}{(1/z)^2+5i} = \lim_{z \to 0} \frac{z^2}{1+5iz^2} = 0$. We leave it to the reader to use Theorem B.9 to check the continuity of g at the points $-2, 0$, and ∞.

EXAMPLE B.12. The function $f(z) = e^z$ on $\mathbb{C}$ cannot be defined at ∞ to make it continuous there. One way to see this is to note that $\lim_{x \to +\infty} e^x = \infty$, but $\lim_{x \to -\infty} e^x = 0$, which implies that the limit $\lim_{z \to \infty} e^z$ does not exist.

B.5. Analyticity in the Riemann Sphere

Just as the rotation $z \mapsto 1/z$ was used in Theorem B.9 to help us to understand the notion of *continuity* in the Riemann sphere $\overline{\mathbb{C}}$, we can also use it to extend our notion of *analyticity* on $\overline{\mathbb{C}}$.

DEFINITION B.13 (Extended notion of analyticity in $\overline{\mathbb{C}}$).
Let $f : \mathrm{domain}(f) \to \overline{\mathbb{C}}$, where $\mathrm{domain}(f) \subset \overline{\mathbb{C}}$.

i) If $f(z_0) = \infty$ where $z_0 \in \mathbb{C}$, then we say that f is *analytic* at z_0 exactly when $\frac{1}{f(z)}$ is analytic at z_0.

425

ii) If $f(\infty) \neq \infty$, then we say that f is *analytic* at ∞ exactly when $f(\frac{1}{z})$ is analytic at 0.

iii) If $f(\infty) = \infty$, then we say that f is *analytic* at ∞ exactly when $\frac{1}{f(\frac{1}{z})}$ is analytic at 0.

One way to summarize the above definition is to say that a map is analytic in this extended sense if after "moving" each instance of ∞ to 0 (by pre composing and/or post composing with the map $z \mapsto 1/z$), we get a map which is analytic in the usual sense.

REMARK B.14. We note that a map f which is analytic at a point $z_0 \in \overline{\mathbb{C}}$ must also be continuous there (as a mapping into $\overline{\mathbb{C}}$). Hence we can easily see that the map $f(z) = e^z$ cannot be analytic at ∞ since there is no way to define f at ∞ in such a way as to make it continuous there.

EXERCISE B.15. Show that $z \mapsto \sin \frac{1}{z}$ is analytic at ∞, but that $z \mapsto \sin z$ is not.

REMARK B.16 (Poles are points of analyticity in the extended sense). When $z_0 \in \mathbb{C}$ is a pole of f of order k, then by Definition A.15, we know that we can express $f(z) = (z - z_0)^{-k} h(z)$ where h is analytic at z_0 (in the usual sense of Section A.2) and $h(z_0) \neq 0$. Thus, $\frac{1}{f(z)} = (z - z_0)^k / h(z)$ is analytic at z_0, which means we can then declare f to be analytic at z_0 by Definition B.13(i), in that extended sense. Since the word *analytic* is commonly used in this text in both the usual sense of Section A.2 and this extended sense, the reader must always be careful to use context to decide in which sense it is being used. Since the context is usually quite clear, no confusion should arise.

We note that a map f with an isolated singularity at $z_0 \in \mathbb{C}$ (see Definition A.15) is analytic in the usual sense in the case of a removable singularity and analytic in the extended sense in the case of a pole. Thus, only when the singularity is essential can the map not be regarded as analytic in either sense.

EXAMPLE B.17. Let $f(z) = \frac{1}{1+z^2}$. Since $f(\frac{1}{z}) = \frac{1}{1+(\frac{1}{z})^2} = \frac{z^2}{z^2+1}$ is analytic at 0 (and takes the value 0 at 0), we can say that f is analytic at ∞ (and $f(\infty) = 0$). Similarly, since $1/f(z) = 1 + z^2$ is analytic at $\pm i$, the map f is is analytic at the poles $\pm i$, a fact we stated more generally in Remark B.16.

EXAMPLE B.18. Let $f(z) = 3z - \frac{1}{z}$ and note that $f(\infty) = \infty$. Since $k(z) = \frac{1}{f(\frac{1}{z})} = \frac{1}{3(\frac{1}{z})-z} = \frac{z}{3-z^2}$ is analytic at zero we pronounce $f(z)$ to be analytic at ∞. Note that, we have $f'(z) = 3 + \frac{1}{z^2}$ and so $f'(\infty) = 3$. We also have $k'(z) = \frac{3+z^2}{(3-z^2)^2}$ yielding $k'(0) = 1/3 = 1/f'(\infty)$. The relationship between $k'(0)$ and $f'(\infty)$ is an important and general property which we state as follows.

LEMMA B.19. If f is analytic at ∞ with $f(\infty) = \infty$, then for $k(z) = \frac{1}{f(\frac{1}{z})}$ we have $k'(0) = 1/f'(\infty)$.

Note, when $f'(\infty) = \infty$ we use the convention that $1/\infty = 0$ (since the map $1/z$ sends ∞ to 0).

PROOF. Note that k is analytic at 0 (by the definition of f being analytic at ∞). Also note that $k(0) = 0$ (since $f(\infty) = \infty$) and so we let N denote the multiplicity of the zero of k at 0. There exists $r > 0$ such that k is analytic on $\Delta(0, r)$. Therefore, the map $f(z) = \frac{1}{k(\frac{1}{z})}$ is analytic on $\{z \in \mathbb{C} : |z| > 1/r\}$. Hence f may be represented by a Laurent series $f(z) = \sum_{n=0}^{\infty} a_n z^n + \sum_{n=1}^{\infty} b_n z^{-n}$ on $\{z \in \mathbb{C} : |z| > 1/r\}$. Since $\frac{1}{k(z)} = f(\frac{1}{z}) = \sum_{n=0}^{\infty} a_n z^{-n} + \sum_{n=1}^{\infty} b_n z^n$ has a pole at 0 of order N, we must have $a_N \neq 0$ and $a_n = 0$ for all $n > N$. Hence we may now factor $f(z) = z^N h(z)$ where $h(z) = \cdots + a_{N-1} z^{-1} + a_N$.

Calling $g(z) = h(1/z) = a_N + a_{N-1} z + \ldots$, we can then express $k(z) = \frac{1}{f(\frac{1}{z})} = \frac{1}{z^{-N} h(1/z)} = \frac{z^N}{g(z)}$. Thus $k'(z) = \frac{N z^{N-1} g(z) - g'(z) z^N}{[g(z)]^2}$ and so

$$k'(0) = \begin{cases} 1/a_N & \text{when } N = 1 \\ 0 & \text{when } N > 1. \end{cases}$$

Note that $h'(z) = \cdots - a_{N-1} z^{-2}$ and so as $z \to \infty$ we have

$$f'(z) = N z^{N-1} h(z) + z^N h'(z) \to \begin{cases} a_N & \text{when } N = 1 \\ \infty & \text{when } N > 1. \end{cases}$$

Thus we have $k'(0) = 1/f'(\infty)$ as promised. $\qquad\square$

Bibliography

1. Ruel V. Churchill and James Ward Brown. *Complex variables and applications.* McGraw-Hill Book Co., New York, eighth edition, 2009.
2. Bruce P. Palka. *An introduction to complex function theory: Undergraduate Texts in Mathematics.* Springer-Verlag, New York, 1991.
3. Edward B. Saff and Arthur David Snider. *Fundamentals of Complex Analysis with Applications to Engineering, Science, and Mathematics.* Prentice Hall, New Jersey, third edition, 2003.